

## Heat capacity prediction of complex molecules by mass connectivity index

O. Koutchoukali, I. Haloui, and M.S. Koutchoukali

Laboratoire d'Ingénierie des Procédés d'Environnement (LIPE), Département de Chimie Industrielle, Université Mentouri, Constantine, Algeria

**Abstract.** Heat capacity prediction and estimation methods of solid organic compounds in terms of temperature are limited, particularly concerning complex molecules with functional groups such as active principles and intermediaries used in pharmaceutical field.

Recently a correlation between heat capacity at constant pressure ( $C_p$ ), temperature and a new concept named mass connectivity index (MCI), for ionic liquids, was published [1-3]. In this predictive method, heat capacity can be calculated at different temperatures, if standard heat capacity at 298.15 K is known. The effect of molecular structure on heat capacity is accounted for in this model by the mass connectivity index, a molecular descriptor, which differentiates between compounds. The Valderrama generalized correlation admits, in addition, two universal coefficients, which are obtained from experimental data regression.

In the present work, a similar approach is used to predict solid state heat capacity of organics and pharmaceutical products. In order to find model parameters, a database was grouped comprising (104) different compounds and a set of more than 5,791 experimental values of solid state  $C_p$ s obtained from literature. These collected data were used in multiple linear regression to find model parameters. It was found that the values of predicted heat capacities of compounds non-included in the database were good; they are quite close to the ones presented in the literature. Moreover, this method is simple to use, since only molecular structure of the component and its solid state heat capacity at 298.15 K should be known.

### 1 Introduction

Heat capacity is one of the basic thermophysical and thermodynamic properties that characterizes a compound. It's evaluation as a function of temperature, with a great accuracy, is fundamental for describing the thermodynamic properties of substances in most thermodynamic and engineering calculations. Particularly, heat capacity is used in the calculation of basic thermodynamic functions such enthalpies and entropies of sublimation at 298.15 K, enthalpies of solvation as well as the partial molar heat capacities of solution at infinite dilution.

Many estimation methods for the heat capacity of pure organic compounds in the liquid state are available. However, there is a lack for estimating the heat capacity of solid organic compounds at constant pressure. Application of existing methods concern specific groups of organic compounds with low molar mass, and in most cases estimation of heat capacity at a single temperature, 298.15 K. Some group contribution methods have been developed for the estimation of solid phase heat capacity and can be divided in two categories: those that predict heat capacity at 298.15 K, such Domalski and Hearing [4, 5], Chickos et al [6] and those that can be used to predict heat capacity over a wide range of temperature Goodman et al [7].

At low temperature, Kabo et al. [8] have developed a predictive method for specific heat capacities from 10 to 150 K for alkanes, alkanols and alkanones. Their correlation was extended and modified later for alkyl and phenyl derivatives of urea [9, 10].

In previous studies, many researchers applied Einstein's model for the description of heat capacities as a function of temperature. As sample model of this type of method, the correlation of Briard et al [11] is applicable for n-alkanes which contains four parameters function of the number of carbons in alkane chain.

More recently a new predictive correlation for solid organics heat capacity [12], valid from 50 K to the melting point temperature, has been developed on the basis of elemental composition. It uses a concept that solid  $C_p$  is a function of the number of fundamental vibrations per mass of molecule, which in this context linked directly to the number of atoms per mass. The correlation has been applied for different pure organic compounds containing C, H, N, O, S, using seven universal coefficients.

The objective of the present study is to provide an extensive survey of the literature for heat capacities of pure complex organics and pharmaceuticals, to correlate experimental data and provide an estimation method for solid heat capacities as a function of temperature, based essentially on the molecular structure.

## 2 Temperature dependence of Cp

Recently, a new descriptor, called mass connectivity index (MCI) was developed by Valderrama et Rojas [1]. This index was then used by Valderrama and al [2, 3] to estimate the heat capacity and density of ionic liquids through a generalized correlation including temperature. This descriptor is based on the molecular connectivity index introduced by Randic [13] in 1975 and subsequently amended by Kier and Hall [14]. It is defined as the sum of the inverse of the square root of the product of the mass of two neighboring functional groups in the molecule. The groups forming the molecule, defined in exactly the same way as a method of classic group contribution, are those used by Valderrama et al. [15] to estimate critical properties. During the summation, it is necessary to consider  $m_i m_j$  different from  $m_j m_i$ .

$$C_p = \sum_{i,j} (1/\sqrt{(m_i m_j)})_{ij} \quad (1)$$

The mass connectivity index is much easier to calculate than other connectivity indices available in the literature. Indeed, only the masses of the groups constituting the molecule are necessary for its determination. In the case of heat capacity, the generalized correlation, quadratic in temperature (second order polynomial), is given by the expression:

$$C_p = C_{p0} + [a(T - T_0) + b(T^2 - T_0^2)] \quad (2)$$

where  $C_{p0}$  is the heat capacity at standard temperature ( $T_0$ ) 298.15 K and  $\lambda$  is the mass connectivity index which discriminates between different ionic liquids. The  $a$  and  $b$  constants, are the universal parameters calculated by multiple linear regression of experimental heat capacities data of ionic liquids available in the literature. It is supposed that standard heat capacity is known experimentally or can be calculated by any group contribution method.

