October 2008 Volume 4 Issue 5

Trade Science Inc.

Organic CHEMISTRY

An Indian Journal

Full Paper

OCAIJ, 4(5), 2008 [369-374]

Topological descriptors based quantitative structure activity relationship study of aromatic amines

P.P.Singh*, Anil Kumar Prajapati

Department of Chemistry, M.L.K.(P.G.)College, Balrampur,U.P., (INDIA)

E-mail: dr_ppsingh@sify.com

Received: 13th August, 2008 ; Accepted: 18th August, 2008

ABSTRACT

QSAR study of 29 aniline derivatives whose carcinogenic activities are reported in terms of log P has been made. QSAR models have been developed with the help of descriptors,connectivity index,valence connectivity index,shape index,molecular weight,accessibility surface area and molar refractivity. Thirty-eight models have been found to have high degree of predictive power with regression coefficient above 0.9 and 12 models above 0.9597. The combination of descriptors providing the best model is log P,valence connectivity index,shape index and molecular weight.

KEYWORDS

QSAR;
Aniline derivatives;
Valence connectivity index;
Molar refractivity;
Shape index.

INTRODUCTION

Aromatic amines[1] are common contaminant in several working environments,including the chemical and mechanical industries. 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 unparallel for other chemical classes. Although the major concern posed by aromatic amines derives from carcinogenic potential,the number of QSAR studies is quite limited[3],hence needs a comprehensive study on QSAR of aromatic amines whose biological activity is reported. Recently sets of quantum chemical descriptors have been used by us for QSAR studies of testosterone and estrogen derivatives[4-8] and topological descriptors[9] for study of toxicity of alcohols. In this paper we report QSAR studies on carcinogencity of aromatic amines with the help of topological descriptors. The quality of QSAR has been evaluated by multi linear regression analysis.

MATERIAL AND METHOD

The study materials of this paper are derivatives of aniline and are presented in TABLE 1 alongwith their carcinogenic activity in term of log P. For QSAR prediction,the 3D modeling and geometry optimization[10-11] of all the compounds have been done with the help of Cache software using the semiempirical PM3 Hamiltonian. The values of various descriptors have been calculated using Cache Software.

Regression analysis has been made by Project Leader Program associated with Pro software of Fuzitsu. The values of descriptors that have been use for QSAR model have been evaluated using the same software by PM3 Hamiltonian method. The descriptors that have been used are defined below:

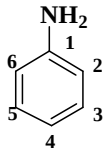
The Vertex-Connectivity Index, $\chi = \chi(G)$, of a (molecular) graph G is defined as[12]:

$$\chi = \sum_{\text{edges}} [d(\nu_i)\, d(\nu_j)]^{-1/2} \quad (1)$$

The sum is taken over all edges of G;$d(\nu_i)$ is the

Full Paper

TABLE 1: Derivatives of aniline (along with their biological activity in term of log P)



Aniline

Compd. no.	Substituents at ring	Substituents at functional	amino group	Activity
1	2-Me	H		1.730
2	2-OEt, 5-NHCOMe	H		0.200
3	3-NO_2, 4-OH	H		0.930
4	H	H		1.260
5	2-Ome	H		1.010
6	4-Cl	H		1.780
7	2Cl, 5-NH_2	H		1.000
8	2NH_2, 4Cl	H		1.000
9	2Me, 4-OMe	H		1.000
10	2-OMe, 5-Me	H		1.480
11	2-OMe,5-NH_2	H		0.230
12	3-NO_2,4-OEt	COMe		0.940
13	2-OMe, 5-NO_2	H		0.960
14	2-NO_2, 4-NH_2	H		0.430
15	2,4,5-Me_3	H		2.670
16	2-OH, 4-NO_2	H		0.930
17	2-OH, 5-NO_2	H		0.200
18	4-OEt	COMe		0.990
19	4-F	Me, NO		1.830
20	H	Me, NO		1.690
21	2-NH_2	H		0.480
22	2,4,5,6-F_4,3-NH_2	H		1.040
23	2,4,6-Me_3	H		2.670
24	H	Me		1.840
25	4-Me	H		1.730
26	2-OH, 5-NO_2	H		0.930
27	2,4,6-Cl_3	H		2.820
28	3-Me	H		1.730
29	2,5-Cl_2, 3-COOH	H		2.000

