IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

QSPR Studies on the Melting Point, Partition Coefficient and Aqueous Solubility.

Madhu Gupta[1], Bhakti Kumari[2,] Monika Sharma[3], Seema Kohli[4], Manish Rao Ambedkar[5],

Department of Chemistry, M.M.H. College, Ghaziabad, Uttar Pradesh, India[1,2,3,4,5],

Abstract:

As we know that melting point has numerous applications in biochemical and environmental sciences due to its relationship with solubility. Sufficient aqueous solubility is essential for a compound to be transferred to the site of action within an organism. In the present research work QSPR models of 29 compounds were developed using QSPR multiple linear regression (MLR).In our works, we calculated many descriptors from the 29 molecular structures by the software Dragon, MLR Methods was used to select descriptor and set up linear model. The square of correlation coefficient(R^2) for best model with penta molecular descriptors is 0.9311. All value's of best penta perametric model are N=29 , SE = (21.0089), R = (0.9649) , R^2 = (0.9311) , R^2A = (0.8162), F RATIO = (8.1029), Q = (0.0459).The results obtained offers excellent regression models that possesses good prediction ability.

Key words : QSPR, multiple linear regression (MLR), Aqueous solubility , Partition coefficient.

Introduction

Melting point is a basic physical property that specifies the transition temperature between solid and liquid phases. Melting point has numerous applications in biochemical and environmental sciences due to its relationship with solubility. Sufficient aqueous solubility is essential for a compound to be transferred to the site of action within an organism. In spite of the huge number of available melting point data, few useful guidelines exist for understanding the relationship between the compound melting point and its chemical structure.

Predicting Melting Points Predicting crystal structures and their physicochemical properties is an important research area. Predicting melting points is one small region of this research area. Melting points are an attractive property as the well established General Solubility Equation (GSE, Equation) 1,2,4 links the melting point to solubility with reference to a thermodynamic cycle via a pure melt: this empirically derived relationship has seen wide usage 3, 4, 5. The GSE has been proposed as a way to accurately predict solubility using only two pieces of empirical data; the MP is the melting point, the second logP. Log P can be reasonably predicted by atom or group 30 For this reason a good prediction of a crystal's melting point could in principle provide a direct useful prediction of a molecule's solubility.

The General Solubility Equation:

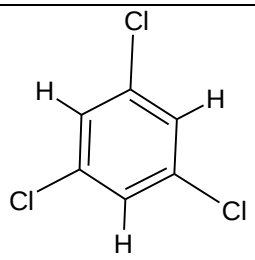
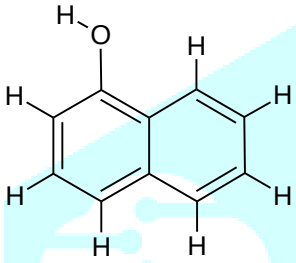
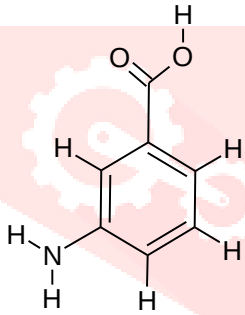
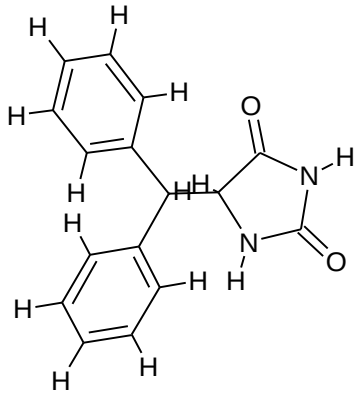
$$\log_{10}S = 0.08 - \log_{10}P - 0.01 \times (MP - 25)$$

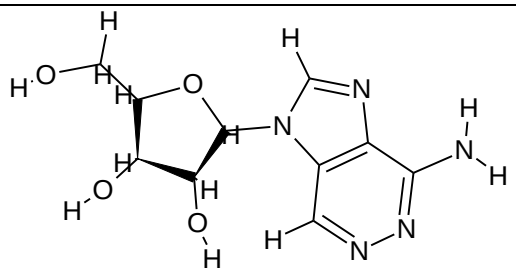
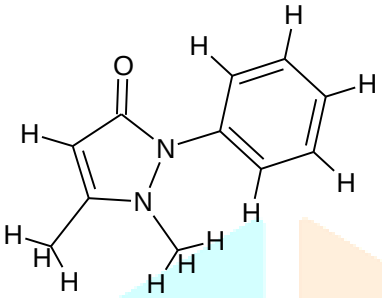
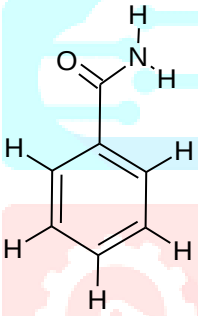
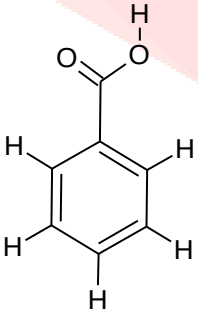
$$\log_{10}S = 0.05 - \log_{10}P - 0.01 \times (MP - 25)$$