Here we have adopted this method for estimation of solid heat capacities of complex organic compounds containing C, H, N, O and S, encountered, particularly in the pharmaceutical industry, for which solid heat capacity data are largely unavailable.

### Database and correlation of Cp

We have investigated the molar heat capacity of various chemical families or individual solid organic compounds acting as biochemical model molecules. A total of 5791 experimental values of solid heat capacity have been found for 104 types of substances with molar masses ranging from 108.14 to 331.43 g.mol<sup>-1</sup>. The majority of the experimental heat capacities of the study compounds collected from the literature are in the temperature range of 70 K to the melting point. Also, we have taken this value as the lowest temperature.

Among the compounds constituting the entire database, 55 compounds and a set of 3238 data points were used in the development of the correlation and determination of parameters of equation (2) (training set). These compounds, their formula, their mass,

connectivity index, the number of experimental Cp data, temperature range and average absolute deviation are given in supplementary data, Table 1.

Experimental heat capacities of the 48 others compounds, also reported in Table 2 of supplementary information with some of their properties, were reserved for prediction and validation of the model.

It should be noted that the Cp data for which we observed the pre-melting effect are not taken into account in the regression of model parameters.

The experimental heat capacities from the training set were correlated by minimizing the objective function (OF) to estimate parameters  $a$  and  $b$  of equation (2). The objective function used is defined as follows:

$$OF = \sum_{i=1}^m \left[ \sum_{j=1}^n \frac{(C_p^{exp} - C_p^{calc})^2}{n} \right] \quad (3)$$

where  $m$  is the number of different compounds and  $n$  the number of experimental data considered, and  $C_p^{exp}$  and  $C_p^{calc}$  are values of the experimental and calculated heat capacities, respectively.

## 3 Results and discussion

The two universal coefficients, appearing in Eq. (2), were obtained using multiple regression and their validity was tested using a heat capacity database (test set). The parameters  $a$  et  $b$  of the generalized correlation (2) :

$$C_p = C_{p0} + [a(T - T_0) + b(T^2 - T_0^2)] \quad (2)$$

are given by:

$$a = 0.31300598 \quad \text{and} \quad b = 1.6997 \times 10^{-4}$$

We have evaluated the correlation of the training set and the prediction of the test set data in terms of an average absolute relative deviation (AARD), and the average absolute deviation (AAD) defined by:

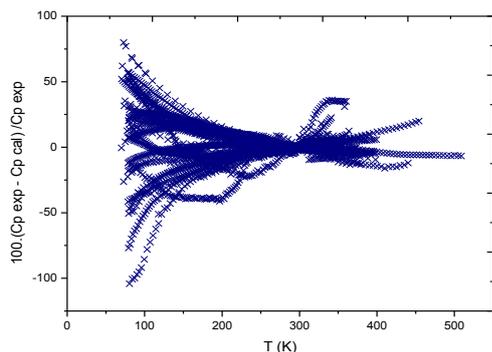
$$AARD = \sum_{i=1}^n \left| \frac{C_{pexp} - C_{cal}}{C_{pexp}} \right| \times \frac{100}{n} \quad (4)$$

$$AAD = \frac{\sum_{i=1}^n |C_{pexp} - C_{cal}|}{n} \quad (5)$$

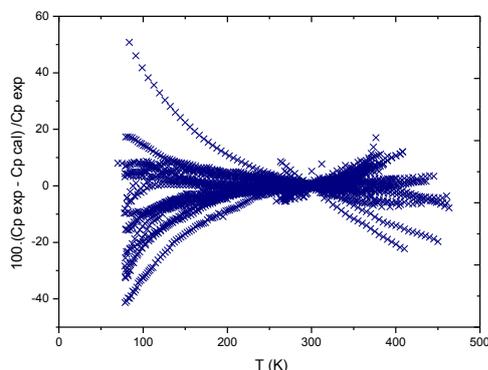
As shown in Table 4, the universal constants obtained here, gives quite reasonable predicted values of the heat capacity for all of the 49 organic compounds considered in the test set.

We note, from Table 3 and Table 4 and figures 1 and 2, that the prediction gives very good results in the range 200-400 K for both correlated Cp as calculated ones. In the low temperature range 50K < T < 250K, we observed an AARD deviation of 21.4% for the training set and 9.18 % for the test set. Some of compounds in the training set database exhibit large deviations, particularly,

benzylsulfide and fluorenone and the correlation clearly degrades at low temperatures.



