The Science of the Total Environment, 109/110 (1991) 59–68
Elsevier Science Publishers B.V., Amsterdam

The QSAR prediction of melting point, a property of environmental relevance

J.C. Dearden

School of Health Sciences, Liverpool Polytechnic, Byrom Street, Liverpool L3 3AF, United Kingdom

ABSTRACT

Melting point is an important environmental parameter, since it affects solubility. The melting point of a compound is controlled primarily by intermolecular forces and molecular symmetry. This study shows that it is possible to predict the melting points of a series of 42 anilines quite well ($r = 0.941$, $s = 24.6$) by an equation containing five parameters; namely, a measure of hydrogen bond donor ability, the hydrophobic substituent constant, molar refractivity, the Sterimol width parameter B_2, and an indicator variable for *m*-substitution.

INTRODUCTION

At first sight, the melting point of a compound may seem to have little to do with its toxicity, but closer study shows that this is not the case. In particular, melting point affects solubility, and solubility controls toxicity in that, if a compound is only poorly soluble, its concentrations in the aqueous environment may be too low for it to exert a toxic effect, and also because aqueous solubility is necessary for a compound to be transported to the active site within an organism.

The solubility of a compound can legitimately be regarded as a partitioning of the compound between its crystal lattice and the solvent. If the forces holding the molecule in the crystal are high, then the solubility will be low. For the same reason the melting point will be high, since melting point is a measure of the energy required to disrupt the crystal lattice. For example, uracil (m.p. > 300°C) shows no toxicity even in saturated solution; however, its 1,3-dimethyl derivative (m.p. 122°C) is appreciably toxic [1]. Similarly, anthracene (m.p. 217°C) is non-toxic, whilst its isomer phenanthrene (m.p. 97°C) is quite toxic [1].

Melting point also affects the toxicity of mixtures. If two or more high-melting compounds form a liquid when mixed (eutectic effect), the mixture is generally more toxic than the individual compounds [2], probably because solubility is increased.

Melting point is determined by the strength of a crystal lattice, which, in

0048-9697/91/$03.50

TABLE 1

Some examples of factors affecting melting point

Compound	Melting point (°C)
$SnCl_2$ (ionic)	246
$SnCl_4$ (covalent)	−33
Ethylbenzene (non-polar)	−95
Anisole (polar)	−33
3-Nitrophenol (H-bonded)	97
3-Nitroanisole (non-H-bonded)	39
2-Nitrophenol (intramolecularly H-bonded)	45
Naphthalene (two fused rings)	80
Anthracene (three fused rings)	217
1,3-Dibromobenzene (unsymmetrical)	−7
1,4-Dibromobenzene (symmetrical)	87
4-Methylaniline (rigid alkyl)	44
4-Ethylaniline (flexible alkyl)	−5

turn, is controlled primarily by three factors: intermolecular forces, molecular symmetry and the conformational degrees of freedom of a molecule [3]. The data in Table 1 illustrate these effects.

Most ionic compounds have very high melting points, because the ionic force holding the ions together are extremely strong. Dipole–dipole forces clearly increase melting point, as the examples of ethylbenzene and anisole demonstrate. Undoubtedly, however, for organic compounds, the most important intermolecular force controlling melting point is hydrogen bonding. By the same token, a compound that is intramolecularly hydrogen bonded, and which, therefore, cannot exert as great an intermolecular attraction, will have a lower melting point.

Melting point tends to increase with size, simply because the molecular surface area available for contact with other molecules increases. Symmetry plays a major role in crystal lattice strength, as the dibromobenzene examples illustrate. Finally, the effect of conformational flexibility is shown by the 4-methyl and 4-ethylaniline melting points. This effect is not, in this case, the packing effect whereby chains with even numbers of carbon atoms have high melting points than chains with one more or fewer carbon atoms, because both 4-n-propylaniline and 4-n-butylaniline are liquids at room temperature.

Because of the importance of melting point, it would be useful to be able to predict it. To date, very little has been published on the prediction of melting point from chemical structure. The melting points of alkanes have long been of interest, from the early work of Kopp [4]. Homologous series were studied by Skau [5], and Charton and Charton [6] utilised the QSAR

approach to correlate differences in melting points between sets of compounds in which the substituents were varied systematically.