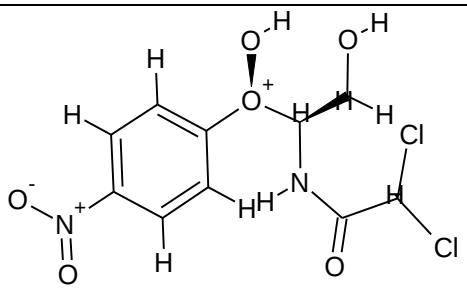
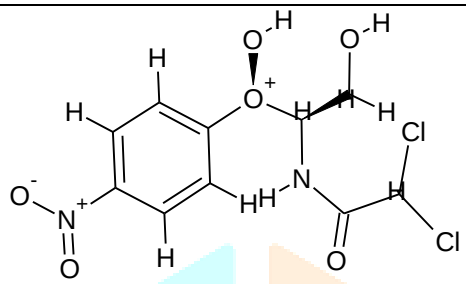
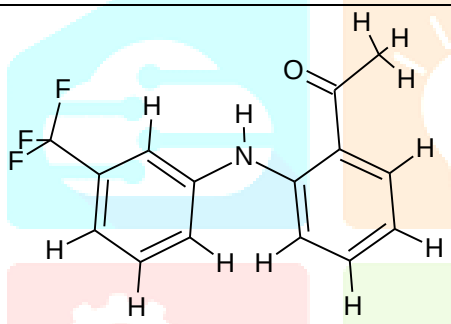
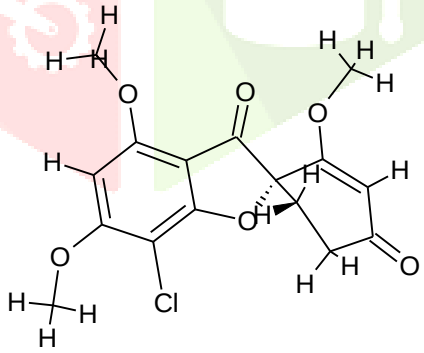
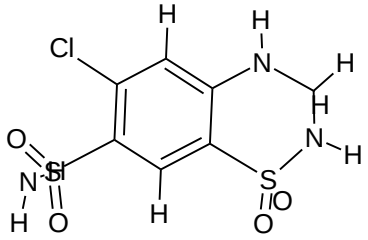
Therefore, methods for estimating the melting point of organic compounds would considerably help medicinal chemists in designing new drugs within a specified range of melting point and solubility. A highly effective tool depending on quantitative structure–property relationship (QSPR) can be utilized to predict melting point for drug-like compounds with no literature values.

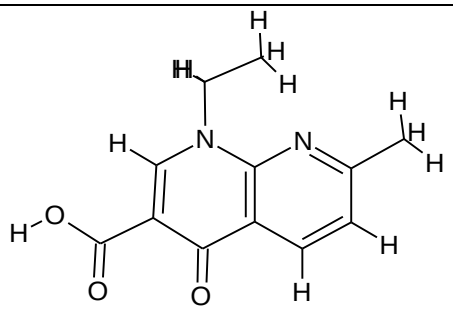
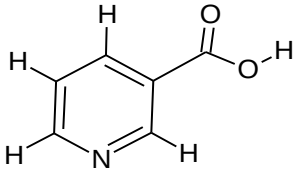
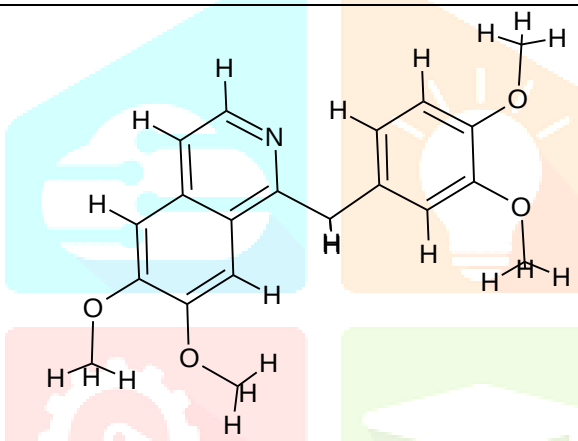
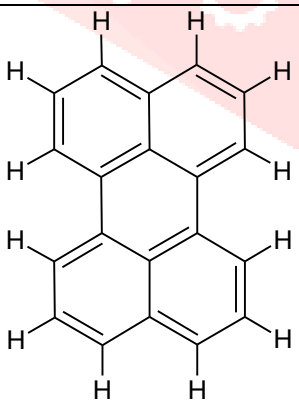
RESULTS AND DISCUSSION:

