Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach

Alan R. Katritzky,*,[⊥] Uko Maran,[⊥,‡] Mati Karelson,[‡] and Victor S. Lobanov[§]

Center for Heterocyclic Compounds, University of Florida, Gainesville, Florida 32611-7200, Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu EE2400, Estonia, and 3-Dimensional Pharmaceuticals, Inc., 665 Stockton Drive, Suite 104, Exton, Pennsylvania 19341

Quantitative structure−property relationships on a large set of descriptors are developed for the melting points of a large set of mono- and disubstituted benzenes (443 compounds). A correlation equation including nine descriptors ($R^2 = 0.8373$) is reported for the whole set of compounds, and six descriptor equations are given for the subsets of *ortho*-, *meta*-, and *para*-substituted compounds, respectively. The importance of the hydrogen bonding descriptor ($HDSA_2$) is demonstrated, and quantum chemical descriptors are successfully applied to obtain predictive models.

INTRODUCTION

Melting point is a fundamental physical property specifying the transition temperature between the solid and liquid phases. Besides its direct utility as an indicator to whether a compound is solid or liquid under normal conditions, melting point has numerous applications in biochemical and environmental sciences due to its relationship with solubility.[1−3] Yalkowsky and Valvini first developed a quantitative relationship to estimate solubility of a compound from the melting point, partition coefficient, and entropy of fusion.[1] Solubility, in turn, is very important in drug design[4] and in an assessment of the effective toxicity of chemicals and materials.[3] For example, melting point has been successfully used as a descriptor in correlations with the aqueous toxicity of the chlorophenols,[5,6] with the release coefficient of benzene and pyridine from polysiloxane matrices,[3] and with the skin corrosivity of the organic acids, bases, and phenols.[7]

However, despite the vast amount of melting point data available and knowledge about the melting phase transition, no general relationship yet relates the melting points of compounds with their chemical structure. Attempts to construct quantitative structure−property relationships (QSPR) to predict melting point from the chemical structure have been mostly confined to limited sets of hydrocarbons and substituted aromatics (see Table 1). To the best of our knowledge the most successful predictions of melting point have been achieved for normal alkanes ($s = 0.51$ K) using topological indices such as the carbon number, the Wiener index, and the Balaban distance sum connectivity index.[8] However, for the combined set of normal and branched alkanes, topological indices have already been shown not to deliver a satisfactory model for predicting melting point ($s = 23.8$ K) in contrast to their success in correlating seven other physical properties.[9] In another study of 303 branched and normal substituted alkanes, the combination of 11 intermolecular forces provided better predictions ($s = 16.4$ K).[10] A comparative study on prediction of physical properties of aldehydes, amines, and ketones using different classes of molecular descriptors produced a moderate correlation with the melting point (Table 1).[11]

Table 1. The Best Previous QSPR Predictions of Melting Points

ref	group of compds	no. (D)	R^2	s	F	n
8	normal alkanes	1	0.9980	0.51	16.7	24
9	normal and branched alkanes	5	0.5700	23.8	13	56
10	branched and unbranched substituted alkanes	11	0.9312	16.4	357.8	303
2	anilines	5	0.8855	24.6	55.6	42
12	pyridines	6	0.8570	36.1	133.6	141
11	aldehydes	5	0.8380	*a*	*a*	27
11	amines	5	0.7950	*a*	*a*	48
11	ketones	5	0.8650	*a*	*a*	30
4	non-hydrogen-bonded benzenes	5	0.8798	22.8	*a*	85

a Data not available from original article.

Correlations of the melting points of aromatic compounds have so far achieved only moderate success. For a set of 42 anilines, Dearden developed a five-parameter correlation equation ($s = 24.6$ K) based on the measure of hydrogen bond donor ability, the hydrophobic substituent constant, the molar refractivity, the Sterimol width parameter B_2, and the indicator variable of *meta*-substitution.[2] Abramowitz and Yalkowsky developed a four-parameter correlation equation ($s = 22.8$ K) with the melting points of 85 non-hydrogen-bonded rigid benzenes.[4] Their equation employed the boiling point of the compound, the logarithm of the symmetry number, the cube of the eccentricity of the molecule, and the number of *ortho* substituents.[4] Our group reported a moderate correlation ($s = 36.1$ K) for melting points of substituted pyridines.[12]

The group contribution approach has also been employed in predicting melting points. Yalkowsky *et al.* developed predictive equations based on group contributions for a set of 1186 rigid aromatic compounds ($s = 37.7$ K)[13] and for a set of 596 aliphatic, non-hydrogen bonded compounds ($s = 34.3$ K).[14] Recently a new group contribution method for estimating the physical properties of pure compounds has been proposed based on the inclusion of second-order group contributions.[15] This method has been used to develop an equation for predicting normal melting points with a standard error of 18.3 K. However, the authors noted that group contribution methods generally have difficulties in providing reliable estimates of melting points, because they depend heavily on nonadditive structural features such as intermolecular interactions and molecular symmetry.[15]

[⊥] University of Florida.
[‡] University of Tartu.
[§] 3-Dimensional Pharmaceuticals.

In the present study we employed our recently reported hydrogen bonding descriptor[16] and other quantum-chemical descriptors[17] to develop correlation equations for predicting melting points of substituted benzenes.

