

Dimitar A. Dobchev · Mati Karelson

## Reparameterized Austin Model 1 for quantitative structure–property relationships in liquid media

Received: 18 April 2005 / Accepted: 18 November 2005 / Published online: 11 January 2006  
© Springer-Verlag 2006

**Abstract** A reparameterization of the quantum-chemical AM1 (Austin Model 1) model has been carried out using a nonlinear optimization based on a modification of the Levenberg–Marquardt technique. The optimum numerical values for the one-electron resonance integral parameters ( $\beta_s$  and  $\beta_p$ ) and core-core repulsion atomic parameters  $\alpha$  were obtained for the elements H, C, N, O, Cl and Br using the statistical fit of a two-parameter QSPR equation for the boiling points of organic compounds. A substantially improved two-parameter correlation ( $R^2=0.9685$ ,  $s=13.48$  K) was obtained by using the new optimized parameters. The QSPR equation employs two molecular descriptors, a bulk cohesiveness descriptor,  $\sqrt[3]{G_I}$  and the area-weighted surface charge of hydrogen-bonding donor atom(s) in the molecule. The model developed shows remarkably accurate predictions of the normal boiling points for nine additional simple inorganic compounds. The new parameters were tested on the critical temperatures of 165 organic compounds. The new QSPR model obtained for this property was found to be statistically significantly better than the original model.

**Keywords** Semiempirical parameterization · QSPR · Boiling point · Critical temperature

### Introduction

Quantum-chemical molecular descriptors have been widely used in the development of quantitative structure-activity/property relationships (QSAR/QSPR) [1–4]. In most cases, especially when large data sets of relatively

large molecules are considered, the descriptors are calculated from the molecular wave function obtained at the semiempirical level of theory. The most common semiempirical parameterizations used are Austin Model 1 (AM1) [5] and Parametric Model 3 (PM3) [6]. These parameterizations were developed by fitting the data on the properties of isolated molecules or those in the gas phase (heats of formation, ionization energies, dipole moments etc.). However, most of the QSPR/QSAR models are built on the data on molecular properties in condensed media (liquids, solutions, and membranes). Indeed, a large variety of quantum chemical models have been developed both at the semiempirical and ab initio level to account for the perturbation of the molecular wave function of a solute in condensed media [7–14]. Unfortunately, these models have only seldom found use in the calculation of molecular descriptors in the liquid phase [15–20]. Therefore, in the present work we wish to examine if a simple reparameterization of the AM1 model could improve the representation of QSPR models for molecular properties in the liquid phase. As a model process for this reparameterization, we have chosen the boiling points of compounds.

First of all, the normal boiling points of compounds provide valuable information in a variety of practical applications [21]. This property has also been applied in various areas of both chemistry and engineering as a powerful parameter that could be used to predict a number of key physical and physicochemical properties. On the other hand, the relationship of the boiling points to the molecular structure, and therefore the molecular properties, is predetermined by the intermolecular interaction in the liquid and by the difference in the molecular internal partition function in the gas phase and in the liquid at the boiling temperature. Our study is based on previously reported results by Katritzky et al. for the boiling points of 298 diverse organic and nine inorganic compounds [22]. Using only a two-descriptor equation, a very significant QSPR model was developed by them for a set of structurally very variable compounds (squared correlation coefficient  $R^2=0.954$  and standard deviation  $s=16.15$  K). The

D. A. Dobchev · M. Karelson (✉)  
Department of Chemistry,  
Tallinn University of Technology,  
Ehitajate tee 5,  
Tallinn 19086, Estonia  
e-mail: mati.karelson@ttu.ee  
Tel.: +372-6-202814  
Fax: +372-6-202819

two descriptors used in this model were Eq. (1) (1) the cubic root of the gravitation index  $\sqrt[3]{G_I}$  and (2) the charged surface area of hydrogen-donor atoms  $HDSA(2)$  [23].

$$T_b = (-170.7 \pm 7.46) + (65.88 \pm 0.86) \sqrt[3]{G_I} + (18470 \pm 540) HDSA(2) \quad (1)$$

Mathematically, these descriptors are defined as follows:

$$G_I = \sum_{ij} (m_i \cdot m_j / r_{ij}^2) \quad (2)$$

where the sum is over  $i$  and  $j$  for all bonded pairs of atoms in the molecule,  $m_i$  and  $m_j$  are the atomic masses and  $r_{ij}$ —the distance between the  $i$ -th and  $j$ -th atoms, and

$$HDSA(2) = \sum \frac{q_D \sqrt{S_D}}{S_{tot}} \quad (3)$$

where  $q_D$  is the partial charge on hydrogen-bonding donor (H) atom(s),  $S_D$  labels the surface area for this atom, and  $S_{tot}$  is the total molecular surface area, calculated from the van der Waals radii of atoms (overlapping spheres). In the last formula Eq. (3), the summation goes over the number of simultaneously possible hydrogen-bonding donor pairs per molecule. The results obtained in the original article using the CODESSA program [24] showed a strong orthogonality of these descriptors (the pair correlation coefficient between  $\sqrt[3]{G_I}$  and  $HDSA(2)$  is only 0.2041). From the physical point of view, both descriptors have obvious meaning;  $G_I$  is connected with the dispersion and cavity-formation effects in liquid,  $HDSA(2)$  reflects the hydrogen-bonding ability of compounds. The theory of condensed media assumes that the energy of cavity formation is proportional to the surface of the molecular cavity. The latter depends on the geometry of the molecule and will thus be directly related to the gravitation index  $G_I$ . The second descriptor,  $HDSA(2)$  is related to the energy of hydrogen bond formation.

The goal of the present work was to adjust the quantum-chemical parameterization for the AM1 semiempirical model to improve the QSPR models of molecular properties in the liquid phase. In general, the new set of semiempirical parameters should produce more adequate characteristics (bond lengths and angles, charges, dipole moments etc.) of the molecules in the liquid phase by using a nonlinear optimization for the AM1(Austin Model) semiempirical parameters  $\alpha$ ,  $\beta_s$  and  $\beta_p$  where  $\alpha$  is core–core repulsion atomic parameter and  $\beta_s$ ,  $\beta_p$ —one-electron resonance integral parameters (for  $s$  and  $p$  states, respectively) [5], a new set of 17 parameters for the atoms H, C, N, O, Cl and Br (Table 1) were derived. The applicability of the new set of parameters was examined using the QSPR model for the critical temperatures of compounds [25].