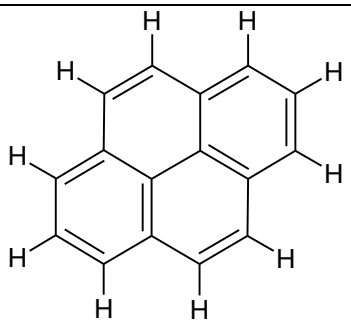
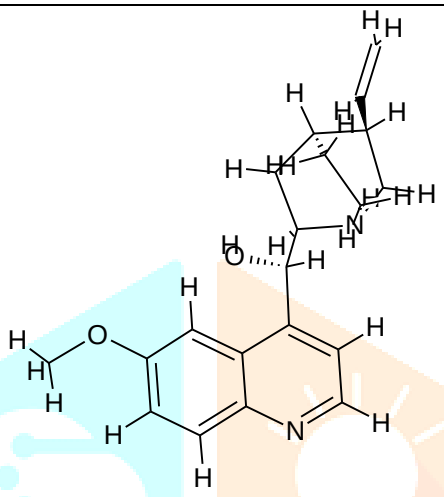
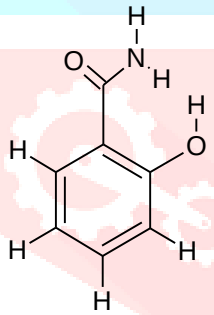
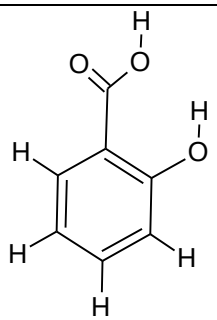
Table 1: Structure

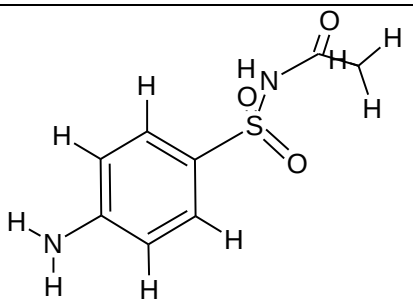
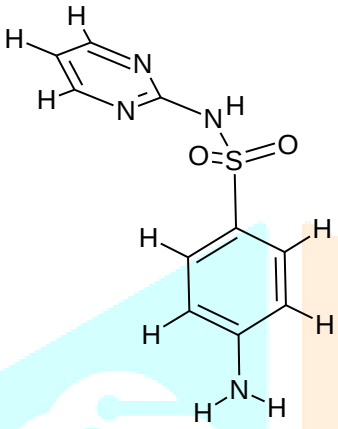
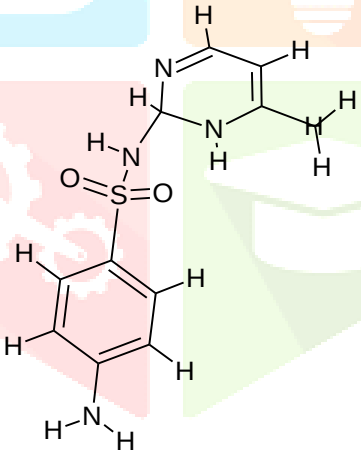
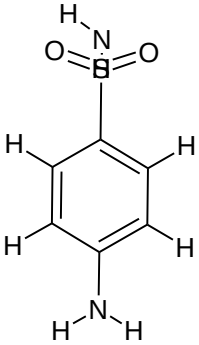
Sr.N.	Structure of The Compounds	M.P
1	 1,3,5-trichlorobenzene	63.5
2	 naphthalen-1-ol	96
3	 3-aminobenzoic acid	187.5
4	 (5*R*)-5-(diphenylmethyl)imidazolidine-2,4-dione	295.5

5	 (2*S*,3*S*,4*R*,5*S*)-2-(4-amino-1*H*-imidazo[4,5-*d*]pyridazin-1-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol	114.5
6	 1,5-dimethyl-2-phenyl-1,2-dihydro-3*H*-pyrazol-3-one	235
7	 benzamide	112.5
8	 benzoic acid	127

9	 1-{(1*S*)-1-[(dichloroacetyl)amino]-2-hydroxyethyl}-1-(4-nitrophenyl)dioxidanium	122.5
10	 1-{(1*S*)-1-[(dichloroacetyl)amino]-2-hydroxyethyl}-1-(4-nitrophenyl)dioxidanium	150.5
11	 1-(2-{[3-(trifluoromethyl)phenyl]amino}phenyl)ethanone	134
12	 (2*S*)-7-chloro-2',4,6-trimethoxy-3*H*,4'*H*-spiro[1-benzofuran-2,1'-cyclohex[2]ene]-3,4'-dione	219
13	 6-chloro-3,4-dihydro-2*H*-1,2,4-benzothiadiazine-7-sulfonamide 1,1-dioxide	269