THEORETICAL BACKGROUND

Melting is the process of transforming a crystal or an amorphous solid into a liquid. At the melting temperature, the free energy of the phase transition is equal to zero, and thus the thermodynamic equation for the melting point is given by eq 1. The enthalpy of the transition ΔH_m depends on the intermolecular interactions which are primarily nonspecific dispersion interactions and specific interactions such as hydrogen bonding. Such interactions in the crystalline phase are intrinsically the same as those in the liquid or in the gas phase but are modified because of the shorter separation distances between molecules in the solid phase and crystal structure geometric constraints.[18]

$$T_m = \frac{\Delta H_m}{\Delta S_m} \qquad (1)$$

The entropy of melting ΔS_m is primarily related to an increase in rotational and conformational degrees of freedom of a molecule in the liquid phase.[4,14] The rotational freedom of a molecule is associated with its rotational symmetry, which also reflects a possible higher packing order of the molecules in a crystal. The increase in a conformational freedom is connected with the higher number of possible conformations a molecule can accept in the liquid phase *versus* the limited number of choices in a crystal. Recently, Yalkowsky *et al.* showed that entropy of melting can be estimated from the rotational symmetry number and the molecular flexibility number.[14,19]

However, there are additional complications. Despite a common belief that molecular crystals are made of frozen chemical entities which can at most undergo very small motions, motions in crystal can be relatively large depending on the crystal's size and shape and the temperature.[20] Moreover, these motions can change the structure of a crystal and affect its melting point.[21] The motions in crystals can be divided into oscillations, reorientations, and phase transitions. Oscillations of atoms occur in all crystals and are highly temperature dependent. The reorientations include rotations of the substituents and rotations in the molecular plane (for instance, that of the benzene ring).[21] When the reorientation motions become relaxed, the rigid crystals transform into the plastic crystals which exhibit preliquid behavior.[22] When part of the translational symmetry of a crystal is lost and the rotation motions become relatively fast, the crystal reaches the liquid−crystalline state. It is estimated that 5−8% of all organic compounds are transformed into a liquid crystal state when close to their melting points.[22] Furthermore, phase transitions are related to polymorphism, which means that molecules are able to form different crystal forms of different structure and heat capacity.[21,23] Apart from the factors mentioned above, measurements of melting points are affected by the purity of a compound and the experimental error.

DATA AND COMPUTATIONAL METHODS

The majority of the melting points of the mono- and disubstituted benzenes used in the present work were

Table 2. Monosubstituted Benzenes and Respective Property Values Used in the Data Set

	compd	mp^a (°C)
1	toluene	−95
2	ethylbenzene	−100
3	*n*-propylbenzene	−92
4	isopropylbenzene	−96
5	*tert*-butylbenzene	−58
6	phenylbenzene	69
7	vinylbenzene	−31
8	benzene	6
9	fluorobenzene	−40
10	chlorobenzene	−45
11	bromobenzene	−31
12	iodobenzene^a	−31.27
13	phenol	43
14	benzaldehyde	−26
15	benzoic acid	122
16	methoxycarbonylbenzene	−13
17	methoxybenzene	−38
18	ethoxybenzene	−33
19	nitrobenzene	5
20	benzyl alcohol	−15
21	ethyl benzoate^a	−34.6
22	methyl cinnamate	36
23	acetophenone	20
24	aniline	−6
25	*N*-methylaniline	−57
26	cyanobenzene^a	−13.00
27	acetanilide	114
28	acetophenone oxime	60
29	benzamide	129

^{a} Property values from CRC Handbook.[24]

provided by Professor A. Gavezzotti (University of Milan) with some (for 31 compounds, see Tables 2 and 3) data being added from the CRC Handbook[24] (Tables 2 and 3). The resulting data set consisted of 443 compounds including benzene, 29 monosubstituted benzenes, and 414 disubstituted benzenes. Disubstituted benzenes were represented by 131 *ortho*-, 128 *meta*-, and 155 *para*-substituted compounds. The structures of the compounds were drawn and optimized using the MMX molecular mechanics force field implemented in the PCMODEL program.[25] The final molecular geometries were refined using the quantum chemical program package MOPAC 6.0.[26] The AM1 parametrization[27] with eigenvector following geometry optimization procedure[28] at a precision level 0.01 kcal/Å gradient norm was used to calculate electronic and thermodynamic parameters for the isolated molecules.

