# Capturing the Crystal: Prediction of Enthalpy of Sublimation, Crystal Lattice Energy, and Melting Points of Organic Compounds

Maryam Salahinejad,[†] Tu C. Le,[‡] and David A. Winkler*[,‡,§]

[†]Faculty of Chemistry, Tarbiat Moallem University, Tehran 15719-14911, Iran
[‡]CSIRO Materials Science & Engineering, Clayton 3168, Australia
[§]Monash Institute of Pharmaceutical Sciences, Parkville 3052, Australia

Ⓢ *Supporting Information*

ABSTRACT: Accurate computational prediction of melting points and aqueous solubilities of organic compounds would be very useful but is notoriously difficult. Predicting the lattice energies of compounds is key to understanding and predicting their melting behavior and ultimately their solubility behavior. We report robust, predictive, quantitative structure−property relationship (QSPR) models for enthalpies of sublimation, crystal lattice energies, and melting points for a very large and structurally diverse set of small organic compounds. Sparse Bayesian feature selection and machine learning methods were employed to select the most relevant molecular descriptors for the model and to generate parsimonious quantitative models. The final enthalpy of sublimation model is a four-parameter multilinear equation that has an r^2 value of 0.96 and an average absolute error of 7.9 ± 0.3 kJ.mol^{-1}. The melting point model can predict this property with a standard error of 45° ± 1 K and r^2 value of 0.79. Given the size and diversity of the training data, these conceptually transparent and accurate models can be used to predict sublimation enthalpy, lattice energy, and melting points of organic compounds in general.
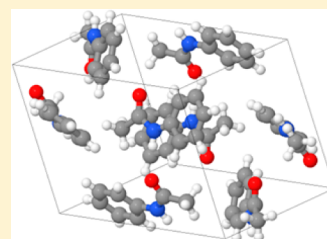
INTRODUCTION

The sublimation enthalpy, ΔH_{sub}, is a thermochemical quantity that represents strength of the intermolecular interactions in a solid and hence the stability of the crystal structure. This property is of interest in diverse applications such as dye dispersion and discoloration of materials, environmental fate, transport of organic contaminants in the atmosphere,[1] the study of the molecular mechanism of solvation of drug molecules,[2,3] and specific packing of the solid state of crystalline materials.[4] It also provides useful information about other thermodynamic quantities such as enthalpies of formation and allows calculation of vapor pressures by using the Clausius−Clapeyron equation.[5]

The sublimation enthalpy is related to the crystal lattice energy E_{latt} by the following simple equation[6]

$$\Delta H_{sub} = -E_{latt} - 2RT \qquad (1)$$

where R and T are the universal gas constant and temperature, respectively. Lattice energies of crystals are of interest in the pharmaceutical and fine chemical industries and in studies of polymorphism and crystal growth.[7,8] An understanding of lattice energies is important for predicting dissolution and melting point behavior.

Many methods have been used to predict sublimation enthalpies of organic compounds including the following: molecular mechanics,[9] quantum mechanical calculations,[10,11] quantitative structure−property/activity relationships (QSPR/ QSAR) approaches,[12−16] and fragment contribution-based methods.[17−21] As reviewed elsewhere,[18] most of these models have been constructed using small sets of specific organic

compounds and thus have limited applicability over a wider range of diverse compounds. Group contribution-based methods can also be used to predict thermodynamic properties of diverse chemical families of pure compounds, but they usually employ many parameters and complex computational procedures to accurately estimate sublimation enthalpies.

Given the ease by which melting points can be measured, there have been fewer models reported for melting points of small organic molecules. Dearden has reviewed recent melting point models.[22] Preiss et al. have summarized the factors affecting accurate measurement of melting point and posed the question whether a simple, universal melting point (T_m) prediction method is possible.[23] One of the main uses of T_m models is to improve the accuracy of aqueous solubility models that employ the General Solubility Equation (GSE).