**Table 1** The original and new optimized AM1 parameters

No	Atom	Parameter	Original values ([5])	Re-optimized values
1	H	$\alpha$ , Å <sup>-1</sup>	2.882324	2.858038
2	C	$\alpha$ Å <sup>-1</sup>	2.648274	2.644260
3	N	$\alpha$ Å <sup>-1</sup>	2.947286	2.911589
4	O	$\alpha$ Å <sup>-1</sup>	4.455371	3.847456
5	Cl	$\alpha$ Å <sup>-1</sup>	2.919368	4.428944
6	Br	$\alpha$ Å <sup>-1</sup>	2.576546	2.631694
7	H	$\beta_s$ , eV	-6.173790	-6.236430
8	C	$\beta_s$ , eV	-15.71580	-15.479268
9	N	$\beta_s$ , eV	-20.29910	-19.73390
10	O	$\beta_s$ , eV	-29.27280	-27.79330
11	Cl	$\beta_s$ , eV	-24.59470	-25.04320
12	Br	$\beta_s$ , eV	-19.39990	-22.20060
13	C	$\beta_p$ , eV	-7.719280	-7.86513
14	N	$\beta_p$ , eV	-18.23870	-19.36820
15	O	$\beta_p$ , eV	-29.27280	-28.05140
16	Cl	$\beta_p$ , eV	-14.63720	-22.78330
17	Br	$\beta_p$ , eV	-8.95720	-8.70759

## Methodology and optimization

The original set of AM1 parameters was produced by optimization of several properties of the molecules (e.g. bonds, angles, heat of formation etc.). Two kinds of parameter were considered in our reparameterization procedure (Table 1). First, within the AM1 and MNDO<sup>26</sup> formalism, the Fock matrix elements  $F_{\rho\sigma}$  are expressed as follows:

$$F_{\rho\sigma} = \beta_{\rho\sigma} - \frac{1}{2} \sum_{\kappa} \sum_{\lambda} P_{\kappa\lambda} \langle \rho\kappa | \sigma\lambda \rangle. \quad (4)$$

where  $\langle \rho\kappa | \sigma\lambda \rangle$  are the two-center two-electron repulsion integrals and  $\beta_{\rho\sigma}$  denote the two-center one-electron core resonance integrals. The latter are approximated to be proportional to the corresponding overlap integrals  $S_{\rho\sigma}$  and a function with an implicit dependence on the interatomic distance  $f_{\text{res}}(R_{AB})$  [26]:

$$\beta_{\rho\sigma} = f_{\text{res}}(R_{AB}) S_{\rho\sigma} \quad (5)$$

In MNDO and AM1 models, the function  $f_{\text{res}}(R_{AB})$  is approximated by:

$$f_{\text{res}}(R_{AB}) = (\beta_p^A + \beta_p^B) / 2 \quad (6)$$

where  $\beta_p^A$ ,  $\beta_p^B$  are adjustable atomic semiempirical resonance parameters. There exist at most two different  $\beta$  parameters for any first-row element i.e.  $\beta_s^A$  and  $\beta_p^A$ , for  $s$  and  $p$  AO's, respectively.

As a successor of MNDO [26], AM1 has different parameterization for the core–core repulsion function (CRF) in order to compensate the excessive repulsion due to atom

interactions at van der Waals distance. It is defined as follows

$$\begin{aligned} \text{CRF}(\text{AB}) = & Z_A Z_B \gamma_{\text{AB}} \{ 1 + \exp(-\alpha_A R_{\text{AB}}) \\ & + \exp(-\alpha_B R_{\text{AB}}) \\ & + R_{\text{AB}}^{-1} \sum_i K_{\text{Ai}} \exp[L_{\text{Ai}}(R_{\text{AB}} - M_{\text{Ai}})] \\ & + R_{\text{AB}}^{-1} \sum_j K_{\text{Bj}} \exp[L_{\text{Bj}}(R_{\text{AB}} - M_{\text{Bj}})] \} \end{aligned} \quad (7)$$

where the parameters  $L$  determine the width of the Gaussians and  $M$  and  $K$  are adjustable parameters for the excessive repulsion between atoms at larger van der Waals distances [5]. As seen from Eq. (7), the parameter  $\alpha$  is directly connected to the interatomic interactions, controlling the appropriate behavior of these repulsions with respect to the interatomic distances  $R_{\text{AB}}$ . Through the above approximations for the electron energy and core–core repulsion energy, the total energy of the molecule can be presented as the parametric function of interatomic distances  $R_{\text{AB}}$ . In this way, all characteristics of the compounds are calculated at the equilibrium geometry for a given parameterization. Therefore, both  $\alpha$  and  $\beta$  parameters were chosen as arguments in our quantum-chemical reparameterization. The optimization of these parameters thus leads to new equilibrium geometries of molecules at which the new values of molecular descriptors (HDSA(2) and  $G_I$ ) will be calculated that would be more adequate for the QSPR model of boiling points.

The set of 298 molecules is structurally sufficiently diverse and includes saturated hydrocarbons, halogenated compounds, and cyano, ester, ether, hydroxyl, amino, carbonyl and carboxyl functionalities. The critical temperatures of 165 organic compounds from the DIPPR database were the same as in the original work [25]. Most of the compounds had the same chemical functionalities and some of them are overlapped with the set of boiling points. The initial geometry of all compounds was optimized using the MOPAC program [27] with the original semiempirical AM1 parameterization. The molecular descriptors were calculated within the CODESSA PRO software [28].

Considering the overall number of parameters optimized (17), the problem of finding the new AM1 parameters was reduced to finding a minimum of the function  $F(\alpha, \beta_s, \beta_p)$  in 17-dimensional space formed by  $\alpha, \beta_s$  and  $\beta_p$  where  $F$  is

$$F(\alpha, \beta_s, \beta_p) = \sum_{k=1}^{298} F_k^2 = \sum_{k=1}^{298} (T_{\text{exp}}^k - T_{\text{calc}}^k)^2 \quad (8)$$

where  $T_{\text{exp}}^k$  denote experimental boiling points and  $T_{\text{calc}}^k$  are

$$T_{\text{calc}}^k = X_1 + X_2 \sqrt[3]{G_I} + X_3 \text{HDSA}(2) \quad (9)$$

the calculated boiling points ( $X_i$  are coefficients for QSPR

equation).  $F_k$  is thus the difference between the experimental and predicted boiling points for a given ( $k$ th) compound. Because of the nature of the task, the most appropriate approach to solve it can be based on an iterative procedure. We have used a modification of the Levenberg–Marquardt technique [29] to minimize the sum of the squares of nonlinear functions  $F_k$ . Our procedure requires the values for the derivatives of the function being minimized with respect to all variables (Jacobian matrix). The main difficulty for the task is the calculation of the derivatives. In the present situation, derivatives of  $F$  were found by finite differences. Because of the complexity of the 17-dimensional surface Eq. (8), care should be taken in the calculation of the Jacobian matrix in order to find the right direction that proceeds to the minimum. As initial guess for the parameters  $\alpha, \beta_s$  and  $\beta_p$ , the original AM1 parameters [5] were chosen (Table 1). As mentioned above, the structure of molecules was initially optimized with these parameters. In this way, each iteration started with the same geometry of the compounds, but with a different set of  $\alpha$  and  $\beta$  values. At each iteration, with different parameters loaded within MOPAC, the new descriptors ( $G_I$  and HDSA(2)) were calculated by CODESSA-PRO. The latter automatically generates the new QSPR equation with its statistical characteristics ( $X_i$  coefficients of Eq. (9), correlation coefficient  $R^2$ , standard deviation  $s$ ,  $t$ -test etc.) of the model and the predicted boiling points  $T_{\text{calc}}$ . The standard deviation of the QSPR model was chosen as the main optimization criterion for the optimization.

