Use of the DIPPR Database for Development of Quantitative Structure−Property Relationship Correlations: Heat Capacity of Solid Organic Compounds[†]

Benjamin T. Goodman, W. Vincent Wilding, John L. Oscarson, and Richard L. Rowley*

Department of Chemical Engineering, 350 CB, Brigham Young University, Provo, Utah 84602

Two group-contribution methods have been developed to estimate heat capacities of organic solids at ambient pressure. The power-law (PL) method utilizes an empirical temperature dependence, while the partition function (PF) method is based on the Einstein−Debye partition function for crystals with a modified frequency distribution function. Both methods have a fixed temperature functionality but utilize group contributions to obtain the compound-specific constants in the predictive equations. The training set for the group-contribution correlations consisted of 455 compounds and 7967 heat capacity data points. Both methods can be used for temperatures above 50 K, and they correlate the training set to within approximately 9.4 $J\cdot mol^{-1}\cdot K^{-1}$, corresponding to an average deviation of 6.8% for the PL method and 8.0% for the PF method. The PL method gives better results at lower temperatures, and the PF method at higher temperatures. Tests on the methods' extrapolation capabilities suggest that at 298 K they have comparable accuracy to two currently available methods that are applicable only at 298 K, but the new methods can be used to estimate solid heat capacities over a wide range of temperatures with an average expected accuracy of approximately 13%.

Introduction

In the interest of completeness, the Design Institute for Physical Property Data (DIPPR) 801 database[1] provides predicted values for physical properties when experimental data are not available. Predicted values are useful estimates that can be very valuable to an engineer for screening and initial design. While accurate prediction methods are available for most of the properties included in the database, a current void in solid property prediction methods is reflected in the number of compounds in the DIPPR database with no estimate of solid properties. The DIPPR database includes temperature-dependent correlations (at 1 atm) for heat capacity (SCP), vapor pressure (SVP), and thermal conductivity (STC) of solids whenever possible. Of the over 1800 compounds in the database, 28% have no SCP data and 38% have a single value listed (usually at the triple point). The situation is much worse for SVP and STC: 87% of the compounds have no SVP data and 96% have no STC values. The purpose of this study was to address the paucity of solid property estimation methods by developing a reliable correlation for SCP as a function of temperature, based principally on the structure of the molecule.

Databases in which the experimental data are evaluated for quality are essential for development of data correlations for use in estimation and prediction of physical properties. An evaluated database provides a reliable training set of values that can be used to determine initially the independent properties and variables that strongly correlate with the property to be estimated and then to regress values for the coefficients in the new correlation. Data from the database not used in the development of a correlation can also be used for testing new correlations or extending current correlations. This rationale has fostered our use of the DIPPR database for development of new prediction methods using QSPR (quantitative structure−property relationship) correlations. Previously, we have used this methodology to develop prediction methods for the normal boiling point[2] and surface tension.[3] In this study, we apply the same methodology to SCP.

While there are methods for predicting SCP for ionic compounds as a function of temperature,[4] current methods for organic solids are yet somewhat limited. For example, both a method based on the sum of elemental contributions[5] and a method based on more complex second-order functional group contributions[6] are available for estimating SCP at a single temperature, 298 K. A low-temperature method for predicting SCP of alkanes, alkenes, alkanols, and alkanones at 10 K increments from 10 to 150 K is also available.[7] This method is quite accurate for the compounds and temperature range for which it has been developed. Recently the method has been modified and extended to alkyl[8] and phenyl[9] derivatives of urea. Here we present two group-contribution methods for estimating the heat capacity of crystalline organic compounds above 50 K.

Temperature Dependence of SCP

A simple, empirical, power-law form for the temperature dependence of SCP,

$$C_P = AT^m \tag{1}$$

has been used previously for solid hydrocarbons.[10] In eq 1, C_P is heat capacity, T is temperature, and A and m are empirical coefficients with m less than 1. To develop a simple, first-order prediction method, we applied this same

---

[†] This contribution will be part of a special print edition containing papers reporting experimental results from the various projects of the Design Institute for Physical Properties of the American Institute of Chemical Engineers.

* To whom correspondence should be addressed. E-mail: rowley@byu.edu.

10.1021/je025656h CCC: $27.50

temperature functionality to all solid organic compounds and treated m as a universal constant. We then used group contributions to obtain compound-specific values for A. Correlation of m for various chemical families or for individual compounds would be a logical starting point for development of a second-order method. However, we felt that at this time the limited number of experimental SCP data was inadequate to simultaneously correlate A and m for individual compounds.

To find an optimum value of m, 455 compounds with a total of 7967 data points at various temperatures were extracted from the DIPPR database (having a DIPPR quality code of estimated accuracy better than 5%) and used to optimize m and regress individual values for A in eq 1. The optimum m in a least-squares sense was found to be 0.793. We therefore use

$$C_P/(\mathrm{J \cdot mol^{-1} \cdot K^{-1}}) = \frac{A/(\mathrm{J \cdot kmol^{-1} \cdot K^{-1}})}{1000}\left(\frac{T}{\mathrm{K}}\right)^{0.793} \qquad (2)$$

as the starting point for development of the first of our two predictive methods, which we designate as the power-law (PL) method.