The CODESSA program[29,30] was used to calculate 639 different molecular descriptors based on the output of MOPAC calculations. These include various constitutional, geometrical,[31] topological,[32−36] electrostatic,[37−40] quantum chemical,[17] and thermodynamic descriptors[18,41,42] (see Table 4). The heuristic multilinear regression procedures available in the framework of the CODESSA program were used to find the best correlation models. These procedures provide colinearity control (i.e., any two descriptors intercorrelated above 0.8 are never involved in the same correlation model) and implement heuristic algorithms for the rapid selection of the best correlation model without testing all possible combinations of the available descriptors. The detailed description of the procedures is available elsewhere.[16,29]

RESULTS AND DISCUSSION

For the set of all 443 substituted benzenes the best correlation model provided a correlation coefficient with R^2

Table 3. Disubstituted Benzenes and Respective Property Values Used in the Data Set

compd[a]	mp (°C) ortho	mp (°C) meta	mp (°C) para	compd[a]	mp (°C) ortho	mp (°C) meta	mp (°C) para
xylene	−25	−47	13	nitrobenzyl iodide	75	86	127
ethyltoluene	−81	−96	−62	nitrobenzoyl chloride	20	35	75
cymene	−71.5[b]	−63.7[b]	−74	nitrophenyl acetate	41	56	83
fluorotoluene	−62[b]	−87.7[b]	−56.8[b]	methoxyphenyl acetate	10	39	54
chlorotoluene	−38	−48	7	ethoxyphenyl acetate	2	34	60
bromotoluene	−28	−40	29	dinitrobenzene	118	89	174
nitrotoluene	−4	16	54	nitrobenzyl alcohol	74	27	93
methylacetophenone	c	c	28	ethyl nitrobenzoate	30	47	57
diethylbenzene	−31.2[b]	−83.89[b]	−42.85[b]	methyl nitrocinnamate	73	124	161
chloroethylbenzene	−82.7[b]	c	−62.6[b]	ethyl nitrocinnamate	44	79	142
diethynylbenzene	c	3	95	nitroacetophenone	28	81	81
bromoethynylbenzene	c	c	65	nitrophenylacetic acid	143	120	153
difluorobenzene	−34[b]	82	−13[b]	nitrobenzyl cyanide	84	62	117
dichlorobenzene	−18	−24	53	methylbenzyl alcohol	36	c	60
bromochlorobenzene	−13	−21	67	chlorobenzyl alcohol	74	c	75
chlorobenzyl bromide	c	22[b]	51	bromobenzyl alcohol	80	c	77
bromoisopropylbenzene	c	c	−23	iodobenzyl alcohol	90	c	72
dibromobenzene	7	−7	89	hydroxybenzyl alcohol	87	73	125
bromoiodobenzene	21	9	92	aminobenzyl alcohol	82	97	65
bromobenzyl bromide	31	41	63	ethyl hydroxybenzoate	c	73	116
iodotoluene	c	−27.2[b]	34	ethyl methoxybenzoate	c	c	7
diiodobenzene	27	40	129	ethyl aminobenzoate	13	c	92
iodobenzyl bromide	57	51	80	methylacetophenone	c	c	28
iodophenyl acetate	c	38	32	chloroacetophenone	c	c	20
iodoacetophenone	c	c	85	bromoacetophenone	c	8	51
cresol	30	12	34	hydroxyacetophenone	4[b]	96	109
ethylphenol	c	−4[b]	47	methoxyacetophenone	c	95[b]	39
n-propylphenol	c	26[b]	21	diacetylbenzene	40	32	113
chlorophenol	9	33	43	aminoacetophenone	20	99	106
bromophenol	5	33	64	phenylenediacetic acid	150	170	244
iodophenol	43	40	94	toluidine	−21	−44	44
dihydroxybenzene	105	111	172	chloroaniline	1	−10	71
hydroxybenzaldehyde	−7[b]	108	116	bromoaniline	32	18	66
hydroxybenzoic acid	159	200	214	iodoaniline	58	25	63
methoxyphenol	32	c	56	hydroxyaniline	174	123	186
nitrophenol	45	97	114	methoxyaniline	6	−12	57
chlorobenzaldehyde	11	18	47	ethoxyaniline	c	c	4
bromobenzaldehyde	22	c	67	nitroaniline	71	114	148
iodobenzaldehyde	37	57	78	diaminobenzene	102	64	141
methoxybenzaldehyde	38	c	2	N-methylphenylenediamine	22	c	36
nitrobenzaldehyde	44	58	106	cyanoaniline	50	53	86
aminobenzaldehyde	40	30[b]	72	N-methylhydroxyaniline	97	c	87
cyanobenzaldehyde	109	78	101	N-methylnitroaniline	37	68	152
methylbenzoic acid	105	112	178	tolunitrile	−14[b]	−23[b]	29
ethylbenzoic acid	68	47	114	chlorobenzonitrile	43	41	96
n-propylbenzoic acid	58	43	141	bromobenzonitrile	53	38	113
isopropylbenzoic acid	64	51	118	α-bromotolunitrile	73	93	116
fluorobenzoic acid	127	124	185	cyanobenzoic acid	190	217	219
chlorobenzoic acid	141	158	240	methyl cyanobenzoate	51	65	62
bromobenzoic acid	150	155	252	nitrobenzonitrile	110	118	149
iodobenzoic acid	162	187	270	ethyl cyanobenzoate	70	56	54
bromomethylbenzoic acid	150	156	229	acetotoluidide	112	66	154
benzenedicarboxylic acid	234	348	c	chloroacetanilide	88	79	179
acetoxybenzoic acid	135	131	187	bromoacetanilide	99	88	167
methoxybenzoic acid	101	110	184	iodoacetanilide	110	119	184
ethoxybenzoic acid	19	137	198	hydroxyacetanilide	209	149	168
nitrobenzoic acid	147	141	239	acetamidobenzoic acid	185[b]	250	251
aminobenzoic acid	145	174	188	methoxyacetanilide	88	80	130
(methylamino)benzoic acid	179	129	168	ethoxyacetanilide	79	96	135
methyl toluate	c	c	34	nitroacetanilide	94	155	216
methyl chlorobenzoate	c	21[b]	44	aminoacetanilide	132	88	163
methyl bromobenzoate	c	32	81	N,N′-diacetylphenylenediamide	186	191	304
methyl iodobenzoate	c	54	114	tolualdehyde oxime	49	60	80
methyl hydroxybenzoate	−8	70	131	iodobenzaldehyde oxime	108	62	c
methyl methoxybenzoate	c	c	49	hydroxybenzaldehyde oxime	63	90	72
methyl nitrobenzoate	−13	79	96	methoxybenzaldoxime	92	40	132
methyl aminobenzoate	24	38	112	nitrobenzaldoxime	103	122	133
chloroanisole	−26.8[b]	c	c	aminobenzaldehyde oxime	135	c	124
bromoanisole	2.5[b]	c	13[b]	toluamide	140	95	159
iodoanisole	c	c	52	chlorobenzamide	141	134	179
chlorophenetole	c	c	21	bromobenzamide	155	155	189
bromophenetole	c	c	4	iodobenzamide	184	186	218
iodophenetole	c	c	29	hydroxybenzamide	139	167	162
nitrophenentole	2.1[b]	34	60	methoxybenzamide	129	c	163
nitrochlorobenzene	33	46	83	ethoxybenzamide	132	139	202
nitrobromobenzene	42	56	127	nitrobenzamide	175	142	200
nitroiodobenzene	54	38	174	aminobenzamide	c	111	114
nitrobenzyl chloride	49	46	71	tolylurea	191	142	183
nitrobenzyl bromide	47	59	100	bromophenyl urea	202	165	227