14	 1-ethyl-7-methyl-4-oxo-1,4-dihydro-1,8-naphthyridine-3-carboxylic acid	229
15	 pyridine-3-carboxylic acid	237.5
16	 1-(3,4-dimethoxybenzyl)-6,7-dimethoxyisoquinoline	146.5
17		278

18		150
19	 (*S*)-[(2*R*,4*S*,5*R*)-5-ethenyl-1-azabicyclo[2.2.2]oct-2-yl](6-methoxyquinolin-4-yl)methanol	170
20	 2-hydroxybenzamide	140
21	 2-hydroxybenzoic acid	159

22	 *N*-[(4-aminophenyl)sulfonyl]acetamide	183
23	 4-amino-*N*-(pyrimidin-2-yl)benzenesulfonamide	254.5
24	 4-amino-*N*-[(2*R*)-6-methyl-1,2-dihydropyrimidin-2-yl]benzenesulfonamide	199.5
25	 4-aminobenzenesulfonamide	165.5

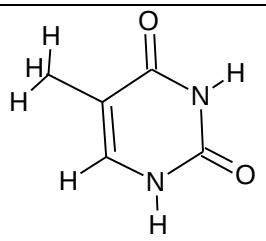
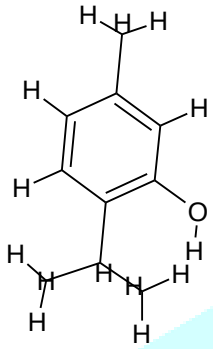
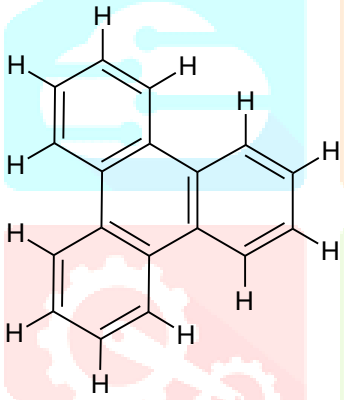
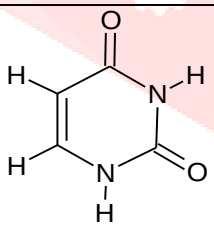
26	 5-methylpyrimidine-2,4(1*H*,3*H*)-dione	316.5
27	 5-methyl-2-(propan-2-yl)phenol	50.5
28		196.5
29	 pyrimidine-2,4(1*H*,3*H*)-dione	330

Table 2: Structural Details of Compounds with their activities and topological parameters used in the present study.

Comp No	M.P	TPSAToT	X0A	MLOGP	Mv	Jhetv
1	63.5	0	0.761	4.063	0.82	3.166
2	96	20.23	0.698	2.637	0.68	2.723
3	187.5	63.32	0.756	1.132	0.64	2.688
4	295.5	58.2	0.697	2.044	0.66	1.887
5	114.5	29.1	0.74	1.628	0.63	2.175
6	235	139.54	0.714	-1.158	0.61	1.466
7	112.5	26.93	0.724	2.306	0.63	1.711
8	127	43.09	0.743	1.294	0.64	2.75
9	122.5	37.3	0.743	1.7	0.65	2.639
10	150.5	118.08	0.764	1.694	0.64	1.885
11	134	29.1	0.738	4.134	0.66	1.889
12	219	71.06	0.729	0.739	0.65	1.671
13	269	135.12	0.756	-0.547	0.67	2.53
14	229	72.19	0.74	1.315	0.64	2.198
15	237.5	50.19	0.743	0.119	0.66	2.536
16	146.5	49.81	0.712	2.443	0.63	1.731
17	278	0	0.655	5.551	0.74	2.318
18	150	0	0.658	4.76	0.73	2.506
19	170	45.59	0.695	2.192	0.62	1.532
20	140	63.32	0.756	1.237	0.64	2.745
21	159	57.53	0.756	1.643	0.65	2.654
22	183	97.64	0.769	0.108	0.63	2.629
23	254.5	106.35	0.724	0.487	0.66	1.878
24	199.5	104.96	0.732	0.588	0.63	1.75
25	165.5	94.56	0.771	-0.115	0.63	2.953
26	316.5	65.72	0.761	-0.302	0.61	2.114
27	50.5	20.23	0.766	2.813	0.59	2.911
28	196.5	0	0.664	5.165	0.72	2.463
29	330	65.72	0.748	-0.712	0.63	1.947

Table 3: Correlation Matrix showing inter- correlation among all the parameters with the activity.