Recently, we attempted [7] to predict melting points of a series of anilines using QSAR. In that study, hydrogen bond donor and acceptor ability was represented only by indicator variables, and symmetry was also modelled by indicator variables for 2-, 3- and 4-substitution. We obtained the following correlation:

$$T_m(\mathrm{K}) = 110H_D + 139F + 38.4^3\chi^v + 42.8H_A + 32.2I_4 + 52.4R - 29.9L + 289 \quad (1)$$

$$(n = 43, r = 0.94, s = 25.3)$$

where H_D is the indicator variable for hydrogen bond donor ability; F the Swain–Lupton field parameter; $^3\chi^v$ the third-order valence-corrected molecular connectivity; H_A the indicator variable for hydrogen bond acceptor ability; I_4 the indicator variable for 4-substitution; R the Swain–Lupton resonance parameter; and L the Sterimol substituent length parameter.

We commented that, if continuous scales of hydrogen bonding ability and better measures of shape and symmetry could be obtained, it should be possible to improve the correlation. We have now obtained such parameters and report here on the application of these and other appropriate parameters to the prediction of the melting points of the same series of anilines.

Melting point is a highly accurate end-point, but this does not mean that its prediction is facile. One of the tenets of QSAR is that all compounds examined should interact with the same receptor. In the case of melting point, the "receptor" is different for each compound, in that the crystal lattice into which each molecule fits is different, comprising, as it does, a cage of different molecules for each compound. Nonetheless, our previous study showed that reasonably good predictions could be made, and this encouraged us to investigate further.

METHOD

Melting points of aniline and 41 derivatives were obtained from a variety of literature sources. Parameters such as π, σ, F, R and MR were obtained from the compilation of Hansch and Leo [8]. Characteristic volumes V_X were calculated as described by Abraham and McGowan [9]. Hydrogen bond donor (α) and acceptor (β) ability values were supplied by Abraham [10]. Molecular connectivities and kappa shape indices were calculated using the MOLCONN-2 program. Sterimol L and B_1–B_4 parameters were supplied by Blaney [11]. These were calculated by taking the L axis in each case to be the bond joining the amino nitrogen atom to the aromatic ring.

TABLE 2

Best sub-sets regression of melting point (K) of 42 anilines

Equation	r	s	F
(2) $T_m = 218 + 277\alpha$	0.819	39.5	81.4
(3) $T_m = 207 + 263\alpha + 69.7F$	0.853	36.3	52.1
(4) $T_m = 145 + 213\alpha - 36.2\pi + 0.799MW$	0.887	32.6	46.5
(5) $T_m = 323 + 180\alpha - 36.1\pi + 8.40MR - 61.2B_2$	0.924	27.5	53.5
(6) $T_m = 329 + 182\alpha - 38.2\pi + 8.91MR - 62.2B_2 - 26.6I_3$	0.941	24.6	55.6
(7) $T_m = 437 + 169\alpha - 40.4\pi + 10.6MR - 63.6B_2 + 47.1I_4 - 19.4L$	0.947	23.7	50.5

Hydrogen bonding indicator variables were given values corresponding approximately to the number of donor or acceptor sites available on the substituent. For example, the hydrogen bond donor value of 2-carboxyl aniline was set at 0.5 because of the likelihood of intramolecular hydrogen bonding. In a similar way, the hydrogen bond acceptor value was set at 1.5 (+2 for the carboxyl oxygens and −0.5 for the amino nitrogen because of intramolecular hydrogen bonding). Because of the high melting point of 2-hydroxyaniline, it was assumed that there was no intramolecular hydrogen bonding in this compound.

Conformational flexibility was accounted for by allocating indicator variable values of 0, 1 or 2 depending on the degrees of freedom of the substituent. For example, Me = 0, Et = 1, OEt = 2.

Best sub-sets regression was carried out using the commercially available statistical software package MINITAB. Best sub-sets regression finds the combination of n parameters that will best describe the variation of biological response in a set of compounds. It differs from step-wise regression in that the latter, having selected the best single parameter, then finds a second parameter which, *together with the first parameter*, gives the best correlation, and so on. This procedure has the weakness that the best two parameters, for example, are not necessarily the best single one plus one other which increases the correlation coefficient of the first by the greatest amount.