The second method developed in this work is based on the Einstein canonical partition function, Q. In the Einstein theory, atomic motions within the solid crystal are modeled as vibrations relative to equilibrium crystalline lattice positions. Expressed in terms of normal frequencies, ν, and the zero of energy for the crystal with all atoms at their equilibrium lattice sites, U_0, the partition function is[11]

$$\ln Q = -\frac{U_0}{kT} - \int_0^{\infty}\left[\ln(1 - \mathrm{e}^{-h\nu/kT}) + \frac{h\nu}{2kT}\right]g(\nu)\,\mathrm{d}\nu \qquad (3)$$

where k is Boltzmann's constant, h is Planck's constant, and $g(\nu)\,\mathrm{d}\nu$ gives the number of normal frequencies between ν and $\nu + \mathrm{d}\nu$. In this model, the partition function and consequently the thermodynamic properties of the crystal are determined by the choice of the frequency distribution function, $g(\nu)$. For crystals of atomic species, the frequency distribution must be constrained to $3N$ total normal frequencies, where N is the number of atoms in the crystal; that is,

$$\int_0^{\infty} g(\nu)\,\mathrm{d}\nu = 3N \qquad (4)$$

It is convenient to use dimensionless frequency, x, and characteristic temperature, Θ, in the SCP expression. These frequency variables are defined by

$$x = \frac{h\nu}{kT} = \frac{\Theta}{T} \qquad (5)$$

The molar heat capacity obtained from eq 3 through standard thermodynamic identities, expressed in terms of dimensionless frequency, is[11]

$$C_P = k\int_0^{\infty}\frac{x^2\mathrm{e}^x g(x)}{(\mathrm{e}^x - 1)^2}\,\mathrm{d}x \qquad (6)$$

where it has been assumed that constant pressure and constant volume heat capacities are approximately equivalent for solids.

The form of the temperature dependence for SCP is thus determined by the distribution function model used to represent the internal frequencies in the crystal. Einstein chose to set all $3N$ frequencies to an identical value of Θ_E, or

$$g(x) = 3N\delta(x - x_E) \qquad (7)$$

where $\delta(x)$ is the Dirac delta function and x_E is related to Θ_E by the definition shown in eq 5. By substituting eq 7 into eq 6, one obtains for the molar SCP

$$C_P = \frac{3R{x_E}^2\mathrm{e}^{x_E}}{(\mathrm{e}^{x_E} - 1)^2} \qquad (8)$$

where R is the gas constant. To improve the performance of the Einstein theory, Debye chose

$$g(x) = \begin{bmatrix} 9N{x_D}^{-3}x^2 & 0 \le x \le x_D \\ 0 & x > x_D \end{bmatrix} \qquad (9)$$

for the frequency distribution, where x_D is defined analogously to x_E but now in terms of Θ_D, the so-called Debye temperature at which the frequency distribution is truncated to conserve modes (eq 4). Using eq 9 in eq 6, one obtains the Debye equation for the molar SCP:

$$C_P = 9R{x_D}^{-3}\int_0^{x_D}\frac{x^4\mathrm{e}^x}{(\mathrm{e}^x - 1)^2}\,\mathrm{d}x \qquad (10)$$

Debye's choice for the vibrational frequency distribution function assumes that the quadratic temperature dependency, known to be true near absolute zero, can be used over the whole temperature domain. However, vibrational frequency distributions vary significantly at higher temperatures and are quite complex.

Here, we relax the quadratic constraint on the frequency distribution used by Debye but retain the power-law relationship shown in eq 9 in the form

$$g(x) = \begin{bmatrix} 3(r+1)NN_a{x_G}^{-(r+1)}x^r & 0 \le x \le x_G \\ 0 & x > x_G \end{bmatrix} \qquad (11)$$

where r is a real number greater than -1. Again, the frequency distribution is truncated at a generalized characteristic temperature, Θ_G, to conserve the total number of modes. In our application of the Einstein theory to complex organic solids, we consider multiatom molecules such that the number of modes is now $3NN_a$ where N is the number of molecules and N_a is the number of atoms per molecule. This approach hybridizes the intra- and intermolecular vibrations within the crystal. In essence, we extend the Einstein model, strictly applicable to monatomic species, to multiatom molecules by viewing the crystal as a collection of atoms located at lattice sites. Spacing between lattice sites occupied by bonded atoms is obviously closer than that between nonbonded atoms, but we treat all vibrations, inter- or intramolecular, as part of a continuous distribution of vibrational frequencies, $g(x)$. One would expect that inclusion of intramolecular vibrations in $g(x)$ may produce a substantially different distribution than that used by Einstein or Debye for monatomic species.

The molar SCP obtained from the vibrational distribution given in eq 11 is

$$C_P = 3N_aR(r+1){x_G}^{-(r+1)}\int_0^{x_G}\frac{x^{r+2}\mathrm{e}^x}{(\mathrm{e}^x - 1)^2}\,\mathrm{d}x \qquad (12)$$

Equation 12 reduces to the Debye equation (eq 10) for monatomic species when $r = 2.0$; likewise, it reduces to the Einstein equation (eq 8) for monatomic species in the limit as r approaches infinity. As we did with the simple empirical equation, we assume that the temperature functionality is the same for all compounds and treat r as a universal constant. This again reduces the SCP equation to an equation in a single variable, Θ_G, which we corrrelate in terms of group contributions using the DIPPR database. The optimum value determined for r was -0.15. The starting point for developing a second predictive equation for SCP, analogous to eq 2, is therefore

$$C_P = 2.55\ N_a R x_G^{-0.85} \int_0^{x_G} \frac{x^{1.85} e^x}{(e^x - 1)^2}\, dx \qquad (13)$$

The units for C_P are determined by the choice of the gas constant R. We designate eq 13 as the partition function (PF) method for calculating SCP. To utilize either of these two methods, PL or PF, for SCP prediction, correlations for the constants A and Θ_G that appear in eqs 2 and 13, respectively, must be developed.

Prediction Equations Based on Group Contributions

By our choice of m in eq 2, we have knowingly deviated from the known T^3 low-temperature limit for SCP. Similarly, the functional form of eq 11 with $r \neq 2$ deviates from the T^3 low-temperature limit. In essence, we have sacrificed agreement at very low temperatures to provide a simple equation that optimizes predictive capability over the temperature range of interest in most engineering applications. The training set that we have used to obtain group contributions for A in eq 2 and Θ_G in eq 13 is a subset of the original data set and includes 455 compounds and 7967 SCP data points.