	M.P	*TPSAToT*	*X0A*	*MLOGP*	*Mv*	*Jhetv*
M.P	1					
TPSAToT	0.423679	1				
X0A	-0.15884	0.449452	1			
MLOGP	-0.41425	-0.81824	-0.60673	1		
Mv	-0.10864	-0.50533	-0.40136	0.639424	1	
Jhetv	-0.4062	-0.32941	0.312199	0.180742	0.373587	1

Table 5 Regression Model

Model No	Parameter Used	AI	B	SE	R	R2	R2A	F Ratio	Q=r/se
24	TPSA(tot)	0.7759 (±0.3192)	139.3364	67.0189	0.4237	**0.1795**	0.1491	5.9069	0.0063
32	TPSA(TOT) X0A	1.1361(±0.3285) -963.9435 (±395.0731)	823.7630	61.6059	0.5765	**0.3324**	0.2810	6.4719	0.0094
34	TPSA(TOT) X0A MLOGP	0.3036(±0.4764) -1 402.4991 (±414.3232) -27.4132 (±12.0516)	1 238.6301	57.1863	0.6685	**0.4469**	0.3805	6.7319	0.0117
49	TPSA(TOT) X0A MLOGP MV	0.2805(±0.4707) -1 393.1164 (±409.1331) -34.2012 (±13.0217) -34.2012 (±299.2836)	993.0849	56.4609	0.6945	**0.4824**	0.3961	5.5911	0.0123
52	TPSA(TOT) X0A MLOGP MV JHETV	0.0995(±0.5066) -1 104.1097 (±506.0181) -35.2393 (±13.0800) -35.2393 (±336.5451) -1 104.1097 (±32.1573)	767.4397	56.5245	0.7091	**0.5028**	0.3947	4.6521	0.0125

Mono Parametric

MP = 0.7759(±0.3192) TPSATOT +139.3364,

N=29 , SE = (67.0189), R = (0.4237) , R2 = (0.1795) , R2A = (0.1491), F RATIO = (5.9069), Q = (0.0063)

BI Parametric

MP = 1.1361(±0.3285) TPSATOT -963.9435 (±395.0731) X0A +823.7630,

N=29 , SE = (61.6059), R = (0.5765) , R2 = (0.3324) , R2A = (0.2810), F RATIO = (6.4719), Q = (0.0094)

Tri Parametric

MP = 0.3036(±0.4764) TPSATOT -1,402.4991 (±414.3232) X0A -27.4132 (±12.0516) MLOGP +1,238.6301,

N=29 , SE = (57.1863), R = (0.6685) , R2 = (0.4469) , R2A = (0.3805), F RATIO = (6.7319), Q = (0.0117)

Tetra Parametric

MP = 0.2805(±0.4707) TPSATOT -1,393.1164 (±409.1331) X0A -34.2012 (±13.0217) MLOGP -34.2012 (±299.2836) MV +993.0849,

N=29 , SE = (56.4609), R = (0.6945) , R2 = (0.4824) , R2A = (0.3961), F RATIO = (5.5911), Q = (0.0123)

Penta Parametric

MP = 0.0995(±0.5066) TPSATOT -1,104.1097 (±506.0181) X0A -35.2393 (±13.0800) MLOGP -35.2393 (±336.5451) MV -1,104.1097 (±32.1573) JHETV +767.4397, N=29 , SE = (56.5245), R = (0.7091) , R2 = (0.5028) , R2A = (0.3947), F RATIO = (4.6521), Q = (0.0125)

Table5: Observed and Estimated Model Topological Indices (MP)

Sn	Observed M.P	*Predicted M.P*	*Residuals*
1	63.5	122.1531	-58.6531
2	96	183.1973	-87.1973
3	187.5	156.2509	31.24907
4	295.5	224.4615	71.03845
5	114.5	163.7486	-49.2486
6	235	313.1318	-78.1318
7	112.5	171.8192	-59.3192
8	127	160.944	-33.944
9	122.5	154.8638	-32.3638
10	150.5	158.1769	-7.67688
11	134	102.5858	31.41422
12	219	237.8215	-18.8215
13	269	243.4948	25.50523
14	229	183.6764	45.3236
15	237.5	220.4119	17.08807
16	146.5	181.8911	-35.3911
17	278	170.6282	107.3718
18	150	183.9788	-33.9788
19	170	209.9796	-39.9796
20	140	150.768	-10.768
21	159	144.0622	14.93781
22	183	177.9108	5.089217
23	254.5	254.5892	-0.08923
24	199.5	230.0694	-30.5694
25	165.5	173.1205	-7.6205
26	316.5	203.4627	113.0373
27	50.5	48.05626	2.443737
28	196.5	159.0962	37.40382
29	330	248.1496	81.85037