## Results and discussion

The optimization according to Eq. (8) resulted in a new set of AM1 (AM1-BP) parameters shown in Table 1. Notably, there is significant difference in  $\alpha$  parameter for oxygen and chlorine atoms as compared with the original parameters by Dewar et al. [5]. The remaining parameters  $\alpha$  were much less changed. The new  $\beta$  parameter values for H and C atoms are close to the original data. A large difference in  $\beta_p$  values was found for chlorine atoms. For the oxygen atom, the values for  $\beta_s$  and  $\beta_p$  were set equal to each other in the original work [5]. However, in our case, these parameters are different. The variance in the parameterization of O and Cl as compared to the original parameters indicates the higher electronegativity of these atoms with respect to the electrostatic molecular interactions in solute–solvent (solute–solute) systems. Thus, this tendency of new optimized AM1 parameters reflects adequately the intermolecular interactions in liquid phase. The statistical characteristics of the final QSPR equation for  $T_b$  is shown in Table 2 (see also Fig. 1 and Table 3 for the predictions). As a rule, the large-sized compounds possess higher boiling temperature. For hydrocarbons, the results of QSPR are very close to the experimental boiling points. This observation indicates that the gravitation index  $G_I$  that accounts simultaneously for both the atomic masses (volumes) and for their distribution within the molecular space describes correctly the dispersion forces and cavity

**Table 2** The QSPR equation for the boiling points (nine), using the descriptors calculated using new AM1 parameterization ( $R^2=0.968$ ,  $s=13.48$  K,  $F=4529.94$ )

<i>n</i>	Descriptor	$X+\Delta X$	<i>t</i> -test
1	Intercept	-178.66±6.20	-28.79
2	$G_T^{1/3}$	66.83±0.72	92.60
3	HDSA(2)	19608.05±454.27	43.16

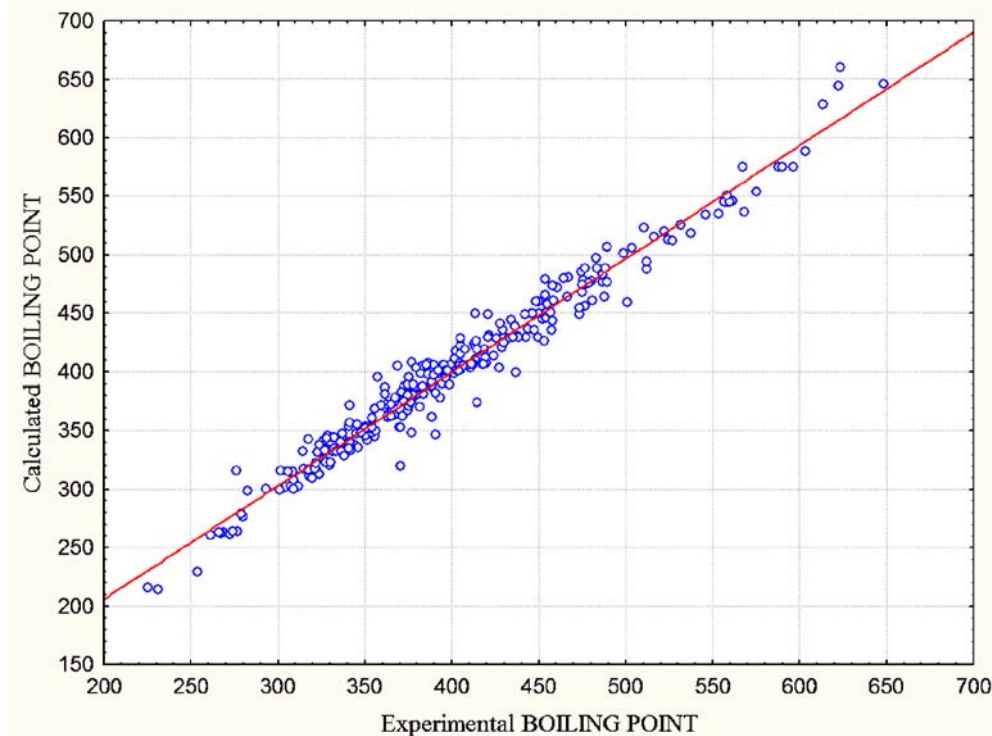
formation processes in the liquid. In general, the larger molecular masses lead to stronger dispersion forces and therefore higher boiling points. These forces also depend on the shape of the molecule, being most important for non-polar species. The energy of cavity formation for a compound is roughly proportional to the molecular volume of the solute, or to the surface area of cavity, and to the microscopic surface tension on the boundary between the solute cavity and the solvent.

The HDSA(2) descriptor combines the surface areas and partial charges of the atoms and successfully addresses the dipole–dipole attractions and hydrogen bonding in liquids. This is also in agreement with the concept that strong dipole–dipole interactions for the formation of hydrogen bonds contribute to the one of the main terms in solvation energy of a molecule [7]. The HDSA(2) descriptor also elucidates the accessibility of the heteroatoms to play an important role in determining the tendency of a molecule to enter the vapor phase.

The optimized semiempirical parameters and the corresponding QSPR model (Eq. (9) and Table 2) were tested on the prediction of the boiling points for a set of nine inorganic compounds (Table 4). Notably, the predictions were significantly better than reported earlier using the descriptors developed from the original AM1 parameterization (average deviation 17 vs. 22 K in the original work).

Another test of the new AM1 parameterization was carried out by the re-development of the earlier QSPR model for the critical temperatures of 165 organic compounds [25]. The new parameters were set up for the MOPAC program and its output files were loaded in CODESSA. Thus, the new molecular descriptors were calculated proceeding from the new geometries and wave functions of the molecules. The new best two-parameter QSPR equation for the critical temperatures found by CODESSA (Table 5), was statistically significantly better than the best two-descriptor equation in the original work [25] ( $R^2=0.868$ ,  $s=28.7$  K,  $F=546$ ). The original model has the same functional dependence on the descriptors as (Table 5). Notably, instead of HDSA(2) descriptor, a similar HASA(2) descriptor was used in the best equation for the critical temperatures (Table 5). Consequently, the AM1 reparameterization also improves the representation of other electrostatic descriptors for the molecules in the liquid phase. All calculated 165 critical temperatures using equation in Table 5 are shown in Table 6 and plotted against the experimental values in Fig. 2.