vertex-degree and $[d(\nu_i)\, d(\nu_j)]^{-1/2}$ is the weight of the i-j edge. The degree of a vertex ν_i,$d(\nu_i)$,is equal to the number of adjacent vertices. Two vertices of graph G are adjacent if there is an edge joining them.

In the case of heterosystems, the connectivity index is given in terms of valence delta values $\delta(\nu_i)$ and $\delta(\nu_j)$ of atom i and j. This kind of connectivity index (valence-connectivity index, $(\chi\nu)$ is defined as:

$$W = (1/2) \sum_{ij} [\delta(\nu_i)\, \delta(\nu_j)]^{-1/2} \quad (2)$$

Valence delta values are available for many kind of atom. The connectivity index can be generalized to include also the weighted paths pe of length l, not only the weight edges (weighted path pl of length one):[13,14]

$$^{1}\chi = \sum_{\text{paths}} [d(\nu_i)\, d(\nu_j) \ldots d(\nu_l + 1)]^{-1/2} \quad (3)$$

Where $d(\nu_i)\, d(\nu_j) \ldots d(\nu_l + 1)$ are valences of vertices $d(\nu_i)\, d(\nu_j) \ldots, \nu_l + 1$ in the considered path of length l. The first-, $^{1}\chi\nu$, second-, $^{2}\chi\nu$, and third-order, $^{3}\chi\nu$, valence-connectivity indices in this paper are calculated by Eq. (3) when valence delta are introduced into it for heterosystems.

Shape index

Shape index quantifying the shape of a chemical sample[15]. The shape index of order 1 (Kappa 1) quantifies the number of cycles in the chemical sample.

Solvent accessible surface (SAS) area

The solvent accessible surface (SAS) area is calculated at an optimized geometry in water, the water geometry is from optimization first using Augmented MM2, then using MOPAC with PM3 parameters and the conductor like screening Model (COSMO)[16].

Molar refractivity

Molar refractivity is calculated using the atom-typing SCHEME of Ghose and Crippen[17].

Log P

Log P is calculated using the atom-typing scheme of Ghose and Crippen[17].

Finally a more general but important property of chemical system is the molecular weight (MW), which has been tested as descriptors

RESULT AND DISCUSSION

The values of the descriptors of aniline derivatives obtained by PM3 calculations are included in TABLE 2. With the help of these values 90 QSAR models using different combination of descriptors have been tried, out of which 38 models provided regression coefficient above 0.9 and 12 models above 0.9597. The predicted activities of the 12 models (PA1-PA12) are included in TABLE 3 and the regression equation developed with the values of descriptors are given below:

PA1 = 0.876556*LPC+0.0820655* VCI -0.168902*SI +0.00591286*MW + 0.00734093
rCV^2 = 0.934529
r^2 = 0.968689

PA2 = 0.901787*LPC+0.129689*SI +0.00591284*MW +0.0044592*MR-0.0155747
rCV^2 = 0.927442

TABLE 2: Values of descriptors of aniline derivatives

S.no. of compd.	Log P calcd.	Connectivity index	Valence connectivity index	Shape index	Molecular weight	Solvent accessibility surface area	Molar Refractivity
1	1.731	3.805	4.887	5.319	107.000	69.000	35.800
2	0.202	6.630	8.333	10.848	194.000	105.000	55.051
3	1.027	5.236	5.520	7.803	154.000	82.000	39.376
4	1.263	3.394	3.964	4.342	93.000	62.000	30.758
5	1.011	4.343	5.295	6.263	123.000	74.000	37.222
6	1.781	3.788	5.021	5.604	128.000	74.000	35.563
7	0.998	4.198	5.521	6.549	143.000	78.000	40.264
8	0.998	4.198	5.521	6.549	143.000	79.000	40.264
9	0.758	5.274	6.626	8.199	153.000	85.000	43.685
10	1.478	4.736	6.218	7.249	390.000	81.000	42.263
11	0.227	4.736	5.795	7.210	138.169	80.000	41.922
12	1.033	7.668	9.019	12.440	224.216	114.000	57.275
13	1.058	5.774	6.481	8.793	168.192	89.000	44.145
14	0.053	5.236	5.650	7.803	153.140	83.000	42.382
15	2.665	4.609	6.732	7.289	135.208	81.000	45.882
16	1.027	5.236	5.520	7.803	154.125	81.000	39.376
17	0.196	4.198	4.834	6.223	124.142	72.000	37.153
18	0.985	6.220	7.833	9.895	179.218	101.000	50.351
19	1.830	5.236	5.990	7.813	154.144	82.000	40.045
20	1.691	4.843	5.689	6.894	136.153	78.000	39.828
21	0.480	3.805	4.464	5.280	108.143	68.000	35.459
22	1.038	5.464	5.667	8.992	180.105	79.000	36.324
23	2.665	4.609	6.732	7.289	135.208	82.000	45.882
24	2.311	4.305	5.834	6.302	121.182	75.000	40.487
25	1.731	3.788	4.887	5.319	107.155	70.000	35.800
26	1.027	5.236	5.520	7.803	154.125	82.000	39.376
27	2.817	4.609	7.134	8.150	196.463	94.000	45.173
28	1.431	3.788	4.887	5.319	107.155	70.000	35.800
29	1.998	5.520	7.355	9.478	206.028	96.000	47.126

TABLE 3: Predicted activities PA1 to PA12 of compound

S. no. of compd.	PA1	PA2	PA3	PA4	PA5	PA6	PA7	PA8	PA9	PA10	PA11	PA12
1	1.661	1.651	1.647	1.639	1.639	1.625	1.667	1.626	1.686	1.683	1.68	1.68
2	0.185	0.159	0.141	0.156	0.152	0.278	0.172	0.19	0.275	0.275	0.283	0.284
3	0.954	0.989	1.011	1.006	1.008	0.947	0.999	1.02	0.932	0.938	0.934	0.934
4	1.257	1.251	1.255	1.244	1.245	1.189	1.239	1.21	1.25	1.252	1.247	1.247
5	0.998	0.981	0.982	0.978	0.978	0.967	1.005	0.974	1.025	1.021	1.02	1.02
6	1.789	1.78	1.785	1.787	1.786	1.733	1.698	1.733	1.71	1.717	1.717	1.717
7	1.072	1.062	1.052	1.062	1.061	1.073	0.995	1.032	1.007	1.007	1.012	1.012
8	1.072	1.062	1.052	1.062	1.061	1.073	0.979	1.026	1.004	1.007	1.012	1.012
9	0.736	0.709	0.706	0.709	0.71	0.749	0.777	0.742	0.799	0.789	0.792	0.793
10	1.253	1.224	1.193	1.167	1.169	1.113	1.21	1.079	1.125	1.117	1.107	1.106
11	0.282	0.262	0.248	0.254	0.253	0.31	0.275	0.264	0.317	0.311	0.315	0.315
12	0.878	0.889	0.901	0.907	0.906	0.946	0.922	0.954	0.947	0.956	0.957	0.957
13	0.976	0.994	1.009	1.006	1.007	0.978	1.018	1.025	0.988	0.992	0.99	0.99
14	0.521	0.547	0.551	0.55	0.55	0.539	0.526	0.551	0.508	0.514	0.513	0.512
15	2.464	2.45	2.429	2.426	2.427	2.501	2.532	2.463	2.559	2.544	2.545	2.545
16	0.954	0.989	1.011	1.006	1.008	0.947	1.001	1.021	0.933	0.938	0.934	0.934
17	0.259	0.257	0.249	0.253	0.253	0.28	0.255	0.258	0.261	0.257	0.259	0.259
18	0.902	0.879	0.875	0.882	0.878	0.938	0.878	0.891	0.979	0.986	0.989	0.99
19	1.695	1.715	1.74	1.73	1.733	1.659	1.77	1.752	1.695	1.694	1.689	1.689
20	1.597	1.601	1.612	1.6	1.601	1.546	1.62	1.59	1.611	1.613	1.608	1.608
21	0.542	0.533	0.523	0.521	0.52	0.529	0.517	0.502	0.545	0.543	0.543	0.543
22	0.935	0.991	1.036	1.038	1.049	0.967	1.164	1.165	0.891	0.866	0.867	0.867
23	2.464	2.45	2.429	2.426	2.427	2.501	2.522	2.458	2.557	2.544	2.545	2.545
24	2.164	2.151	2.144	2.135	2.136	2.142	2.208	2.142	2.225	2.217	2.215	2.215
25	1.661	1.651	1.647	1.64	1.639	1.628	1.652	1.621	1.683	1.683	1.68	1.68
26	0.954	0.989	1.011	1.006	1.008	0.947	0.997	1.02	0.932	0.938	0.934	0.934
27	2.848	2.836	2.844	2.867	2.865	2.816	2.674	2.801	2.638	2.646	2.654	2.654
28	1.661	1.651	1.647	1.64	1.639	1.628	1.653	1.621	1.684	1.683	1.68	1.68
29	1.98	1.99	2.005	2.022	2.023	1.993	1.933	2.023	1.848	1.848	1.854	1.854