[a] Compounds in the table are grouped in the following order: toluenes, ethynylbenzenes, *chloro-*, *bromo-*, *iodo*benzenes, phenols, benzaldehydes, benzoic acids, methyl benzoates, anisoles, phenentoles, nitrobenzenes, benzyl alcohols, ethyl benzoates, acetophenones, anilines, nitriles, acetanilides, oximes, amides, and ureas. [b] Property values from CRC Handbook.[24] [c] Property values are not available.

Table 4. Calculated Descriptors

subset of the descriptors	no. of descriptors
constitutional	38
electrostatic	82
geometrical	12
topological	37
quantum chemical	433
thermodynamical	37
total	639
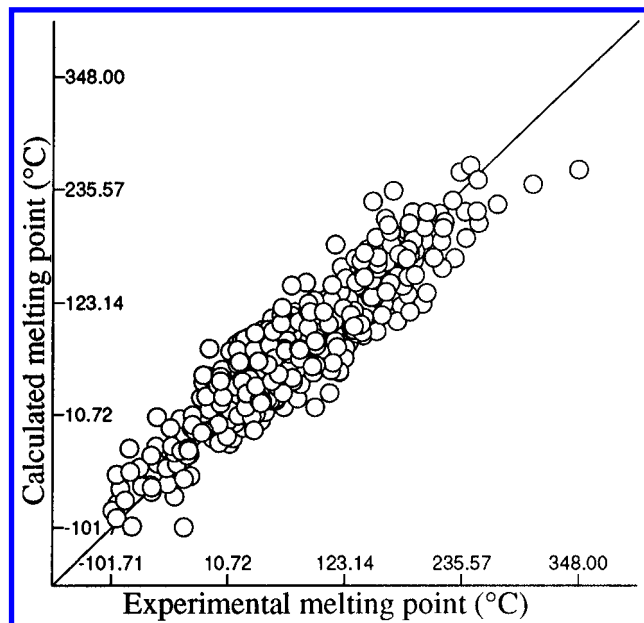
Figure 1. Nine-parameter correlation of the calculated *vs* experimental melting points for the full set of compounds.

$= 0.8373$ (Figure 1). The nine descriptors involved in the model are listed in Table 5 and include quantum-chemical, electrostatic, and topological descriptors. According to the *t*-test the most significant descriptor in the model is our recently reported hydrogen bonding descriptor HDSA$_2$ (*Hydrogen Donor Charged Surface Area*).[16] This descriptor was found to be most successful in quantifying hydrogen bonding for the prediction of the boiling point,[16] remarkably the same descriptor is also highly significant for the melting point. We calculated two variations of the HDSA$_2$ descriptor based on the Mulliken charges from MOPAC and based on the empirical charges derived from Zefirov's electronegativity equalization scheme.[37] For the whole set of 443 substituted benzenes Zefirov's charges delivered a slightly better correlation model than Mulliken charges ($R^2 = 0.8373$ *vs* $R^2 = 0.8142$).

The second most significant descriptor in the correlation model is the *Average Valency of a Hydrogen Atom*. On its own, this descriptor exhibited the highest single parameter equation correlation coefficient for melting point: $R^2 = 0.4551$. The descriptor is calculated as an average value of the quantum chemical valencies of all the hydrogen atoms in a molecule. The highest values of the quantum chemical valencies of the hydrogen atoms are usually associated with the aliphatic hydrogens. Thus, the negative regression coefficient for this descriptor reflects the fact that benzenes with aliphatic substituents have lower melting points than those with non-aliphatic substituents.