For compounds that have more than one solid phase, the crystalline phase stable at the lowest temperature was chosen. This lowest-temperature phase is the most influenced by the interactions of functional groups, and it is consistent with Bondi's definition for a "standard" heat of sublimation. Bondi uses the lowest first-order phase transition as the standard because "solids above this transition often exhibit sharply reduced lattice energy and are unrepresentative of the 'typical' solid."[12] Because we have developed the correlations based only on this "standard" phase, it should not be applied to other crystalline phases.

The functional group definitions chosen are similar to those used in the Joback method for boiling points.[13] These group definitions are common, simple to use, and available in many automated software prediction packages. A limited QSPR analysis indicated that group contributions adequately correlated the values of A in eq 2. However, the QSPR analysis indicated that the radius of gyration, RG, was statistically significant in addition to group contributions for the correlation of Θ_G values in eq 13. In addition to the standard Joback groups, quadratic terms were found to be statistically important for the two most common groups: methylene and aromatic carbon groups. A correction for multiple halogen groups was also necessary and is included as a correction term based on the fraction of C or Si terminal valences occupied by halogen atoms. The final correlations obtained were

$$A/(\mathrm{J{\cdot}kmol^{-1}{\cdot}K^{-1}}) = \exp\left(6.7796 + \sum_i^{NG} a_i n_i + \sum_i^{NG} b_i n_i^2\right) \qquad (14)$$

$$\Theta_G/\mathrm{K} = 1886.2 + 3.3626 \times 10^{12}\mathrm{RG} + \sum_i^{NG} \alpha_i n_i + \sum_i^{NG} \beta_i n_i^2 + \sum_i^{NG} \gamma_i \frac{n_i}{n_X} \qquad (15)$$

where a_i, b_i, α_i, β_i, and γ_i are values for group i regressed from the training set, n_i is the number of times that group i appears in the molecule, n_X is the total number of all halogen and hydrogen atoms attached to C and Si atoms in the molecule, NG is the total number of groups in the molecule, and RG is the radius of gyration of the molecule in meters. Values of the radius of gyration are obtainable from several sources including the DIPPR 801 database. These equations should not be used for temperatures below 50 K for reasons mentioned above.

The parameters for eqs 14 and 15 were obtained using the multiple regression package in Oxford Molecular Tsar 3.2.[14] Tables 1 and 2 contain the values of the group contributions obtained. Linear group terms are given in Table 1; the nonlinear terms for methylene and aromatic carbon groups and the correction terms for the halogen fractions are given in Table 2. Table 1 also illustrates group definitions. The designated group is highlighted with a bold typeface in the SMILES formula[15] for the compound. SMILES (Simplified Molecular Input Line Entry Specification) is a simple in-line chemical notation for the structure of a compound. SMILES formulas are compiled in the DIPPR database and are very convenient for software that automates the parsing of molecular structure into groups. For example, DIADEM,[16] which is a front-end interface and analysis tool for the DIPPPR 801 database, utilizes a structure parser based on the SMILES formulas stored in the DIPPR 801 database. Simple SMILES tutorials can be found in our previous work[2] or on the world wide web.[17] Included as an appendix to this paper are computational examples that illustrate the group definitions, the use of eqs 14 and 15, and the procedures for obtaining SCP.

We find it useful to evaluate correlation of the training set data in terms of an average absolute percent deviation (AAPD), the standard deviation (SD), and the average absolute deviation (AAD). As SCP values may range over more than an order of magnitude, these three measures provide useful statistics in different parts of the temperature range. AAPD emphasizes absolute errors in the region where the magnitude of SCP is small, at low temperatures, while SD weights more heavily the larger absolute errors expected at higher temperatures where SCP is larger.

Table 3 shows the AAPD, AAD, and SD results of the correlation for the training sets used to obtain the values in Tables 1 and 2. Fractional deviations of the correlated values from the training set values are shown for the PL method (eqs 2 and 14) in Figure 1. The PL method performs well at lower temperatures, but there is a negative bias to the residuals at higher temperatures. A similar plot for the PF method (eqs 13 and 15) is shown in Figure 2b. There is no noticeable bias in this correlation. Most of the deviation from experimental values is due to inadequacies in the assumed temperature dependence of the model, that is, the assumption of a universal constant m in eq 1 and r in eq 12. This can be observed by comparing parts a and b of Figure 2. Figure 2a utilizes the values of Θ_G regressed from the experimental data using eq 13 rather than the values correlated in terms of groups using eq 15. Little degradation of prediction values occurs in the estimation of Θ_G from the group-contribution correlation. Table 3 indicates that