**Fig. 1** The calculated vs experimental normal boiling points according to the best two-descriptor equation (Table 2) and new AM1 parameters





**Table 3** Experimental boiling points vs. calculated boiling points (equation of Table 2) for 298 organic compounds

No	Compound	$T_{exp}$ , K	$T_{calc}$ , K
1	1,1-dichloropropane	361.25	370.60
2	1,1-diphenylethane	545.78	534.34
3	1,2,3,4-tetrahydronaphtalene	480.77	460.98
4	1,2,3-trimethylbenzene	449.27	429.56
5	1,2,4-trimethylbenzene	442.53	429.76
6	1,2-dichloropropene	361.25	386.77
7	1,2-diphenylethane	553.65	534.66
8	1,2-propylene glycol	460.75	471.65
9	1,3-butadiene	268.74	263.53
10	1,3-butanediol	480.15	477.29
11	1,3-cyclohexadiene	353.49	351.43
12	1,3-dichloropropane	393.55	378.04
13	1,3-propylene glycol	487.55	463.82
14	1,4-butanediol	501.15	459.52
15	1,4-dichlorobutane	427.05	403.14
16	1,5-dichloropentane	453.15	426.31
17	1,5-hexadiene	332.61	334.84
18	1,5-pentanediol	512.15	487.71
19	1,6-hexanediol	516.15	515.23
20	1-bromobutane	374.75	366.81
21	1-bromopropane	344.15	337.93
22	1-butene	266.90	262.52
23	1-chlorobutane	351.58	341.94
24	1-chloropentane	381.54	370.42
25	1-decanol	503.35	505.57
26	1-decene	443.75	436.29
27	1-dodecane	486.50	476.66
28	1-heptene	366.79	363.34
29	1-hexadecene	558.02	545.13
30	1-hexanal	401.45	398.22
31	1-hexanol	430.15	429.46
32	1-hexene	336.63	334.05
33	1-octadecene	587.97	574.99
34	1-octene	394.44	389.78
35	1-pentanol	410.95	403.55
36	1-pentene	303.11	300.95
37	1-tetradecene	524.25	512.63
38	2,2,3,3-tetramethylpentane	413.44	410.72
39	2,2,3-trimethylbutane	354.03	360.84
40	2,2,3-trimethylpentane	383.00	387.36
41	2,2,4-trimethylpentane	372.39	387.52
42	2,2-dimethyl-1-propanol	386.25	406.93
43	2,2-dimethylbutane	322.88	332.00
44	2,3,3-dimethylpentane	387.92	387.34
45	2,3-butanediol	453.85	479.22
46	2,3-dimethyl-1-butene	328.76	333.92
47	2,3-dimethyl-2-butene	346.35	335.78
48	2,3-dimethyl-3-butadiene-1-ol	341.93	335.53
49	2,3-dimethylbutane	331.13	332.31
50	2,3-dimethylhexane	388.76	388.22
51	2,3dimethylpentane	362.93	361.66
52	2,4,4-trimethyl-1-pentene	374.59	389.01
53	2,4,4-trimethyl-2-pentene	378.06	389.52

**Table 3** (continued)

No	Compound	$T_{exp}$ , K	$T_{calc}$ , K
54	2,6-xylenol	474.22	485.65
55	2-bromobutane	364.37	364.15
56	Bromopropane	332.56	335.11
57	2-ethylbutyric acid	466.95	480.82
58	2-ethyl-1-butanol	419.65	407.26
59	2-ethyl-1-butene	337.82	334.28
60	2-ethyl-1-hexanol	457.75	473.15
61	2-ethylhexyl acrylate	489.15	506.49
62	2-hexanol	413.04	423.96
63	2-hexanone	400.85	403.59
64	2-methyl-1-butanol	401.85	411.39
65	2-methyl-1-butene	304.30	301.40
66	2-methyl-1-pentanol	421.15	430.79
67	2-methyl-1-pentene	335.25	334.43
68	2-methyl-2-butanol	375.15	395.58
69	2-methyl-2-pentene	340.45	335.41
70	2-methyl-3-ethylpentene	388.80	388.13
71	2-methylbutyric acid	450.15	459.79
72	2-methylhexane	363.20	362.17
73	2-methylpentane	333.41	332.77
74	2-methylpyridine	402.55	417.62
75	2-pentanol	392.15	398.85
76	2-pentanone	375.46	377.82
77	2-propanol	355.41	344.97
78	3,3-diethylpentane	419.34	411.69
79	3,3-dimethyl-1-butene	314.40	332.77
80	3-chloropropene	318.11	310.86
81	3-hexanone	396.65	394.74
82	3-methyl-1-butanol	404.35	407.36
83	3-methyl-1-butene	293.21	300.41
84	3-methyl-1-pentene	327.33	333.51
85	3-methyl-2-butanol	384.65	403.43
86	3-methyl-2-butene	311.71	302.56
87	3-methylhexane	365.00	362.19
88	3-methylpentane	336.42	332.67
89	3-methylpyridine	417.29	417.60
90	3-pentanol	388.45	391.41
91	4-methyl-1-pentene	327.01	333.63
92	4-methyl-2-pentanol	404.85	422.61
93	4-methylpyridine	418.50	418.72
94	Acetal	376.75	407.88
95	Acetone	329.44	320.57
96	Acetophenone	475.15	467.63
97	Acetylacetone	413.55	449.81
98	Acrylaldehyde	325.84	340.76
99	Acrylonitrile	350.50	345.85
100	Adiponitrile	568.15	536.84
101	$\alpha$ -methylstyrene	438.65	429.56
102	Allyl acetate	377.15	381.55
103	Allyl alcohol	370.23	352.62
104	Allyl amine	326.45	325.79
105	Aniline	457.60	435.91
106	Benzaldehyde	451.90	444.85