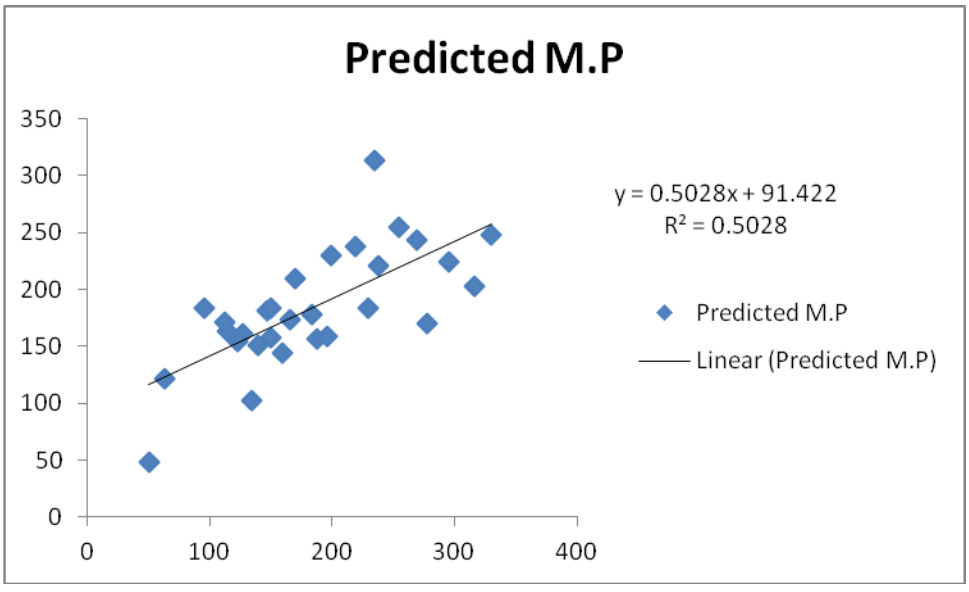


Table6: Cross validation of Topological Indices (MP)

Model No	Parameter Used	press	ssy	Press/SSY	R²CV	PSE	Spress
24	TPSA(tot)	121 271.27	147 802.33	0.820496	0.179504	64.66659	67.01887
32	TPSA(TOT) X0A	98 677.32	147 802.33	0.66763	0.33237	58.33238	61.60586
34	TPSA(TOT) X0A MLOGP	81 756.77	147 802.33	0.553149	0.446851	53.09613	57.18628
49	TPSA(TOT) X0A MLOGP MV	76 508.01	147 802.33	0.517637	0.482363	51.36348	56.46091
52	TPSA(TOT) X0A MLOGP MV JHETV	73 485.35	147 802.33	0.497187	0.502813	50.33863	56.52447

Table7: Topological Indices - Regression of Training Set (MP)

Model No	Parameter Used	AI	B	SE	R	R2	R2A	F Ratio	Q=r/se
24	TPSA(TOT)	0.3436 (±0.4499)	183.3617	72.4033	0.1772	**0.0314**	-0.0224	0.5832	0.0024
32	TPSA(TOT) X0A	1.0381(±0.4676) -1 426.2718 (±534.2383)	1 191.0079	62.5372	0.5635	**0.3175**	0.2372	3.9546	0.0090
34	TPSA(TOT) X0A MLOGP	-0.1331(±0.4175) -2 434.7557 (±433.7103) -45.0077 (±10.1631)	2 077.2142	43.2082	0.8327	**0.6934**	0.6359	12.0600	0.0193
49	TPSA(TOT) X0A MLOGP MV	-0.1971(±0.4505) -2 197.4258 (±683.3182) -46.3021 (±10.8010) -46.3021 (±616.1624)	1 724.7748	44.3170	0.8352	**0.6976**	0.6169	8.6504	0.0188
52	TPSA(TOT) X0A MLOGP MV JHETV	-0.6018(±0.4913) -1 335.4615 (±829.7143) -47.5611 (±10.2496) -47.5611 (±660.4289) -1 335.4615 (±31.7660)	904.3283	41.9392	0.8644	0.7472	0.6570	8.2771	0.0206

Mono Parametric

MP = 0.3436(±0.4499) TPSATOT +183.3617,

N=20 , SE = (72.4033), R = (0.1772) , R2 = (0.0314) , R2A = (-0.0224), F RATIO = (0.5832), Q = (0.0024)

BI Parametric

MP = 1.0381(±0.4676) TPSATOT -1,426.2718 (±534.2383) X0A +1,191.0079,

N=20 , SE = (62.5372), R = (0.5635) , R2 = (0.3175) , R2A = (0.2372), F RATIO = (3.9546), Q = (0.0090)