Table 1. Linear Group Values for Equations 14 and 15

group	description	example	SMILES formula	*a* (eq 14)	α (eq 15)
$-CH_3$	methyl	*n*-butane	**C**CCC	0.201 84	−241.7
$>CH_2$	methylene	1-heptanol	OC**C**CCCCC	0.116 44	17.929
>CH−	secondary carbon	2,3-dimethylpentane	C**C**(C)C(C)CC	0.030 492	229.47
>C<	tertiary carbon	2,2-dimethylbutane	C**C**(C)(C)CC	−0.040 64	529.76
$CH_2=$	terminal alkene	1-octene	**C**=CCCCCCC	0.185 11	−387.3
−CH=	alkene	1,3-butadiene	C=**C**C=C	0.112 24	−118.89
>C=	substituted alkene	isobutene	C=**C**(C)C	0.028 794	191.3
=C=	allene	1,2-butadiene	C=**C**=CC	0.053 464	−154.12
#CH	terminal alkyne	ethylacetylene	CCC#**C**	−0.029 14	−538.85
#C−	alkyne	dimethylacetylene	C**C**#CC	0.132 98	−225.13
Ar −CH=	aromatic carbon	biphenyl	c1**c**cccc1(c2ccccc2)	0.082 478	−36.615
Ar >C=	substituted aromatic C	toluene	c1ccccc**c**1(C)	0.012 958	148.32
Ar −O−	furan oxygen	furan	C1=C**O**C=C1	0.066 027	−70.693
Ar −N=	pyridine nitrogen	quinoline	c1(ccc**n**2)c2cccc1	0.056 641	−229.57
Ar >N−	substituted pyrrole N	*N*-methylpyrrole	**n**1(C)cccc1	0.008 938	215.18
Ar −NH−	pyrrole nitrogen	pyrrole	C1=CC=C**N**1	−0.052 46	178.85
Ar −S−	thiophene sulfur	thiophene	**S**1C=CC=C1	0.090 926	−492.78
−O−	ether	dimethyl ether	C**O**C	0.064 068	−154.96
−OH	alcohol	1-pentanol	CCCCC**O**	0.103 41	−286.75
−COH	aldehyde	1-butanal	CCC**C=O**	0.156 99	−451.8
>C=O	ketone	3-hexanone	CC**C(=O)**CCC	0.129 39	−252.22
−COO−	ester	methyl methacrylate	C=C(C)**C(=O)O**C	0.136 86	−530.27
−COOH	acid	*n*-butyric acid	CCC**C(=O)O**	0.210 19	−498.54
−COOCO−	anhydride	maleic anhydride	**O1C(=O)**C=CC1**(=O)**	0.330 91	−1321.5
$-CO_3-$	carbonate	ethylene carbonate	C1**OC(=O)O**C1	0.251 7	−639.94
$-NH_2$	primary amine	methylamine	C**N**	0.056 138	−53.298
>NH	secondary amine	piperidine	C1CCCC**N**1	−0.007 17	363.75
>N−	tertiary amine	trimethylamine	C**N**(C)C	−0.016 61	377.78
=NH		dicyandiamide	N#CNC**(=N)**N	0.176 89	−568.75
#N	nitrile	acetonitrile	CC**#N**	0.015 355	−515.66
−N=N−	diazide	*p*-aminoazobenzene	Nc1ccc(cc1)**N=N**c2ccccc2	0.368 7	−761.63
$-NO_2$	nitro	nitrobenzene	c1(**N(=O)=O)**ccccc1	0.233 27	−619.91
−N=C=O	isocyanate	phenyl isocyanate	c1**(N=C=O)**ccccc1	0.269 8	−703.05
−SH	thiol/mercaptan	*n*-hexyl mercaptan	CCCCCC**S**	0.211 23	−594.12
−S−	sulfide	diethyl sulfide	CC**S**CC	0.142 32	−391.13
−SS−	disulfide	di-*n*-propyl disulfide	CCC**SS**CCC	0.314 57	−734.81
=S	sulfur double bond	thiourea	NC**(=S)**N	0.137 53	−949.61
>S=O	sulfoxide	dimethyl sulfoxide	C**S(=O)**C	0.040 002	−251.27
−F	fluoride	benzotrifluoride	c1(C(F)**(F)**F)ccccc1	0.155 11	−320.76
−Cl	chloride	ethyl chloride	CC**[Cl]**	0.169 95	−429.06
−Br	bromide	bromobenzene	c1**(Br)**ccccc1	0.191 12	−70.347
−I	iodide	iodobenzene	c1**(I)**ccccc1	0.113 18	−589
>Si<	silane	tetramethylsilane	C**[Si]**(C)(C)C	0.122 13	140.96
>Si(O−)−	siloxane	hexamethyldisiloxane	C[Si](C)(C)**O[Si]**(C)(C)C	0.101 25	77.804
cyc >Si(O−)−	cyclic siloxane	octamethylcyclotetrasiloxane	[Si]1(C)(C)**O[Si]**(C)(C)O-[Si](C)(C)O[Si](C)(C)O1	0.063 438	77.804
$P(=O)(O-)_3$	phosphate	triphenyl phosphate	c1ccccc1(**O[P](=O)(O**c2ccccc2)**O**c3ccccc3)	0.150 16	−520.71
>P−	phosphine	triphenylphosphine	**P**(c1ccccc1)(c2ccccc2)(c3ccccc3)	0.069 602	489.97
>P(=O)−	phosphine oxide	triphenylphosphine oxide	**P(=O)**(c1ccccc1)(c2ccccc2)(c3ccccc3)	0.218 75	−242.12

Table 2. Nonlinear and Halogen Group Values for Equations 14 and 15

A. Nonlinear Terms

group	description	eq 14, *b*	eq 15, β
$-CH_2$	methylene	−0.001 88	−2.9045
Ar =CH−	aromatic carbon	−0.000 33	−2.9616

B. Halogen Fraction Terms

group	description	eq 15, γ
−Cl	Cl fraction	−1361.4
−F	F fraction	−1231.3
−Br	Br fraction	−3864.5

Table 3. Quality of Training Set Correlation

	PL method	PF method
training set compounds	455	455
training set SCP values	7967	7967
AAPD/%	6.84	7.96
AAD/(J·mol^{-1}·K^{-1})	9.30	9.43
SD/(J·mol^{-1}·K^{-1})	19.2	16.5

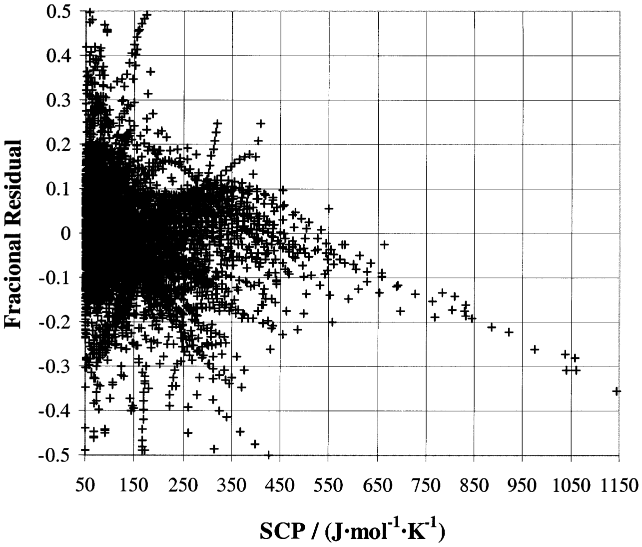

Figure 1. Fractional deviations of correlated SCP values from the training set values when correlated with the PL method.