**Table 3** (continued)

No	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K
107	Benzene	353.24	353.02
108	Benzoic acid	522.40	519.90
109	Benzyl acetate	486.65	482.88
110	Benzyl alcohol	477.85	474.70
111	Benzyl benzoate	596.65	574.98
112	Bicyclohexyl	512.19	494.28
113	Bromobenzene	429.24	435.64
114	Butyl vinyl ether	366.97	374.40
115	Chlorobenzene	404.87	415.27
116	<i>Cis</i> -1,2-dimethylcyclohexane	402.94	401.10
117	<i>Cis</i> -1,3-dimethylcyclohexane	393.24	401.18
118	<i>Cis</i> -1,4-dimethylcyclohexane	397.47	401.18
119	<i>Cis</i> -2-butene	276.87	263.97
120	<i>Cis</i> -2-hexene	342.03	334.95
121	Cumene	425.56	428.39
122	Cyclohexane	353.87	348.51
123	Cyclohexanol	434.00	444.01
124	Cyclohexanone	428.90	420.86
125	Cyclohexylamine	407.65	419.17
126	Cyclopentadiene	314.65	317.27
127	Cyclopentane	322.40	316.50
128	Cyclopentene	317.38	316.99
129	Cyclohexene	356.12	350.12
130	Di- <i>n</i> -butyl ether	413.44	421.69
131	Di- <i>n</i> -hexyl ether	498.85	501.15
132	Di- <i>n</i> -propyl ether	362.79	372.57
133	Di- <i>n</i> -propylamine	382.00	398.24
134	Dibutyl phthalate	613.15	628.67
135	Dibutyl sebacate	622.15	644.70
136	Diethyl ether	308.58	312.63
137	Diethyl ketone	375.14	371.05
138	Diethyl phthalate	567.15	575.18
139	Diethylamine	328.60	345.50
140	Diisopropyl ether	341.45	371.83
141	Diisopropylamine	357.05	395.52
142	Dimethylphthalate	556.85	545.55
143	Dimethyl terephthalate	561.15	545.78
144	Diphenyl ether	531.46	525.41
145	Diphenyl amine	575.15	553.96
146	Diphenylmethane	537.42	518.36
147	Divinyl ether	301.45	315.93
148	Ethyl acetate	350.21	353.47
149	Ethyl acrylate	372.65	381.17
150	Ethyl benzoate	486.55	482.85
151	Ethyl formate	327.46	322.59
152	Ethyl isobutyrate	383.00	405.24
153	Ethyl isopropyl ketone	386.55	398.14
154	Ethyl- <i>n</i> -butyrate	394.65	405.66
155	Ethyl propionate	372.25	380.75
156	Ethyl propyl ether	337.01	344.32
157	Ethyl vinyl ether	308.70	314.95
158	Ethylbenzene	409.35	405.81
159	Ethylcyclohexane	404.95	401.46
160	Ethylcyclopentane	376.62	375.34

**Table 3** (continued)

No	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K
161	Ethylenecarboxylic acid	414.15	426.41
162	Hexamethylene imine	404.85	404.81
163	Hexanenitrile	436.75	399.81
164	Isopentane	300.99	299.50
165	Isobutane	261.43	260.95
166	Isobutanol	380.81	381.27
167	Isobutene	266.25	263.26
168	Isobutyl acetate	389.80	405.39
169	Isobutyl acrylate	405.15	428.66
170	Isobutyl formate	371.22	379.94
171	Isobutyl	420.65	449.30
172	Isobutylamine	340.88	353.39
173	Isobutylbenzene	445.94	449.77
174	Isobutyraldehyde	337.25	347.14
175	Isobutyric acid	427.85	441.39
176	Isobutyronitrile	376.76	348.42
177	Isophorone	488.35	488.74
178	Isoprene	307.21	302.23
179	Isopropyl acetate	361.65	380.90
180	Isopropyl chloride	308.85	307.88
181	Isopropyl amine	305.55	315.01
182	Isovaliric acid	448.25	460.20
183	<i>m</i> -cresol	475.43	478.54
184	<i>m</i> -diethylbenzene	454.29	450.45
185	<i>m</i> -diethylbenzene	476.33	488.61
186	<i>m</i> -ethyltoluene	434.48	429.21
187	<i>m</i> -toluidine	476.55	456.36
188	<i>m</i> -xylene	412.27	406.42
189	Mesityl oxide	402.95	402.99
190	Mesitylene	437.89	429.82
191	Methacrolein	341.15	356.49
192	Methyl acrylate	353.35	354.06
193	Methyl ethyl ketone	352.79	352.04
194	Methyl acetate	330.09	323.20
195	Methyl ethyl ether	280.05	276.51
196	Methyl isobutyl ketone	389.65	396.98
197	Methyl isobutyl ether	331.70	344.10
198	Methyl isopropenyl ketone	371.15	383.16
199	Methyl isopropyl ether	323.75	312.44
200	Methyl isopropyl ketone	367.55	378.17
201	Methyl <i>n</i> -butyrate	375.90	380.93
202	Methyl propionate	352.60	353.59
203	Methyl <i>sec</i> -butyl ether	332.15	344.08
204	Methyl <i>tert</i> -butyl ether	328.35	343.60
205	Methyl <i>tert</i> -pentyl ether	359.45	371.83
206	Methyl vinyl ether	278.65	279.19
207	Methanal	253.65	228.95
208	Methylcyclohexane	374.08	376.11
209	Methylcyclopentadiene	345.93	348.31
210	Methylcyclopentane	344.96	347.47
211	<i>N,N</i> -dimethylaniline	466.69	463.90
212	<i>n</i> -butane	272.65	261.45
213	Butanol	390.81	382.02
214	<i>n</i> -butyl acetate	399.15	405.72

**Table 3** (continued)

No	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K
215	<i>n</i> -butyl acrylate	421.00	428.97
216	<i>n</i> -butyl ethyl ether	365.35	372.56
217	<i>n</i> -butyl formate	379.25	380.29
218	<i>n</i> -butyl stearate	623.15	660.38
219	<i>n</i> -butylamine	350.55	352.03
220	<i>n</i> -butylbenzene	456.46	450.08
221	<i>n</i> -butylcyclohexane	454.13	446.21
222	<i>n</i> -butyraldehyde	347.95	348.42
223	<i>n</i> -butyric acid	436.42	439.01
224	<i>n</i> -butyronitrile	390.75	346.41
225	<i>n</i> -decane	447.30	435.76
226	<i>n</i> -dodecane	489.47	476.21
227	<i>n</i> -heptane	371.58	362.63
228	<i>n</i> -hexadecane	560.01	544.79
229	<i>n</i> -hexane	341.88	333.25
230	<i>n</i> -hexanoic acid	478.85	475.51
231	<i>n</i> -hexylamine	404.65	404.74
232	<i>n</i> -nonadecane	603.05	588.78
233	<i>n</i> -nonane	423.97	413.36
234	<i>n</i> -octadecane	589.86	574.68
235	<i>n</i> -octane	398.83	389.13
236	<i>n</i> -pentane	309.23	300.04
237	<i>n</i> -pentyl formate	406.60	405.24
238	<i>n</i> -pentylamine	377.65	380.12
239	Propanol	370.35	352.65
240	<i>n</i> -propionaldehyde	321.15	317.36
241	<i>n</i> -propyl acetate	374.65	380.81
242	<i>n</i> -propyl chloride	319.67	309.94
243	<i>n</i> -propyl formate	353.97	353.13
244	<i>n</i> -propyl propionate	395.65	405.66
245	<i>n</i> -propyl amine	321.65	322.04
246	<i>n</i> -propylbenzene	432.39	428.72
247	<i>n</i> -propylcyclohexane	429.90	424.59
248	<i>n</i> -propylcyclopentane	404.11	400.72
249	<i>n</i> -tetradecane	526.73	512.20
250	Neopentane	282.65	298.79
251	Neopentyl glycol	483.00	496.79
252	<i>o</i> -cresol	464.15	479.63
253	<i>o</i> -dichlorobenzene	453.57	465.94
254	<i>o</i> -diethylbenzene	456.61	450.27
255	<i>o</i> -ethyltoluene	438.33	429.09
256	<i>o</i> -toluidine	473.55	449.39
257	<i>o</i> -xylene	417.58	406.39
258	<i>p</i> -cresol	475.13	474.14
259	<i>p</i> -cymene	450.28	450.21
260	<i>p</i> -diethylbenzene	456.94	450.43
261	<i>p</i> -diisopropylbenzene	483.65	488.64
262	<i>p</i> -ethyltoluene	435.16	429.27
263	<i>p</i> -hydroquinone	558.15	550.36
264	<i>p</i> -tuluidine	473.40	454.67
265	<i>p</i> -xylene	411.51	406.48
266	Phenol	454.99	457.86
267	Piperidine	379.55	403.52
268	Propane	231.11	214.55