Difficulties in experimental determination of sublimation enthalpies and lattice energies, along with need for significant computational resources and knowledge of the crystal structure for calculation of these parameters,[24] make QSPR models an alternative, practical approach. We report a simple, robust QSPR model for enthalpies of sublimation and crystal lattice energies of a large, chemically diverse set of organic components. This model differs from previous models of enthalpy of sublimation in its applicability to more chemically diverse molecules and in its higher accuracy. We also report a robust melting point model that uses the enthalpy of sublimation as a descriptor, which is derived from an even
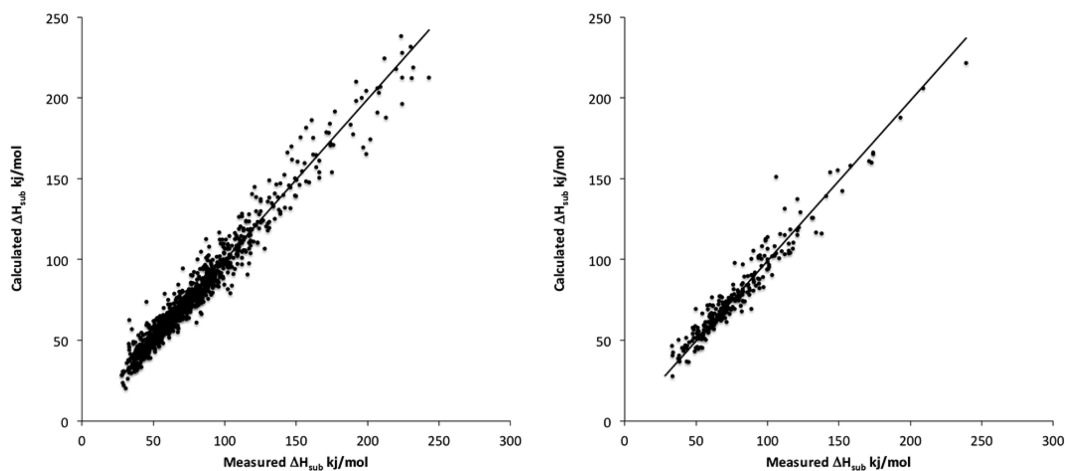
Figure 1. Plot of the predicted versus experimental values of sublimation enthalpy for training (left) and test set (right) for the MLR model.

larger and more chemically diverse set of experimental melting point data.

2. MATERIALS AND METHODS

2.1. Data Set. A heterogeneous set of 1302 organic molecules was taken from the work of Gharagheizi,[18] compiled from DIPPR 801 database.[25] The CAS number, chemical name, SMILES representation, experimental enthalpy of sublimation, and molecular descriptors for organic molecules in the data set are listed in the Supplementary Table 1. The data set was randomly divided multiple times into a training set (80% of the data, 1042 molecules) and a test set (20% of the data, 260 molecules). The melting point data were taken from the PhysProp database[26] and consisted of 8241 data points after excluding compounds not unambiguously defined or those containing metals or elements not commonly found in small organic, drug-like molecules. Due to the very large size of the data sets, and the impracticability of determining or predicting the pK_as, compounds such as acids and bases that may be ionized in solution were used in their neutral form for the models. Zhao et al.[27] and others have shown that if indicator variables are used to account for ionization of strong acids with pK_a <4.5 and bases with pK_a > 8.5 the additional indicator variable is often barely significant. It is noteworthy that previous reports of melting point or enthalpy of sublimation models with moderate to large data sets have generally not included ionization either.

2.2. Molecular Optimization and Descriptor Calculation. The 3D structure of the organic molecules was constructed using the standard tools available in the SYBYL-X1.1 molecular modeling package (Tripos Inc., St. Louis, USA). Energy minimization was performed using the Tripos force field with a distance dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/mol Å. Partial atomic charges were calculated using the Gasteiger-Hückel method.

Subsets of descriptors selected from a pool of 111 VolSurf and charged partial surface area (CPSA) descriptors were used to generate the models. These descriptors were chosen to span a range of molecular properties that are likely to be relevant to crystal lattice interactions. Other choices could have been made (e.g., use of Dragon descriptors). However, a substantial number of members in these descriptor sets correlate strongly with those in other sets (i.e., they capture similar molecular

information). As long as a sufficiently representative number of interactions are captured by the descriptors used in the model, the model should not be sensitive to the exact set used. VolSurf is a computational procedure that generates useful quantitative 2D descriptors from the 3D maps of molecular interaction fields (MIF) between different probes and all the atoms in a target molecule.[28] In the present work, VolSurf+ 4.0.1 (Molecular Discovery Ltd., Oxford, UK) was used. Four probes, water (OH_2); hydrophobic (DRY); hydrogen bond donor (NH); and H-bond acceptor (=O), were used to generate the VolSurf descriptors. A grid spacing of 1 Å was used for the MIF calculations. CPSA descriptors such as molecular surface area, molecular volume, partial positive and negative surface area, weighted parameters related to these properties, that described polar intermolecular interactions[29,30] were calculated by the SYBYL software.