Full Paper

TABLE 4: Good QSAR models with regression coefficient (r^2) cross-validation coefficient (rcv^2) and descriptors used

S. no.	Predicted activity	rCV^2	r^2	Descriptors used
1	PA1	0.934529	0.968689	Log P calcd. , valence connectivity index, shape index, Molecular weight
2	PA2	0.927442	0.966753	Log P calcd., shape index, Molecular weight, molar refractivity
3	PA3	0.921406	0.966175	Log P calcd., connectivity index, shape index, Molecular weight
4	PA4	0.925457	0.965999	Log P calcd., shape index, Molecular weight, solvent accessibility surface area
5	PA5	0.926311	0.965985	Log P calcd., shape index, Molecular weight
6	PA6	0.91164	0.96217	Log P calcd., connectivity index, Molecular weight, molar refractivity
7	PA7	0.906737	0.96158	Log P calcd., valence connectivity index, shape index, Molecular weight, solvent accessibility surface area
8	PA8	0.901344	0.960846	Log P calcd., connectivity index, Molecular weight, solvent accessibility surface area
9	PA9	0.891479	0.960463	Log P calcd., valence connectivity index, shape index, Molecular weight, solvent accessibility surface area
10	PA10	0.899623	0.960463	Log P calcd., valence connectivity index, shape index, solvent accessibility surface area
11	PA11	0.908388	0.960287	Log P calcd., valence connectivity index, shape index
12	PA12	0.889986	0.960287	Log P calcd., valence connectivity index, shape index, molar refractivity