**Table 3** (continued)

No	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K
269	Propionic acid	414.32	419.14
270	Propionitrile	370.50	319.73
271	Propylene	225.43	216.03
272	Pyridine	388.41	361.34
273	Quinoline	510.75	522.79
274	<i>Sec</i> -butyl acetate	385.15	405.41
275	<i>Sec</i> -butyl alcohol	372.70	374.97
276	<i>Sec</i> -butyl chloride	341.25	340.18
277	<i>Sec</i> -butyl amine	336.15	340.01
278	<i>Sec</i> -butyl benzene	446.48	449.64
279	Stearic acid	648.35	645.59
280	Styrene	418.31	406.43
281	<i>Tert</i> -butyl acetate	369.15	405.04
282	<i>Tert</i> -butyl alcohol	355.57	368.90
283	<i>Tert</i> -butyl chloride	323.75	338.03
284	<i>Tert</i> -butyl amine	317.55	342.45
285	<i>Tert</i> -butylbenzene	442.30	449.01
286	Tetrahydrofuran	338.00	328.78
287	Toluene	383.78	381.00
288	<i>Trans</i> -1,2-dimethylcyclohexane	396.58	401.06
289	<i>Trans</i> -1,3-dimethylcyclohexane	397.61	401.17
290	<i>Trans</i> -1,4-dimethylcyclohexane	392.51	401.19
291	<i>Trans</i> -2-butene	274.03	263.84
292	<i>Trans</i> -2-hexene	341.02	334.94
293	<i>Trans</i> -crotonic acid	458.15	443.28
294	Trimethylamine	276.02	315.52
295	Valeraldehyde	376.15	373.82
296	Valeric acid	458.65	460.87
297	Valeronitrile	414.45	374.04
298	Vinyl acetate	345.65	354.95

**Table 4** Experimental vs calculated boiling points for NINE inorganics using Eq. (9) with new AM1 parameters

No	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K	$F=T_{\text{exp}}-T_{\text{calc}}$
1	H <sub>2</sub> O	373	369.17	3.83
2	H <sub>2</sub> O <sub>2</sub>	425	436.24	-11.24
3	NH <sub>3</sub>	240	256.73	-16.73
4	N <sub>2</sub> H <sub>4</sub>	387	357.04	29.96
5	NH <sub>2</sub> OH	330	343.62	-13.62
6	HCN	299	308.84	-9.84
7	CH <sub>3</sub> F	195	194.31	0.69
8	CH <sub>3</sub> NH <sub>2</sub>	267	262.28	4.72
9	HF	293	280.77	12.23

**Table 5** The QSPR equation for the critical temperatures of 165 organic compounds obtained with the re-optimized AM1 parameters ( $R^2=0.902$ ,  $s=25.04$  K,  $F=742.73$ )

$N$	Descriptor	$X+\Delta X$	$t$ -test
1	Intercept	-173±18.3	-6.40
2	$G_1^{1/3}$	81±2.2	36.81
3	HASA(2)	12±0.76	16.71

**Table 6** The experimental vs. predicted critical temperatures for 165 organic compounds (according to the QSPR model from Table 5)