2.3. Variable Selection and Model Construction. Multiple linear regression (MLR) and multiple linear regression with expectation maximization (MLREM),[31−33] implemented in an in-house modeling package CSIRO-Biomodeller, were used to choose the most informative variables and to generate the linear QSPR models. To generate sparse nonlinear models we used Bayesian regularized artificial neural networks employing a Gaussian (BRANNGP)[34] or Laplacian prior (BRANNLP).[35] These implement Bayesian regularization in two different ways that optimize the complexity of the models, automatically prune the number of weights in the neural network, and, in the case of the BRANNLP method, automatically prune out the less relevant molecular descriptors using nonlinear feature selection. These neural networks generate optimally sparse and predictive models of properties and minimize overfitting even for relatively small data sets, as has been described previously.[36,37]

3. RESULTS AND DISCUSSION

3.1. Enthalpy of Sublimation Models. Intermolecular interactions such as van der Waals and hydrogen bonding forces control the lattice energy[38] and hence the sublimation enthalpy in a molecule. The highly ordered crystal lattice originates from these strong cohesive interactions. Compounds that contain a large number of hydrogen-bond acceptor/donor groups, are relatively planar, are relatively rigid, or exhibit a high degree of symmetry,[39,40] show tight packing within the crystal lattice.

We initially used simple multiple linear regression (MLR) to identify a relationship between the enthalpy of sublimation and molecular descriptors. The most important VolSurf descriptor for predicting sublimation enthalpy was the hydrophilicity W1. Equation 2 summarizes the relationship between W1 and ΔH_{sub}. W1 is computed from the MIFF molecular fields at several energy levels and is defined as the volume of molecular envelope at the selected energy level which is accessible to and interacts favorably with water molecules.[28] This variable also represents the polarizability and dispersion forces in a molecule.

$$\Delta H_{sub} = (-16.96 \pm 0.62) + (0.150 \pm 0.002) \times \text{W1} \tag{2}$$

This single descriptor QSPR model gives an r^2 for both the training and test sets of 0.90 (i.e., the model accounts for 90% of the variance in the data) and a standard error of estimation (SEE) and prediction (SEP) of 10.9 and 11.4 kJ.mol^{-1} for the training and test sets respectively. This simple equation could be used as a general equation to predict sublimation enthalpy as well as lattice energy of organic compounds by purely computational means.

The simple linear model for the enthalpy of sublimation can be improved by adding additional relevant molecular descriptors. Our context sensitive sparse Bayesian feature selection methods chose a small number of additional relevant descriptors, FPSA3, V, and PSA, for the model. The fractional charged partial surface area (FPSA3) is obtained by dividing the atomic charge weighted positive surface area by the total molecular surface area.[29,30] This descriptor contains molecular information relating to polar and electrostatic intermolecular interactions. V is the water-excluded molecular volume (Å^3), the volume enclosed by the water-accessible surface computed at a repulsive value of 0.2 kcal/mol.[28] PSA, the polar surface area, represents the sum of polar contributions to the molecular surface. All four descriptors involved in the MLR model showed positive correlations with sublimation enthalpy (see Figure 1). These descriptors mainly represent dispersion, van der Waals, and electrostatic interactions in a molecule that contribute to crystal packing and crystal structure stability,[41] thus increasing the sublimation enthalpy of organic compounds. The statistics of the four-descriptor MLR model (eq 3) are summarized in Table 1. The significant improvement in

Table 1. Statistical Results for ΔH_{sub} MLR Model[a]

r^2_{train}	r^2_{test}	SEE (kJ/mol)	SEP (kJ/mol)
0.954 ± 0.002	0.950 ± 0.01	7.27 ± 0.06	7.88 ± 0.35

[a]Errors were calculated by repeated calculations of the model using different random selections of training and test sets.

model statistics shows that these descriptors make important contributions to the predictivity of the model.

$$\Delta H_{sub} = -(23.5 \pm 0.3) + (386.7 \pm 12.8)\text{FPSA3}$$
$$+ (0.385 \pm 0.019)\text{PSA} + (0.114 \pm 0.003)\text{V}$$
$$+ (0.0495 \pm 0.0027)\text{W1} \tag{3}$$

A sublimation enthalpy model using a neural network with a structure of 4−2−1 (four input modes, two hidden layer nodes and one output node) was also generated. There was no significant improvement in statistical results of ANN model compared to four variable MLR models suggesting that the relationship between the molecular descriptors and the

enthalpy of sublimation is quite linear. The compounds used to generate the enthalpy of sublimation model and the four descriptors are listed in Supplementary Table 1.