r^2 = 0.966753

PA3 = 0.908926*LPC+0.0323916 *CI - 0.139629*SI +0.00647117* MW + 0.000242272
rCV^2 = 0.921406
r^2 = 0.966175

PA4=0.902819*LPC-0.127606SI+ 0.00657045*MW +0.000832663* SASA-0.00595157
rCV^2 = 0.925457
r^2 = 0.965999

PA5 = 0.904117*LPC-0.12361*SI+ 0.00663844*MW +0.0209738
rCV^2 = 0.926311
r^2 = 0.965985

PA6 = 0.876544*LPC-0.170017*CI + 0.00308498*MW +0.0106198*MR+0.0443503
rCV^2 = 0.91164
r^2 = 0.96217

PA7 = .910534*LPC+0.0818986* VCI+ 0.00340623MW +0.0187376*SASA+0.615617
rCV^2 = 0.906737
r^2 = 0.96158

PA8=0.909616*LPC-0.0660378* CI+ 0.00486398* MW-0.00776071*SASA+ 0.315788
rCV^2 = 0.901344
r^2 = 0.960846

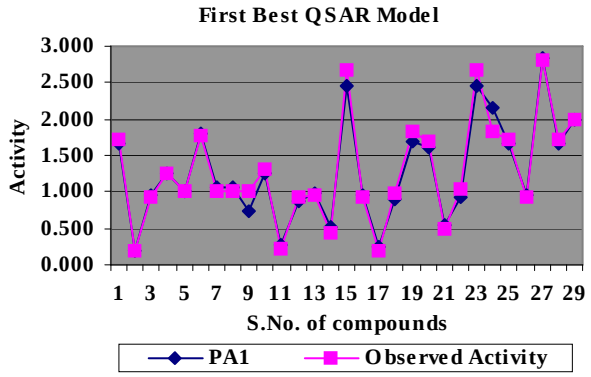
PA9 = 0.874749*LPC+ 0.128346*VCI-0.0702169*SI-0.0035159* SASA+ 0.160653
rCV^2 = 0.891479
r^2 = 0.960463

PA10 = 0.877748*LPC+ 0.0160929*CI+0.109343VCI-0.0883073*SI+ 0.0381659
rCV^2 = 0.899623
r^2 = 0.960329

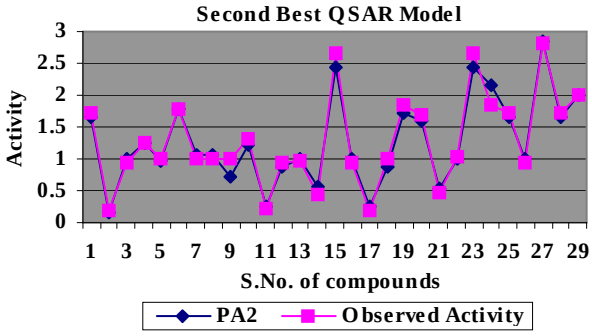
PA11 = 0.874079*LPC+ 0.113356*VCI - 0.081681*SI+

TABLE 5: Values of observed activity, predicted activity and the difference between these two activities

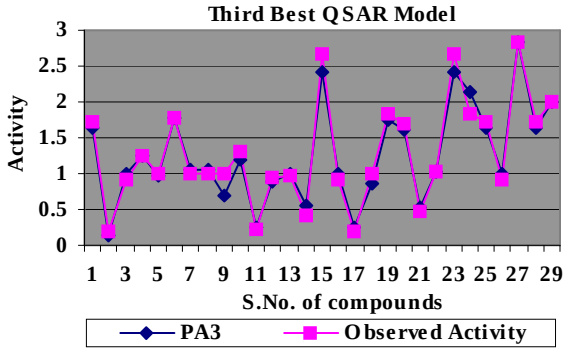
S. no. of compound	Predicted activity PA1	Observed activity	Difference
1	1.661	1.730	0.069
2	0.185	0.200	0.015
3	0.954	0.930	-0.024
4	1.257	1.260	0.003
5	0.998	1.010	0.012
6	1.789	1.780	-0.009
7	1.072	1.000	-0.072
8	1.072	1.000	-0.072
9	0.736	1.000	0.264
10	1.253	1.310	0.057
11	0.282	0.230	-0.052
12	0.878	0.940	0.062
13	0.976	0.960	-0.016
14	0.521	0.430	-0.091
15	2.464	2.670	0.206
16	0.954	0.930	-0.024
17	0.259	0.200	-0.059
18	0.902	0.990	0.088
19	1.695	1.830	0.135
20	1.597	1.690	0.093
21	0.542	0.480	-0.062
22	0.935	1.040	0.105
23	2.464	2.670	0.206
24	2.164	1.840	-0.324
25	1.661	1.730	0.069
26	0.954	0.930	-0.024
27	2.848	2.820	-0.028
28	1.661	1.730	0.069
29	1.980	2.000	0.020