A further four descriptors: the *Total Molecular Surface Area* (TMSA), the *Average Information Content (second order)*,[32] the *Average Structural Information Content (first order)*,[32] and the *Symmetry Number*[20] are clearly related to the molecular shape, size, and symmetry and can be attributed to the effects of molecular packing in a crystal. The remaining descriptors (Table 5) are most likely connected to the influence of other intermolecular interactions, such as charge-transfer and dipole−dipole interaction in a solid phase.

We also investigated the correlations of the melting point for each of the individual subsets of *ortho-, meta-, and para-*substituted benzenes. The monosubstituted benzenes were included in all three subsets considering hydrogen atom as a second substituent. The application of the nine-parameter correlation model (Table 5), derived for the whole set, to

Table 5. Nine-Parameter Correlation Equation for the Full Set of the Descriptors and Compounds[a]

	$X \pm \Delta X$	*t*-test	$R^{2\ b}$	F^c	name of the descriptor
0	6738.7 ± 460.5	14.6348			intercept
1	42290 ± 2384.2	17.7373	0.3829	273.63	HDSA$_2$ [Zefirov's PC]
2	-5242.5 ± 339.6	-15.4388	0.5546	273.95	av valency of a H atom
3	0.95 ± 0.085	11.1742	0.7477	433.67	total molecular surface area [semi-MO PC]
4	419.96 ± 42.69	9.8371	0.7504	329.17	av structural information content (first)
5	-79.2 ± 8.64	-9.1661	0.8061	363.41	av information content (second)
6	-147.83 ± 23.16	-6.3840	0.8155	321.26	max. total interaction for a C−H bond
7	-8350.4 ± 1814.4	-4.6023	0.8256	294.27	av nucleoph. react. index for a C atom
8	0.004 ± 0.00099	4.0310	0.8315	267.66	BETA polarizability
9	10.63 ± 2.69	3.9432	0.8373	247.62	symmetry number

[a] Figure 1: $R^2 = 0.8373$, $F = 247.62$, $s = 30.19$, $n = 443$. [b] R^2 for the correlation involving the given descriptor and those above it in table. [c] F for the correlation involving the given descriptor and those above it in the table.

Table 6. Six-Parameter Correlation Equations for the *Ortho* Subsets[a]

	$X \pm \Delta X$	*t*-test	$R^{2\ b}$	F^c	name of the descriptor
0	9980.3 ± 837.8	11.9130			intercept
1	15372 ± 1341.1	11.4623	0.5588	184.55	HDSA$_2$ [semi-MO PC]
2	3.33 ± 0.35	9.4805	0.6342	136.10	ALFA polarizability (DIP)
3	-4493.3 ± 481	-9.3409	0.7784	185.94	av valency of a H atom
4	-452.2 ± 61.67	-7.3330	0.8148	170.52	max. total interaction for a C−H bond
5	-4.32 ± 0.87	-4.9414	0.8393	160.82	ZX shadow
6	-153.45 ± 44.3	-3.4640	0.8510	145.59	max. partial charge for a O atom [Zefirov's PC]

[a] Figure 2: $R^2 = 0.8510$, $F = 145.59$, $s = 28.30$, $n = 160$. [b] Cf. footnote b in Table 5. [c] Cf. footnote c in Table 5.

PREDICTION OF MELTING POINTS FOR THE SUBSTITUTED BENZENES

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **917**

Table 7. Six-Parameter Correlation for the *Meta* Subset[a]

	$X \pm \Delta X$	t-test	$R^{2\ b}$	F^c	name of the descriptor
0	5576.3 ± 615.1	9.0647			intercept
1	-5568.3 ± 622.18	-8.9497	0.5330	177.45	av valency of a H atom
2	-14626 ± 1992	-7.3429	0.7287	206.85	av nucleoph. react. index for a C atom
3	36243 ± 5125	7.0718	0.8106	218.25	HDSA$_2$ [Zefirov's PC]
6	11.26 ± 2.55	4.4161	0.8251	180.06	symmetry number (AMPAC)
4	-20.9 ± 5.67	-3.6923	0.8521	173.99	LUMO energy
5	0.28 ± 0.085	3.3257	0.8623	156.50	DPSA$_2$ difference in CPSAs (PPSA$_2$-PNSA$_2$) [Zefirov's PC]

[a] Figure 3: $R^2 = 0.8623$, $F = 156.50$, $s = 28.29$, $n = 157$. *[b]* Cf. footnote b in Table 5. *[c]* Cf. footnote c in Table 5.