3.2. Melting Point Models. We used the enthalpy of sublimation model to calculate ΔH_{sub} for all 8241 compounds in the melting point data set. We then used ΔH_{sub} as a descriptor in the melting point models so we could capture the important crystal lattice interactions that modulate the melting point. We varied the sparsity of the linear and nonlinear models to determine the optimum number of descriptors, varied the number of hidden layer nodes in the neural nets over a small range (1−3) to optimize the network architecture, and ran multiple random selections of the training and test sets to provide estimates of the robustness and range of predictions of the models. We found that, apart from the enthalpy of sublimation, a significant number of additional descriptors were required to generate the most predictive melting point model. Table 2 summarizes the performance of the linear and

Table 2. Statistical Results for Melting Point Models[a]

	r^2_{train}	r^2_{test}	SEE (K)	SEP (K)
MLR	0.77 ± 0.00	0.75 ± 0.01	48.0 ± 0.1	50.3 ± 0.7
BRANNGP	0.79 ± 0.01	0.77 ± 0.01	45.0 ± 0.4	46.3 ± 0.6
BRANNLP	0.79 ± 0.01	0.79 ± 0.01	42.9 ± 0.3	45.1 ± 1.0

[a]Errors were calculated by repeated calculations of the model using different random selections of training and test sets.

nonlinear melting point models. The linear MLREM model was less accurate than the two neural network models, with the sparse Bayesian network employing the Laplacian prior (BRANNLP) having the best predictive performance and robustness. The BRANNLP model (see Figure 2) employed 97 descriptors from the original pool, including the enthalpy of sublimation, and the neural network had 290 effective weights, much fewer than the number of molecules in the training set. This model had an average melting point prediction error (SEP) that was 5° lower than the linear model (Table 2). This suggests that the relationship between molecular structure and melting point is complex (many descriptors required) and significantly nonlinear. The descriptors used in the linear and nonlinear T_m models are listed and described in Supplementary Table 2.

3.3. Relationship between ΔH_{sub} and T_m. Gopal[42] reported an approximately constant relationship between the melting point and the lattice or binding energy of crystals of alkali halides. We were interested in whether such a relationship exists for small organic molecules. For the 8241 compounds in the melting point data set the ratio of enthalpy of sublimation to melting point in Kelvin was 0.32 ± 0.10 kJ/mol·K. A plot of the relationship between these two properties revealed reasonable linearity for compounds with lower molecular size and complexity (Figure 3, ΔH_{sub} <150 kJ/mol, black markers). However, larger more complex compounds deviate markedly from this relatively constant ratio (see Figure 3 ΔH_{sub} > 150 kJ/mol, gray markers). This observation is consistent with a significant contribution to T_m from conformational entropy. Preiss et al.[23] recently reported a melting point model which, although of limited predictive accuracy, was grounded in thermodynamic theory.

According to the Gibbs-Helmholz eq 4
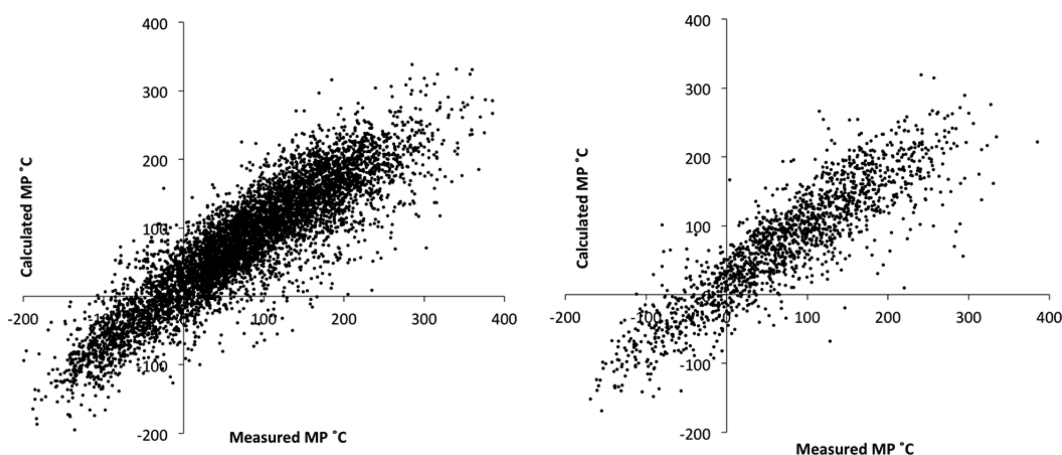
$$\Delta G = \Delta H - T\Delta S \tag{4}$$