the PL equations correlate the data overall slightly better, but the SD for the PF method is lower, suggesting that it may be preferred at higher temperatures (above 250 K). The high-temperature bias noted in Figure 1 is consistent with this suggestion. The lower SD may also suggest that the PF method is less susceptible to larger errors. We note,

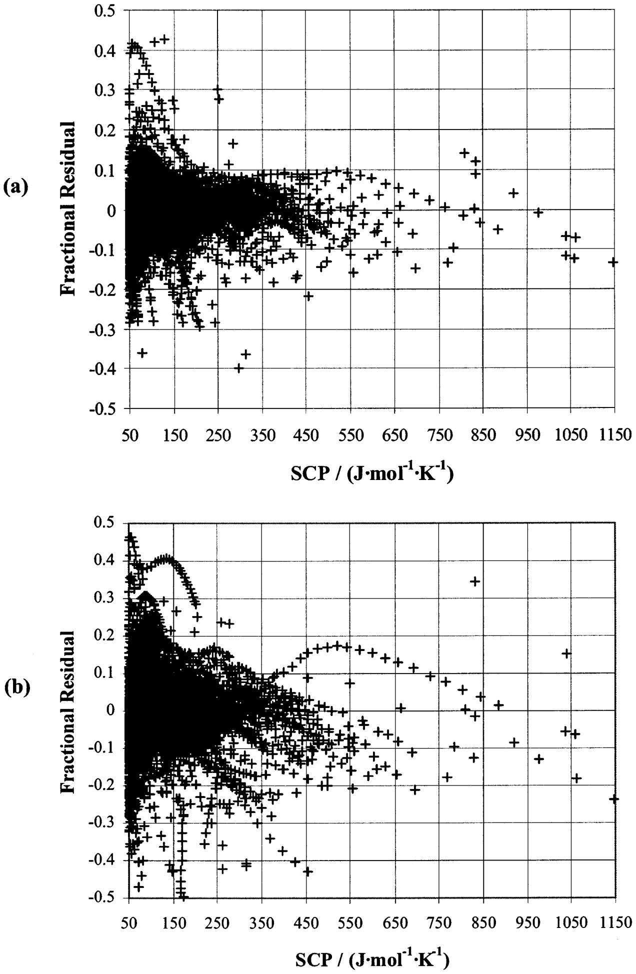

Figure 2. Fractional deviations of SCP values from the training set values when the PF method (eq 13) is used with (a) regressed values of Θ_G and (b) correlated values of Θ_G using eq 15.

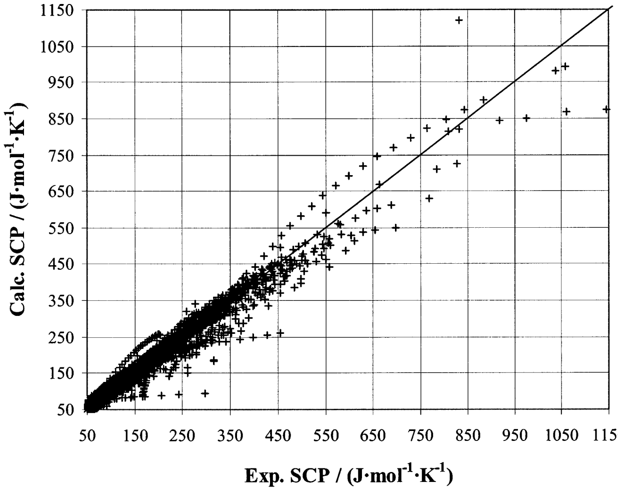

Figure 3. Comparison of correlated and experimental SCP values obtained in the regression of eq 15 from the training set for the PF method.

for example, that some diester compounds produced a noticeably larger error with PL than with PF.

Discussion of the Correlations

Figure 3 shows the overall correlation for the PF method in terms of absolute deviation of the correlated SCP values from the accepted experimental data in the DIPPR 801 database. There is no apparent bias for the PF method, and the correlation reproduces reasonably well the experimental data for the compounds in the training set. Example comparisons of SCP values predicted by the PL and PF methods to experimental data for three compounds are shown in Figure 4.

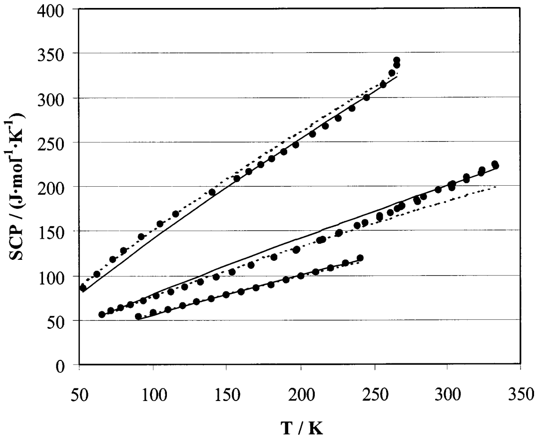

Figure 4. Comparison of experimental (●) SCP data to values predicted using the PL (dashed line) and PF (solid line) methods for *n*-decylcyclohexane (top series), biphenyl (middle series), and iodobenzene (bottom series).