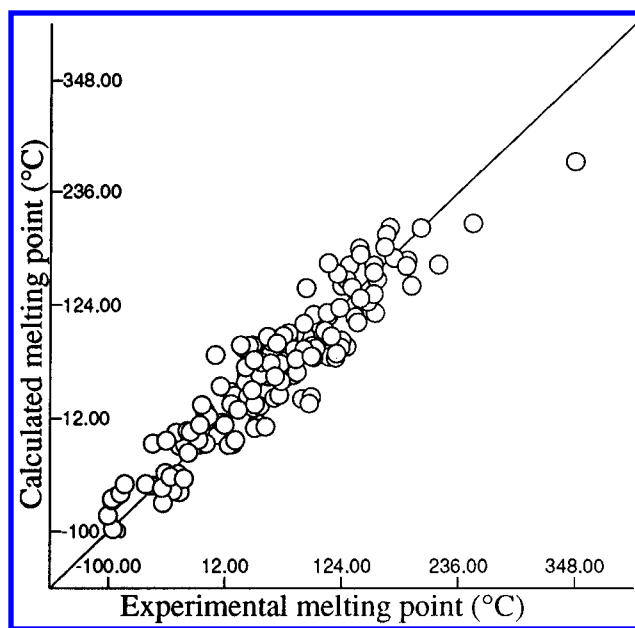**Table 8.** Six-Parameter Correlation Equations for the *Para* Subset[a]

	$X \pm \Delta X$	t-test	$R^{2\ b}$	F^c	name of the descriptor
0	2587.7 ± 674.9	3.8345			intercept
1	5.45 ± 0.52	10.4684	0.2140	49.55	XY shadow
2	0.54 ± 0.072	7.5439	0.3775	54.98	new H-acceptors PSA
3	-155.6 ± 24.4	-6.3691	0.7873	222.04	min. e$-$e repulsion for a H atom
4	9468.2 ± 1629.7	5.8099	0.8223	207.15	HDSA$_2$ [semi-MO PC]
5	-41.2 ± 8.9	-4.6183	0.8548	209.65	av information content (second)
6	-2448.7 ± 722.0	-3.3916	0.8637	186.54	av valency of a H atom

[a] Figure 4: $R^2 = 0.8637$, $F = 186.94$, $s = 30.50$, $n = 184$. *[b]* Cf. footnote b in Table 5. *[c]* Cf. footnote c in Table 5.
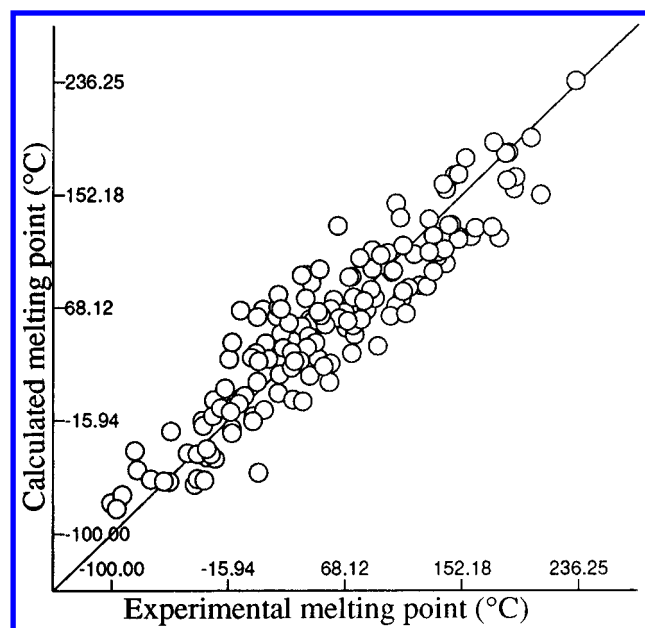
Figure 2. Six-parameter correlation of the calculated *vs* experimental melting points for the *ortho* set.

Figure 3. Six-parameter correlation of the calculated *vs* experimental melting points for the *meta* set.

each of the three subsets gave a similar correlation coefficient for the *ortho* subset ($R^2 = 0.839$) and higher correlation coefficients for the *meta* ($R^2 = 0.862$) and *para* ($R^2 = 0.864$) subsets. However, we observed that some of the descriptors from the whole set model were less significant for some of the subsets. By carrying out individual correlation analyses for the subsets we were able to develop predictive equations of a better quality.

For the *ortho* subset, the best correlation model involved six descriptors and delivered correlation coefficient $R^2 = 0.8510$ (Table 6 and Figure 2). The two most significant descriptors remained the HDSA$_2$ and the *Average Valency of a Hydrogen Atom*. However, in this case the HDSA$_2$ based on Mulliken charges outperformed the corresponding descriptor calculated from Zefirov charges. Additional descriptors are *Alpha Polarizability*,[17] *ZX shadow*,[31] and *Maximum partial charge for a O atom*. The last descriptor is derived again from Zefirov's electronegativity equalization scheme.[37,38]

For the *meta* subset, four descriptors were retained from the original equation (Table 7 and Figure 3). The order of their importance was changed, and the additional descriptors were the *LUMO energy*[17] and *difference between the partial positive and negative surface areas* (CPSA$_2$).[39]

For the *para* subset, the changes in QSAR model are more significant (Table 8 and Figure 4). New descriptors are more important in comparison with the three descriptors retained from the QSAR model for the whole set of compounds. New descriptors involved in the *para* model are *XY shadow*,[31] *H-acceptor partial surface area*, and *minimum electron$-$electron repulsion for a H atom*.[17]