Tri Parametric

MP = -0.1331(±0.4175) TPSATOT -2,434.7557 (±433.7103) X0A -45.0077 (±10.1631) MLOGP +2,077.2142,

N=20 , SE = (43.2082), R = (0.8327) , R2 = (0.6934) , R2A = (0.6359), F RATIO = (12.0600), Q = (0.0193)

Tetra Parametric

MP = -0.1971(±0.4505) TPSATOT -2,197.4258 (±683.3182) X0A -46.3021 (±10.8010) MLOGP -46.3021 (±616.1624) MV +1,724.7748,

N=20 , SE = (44.3170), R = (0.8352) , R2 = (0.6976) , R2A = (0.6169), F RATIO = (8.6504), Q = (0.0188)

Penta Parametric

MP = -0.6018(±0.4913) TPSATOT -1,335.4615 (±829.7143) X0A -47.5611 (±10.2496) MLOGP -47.5611 (±660.4289) MV -1,335.4615 (±31.7660) JHETV +904.3283, N=20 , SE = (41.9392), R = (0.8644) , R2 = (0.7472) , R2A = (0.6570), F RATIO = (8.2771), Q = (0.0206)

Table8 Topological Indices- Observed and Estimated of Training Set (MP)

Sn	Observed M.P	*Predicted M.P*	*Residuals*
1	187.5	170.6868	16.81322
2	295.5	267.2943	28.20569
3	122.5	187.2337	-64.7337
4	150.5	142.6122	7.887827
5	134	130.545	3.455003
6	219	282.3034	-63.3034
7	269	239.536	29.46404
8	229	203.8205	25.17954
9	237.5	268.0563	-30.5563
10	278	232.5978	45.40224
11	140	162.6907	-22.6907
12	159	159.619	-0.61905
13	183	176.521	6.47904
14	254.5	276.7864	-22.2864
15	199.5	244.9949	-45.4949
16	165.5	169.2448	-3.74483
17	316.5	237.1178	79.38222
18	50.5	51.76483	-1.26483
19	196.5	215.3786	-18.8786
20	330	298.6962	31.30383

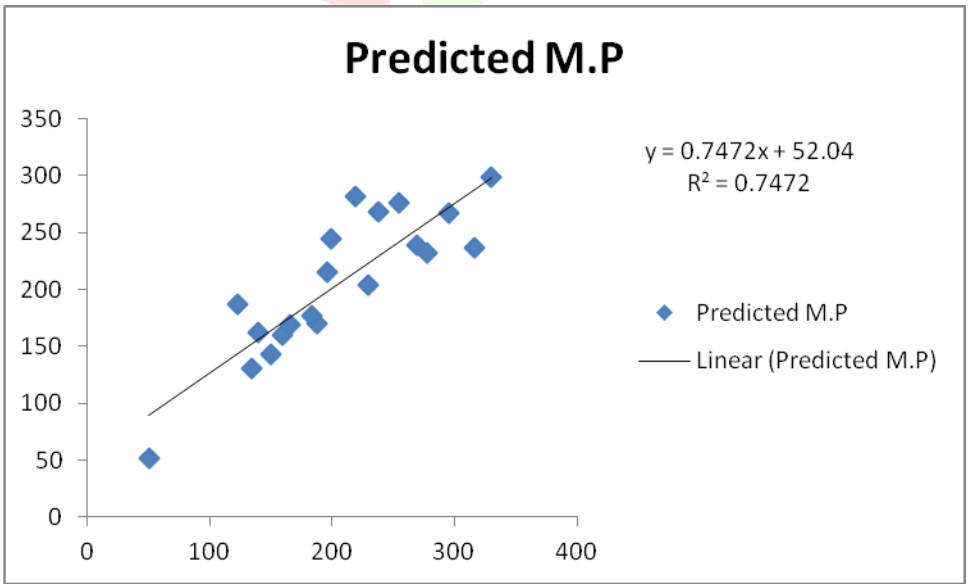


Table9: Topological Indices- Crossed validation of Training Set (MP)

Model No	Parameter Used	press	ssy	Press/SSY	R²CV	PSE	Spress
24	TPSA(TOT)	94 360.20	97 417.44	0.968617	0.031383	68.68777	72.40327
32	TPSA(TOT) X0A	66 485.39	97 417.44	0.682479	0.317521	57.65648	62.53723
34	TPSA(TOT) X0A MLOGP	29 871.13	97 417.44	0.30663	0.69337	38.64656	43.20817
49	TPSA(TOT) X0A MLOGP MV	29 459.95	97 417.44	0.302409	0.697591	38.37965	44.31701
52	TPSA(TOT) X0A MLOGP MV JHETV	24 624.55	97 417.44	0.252774	0.747226	35.08885	41.9392

Table10: Topological Indices- Regression of Test Set (MP)