Table 4. Comparison of Predicted SCP Values to Those from a Test Set

T	no. of compds	no. of SCP values	statistic	PL	PF
>50 K	45	948	AAPD/%	13.0	20.7
			AAD/(J·mol^{-1}·K^{-1})	25.2	21.9
			SD/(J·mol^{-1}·K^{-1})	54.2	48.1
>50 K & <250 K	45	788	AAPD/%	12.8	22.6
			AAD/(J·mol^{-1}·K^{-1})	19.5	19.5
			SD/(J·mol^{-1}·K^{-1})	42.3	49.0
>250 K	22	160	AAPD/%	14.3	11.5
			AAD/(J·mol^{-1}·K^{-1})	53.2	33.3
			SD/(J·mol^{-1}·K^{-1})	92.9	43.8

To test the extrapolative capability of the correlations developed here, the PL and PF methods were used to predict SCP values for 45 compounds available in the TRC handbooks[18] but not in the DIPPR 801 database from which the training set was developed. The results of this test are shown in Table 4. While the AAPD is approximately 8% for correlation of the training set, the test set results suggest an expected average accuracy for new predictions of about 13%. Unfortunately, we do not know the accuracy of the experimental values compiled in the TRC handbook for the test set compounds, but we do not expect experimental error to be a large component of this error. We have also divided the comparison in Table 4 into two different temperature ranges. In the low-temperature region, defined as 50 K < *T* < 250 K, the PL method is a better predictor of SCP in terms of percent error than the PF method with AAPD values of 12.8% and 22.7%, respectively. However, in terms of absolute deviations, the PF method is equivalent to the PL method in the low-temperature range but superior overall and particularly at higher temperatures. The PF method is a better predictor of SCP above 250 K with an AAPD of 11.5% compared to 14.3% for the PL method, and we recommend this method for predictions above 250 K.

It is useful to compare the two methods developed here with other available methods. The intent of this comparison of methods is not to identify one technique as superior but

Table 5. Comparison of Predictive Methods at 298 K

method	no. of compds	AAPD %	AAD $J \cdot mol^{-1} \cdot K^{-1}$	SD $J \cdot mol^{-1} \cdot K^{-1}$
A. Common Test Set				
Domalski–Hearing	83	8.95	20.5	38.2
modified Kopp	83	7.13	20.6	32.3
PL method	83	9.53	26.9	43.6
PF method	83	7.72	23.8	43.9
B. Larger Test Set				
modified Kopp	127	9.52	23.4	37.5
PL method	127	10.7	26.4	41.6
PF method	127	8.47	22.3	40.3

Table 6. Comparison of Predictive Methods from 50 K to 150 K

method	compds	points	AAPD %	AAD $J \cdot mol^{-1} \cdot K^{-1}$	SD $J \cdot mol^{-1} \cdot K^{-1}$
A. Alkanes					
Kabo et al. no. 1	46	627	3.34	4.57	10.7
PL method	46	627	5.14	6.01	11.2
PF method	46	627	4.69	5.78	11.0
B. Alkanes, Alkenes, Alkanols, and Alkanones					
Kabo et al. no. 2	87	962	3.33	3.45	10.7
PL method	87	962	5.54	5.51	9.89
PF method	87	962	5.64	5.63	9.99

rather to identify the classes of compounds and temperature domain most appropriate for each method. Kopp's rule[5] uses atomic increments to SCP with unique values for each element. It can be applied to virtually any organic compound because of its simplicity, but it is valid only at 298 K. The Domalski–Hearing[6] method uses a more complex second-order group-contribution method which theoretically should be more accurate because its more specific group definitions account for neighbor effects, but it is also only applicable at 298 K. We have compared estimations from these two methods to those made at 298 K using the PL and PF methods in Table 5. The comparison test set included 127 compounds obtained from the DIPPR 801 database for compounds with experimental data within 10 K of 298 K. Although these compounds were in our training set, it is highly likely that they were also in the training set used to obtain the Domalski–Hearing group values owing to the limited SCP data available. The test set for part A of Table 5 is a subset of the available data which includes 83 compounds for which Domalski–Hearing group values are available; part B shows the results for all 127 compounds for the three methods that can be used for this entire test set. The PL and PF methods are of comparable accuracy to Kopp's rule and the Domalski–Hearing method at 298 K, but the new methods may be used over a large temperature range. With respect to the domain of compounds to which the methods can be applied, Kopp's rule is the most universal with the PL, PF, and Domalski–Hearing methods all about equivalent.

Currently the only other temperature-dependent methods of which we are aware are the simple additive correlations by Kabo et al.[7] for specific classes of solid organic molecules. These methods utilize constants compiled at 10 K increments from 10 to 150 K. One such method is available for estimating the heat capacity of solid alkanes based upon the number of adjacent carbon atoms to each carbon center. A second correlation applicable to alkanes, alkenes, alkanols, and alkanones is based upon "effective bonds", where each carbon–carbon bond pair (influenced by its neighbors) is the basic additive unit. A high degree of accuracy is claimed for these methods, and our calculations shown in Table 6 support these methods as very accurate for the substances to which they apply. When applied specifically to these families and to temperatures below 150 K, the PL and PF methods are not as accurate. However, the PL and PF methods can be used for a much broader range of compounds and are generally applicable at much higher temperatures.

Summary

Two group-contribution methods have been developed to predict heat capacities of organic solids at ambient pressure. The PL method utilizes an empirical temperature dependence based on a power-law expression observed for solid hydrocarbons and ionic crystals. The PF method is based on the Einstein partition function for crystals and the Debye idea of using a temperature-dependent vibration distribution function. Whereas Debye used a quadratic temperature dependence, we have allowed the power to which the temperature in the frequency distribution is raised to be optimized for the whole training set of organic compounds available to us from the DIPPR 801 database. Both methods then have a fixed or universal temperature functionality. We viewed this as a necessary constraint at this time because of the relatively small amount of SCP data available. Results suggest that some flexibility in this temperature dependence for families of compounds, perhaps correlation of it with molecular descriptors, might be an avenue for improvement of the methods when additional data are available. We have correlated the compound-specific constants in the PL and PF methods primarily in terms of first-order structural group contributions, but radius of gyration was also found to be a significant correlating property for the PF method.