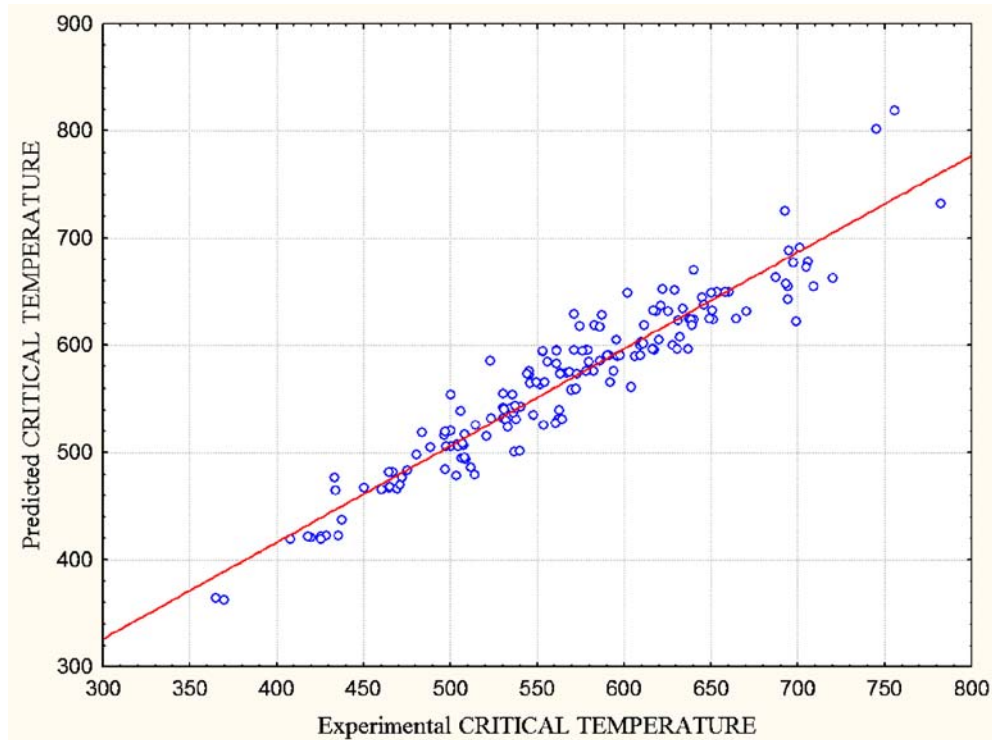
Number	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K	Number	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K
1	3-chloropropene	514.2	479.5	83	Ethylcyclopentane	569.5	558.2
2	Propylene	364.8	363.9	84	2,3-dimethylpentane	537.4	541.5
3	1,2-dichloropropene	545.1	572.1	85	<i>n</i> -heptane	540.3	542.7
4	Ethylformate	508.4	493.9	86	2-methylhexane	530.4	542.2
5	Methyl acetate	506.8	494.6	87	3-methylhexane	535.3	542.0
6	Propionic acid	604.0	561.2	88	Ethylbenzene	617.2	595.4
7	<i>n</i> -propyl chloride	503.2	478.4	89	<i>m</i> -xylene	617.1	596.1
8	Propane	369.8	362.1	90	<i>p</i> -xylene	616.3	596.1
9	Methyl ethyl ether	437.8	437.6	91	2,6-xylenol	701.1	691.0
10	<i>n</i> -propanol	536.7	500.4	92	<i>N,N</i> -dimethylaniline	687.2	663.6
11	Methylal	480.6	498.1	93	<i>cis</i> -1,2-dimethylcyclohexane	606.2	589.6
12	Isopropylamine	471.9	476.3	94	<i>trans</i> -1,2-dimethylcyclohexane	596.2	589.5
13	<i>n</i> -propylamine	497.0	484.4	95	<i>cis</i> -1,3-dimethylcyclohexane	591.2	589.7
14	Trimethylamine	433.3	476.9	96	<i>trans</i> -1,3-dimethylcyclohexane	598.0	589.7
15	1,3-butadiene	425.4	421.9	97	<i>trans</i> -1,4-dimethylcyclohexane	590.2	589.7
16	Butyronitrile	582.3	575.3	98	ethylcyclohexane	609.2	589.8
17	1-butene	419.6	420.6	99	isobutyl isobutyrate	602.0	648.4
18	<i>cis</i> -2-butene	435.6	422.4	100	2,3-dimethylhexane	563.4	573.9
19	<i>trans</i> -2-butene	428.6	422.2	101	2-methyl-3-ethylpentane	567.0	573.8
20	Isobutene	417.9	421.5	102	<i>n</i> -octane	568.8	575.0
21	Methyl ethyl ketone	535.5	553.7	103	2,2,3-trimethylpentane	563.5	572.6
22	Tetrahydrofuran	540.2	501.4	104	2,2,4-trimethylpentane	544.0	573.0
23	<i>n</i> -butyric acid	628.0	599.4	105	2,3,3-trimethylpentane	573.0	572.8
24	Ethyl acetate	523.3	531.5	106	2-ethyl-1-hexanol	640.0	670.3
25	Isobutyric acid	609.2	600.0	107	Quinoline	782.2	731.9
26	Methyl propionate	530.6	531.6	108	Cumene	631.2	622.9
27	<i>n</i> -propyl formate	538.0	531.1	109	<i>o</i> -ethyltoluene	651.2	623.8
28	<i>n</i> -butane	425.2	419.3	110	<i>p</i> -ethyltoluene	640.2	624.0
29	Isobutane	408.1	418.7	111	Mesitylene	637.4	624.6
30	Butanol	562.9	539.9	112	<i>n</i> -propylbenzene	638.4	623.3
31	<i>sec</i> -butyl alcohol	536.0	537.0	113	1,2,3-trimethylbenzene	664.5	624.2
32	<i>tert</i> -butyl alcohol	506.2	538.4	114	1,2,4-trimethylbenzene	649.1	624.4
33	Diethyl ether	466.7	481.7	115	<i>n</i> -propylcyclohexane	639.2	618.2
34	Isobutanol	547.7	535.5	116	3,3-diethylpentane	610.1	602.9
35	<i>n</i> -butylamine	531.9	529.6	117	<i>n</i> -nonane	595.7	604.5
36	Diethylamine	496.6	516.1	118	2,2,3,3-tetramethylpentane	610.9	601.3
37	Pyridine	620.0	604.5	119	1,2,3,4-tetrahydronaphthalene	720.2	662.6
38	Cyclopentane	511.8	486.3	120	<i>n</i> -butylbenzene	660.6	649.3
39	1-pentane	464.8	467.5	121	<i>p</i> -cymene	653.2	649.5
40	Diethyl ketone	561.0	582.5	122	Isobutylbenzene	650.2	649.0
41	2-pentanone	561.1	595.4	123	<i>n</i> -decane	618.2	631.8
42	Ethyl propionate	546.0	564.7	124	<i>n</i> -tetradecane	692.4	725.1
43	Isobutyl formate	551.4	563.8	125	<i>n</i> -octadecane	745.3	801.3
44	<i>n</i> -propyl acetate	549.4	564.9	126	<i>n</i> -nonadecane	755.9	818.5
45	Valeric acid	651.0	632.2	127	<i>sec</i> -butylchloride	520.6	515.3
46	Isopentane	460.4	465.7	128	Methyl isopropyl ether	464.5	481.5
47	Neopentane	433.8	464.8	129	<i>sec</i> -butylamine	514.3	525.4
48	<i>n</i> -pentane	469.7	466.3	130	<i>tert</i> -butylamine	483.9	518.5
49	2-methyl-2-butanol	545.2	575.4	131	2-methyl-1-butene	465.0	468.1
50	3-methyl-1-butanol	579.5	584.3	132	2-methyl-2-butene	471.0	469.4
51	1-pentanol	586.2	584.9	133	3-methyl-1-butene	450.4	466.9
52	Bromobenzene	670.2	631.7	134	Methyl isopropyl ketone	553.0	593.4
53	Chlorobenzene	632.4	606.9	135	Methyl <i>n</i> -butyrate	554.5	565
54	Benzene	562.2	531.1	136	Piperidine	594.1	575.7