RESULTS AND DISCUSSION

Regression analysis, using the best sub-sets algorithm, showed that it was possible to model melting point in this series of anilines with relatively few parameters. Table 2 gives the correlations found with up to six parameters. It will be seen that T_m, rather than log T_m, is used as the dependent variable. It is, of course, customary to use logarithmic terms in QSAR, but in this case log T_m gave much poorer correlation than did T_m. Because of this and because the

range of T_m of organic compounds is relatively small, it is believed that the use of T_m is justified. The results in Table 2 clearly show that the correlation involving five parameters is the most appropriate, for it gives a good quantitative description of melting point. The standard error of 24.6° is rather high, but is acceptable considering the range of melting points covered (> 250°). Table 3 gives the correlation matrix for the parameters used in Eqns (2)–(7) and it can be seen that there is no pronounced collinearity amongst the parameters used in Eqn (6). The highest collinearity is between two steric terms (MR and B_2), as might be expected, but even here the r^2 value of 0.56 is sufficiently low not to cause concern.

Table 4 gives the melting points predicted by Eqn (6), together with the experimental values, and Fig. 1 shows the relationship between predicted and observed values. By far the most poorly predicted melting point is that of 4-methylaniline (ΔT_m 47.8°). Removing this compound from the data set improves the correlation slightly, but, as there is no obvious reason for this compound being an outlier, it has been kept in.

Equation (6) is an improvement over Eqn (1), in that only five parameters are required to give about the same correlation coefficient and a slightly improved standard error.

The most important single parameter is α, a measure of hydrogen bond donor ability. It is remarkable, particularly when numerous assumptions had to be made regarding additivity and the effects of intramolecular hydrogen bonding and steric factors, that this term alone accounts for 67% of the variation of melting point. One might perhaps have expected β, a measure of hydrogen bond acceptor ability, to play an important role also. That it does not is probably a consequence of the fact that the presence of a hydrogen bond donor automatically means that a hydrogen bond acceptor is present also, whereas the converse is not true.

The second most important parameter is the hydrophobic substituent constant π; the negative sign on this term indicates that it is probably modelling polarity. Although log P (and hence π) contains polarity, hydrogen bonding and bulk bonding and bulk components [12], the last two factors have positive coefficients in Eqn (6).

Molar refractivity is essentially a measure of bulk volume [13] and, as expected, melting point in this series of anilines increases with MR. The Sterimol width term, B_2, is a shape parameter; its negative coefficient indicates that it is probably modelling asymmetry here, since lack of molecular symmetry lowers melting point. In a similar fashion the indicator variable I_3 (for 3-substitution) also has a negative coefficient and again probably models asymmetry. In this connection it may be noted that I_4 (for 4-substitution) is included in Eqn (7), its positive coeffcient indicating that it models symmetry.

TABLE 3

Correlation matrix of parameters used in Eqns (2)–(7)

	T_m	α	π	MR	B_2	I_3	I_4	MW	L
α	0·819								
π	−0.593	−0.564							
MR	−0.218	−0.332	0.697						
B_2	−0.551	−0.479	0.582	0.747					
I_3	−0.081	0.071	−0.041	0.073	0.015				
I_4	0.161	−0.011	0.116	0.167	0.124	−0.552			
MW	0.009	−0.172	0.571	0.685	0.288	0.083	0.018		
L	−0.025	−0.185	0.405	0.684	0.488	−0.211	0.697	0.300	
F	0.375	0.170	−0.246	−0.321	−0.655	0.071	−0.116	0.239	−0.289

TABLE 4

Observed melting points of substituted anilines, those predicted from Eqn (6) and parameters used in Eqn (6)