The methods developed in this study fill an important gap in predictive capabilities for organic solid properties. Both methods correlate the training set within an AAPD of about 8%. From the limited evaluations on other test data sets that we were able to perform, we estimate an average accuracy of approximately 13% for the two methods. However, the PL method is expected to be slightly more accurate than this at lower temperatures. The PF method, on the other hand, is expected to have this accuracy or better at temperatures above 250 K, with reduced accuracy below 250 K. Neither method is recommended below 50 K. We recommend use of the simpler PL method for temperatures between 50 and 250 K or when quick estimates are needed, but use of the PF method for temperatures above 250 K.

Appendix: Sample Calculations

Example 1. Calculation of SCP for 2-Methylheptane:

SMILES formula: CCC(C)CCCC

RG = 4.490×10^{-10} m
$N_a = 26$ (26 atoms in the molecule)

linear groups (Table 1)				nonlinear groups/corrections (Table 2)				
group	n_i	a_i	α_i	group	n_i	b_i	β_i	γ_i
$-CH_3$	3	0.201 84	−241.7	$>CH_2$	4	−0.001 88	−2.9045	0
$>CH_2$	4	0.116 44	17.929					
$>CH-$	1	0.030 49	229.47					

(a) PL Method:

eq 14: $$A = \exp[6.7796 + (3)(0.20184) + (4)(0.11644) + 0.030492 + (4^2)(-0.00188)] = 2569.1$$

eq 2: $$C_P = \frac{2569.1}{1000}T^{0.79267}$$

(b) PF Method:

eq 15: $$\Theta_G = 1886.2 + (3.3626 \times 10^{12}) \times (4.490 \times 10^{-10}) + (3)(-241.7) + (4)(17.929) + 229.47 + (4^2)(-2.9045) = 2925.6$$

eq 13: $$C_P = (2.55)(26)(8.314)\left(\frac{T}{2925.6}\right)^{0.85} \int_0^{2925.6/T} \frac{x^{1.85}e^x}{(e^x - 1)^2}\,dx$$

(c) Results:

T / K	C_P (expt) / J·mol^{-1}·K^{-1}	C_P (PL) / J·mol^{-1}·K^{-1}	C_P (PF) / J·mol^{-1}·K^{-1}
100	102.1	108.4 (6.2%)	99.1 (−2.9%)
152	140.4	138.2 (−1.6%)	141.9 (1.1%)

Example 2. Calculation of SCP for p-Cresol:

SMILES formula: c1(C)ccc(O)cc1

$RG = 3.762 \times 10^{-10}$ m
$N_a = 16$ (16 atoms in the molecule)

linear groups (Table 1)				nonlinear groups/corrections (Table 2)				
group	n_i	a_i	α_i	group	n_i	b_i	β_i	γ_i
$-CH_3$	1	0.201 84	−241.7	Ar−CH═	4	−0.000 33	−2.9616	0
Ar −CH═	4	0.082 478	−36.615					
Ar >C═	2	0.012 958	148.32					
−OH	1	0.103 41	−286.75					

(a) PL Method:

eq 14: $$A = \exp[6.7796 + 0.20184 + (4)(0.082478) + (2)(0.012958) + 0.10341 + (4^2)(-0.00033)] = 1694.9$$

eq 2: $$C_P = \frac{1694.9}{1000}T^{0.79267}$$

(b) PF Method:

eq 15: $$\Theta_G = 1886.2 + (3.3626 \times 10^{12}) \times (3.762 \times 10^{-10}) - 241.7 + (4)(-36.615) + (2)(148.32) - 286.75 + (4^2)(-2.9616) = 2725.6$$

eq 13: $$C_P = (2.55)(16)(8.314)\left(\frac{T}{2725.6}\right)^{0.85} \int_0^{2725.6/T} \frac{x^{1.85}e^x}{(e^x - 1)^2}\,dx$$

(c) Results:

T / K	C_P (expt) / J·mol^{-1}·K^{-1}	C_P (PL) / J·mol^{-1}·K^{-1}	C_P (PF) / J·mol^{-1}·K^{-1}
110	67.8	70.4 (3.8%)	70.2 (3.5%)
307.93	155.2	159.1 (2.5%)	167.9 (8.2%)

Example 3. Calculation of SCP for 1,1,2-Trichlorotrifluoroethane:

SMILES formula: ClC(Cl)(F)C(F)(F)Cl

$RG = 3.791 \times 10^{-10}$ m
$N_a = 8$ (8 atoms in the molecule)

linear groups (Table 1)				nonlinear groups/corrections (Table 2)				
group	n_i	a_i	α_i	group	n_i	b_i	β_i	γ_i
>C<	2	−0.040 64	529.76	−F	3	0	0	−1231.3
−F	3	0.155 11	−320.76	−Cl	3	0	0	−1361.4
−Cl	3	0.169 95	−429.06	n_X	6			

(a) PL Method:

eq 14: $$A = \exp[6.7796 + (2)(-0.04064) + (3)(0.15511) + (3)(0.16995)] = 2150.6$$

eq 2: $$C_P = \frac{2150.6}{1000}T^{0.79267}$$

(b) PF Method:

eq 15: $$\Theta_G = 1886.2 + (3.3626 \times 10^{12}) \times (3.791 \times 10^{-10}) + (2)(529.76) + (3)(-320.76) + (3)(-429.06) + \left(\frac{3}{6}\right)(-1231.3) + \left(\frac{3}{6}\right)(-1361.4) = 674.7$$

eq 13: $$C_P = (2.55)(8)(8.314)\left(\frac{T}{674.7}\right)^{0.85} \int_0^{674.7/T} \frac{x^{1.85}e^x}{(e^x - 1)^2}\,dx$$

(c) Results:

T / K	C_P (expt) / J·mol^{-1}·K^{-1}	C_P (PL) / J·mol^{-1}·K^{-1}	C_P (PF) / J·mol^{-1}·K^{-1}
130.8	127.3	102.4 (−19.5%)	126.1 (−0.9%)
230.75	158.3	160.6 (1.5%)	165.7 (4.7%)