Figure 2. Observed versus calculated melting points from BRANNLP model for training set (left) and test set (right).
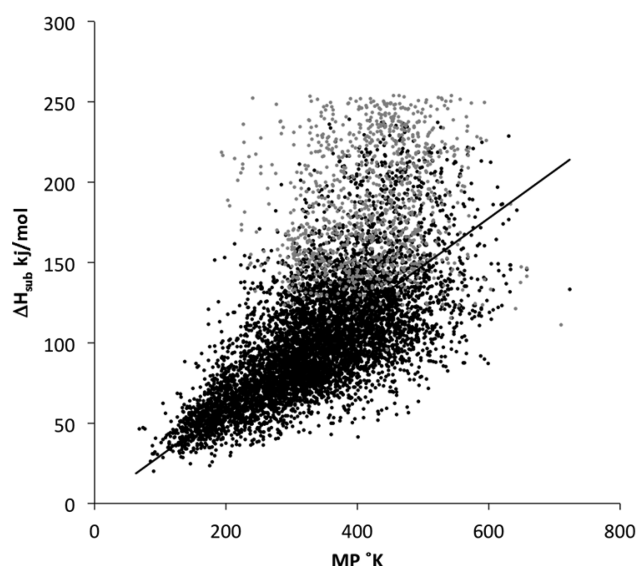

Figure 3. Plot of ΔH_{sub} vs MP showing relatively constant ratio except for larger, more flexible molecules (gray).

Figure 4. Linear relationship between $\Delta H_{sub}/T_m$ and the number of rotatable bonds N_{rot} for the 8241 compounds in the data set.

At melting, $T_m = T_{fus}$, and the solid and liquid are in equilibrium, $\Delta G_{fus} = 0$ and

$$T_m = \Delta H_{fus}/\Delta S_{fus} \tag{5}$$

The melting entropy can be approximated by eq 6

$$\Delta H_{fus} = a\ln\sigma + bN_{rot} + c \tag{6}$$

where a, b, and c are coefficients fitted to experimental data, σ is the rotational symmetry number representing the number of rotations of the molecule which are equivalent, and N_{rot} is the conformational freedom of the molecule derived from the number of rotatable bonds. Equating ΔH_{latt} to ΔH_{fus}, combining eqs 1, 5, and 6 and rearranging, we obtain eq 7

$$\Delta H_{sub}/T_m = a\ln\sigma + bN_{rot} + c' \tag{7}$$

where c' is $c + 2R$. Assuming the term in N_{rot} dominates, we can see that there is a linear relationship between $\Delta H_{sub}/T_m$ and N_{rot}. Figure 4 is a plot of this relationship for the 8241 compounds in our data set. The r^2 for the relationship is 0.55. Extrapolating the graph to zero conformational degrees of freedom, the ratio of enthalpy of sublimation to melting point is 0.25. Other parameters related to molecular size and flexibility
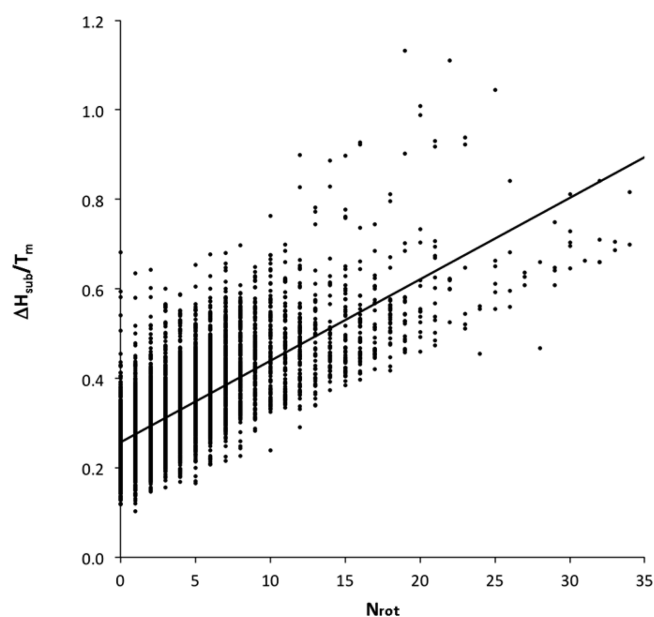
such as number of bonds, polarizability, and molar refractivity showed a similar relationship, albeit with significantly smaller r^2 values.