Substituent	T_m (K)		ΔT_m	α	π	MR	B_2	I_3
	Observed	Predicted						
H	266.7	290.2	−23.5	0.26	0.00	1.03	1.533	0
2-F	244.5	281.0	−36.5	0.30	0.41	0.92	1.530	0
2-Cl	271.1	300.5	−29.4	0.30	0.71	6.03	1.764	0
2-Br	305	303.8	1.2	0.30	0.86	8.88	2.027	0
2-I	333.5	316.0	17.5	0.31	1.12	13.94	2.425	0
2-Me	249.3	258.3	−9.0	0.23	0.56	5.65	2.276	0
2-Et	230	222.8	7.2	0.23	1.02	10.30	3.230	0
2-OMe	279.2	292.7	13.5	0.23	−0.02	7.87	2.397	0
2-OH	447	441.6	5.4	0.86	−0.67	2.85	1.526	0
2-NH_2	375.5	373.3	2.0	0.39	−1.23	5.42	1.958	0
2-COOH	419.5	438.7	−19.2	0.72	−0.32	6.93	1.532	0
2-NO_2	344.5	342.3	2.2	0.19	−0.28	7.36	1.569	0
3-Cl	262.7	280.1	−17.4	0.33	0.71	6.03	1.752	1
3-Br	291.5	282.9	8.6	0.33	0.86	8.88	2.023	1
3-I	306	290.2	15.8	0.33	1.12	13.94	2.471	1
3-Me	242.6	232.8	9.8	0.23	0.56	5.65	2.258	1
3-Et	209	197.4	11.6	0.23	1.02	10.30	3.211	1
3-$CH_2C_6H_5$	312	297.1	14.9	0.24	1.96	25.36	3.217	1
3-OMe	272	270.4	1.6	0.25	−0.02	7.87	2.387	1
3-OH	395	414.7	−19.7	0.86	−0.67	2.85	1.531	1
3-NH_2	336.5	369.4	−32.9	0.52	−1.23	5.42	1.976	1
3-CN	326.5	354.8	−28.3	0.38	−0.57	6.33	1.527	1
3-COOH	447	425.6	21.4	0.80	−0.32	6.93	1.549	1
3-COOMe	312	328.1	−16.1	0.30	−0.01	12.87	2.315	1
3-NO_2	387	356.3	30.7	0.40	−0.28	7.36	1.531	1
4-F	272.2	277.0	−4.8	0.28	0.41	0.92	1.535	0
4-Cl	345.5	305.7	39.8	0.30	0.71	6.03	1.680	0
4-Br	339.4	317.7	21.7	0.31	0.86	8.88	1.833	0
4-I	340.5	337.4	3.1	0.31	1.12	13.94	2.081	0
4-Me	316.7	268.9	47.8	0.23	0.56	5.65	2.106	0
4-Et	268.1	295.5	−27.4	0.23	1.02	10.30	2.061	0
4-iPr	210	255.2	−45.2	0.23	1.53	14.96	3.064	0
4-tBu	290	276.1	13.9	0.23	1.98	19.62	3.118	0
4-$CH_2C_6H_5$	307.5	345.6	−38.1	0.34	1.96	25.36	3.158	0
4-OMe	330.2	296.1	34.1	0.23	−0.02	7.87	2.342	0
4-OEt	275.4	266.7	8.7	0.23	0.38	12.47	3.228	0
4-OH	457	441.4	15.6	0.86	−0.67	2.85	1.529	0
4-NH_2	419	392.4	26.6	0.52	−1.23	5.42	2.034	0
4-NHMe	309	320.3	−11.3	0.43	−0.47	10.33	3.168	0
4-CN	359	383.3	−24.3	0.40	−0.57	6.33	1.555	0
4-COOH	461.5	460.5	1.0	0.85	−0.32	6.93	1.562	0
4-NO_2	421.5	387.0	34.5	0.42	−0.28	7.36	1.523	0

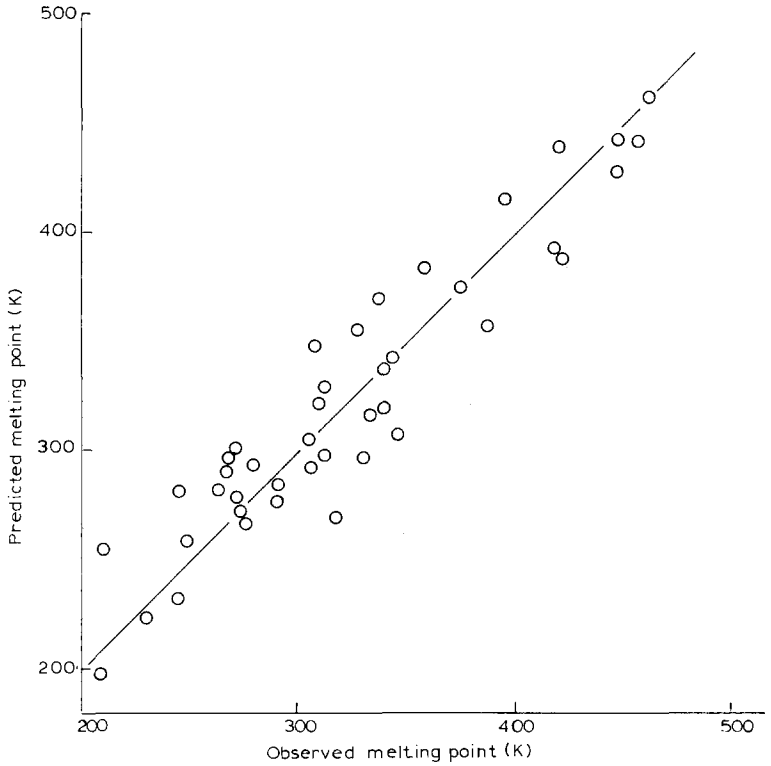

Fig. 1. Predicted versus observed melting points of anilines.