Example 4. Calculation of SCP for Biphenyl:

SMILES formula: c1ccccc1c2ccccc2

$RG = 4.834 \times 10^{-10}$ m
$N_a = 22$ (22 atoms in the molecule)

linear groups (Table 1)				nonlinear groups/corrections (Table 2)				
group	n_i	a_i	α_i	group	n_i	b_i	β_i	γ_i
Ar ═CH−	10	0.082 478	−36.615	Ar ═CH−	10	−0.000 33	−2.9616	0
Ar ═C<	2	0.012 958	148.32					

(a) PL Method:

$$\text{eq 14:}\quad A = \exp[6.7796 + (10)(0.082478) + (2)(0.012958) + (10^2)(-0.00033)] = 1992.8$$

$$\text{eq 2:}\quad C_P = \frac{1992.8}{1000}T^{0.79267}$$

(b) PF Method:

$$\text{eq 15:}\quad \Theta_G = 1886.2 + (3.3626 \times 10^{12}) \times (4.834 \times 10^{-10}) + (10)(-36.615) + (2)(148.32) + (10^2)(-2.9616) = 3146.0$$

$$\text{eq 13:}\quad C_P = (2.55)(22)(8.314)\left(\frac{T}{3146.0}\right)^{0.85} \int_0^{3146.0/T} \frac{x^{1.85}e^x}{(e^x - 1)^2}\,dx$$

(c) Results:

T/K	C_P (expt)/J·mol^{-1}·K^{-1}	C_P (PL)/J·mol^{-1}·K^{-1}	C_P (PF)/J·mol^{-1}·K^{-1}
197.25	129.3	131.4 (1.6%)	140.4 (8.6%)
302.25	201.3	184.3 (−8.5%)	201.7 (0.2%)

Literature Cited

(1) Rowley, R. L.; Wilding, W. V.; Oscarson, J. L.; Zundel, N. A.; Marshall, T. L.; Daubert, T. E.; Danner, R. P. *DIPPR Data Compilation of Pure Compound Properties*; Design Institute for Physical Properties, AIChE: New York, 2002.

(2) Ericksen, D.; Wilding, W. V.; Oscarson, J. L.; Rowley, R. L. Use of the DIPPR Database for the Development of QSPR Correlations: Normal Boiling Point. *J. Chem. Eng. Data* **2002**, *47*, 1293−1302.

(3) Knotts, T. A.; Wilding, W. V.; Oscarson, J. L.; Rowley, R. L. Use of the DIPPR Database for the Development of QSPR Correlations: Surface Tension. *J. Chem. Eng. Data* **2001**, *46*, 1007−1012.

(4) Mostafa, A. T. M. G.; Eakman, J. M.; Montoya, M. M.; Yarbro, S. L. Prediction of Heat Capacities of Solid Inorganic Salts from Group Contributions. *Ind. Eng. Chem. Res.* **1996**, *35*, 343−348.

(5) Hurst, J. E., Jr.; Harrison, B. K. Estimation of Liquid and Solid Heat Capacities using a Modified Kopp's Rule. *Chem. Eng. Commun.* **1992**, *112*, 21−30.

(6) Domalski, E. S.; Hearing, E. D. Estimation of the Thermodynamic Properties of C−H−N−O−S−Halogen Compounds at 298.15 K. *J. Phys. Chem. Ref. Data* **1993**, *22*, 805−1159.

(7) Kabo, G. J.; Kozyro, A. A.; Diky, V. V. Additivity of Thermodynamic Properties of Organic Compounds in the Crystalline State. 1. Additive Calculations for Thermodynamic Properties of Alkanes, Alkenes, Alkanols, and Alkanones. *J. Chem. Eng. Data* **1995**, *40*, 160−166.

(8) Kabo, G. J.; Kozyro, A. A.; Diky, V. V.; Simirski, V. V. Additivity of Thermodynamic Properties of Organic Compounds in the Crystalline State. 2. Heat Capacities and Enthalpies of Phase Transition of Alkyl Derivatives of Urea in Crystalline State. *J. Chem. Eng. Data* **1995**, *40*, 371−393.

(9) Kabo, G. J.; Kozyro, A. A.; Diky, V. V.; Simirski, V. V. Additivity of Thermodynamic Properties of Organic Compounds in the Crystalline State. 3. Heat Capacities and Related Properties of Urea Phenyl Derivatives. *J. Chem. Eng. Data* **2002**, *47*, 239−244.

(10) Guilyazetdinov, L. P. Structural Group Composition and Thermodynamic Properties of Petroleum and Coal Tar Fractions. *Ind. Eng. Chem. Res.* **1995**, *34*, 1352−1363.

(11) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1973.

(12) Bondi, A. *Physical Properties of Molecular Crystals, Liquids, and Glasses*; John Wiley and Sons: New York, 1968.

(13) Joback, K. G.; Reid, R. C. Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.* **1987**, *57*, 223−243.

(14) *TSAR Version 3.2*; Oxford Molecular Group, Oxford Molecular Limited: Oxford, 1998.

(15) Wininger, D. SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.

(16) Rowley, J. R.; Wilding, W. V.; Oscarson, J. L.; Rowley, R. L. *DIADEM Version 2.0, DIPPR Information and Data Evaluation Manager*; BYU-DIPPR TPL; Brigham Young University: Provo, UT, 2002.

(17) *SMILES Tutorial*; www.daylight.com, 1998.

(18) *TRC Thermodynamic Tables, Hydrocarbons*; Thermodynamics Research Center, The Texas A & M University System: College Station, TX, 1998; Vol. XIII, Table vc. *TRC Thermodynamic Tables, Non-Hydrocarbons*; Thermodynamics Research Center, The Texas A & M University System: College Station, TX, 1998.

Received for review December 12, 2002. Accepted March 20, 2003. Financial support for this work through the DIPPR 801 project is gratefully acknowledged.

JE025656H