3.4. Factors Limiting Model Accuracy. The crystal form affects melting of a compound, which limits the ability of models to predict this property. Polymorphism occurs when a molecule can crystallize in more than one crystal structure under seemingly identical conditions. Although there is only one thermodynamically stable phase, persistent metastable crystalline phases can form that eventually transform to the stable phase. If the compound exhibits polymorphism, the various polymorphic forms will have different melting points. Melting points of different crystalline polymorphic forms or hydrates can differ in extreme cases by 2- to 10-fold. Preiss et al. recently reviewed the factors affecting reliable measurement of melting points.[23] Some of the more important factors were impurity or contamination (including hygroscopy), phase behavior (including polymorphism, errors up to 20K), disorder and defects (errors up to 40K), decomposition, and thermal history (annealing out of point defects on repeated thermal cycling). Clearly the choice to use experimental solubility data

Table 3. Comparison between Recent Melting Point Models

study	data set size	training		test	
		r^2	SEE (K)	r^2	SEP (K)
organosilicons (Liu and Holder 2011)[43]	97 train	0.79	50.5	0.71	42.4
	32 test				
anilines (Dearden 1991)[44]	42 train	0.89	23.7		
perfluorinated compounds (Bhhatarai et al. 2011)[45]	94 train	0.81	40.2		
	15 test			0.85	37.0
organic compounds (Karthikeyan et al. 2005)[46]	2089 train	0.66	48.0		
	1042 test			0.66	49.3
	1042 validation			0.65	50.4
	277 external test			0.66	41.4
fatty acids (Guendouzia et al. 2012)[47]	62 train	0.95	10.0		
	42 train	0.95			
	20 test			0.96	
drug-like compounds (Modarresi et al. 2006)[48]	278 train	0.66	40.9		
	45 test			0.79	42.0
organic compounds (Nigsch et al. 2006)[49]	3119 train	0.33	53.3		
	1000 test			0.44	48.8
	277 test			0.19	50.4
organic compounds (Hughes et al 2008)[50]	150 train	0.84	27.5		
	50 test			0.50	48.1
	87 external test	0.46	52.8	0.46	52.8
drug-like compounds (Habibi-Yangjeh et al 2008)[51]	195 train	0.97	12.6		
	64 external test			0.97	12.8
organic compounds (Lazzús 2009)[52]	260 training	0.96	11.5		
	73 test			0.98	
this work	6593 training	0.79 ± 0.01	42.9 ± 0.3		
	1648 test			0.78 ± 0.01	45.1 ± 1.0

within the range 20−30 °C, the range commonly encountered in laboratory environments, will also affect the accuracy of the models as solubility varies with temperature. Given these factors, the performance of the melting point model is high (see Table 3).

4. CONCLUSION

We have developed simple QSPR models based on Volsurf and CPSA descriptors that make good quantitative predictions of the sublimation enthalpy and lattice energy of organic compounds. These models provide general equations for quickly estimating these properties for new organic compounds. The results elucidate the contributions of hydrophilicity, molecular volume, and electrostatic interactions to lattice energy and sublimation enthalpy. We have also developed a simple, accurate model for prediction of melting points of organic molecules that employs calculated values of ΔH_{sub} as a descriptor, allowing an affirmative answer to be given to the question posed by Preiss et al. in the Introduction. We also explored the relationships between melting point, enthalpy of sublimation, and conformational entropy in this large data set and identified the key contribution of conformational entropy.