In Tables 5 and 6, the correlations at the breakpoint on the F-values as a function of the number of descriptors are given in bold. It has been argued that the correlations at the breakpoint according to the F-value correspond to the statistically most significant model.[43] This may be the case for the whole set of substituted benzenes (Table 5). However, in the case of subsets of compounds (Table 6$-$8),
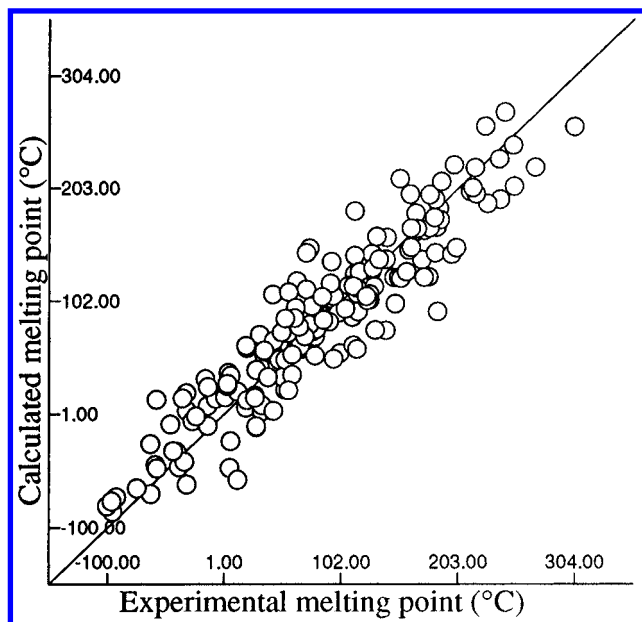
Figure 4. Six-parameter correlation of the calculated *vs* experimental melting points for the *para* set.

Table 9. Validation of the Melting Point Correlation Equation

subset	fit $R^{2 \, a}$	fit error[a]	prediction $R^{2 \, b}$	prediction error[b]
1	0.8148	30.3	0.8089	31.9
2	0.8649	28.8	0.7938	31.3
3	0.8178	30.8	0.8216	30.5

[a] Correlation coefficient and standard error of the fit of the melting point correlation equation to the randomly selected subset. [b] Correlation coefficient and standard error between the experimental and predicted melting points for the two other subsets.

additional descriptors had substantial improvement to the predicted melting points.

In order to validate the equation developed for the prediction of melting points, we divided the whole set of compounds into three subsets randomly and obtained correlation models for each of the subsets. The resulting new correlation equations were used in turn to predict melting points for the other two subsets. Predicted melting points were then compared with the experimental ones. Correlation results are summarized in Table 9 and demonstrate that in all three tests there is a good correlation between the predicted and experimental values, whereas the standard error does not deviate by more than two degrees.

CONCLUSIONS

Equations have been developed for predicting melting points of substituted benzenes based on the correlation analysis of an extensive collection of the melting point data for the disubstituted benzenes. Correlation equations are reported for the *ortho*, *meta*, and *para* subsets and for the collective set of all substituted benzenes. In all correlations, the hydrogen bonding descriptor $HDSA_2$ has been shown to be very successful in quantifying hydrogen bonding. In all models, the quantum chemical descriptors were found to be of a significant utility in developing successful correlations with the melting point. They were successfully applied together with the other types of descriptors and their use led to much more improved results in comparison to those obtained by using only topological, geometrical, and conventional descriptors.