Model No	Parameter Used	AI	B	SE	R	R2	R2A	F Ratio	Q=r/se
52	TPSA(TOT) X0A MLOGP MV JHETV	1.3144(±0.9648) -519.5762 (±442.6204) 12.1097 (±29.3399) 12.1097 (±394.2575) -519.5762 (±31.4719)	497.6368	21.0089	0.9649	0.9311	0.8162	8.1029	0.0459

MP = 1.3144(±0.9648) TPSATOT -519.5762 (±442.6204) X0A 12.1097 (±29.3399) MLOGP 12.1097 (±394.2575) MV -519.5762 (±31.4719) JHETV +497.6368, N=29 , SE = (21.0089), R = (0.9649) , R2 = (0.9311) , R2A = (0.8162), F RATIO = (8.1029), Q = (0.0459)

Table11: Topological Indices- Observed and Estimated of Test Set (MP)

Sn	Observed M.P	*Predicted M.P*	*Residuals*
1	63.5	64.04558	-0.54558
2	96	120.6303	-24.6303
3	114.5	104.9991	9.500865
4	235	234.5918	0.408173
5	112.5	120.53	-8.02995
6	127	114.5684	12.43157
7	146.5	158.4163	-11.9163
8	150	136.8469	13.15308
9	170	160.3715	9.62848

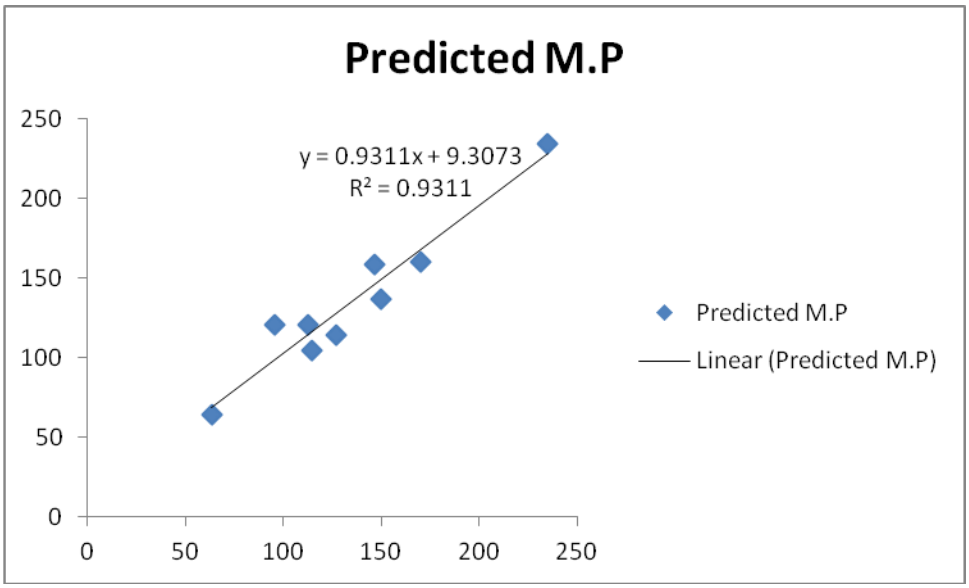


Table12: Topological Indices- Crossed validation of Test Set (MP)

Model No	Parameter Used	press	ssy	Press/SSY	R²CV	PSE	Spress
52	TPSA(TOT) X0A MLOGP MV JHETV	1 324.12	19 206.00	0.068943	0.931057	12.12947	13.75353

Conclusions

QSPR models were developed using QSPR multiple linear regression (MLR). Analyses Predictive linear QSPR models were developed for the relevant descriptors.

The results obtained offers excellent regression models that possesses good prediction ability.

References

1. Yalkowsky, S. H. & Valvani, S. C. Solubility and partitioning I: solubility of nonelectrolytes in water. Journal of pharmaceutical sciences 69, 912922 (1980).

2. Jain, N. & Yalkowsky, S. H. Estimation of the aqueous solubility I: application to organic nonelectrolytes. Journal of pharmaceutical sciences 90, 234252 (2001)

3. Sanghvi, T., Jain, N., Yang, G. & Yalkowsky, S. H. Estimation of Aqueous Solubility By The General Solubility Equation (GSE) The Easy Way. QSAR & Combinatorial Science 22, 258262. issn: 1611-0218 (2003).

4. Ran, Y. & Yalkowsky, S. H. Prediction of Drug Solubility by the General Solubility Equation (GSE). Journal of Chemical Information and Computer Sciences 41, 354357. issn: 0095-2338 (2001).

5. Ran, Y., Jain, N. & Yalkowsky, S. H. Prediction of Aqueous Solubility of Organic Compounds by the General Solubility Equation (GSE). Journal of Chemical Information and Computer Sciences 41, 12081217 (2001).