Graph 1: Graph between predicted activity PA1 and observed activity



Graph 2: Graph between predicted activity PA2 and observed activity



Graph 3: Graph between predicted activity PA3 and observed activity

0.0481882
rCV^2 = 0.908388
r^2 = 0.960287

PA12 =0.873820*LPC+0.114395* VCI- 0.0819386*SI0.00011381 *MR +0.0490193
rCV^2 = 0.889986
r^2 = 0.960287

On the basis of the values of regression coefficient, the QSAR models have been arranged in decreasing order of quality of prediction and are included in TABLE 4.

The predicted activity PA1 gives best QSAR models with regression coefficient 0.968689 and cross validation coefficient 0.934529. With the help of these MLR equations, the activity of aniline derivatives can be best predicted. In all the above QSAR models, the first descriptor is Log P. Even the single descriptor LPC provides good predictive power. The descriptor LPC is supposed to be the prime descriptor for QSAR models. The predicted activities and the observed activity are included in TABLE 5. The closeness between the predicted activity PA1-PA3 and the observed activity is well demonstrated by the graphs drawn between the two activities in Graph1-3.

In the second best QSAR model the descriptors are log P calcd., shape index, molecular weight and molar refractivity. In this case the predicted activity is PA2 and is given by:

PA2 = 0.901787*LPC+0.129689*SI+ 0.00591284*MW+ 0.0044592*MR-0.0155747
rCV^2 = 0.927442
r^2 = 0.966753

Graph between predicted activity and observed activity PA2 of the compounds of aniline derivatives is shown in the Graph-2.

In the third best QSAR model the combination of descriptor are Log P, connectivity index, shape index and molecular weight. In this case the predicted activity is given by PA3.

PA3 = 0.908926*LPC+0.0323916*CI-0.139629*SI+ 0.00647117*MW+ 0.000242272
rCV^2 = 0.921406
r^2 = 0.966175

REFERENCES

[1] D.Y.Lai, Y.T.Woo, M.F.Argus, J.C.Arcos, S.C. DeVito, R.L.Garrett; American Chemical Society: Washington, D.C., 1-13 **(1996)**.

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

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

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

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

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

[7] F.A.Pasha, H.K.Srivastava, P.P.Singh; Mol.Diver.,

9(1-3), 215 **(2005)**.
[8] H.K.Srivastava, F.A.Pasha, P.P.Singh; Int.J. Quan.Chem., **103**, 237 **(2005)**.
[9] R.K.Singh, A.K.R.Khan, V.K.Sahu, P.P.Singh; Int.J.Quan.Chem., **108**, **(2008)**.
[10] M.C.Flanigan;, A.Komornicki, J.W.Mciver; 'Modern Theoritical Chemistry Plenum Press', New York, 8 **(1977)**.
[11] M.J.S.Dewar, H.S.Rzepa, J.Am.Chem.Soc., **100**, 777 **(1978)**.
[12] L.B.Kier, L.H.Hall, W.J.Murray, M.Randic; J. Pharm.Sci., 65 **(1976)**.
[13] M.Randic, S.Sabljic, S.Nikolic, N.Trinajstic; Int.J. Quantum.Chem., Quantum Biol.Symp., **15**, 267 **(1988)**.
[14] S.Nikolic, M.Medic-Saric, Matijevic-Soca; Croat. Chem.Acta, 66 **(1993)**.
[15] L.H.Hall, L.B.Kier, K.B.Lipkowitz, D.B.boyd; 'Reviews in Computational Chemistry', Ch.9, **(1992)**.
[16] A.Klamt, G.Schuurmann; J.Chem.Soc.Perkin Trans., **2**, 799 **(1993)**.
[17] A.K.Ghose et al.; J.Comput.Chem., **9**, 80 **(1988)**.