**Figure 1.** Relative deviation of the correlated Cp from the training set.



**Figure 2.** Relative deviation of the correlated Cp from the test set.

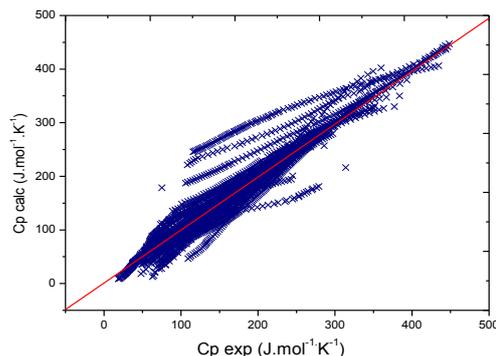
**Table 3.** Deviations of correlated Cp values from training set.

T	no of compds	no of Cp values	deviation	statistic
70 K < T < 200 K	33	1404	AARD(%)	21.4
			AAD (J.mol <sup>-1</sup> .K <sup>-1</sup> )	25.02
T > 200 K	56	1942	AARD(%)	4.87
			AAD (J.mol <sup>-1</sup> .K <sup>-1</sup> )	9.69

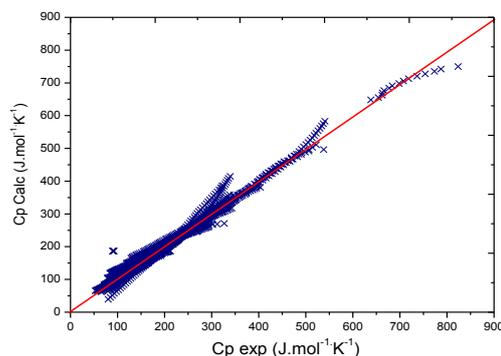
**Table 4.** Deviations of estimated Cp values from the test set.

T	no of compds	no of Cp values	deviation	statistic
70 K < T < 200 K	23	772	AARD(%)	9.185
			AAD (J.mol <sup>-1</sup> .K <sup>-1</sup> )	10.55
T > 200 K	50	1768	AARD(%)	2.238
			AAD (J.mol <sup>-1</sup> .K <sup>-1</sup> )	5.39

To test the capability of the correlation developed here we compared heat capacity experimental values from predicted and correlated values by mean of the figures 3 and 4. These figures shows that deviations are acceptable.



**Figure 3.** Predicted Cp values function of experimental values from the training set.



**Figure 4.** Predicted Cp values function of experimental values from the test set.

## 4 Conclusion

Thermo-physical properties models play a major role in process modeling. In the case of heat capacity, it appears that while there is abundant information on liquid heat capacity and gas heat capacity both experimental data or generalized correlations (or models), there was less attention given to the prediction of heat capacity of complex organics in the solid state. As a result, much effort is required to fill in this gap, especially for hetero-compounds.

We have adopted here generalized correlation method of Valderrama et al [], including mass connectivity index and standard heat capacity at 298.15 K for estimating the solid heat capacities of organic compounds, commonly used as derivatives in pharmaceutical industry as a function of temperature. The generalized correlation allows the prediction of solid heat capacity of complex organic compounds from 70 K to their melting point, but with better accuracy above 200 K.

## References

1. J.O. Valderrama, R.E. Rojas, Fluid Phase Equilibria **297** 107 (2010)
2. J.O. Valderrama, A. Taro, R.E. Rojas, J. Chem. Thermodynamics **43** 1068 (2010)
3. J.O. Valderrama, G. Martinez, R.E. Rojas, Thermochim. Acta **83** 513 (2011)

4. E.S. Domalski, E.D. Hearing, J. Phys. Chem. Ref. Data. **17** 1637 (1988)
5. E.S. Domalski, E.D. Hearing, J. Phys. Chem. Ref. Data. **22** 805 (1993)
6. J.S. Chickos, D. G. Hesse, J.F. Liebman, Struct. Chem. **4** 261 (1993)
7. B.T. Goodman, W.V. Wilding, J.L. Oscarson, R.L. Rowley, J. Chem. Eng. Data. **49** 24 (2004)
8. G.J. Kabo, A.A. Kozyro, V.V. Diky, J. Chem. Eng. Data **40** 160 (1995)
9. G.J. Kabo, A.A. Kozyro, V.V. Diky, V.V. Simirsky, J. Chem. Eng. Data **40** 371 (1995)
10. V.V. Diky, A.A. Kozyro, G.J. Kabo, J. Chem. Eng. Data **47** 239 (2002)
11. A.J. Briard, M. Bouroukba, D. Petitjean, M. Dirand, J. Chem. Eng. Data **48** 1508 (2003)
12. V. Lastovska and J.M Shaw, J. Chem. Eng. Data **52** 1160 (2007)
13. M. Randic, J. Am. Chem. Soc. **97** (23) 6609 (1975)
14. L.B. Kier, L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic press, New York, 1986
15. J.O. Valderrama, W.W. Sanga, J.A. Lazzús, Ind. Eng. Chem. Res. **47** 1318 (2008)