One of the tests of a predictive equation like Eqn (6) is how well it predicts values of compounds not included in the training set. Five compounds (the same as those used earlier by Dearden and Rahman [7]) were removed from the training set, and the best sub-sets equation using five parameters were re-calculated as:

$$T_m = 331 + 181\alpha - 38.7\pi + 8.62MR - 62.1B_2 - 27.4I_3 \quad (8)$$

$$(n = 37, r = 0.931, s = 26.2, F = 40.6)$$

Table 5 gives the melting points, calculated from Eqn (8), of the five compounds deleted from the training set. As would be expected, the melting points predicted by Eqn (8) are slightly less accurate than those predicted by Eqn (6), but they indicate that the correlations obtained here are robust and can be used for reasonably good predictions of the melting points of substituted anilines.

Further refinement of the parameters will almost certainly improve the predictive ability of the correlations and work is continuing on this.

TABLE 5

Melting points of the five components omitted from Eqn (8), and predicted by that equation

Substituent	T_m(K)		ΔT_m
	Observed	Predicted	
2-Br	305	302.7	2.3
2-Et	230	221.4	8.6
3-$CH_2C_6H_5$	312	290.0	22.0
4-I	340.5	334.7	5.8
4-COOH	461.5	460.0	1.5

ACKNOWLEDGEMENTS

I am grateful to Dr M.H. Abraham of University College London for supplying α and β values, and to Dr F.E. Blaney of SmithKline Beecham, Harlow, for supplying Sterimol parameters for the compounds used in this work.

REFERENCES

1 R.L. Lipnick, Narcosis: fundamental and baseline toxicity mechanism for nonelectrolyte organic chemicals, in W. Karcher and J. Devillers (Eds), Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Kluwer Academic Publishers, Dordrecht, 1990, pp. 281–293.
2 D. Mackay, University of Toronto, personal communication, 1990.
3 R. Abramowitz and S.H. Yalkowsky, Melting point, boiling point and symmetry. Pharm. Res., 7 (1990) 942–947.
4 H. Kopp, Über die Vorausbestimmung einiger physikalischen Eigenschaften bei mehreren Reihen organischer Verbingungen. Annu. Rev. Chem. Pharm., 41 (1842) 79–89.
5 E.L. Skau, Melting and freezing temperatures, in H.F. Mark, J.J. McKetta and D.F. Othmer (Eds), Kirk–Othmer Encyclopedia of Chemical Technology, Vol. 13, 2nd edn, Interscience, New York, 1967, pp. 198–217.
6 M. Charton and B.I. Charton, The estimation of melting points, in M. Kuchar (Ed.), QSAR in Design of Bioactive Compounds, J.R. Prous, Barcelona, 1984, pp. 41–51.
7 J.C. Dearden and M.H. Rahman, QSAR approach to the prediction of melting points of substituted anilines. Math. Comput. Model., 11 (1988) 843–846.
8 C. Hansch and A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979, p. 49.
9 M.H. Abraham and J.C. McGowan, The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. Chromatographia, 23 (1987) 243–246.
10 M.H. Abraham, University College London, personal communication, 1990.
11 F.E. Blaney, SmithKline Beecham, personal communication, 1990.

12 R.W. Taft, M.H. Abraham, G.R. Famini, R.M. Doherty, J.-L.M. Abboud, and M.J. Kamlet, Solubility properties in polymers and biological media 5: an analysis of the physico-chemical properties which influence octanol–water partitition coefficients of aliphatic and aromatic solutes. J. Pharm. Sci., 74 (1985) 807–814.

13 J.C. Dearden, S.J.A. Bradburne and M.H. Abraham, The nature of molar refractivity, in C. Silipo and A. Vittoria (Eds), QSAR: Rational Approaches to the Design of Bioactive Compounds, Elsevier, Amsterdam, 1991, pp. 143–150.