**Table 6** (continued)

Number	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K	Number	Compound	$T_{\text{exp}}$ , K	$T_{\text{calc}}$ , K
55	Phenol	694.3	642.4	137	Ethyl propyl ether	500.2	520.4
56	Aniline	699.0	622.0	138	Methyl tert-butyl ether	497.1	519.5
57	2-methylpyridine	621.0	636.2	139	3-methylpyridine	645.0	643.9
58	Cyclohexene	560.4	527.4	140	1,5-hexadiene	507.0	508.8
59	Cyclohexanone	629.2	651.3	141	Hexanenitrile	622.1	651.6
60	Cyclohexane	553.5	525.5	142	3-hexanone	582.8	618.9
61	1-hexene	504.0	507.8	143	Ethyl isobutyrate	553.2	594.7
62	Methylcyclopentane	532.8	524.2	144	<i>n</i> -propyl propionate	578.0	595.1
63	Cyclohexanol	625.2	631.1	145	1-hexanol	611.4	618.9
64	2-hexanone	587.1	627.9	146	2-hexanol	586.2	616.4
65	Methyl isobutyl ketone	571.4	629.2	147	4-methyl-2-pentanol	574.4	617.4
66	Ethyl <i>n</i> -butyrate	571.0	595.2	148	Acetone	508.2	517.5
67	Isobutyl acetate	561.0	594.9	149	2-propanol	508.3	495.9
68	2,2-dimethylbutane	488.8	505.3	150	Ethyl vinyl ether	475.2	483.7
69	2,3-dimethylbutane	500.0	505.6	151	Isovaleric acid	634.0	634.1
70	<i>n</i> -hexane	507.4	506.8	152	<i>n</i> -butyl acetate	579.2	595.2
71	2-methylpentane	497.5	506.3	153	<i>p</i> -cresol	704.7	673.0
72	3-methylpentane	504.4	506.1	154	<i>p</i> -toluidine	693.2	657.4
73	Diisopropyl ether	500.1	553.9	155	1-heptene	537.3	543.5
74	Di- <i>n</i> -propyl ether	530.6	554.8	156	Methylcyclohexane	572.2	559.1
75	Diisopropylamine	523.1	585.0	157	2,2,3-trimethylbutane	531.2	540.5
76	Di- <i>n</i> -propylamine	555.8	583.9	158	<i>o</i> -xylene	630.4	596.0
77	Benzaldehyde	695.0	688.2	159	1-octene	578.2	575.8
78	Toluene	591.8	565.1	160	<i>m</i> -methyltoluene	637.2	596.1
79	<i>m</i> -cresol	705.9	678.0	161	<i>p</i> -diethylbenzene	658.0	649.9
80	<i>o</i> -cresol	697.6	676.7	162	1-decene	617.1	632.5
81	<i>m</i> -toluidine	709.2	654.3	163	Propionitrile	564.4	531.0
82	<i>o</i> -toluidine	694.2	654.5	164	4-methylpyridine	646.2	637.7
				165	<i>n</i> -pentyl formate	576.0	594.6

**Fig. 2** The predicted vs experimental critical temperatures according to the best two-descriptor equation (Table 5) and new AM1 parameters



## Conclusions

The original AM1 quantum-chemical parameterization was improved to be more adequate for developing molecular descriptors and the related QSPR models in the liquid phase. A nonlinear optimization technique based on the Levenberg–Marquardt algorithm was used to optimize 17 parameters for the core–core repulsion function and one-electron resonance integrals for H, C, O, Cl, N and Br atoms. The results of the QSPR treatment of normal boiling points demonstrate that the new parameters describe the intermolecular interactions in liquid phase better. The descriptors appearing in the two-parameter QSPR equation are connected with dispersion and cavity formation processes, hydrogen bonding ability of the compounds and electrostatic intermolecular interactions in liquid media.

The new set of parameters also enabled the better prediction of the normal boiling points of nine simple inorganic compounds as compared to the original QSPR model.

The results obtained for the critical temperature of 165 organic compounds indicate that similar processes govern this property as of the boiling points. A two-descriptor original QSPR equation was improved by using the re-optimized parameters  $\alpha$ ,  $\beta_s$  and  $\beta_p$  from the treatment of boiling points. The fact that the new QSPR model involves two similar descriptors ( $\sqrt[3]{G_I}$  and HASA(2)) reveals that the dispersion and cavity formation phenomena as well as the hydrogen bonding ability are the most important interactions determining the critical temperatures of compounds. A more extensive validation of the new AM1 parameterization will be carried out on a series of other properties of compounds in the liquid phase (density, vapor pressure, distribution coefficients etc.).

**Acknowledgements** This work was supported by a grant from the EU Research and Training Network: (IMAGETOX, Contract No. HPRNCT-1999-00015). Support from Estonian Science Foundation is also kindly acknowledged (Grant No. 4548).

## References

1. Karelson M, Lobanov VS, Katritzky AR (1996) *Chem Rev* 96:1027–1043

2. Karelson M (2000) *Molecular Descriptors in QSAR/QSPR*. Wiley, New York
3. Todeschini R, Consonni V (2000) *Handbook of Molecular Descriptors*. Wiley, Weinheim
4. Karelson M (2004) Chapter 24 In: Tollenaere JP, Bultnick P, de Winter H, Langenaeker W (eds) *Computational medicinal chemistry and drug discovery*. Dekker, New York, pp 641–668
5. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107:3902–3909
6. Stewart JJP (1989) *J Comput Chem* 10:209–220
7. Tomasi J, Persico M (1994) *Chem Rev* 94:2027–2094
8. Karelson M (1997) *Adv Quant Chem* 28:141–157
9. Cramer CJ, Truhlar DG (1999) *Chem Rev* 99:2161–2200
10. Orozco M, Luque FJ (2000) *Chem Rev* 100:4187–4225
11. Klamt A, Eckert F (2000) *Fluid Phase Equil* 172:43–72
12. Jorgensen S, Ratner MA, Mikkelsen KV (2001) *J Chem Phys* 115:3792–3803
13. Curutchet C, Cramer CJ, Truhlar DG, Ruiz-Lopez MF, Rinaldi D, Orozco M, Luque FJ (2003) *J Comput Chem* 24:284–297
14. Luque FJ, Curutchet C, Munoz-Muriedas J, Bidon-Chanal A, Soteras I, Morreale A, Gelpi JL, Orozco M (2003) *Phys Chem Chem Phys* 5:3827–3836
15. Katritzky AR, Lan M, Karelson M (1996) *J Chem Inf Comput Sci* 36:1162–1168
16. Karelson M, Sild S, Maran U (2000) *Mol Simulat* 24:229–242
17. Huang MJ, Leszczynski J (2001) *J Comput-Aid Mol Design* 15:323–333
18. Borges EG, Takahata Y (2002) *Quim Nova* 25:1061–1066
19. Frecer V, Kabelac M, de Nardi P, Priel S, Miertus S (2004) *J Mol Graph Modell* 22:209–220
20. Padmanabhan J, Parthasarathi R, Sarkar U, Subramanian, V, Chattaraj PK (2004) *Chem Phys Lett* 383:122–128
21. Shriner RL, Curtin DY, Fuson RC, Morrill TC (1980) *The systematic identification of organic compounds*, 6th edn. Wiley, New York
22. Katritzky AR, Mu L, Lobanov VS, Karelson M (1996) *J Phys Chem* 100:10400–10407
23. Stanton DT, Egolf LM, Jurs PC, Hicks MG (1992) *J Chem Inf Comput Sci* 32:306–316
24. Katritzky AR, Lobanov VS, Karelson M (1995) *CODESSA Training Manual*. University of Florida
25. Katritzky AR, Lan M, Karelson M (1998) *J Chem Inf Comput Sci* 38:293–299
26. Dewar MJS, Thiel W (1977) *J Am Chem Soc* 99:4899–4907
27. Stewart JJP (1989) *MOPAC 6.0 QCPE No. 455*
28. <http://www.codessa-pro.com>
29. Moore JJ (1977) *The Levenberg–Marquardt Algorithm: Implementation and Theory, Numerical Analysis*. In: Watson GA (ed) *Lecture Notes in Mathematics* 630. Springer, pp 105–116