REFERENCES AND NOTES

(1) Yalkowsky, S. H.; Valvani, S. C. Solubility and Partitioning I: Solubility of Nonelectrolytes in Water. *J. Pharm. Sci.* **1980**, *69*, 912−922.
(2) Dearden, J. C. The QSAR Prediction of Melting Point, a Property of Environmental Relevance. *Sci. Total Environ.* **1991**, *109/110*, 59−68.
(3) Matheson, L. E.; Chen, Y. A Quantitative Structure-Transportability Relationship for the Release of a Series of Substituted Benzenes and Pyridines from a Planar Polydimethylsiloxane Matrix. *Int. J. Pharm.* **1995**, *125*, 297−307.
(4) Abramowitz, R.; Yalkowsky, S. H. Melting Point, Boiling Point, and Symmetry. *Pharm. Res.* **1990**, *7*, 942−947.
(5) Benoit-Guyod, J-L; Andre, C.; Taillandier, G.; Rochat, J.; Boucherle, A. Toxicity and QSAR of Chlorophenols on Lebistes reticulatus. *Ecotoxico. Environ. Saf.* **1984**, *8*, 227−235.
(6) Devillers, J.; Chambon, P. Acute Toxicity and QSAR of Chlorophenols on Daphnia magna. *Bull. Environ. Contam. Toxicol.* **1986**, *37*, 599-605.
(7) Barratt, M. D. Quantitative Structure Activity Relationships for Skin Corrosivity of Organic Acids, Bases and Phenols. *Toxicol. Let.* **1995**, *75*, 169−176.
(8) Hanson, M. P.; Rouvray, D. H. In *Graph Theory and Topology in Chemistry*; King, R. B., Rouvray, D. H., Eds.; Elsevier Science Publishers B. V.: Amsterdam, 1987; Vol. 51, pp 201−208.
(9) Needham, D. E.; Wei, I-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, *110*, 4186−4194.
(10) Charton, M.; Charton, B. Quantitative Description of Structural Effects on Melting Points of Substituted Alkanes. *J. Phys. Org. Chem.* **1994**, *7*, 196−206.
(11) Katritzky, A. R.; Gordeeva, E. V. Traditional Topological Indices *vs* Electronic, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR Research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835−857.
(12) Katritzky, A. R.; Lobanov, V. S.; Karelson, M.; Murugan, R.; Grendze, M. P.; Toomey, J. E., Jr. Comprehensive Descriptors for Structural and Statistical Analysis. 1. Correlations between Structure and Physical Properties of Substituted Pyridines. *Rev. Roum. Chim.* **1996**, *41*, 851−867.
(13) Simamora, P.; Miller, A. H.; Yalkowsky, S. H. Melting Point and Normal Boiling Point Correlations: Applications to Rigid Aromatic Compounds. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 437−440.
(14) Krzyzaniak, J. F.; Myrdal, P. B.; Simamora, P.; Yalkowsky, S. H. Boiling Point and Melting Point Prediction for Aliphatic, Non-Hydrogen-Bonding Compounds. *Ind. Eng. Chem. Res.* **1995**, *34*, 2530−2535.
(15) Constantinou, L.; Gani, R. New Group Contribution Method for Estimating Properties of Pure Compounds. *AIChE J.* **1994**, *40*, 1697−1710.
(16) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem.* **1996**, *100*, 10400−10407.
(17) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1043.
(18) Atkins, P. W. In *Physical Chemistry*, 5th ed.; W. H. Freeman and Co.: New York, 1994; Chapter 19−20, pp 663−717.
(19) Dannenfelser, R. M.; Surendran, N.; Yalkowsky, S. H. Molecular Symmetry and Related Properties. *SAR QSAR Environ. Res.* **1993**, *1*, 273−292.
(20) Kitaigorodsky, A. I. *Molecular Crystals and Molecules*; Loebl, E. M., Ed.; Academic Press: New York, 1973.
(21) Gavezzotti, A.; Simonetta, M. Crystal Chemistry in Organic Solids. *Chem. Rev.* **1982**, *82*, 1−13.
(22) Steinstrasser, R.; Pohl, L. Chemistry and Applications of Liquid Crystals. *Angew. Chem., Int. Ed. Engl.* **1973**, *12*, 617−630.
(23) Brittain, H. G. Crystallographic Consequences of Molecular Dissymmetry. *Pharm. Res.* **1990**, *7*, 683−690.
(24) *CRC Handbook of Chemistry and Physics*, 65th ed.; Weast, R. C., Ed.; CRC Press, Inc.: Boca Raton, FL, 1984; pp C65−C576.
(25) PCMODEL, *Molecular Modelling Package*; Serena Software: Bloomington, IN, 1992.
(26) Stewart, J. J. P. *MOPAC Program Package*; QCPE, No. 455, 1989.
(27) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

PREDICTION OF MELTING POINTS FOR THE SUBSTITUTED BENZENES

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **919**

(28) Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem*. **1986**, *7*, 385−395.

(29) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279−287.

(30) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA: Reference Manual, version 2.0*; Gainesville, FL, 1994.

(31) Rohrbaugh, R. H.; Jurs, P. C. Descriptions of Molecular Shape Applied in Studies of Structure/Activity and Structure/Property Relationships. *Anal. Chim. Acta.* **1987**, *199*, 99−109.

(32) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topological Indices in Organic Chemistry. *Russ. Chem. Rev.* **1988**, *57*, 191−208.

(33) Kier, L. B.; Hall L. H. In *Molecular Connectivity in Chemistry and Drug Research*; Stevens, G., Ed.; Academic Press: New York, 1976; pp 27−39.

(34) Kier, L. B.; Hall L. H. In *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, 1986.

(35) Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

(36) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(37) Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I. Scheme for the Calculation of the Electronegativities of Atoms in a Molecule in the Framework of Sanderson's Principle. *Dokl. Akad. Nauk SSSR* **1987**, *296*, 883−887.

(38) Kirpichenok, M. A.; Zefirov, N. S. Electronegativity and Molecular Geometry. I. General Principles of the Method and Analysis of the Effect of Short-Range Electrostatic Interactions on Bond Lengths in Organic Molecules. *Zh. Org. Khim.* **1987**, *23*, 607−623.

(39) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Realtionship Studies. *Anal. Chem.* **1990**, *62*, 2323−2329.

(40) Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Pyrans and Pyrroles. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306−316.

(41) McQuarrie, D. A. In *Statistical Thermodynamics;* Woods, J. A., Ed.; Harper & Row, Publishers: New York, 1973.

(42) Akhiezer, A. I.; Peletminskii, S. V. *Methods of Statistical Physics*; Pergamon Press: Oxford, 1981.

(43) Myers, R. H. *Classical and Modern Regression With Applications*; PWS-KENT Publishing Co.: Boston, 1990.

(44) Gavezzotti, A. Molecular Symmetry, Melting Temperatures and Melting Enthalpies of Substituted Benzenes and Naphthalenes. *J. Chem. Soc., Perkin Trans. 2* **1995**, 1399−1404.