ASSOCIATED CONTENT

S Supporting Information

Supplementary Tables 1 and 2. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dave.winkler@csiro.au.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Nakajoh, K.; Shibata, E.; Todoroki, T.; Ohara, A.; Nishizawa, K.; Nakamura, T. Measurement of temperature dependence for the vapor pressures of twenty-six polychlorinated biphenyl congeners in commercial Kanechlor mixtures by the Knudsen effusion method. *Environ. Toxicol. Chem.* **2006**, *25* (2), 327−336.

(2) Perlovich, G. L.; Rodionov, S. V.; Bauer-Brandl, A. Thermodynamics of solubility, sublimation and solvation processes of parabens. *Eur. J. Pharm. Sci.* **2005**, *24* (1), 25−33.

(3) Perlovich, G. L.; Volkova, T. V.; Bauer-Brandl, A. Towards an understanding of the molecular mechanism of solvation of drug molecules: A thermodynamic approach by crystal lattice energy, sublimation, and solubility exemplified by paracetamol, acetanilide, and phenacetin. *J. Pharm. Sci.* **2006**, *95* (10), 2158−2169.

(4) Gavezzotti, A. Molecular packing and other structural properties of crystalline oxohydrocarbons. *J. Phys. Chem.* **1991**, *95* (22), 8948−8955.

(5) Chickos, J. S. A protocol for correcting experimental fusion enthalpies to 298.15K and it's application in indirect measurements of sublimation enthalpy at 298.15K. *Thermochim. Acta* **1998**, *313* (1), 19−26.

(6) Ouvrard, C.; Mitchell, J. B. O. Can we predict lattice energy from molecular structure? *Acta Crystallogr., Sect. B: Struct. Sci.* **2003**, *59*, 676−685.

(7) Price, S. L. Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Acc. Chem. Res.* **2008**, *42* (1), 117−126.

(8) Day, G. M.; Chisholm, J.; Shan, N.; Motherwell, W. D. S.; Jones, W. An assessment of lattice energy minimization for the prediction of molecular organic crystal structures. *Cryst. Growth Des.* **2004**, *4* (6), 1327−1340.

(9) Osborn, J. C.; York, P. A comparison of sublimation enthalpies with lattice energies calculated using force fields. *J. Mol. Struct.* **1999**, *474* (1−3), 43−47.

(10) Perlovich, G. L.; Raevsky, O. A. Sublimation of molecular crystals: prediction of sublimation functions on the basis of HYBOT physicochemical descriptors and structural clusterization. *Cryst. Growth Des.* **2010**, *10* (6), 2707−2712.

(11) Feng, S.; Li, T. Predicting lattice energy of organic crystals by density functional theory with empirically corrected dispersion energy. *J. Chem. Theory Comput.* **2005**, *2* (1), 149−156.

(12) Puri, S.; Chickos, J. S.; Welsh, W. J. Three-dimensional quantitative structure−property relationship (3d-qspr) models for prediction of thermodynamic properties of polychlorinated biphenyls (PCBs): enthalpy of sublimation. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (1), 109−116.

(13) Welsh, W. J.; Tong, W.; Collantes, E. R.; Chickos, J. S.; Gagarin, S. G. Enthalpies of sublimation and formation of polycyclic aromatic hydrocarbons (PAHs) derived from comparative molecular field analysis (CoMFA): application of moment of inertia for molecular alignment. *Thermochim. Acta* **1997**, *290* (1), 55−64.

(14) Ruoff, R. S. Prediction of enthalpies of sublimation of fullerenes from first-order molecular connectivity theory. *Chem. Phys. Lett.* **1993**, *208* (3−4), 256−258.

(15) Charlton, M. H.; Docherty, R.; Hutchings, M. G. Quantitative structure-sublimation enthalpy relationship studied by neural networks, theoretical crystal packing calculations and multilinear regression analysis. *J. Chem. Soc., Perkin Trans. 2* **1995**, No. 11, 2023−2030.

(16) Mathieu, D.; Bougrat, P. Model equations for estimating sublimation enthalpies of organic compounds. *Chem. Phys. Lett.* **1999**, *303* (5−6), 601−606.

(17) Gharagheizi, F. A new molecular-based model for prediction of enthalpy of sublimation of pure components. *Thermochim. Acta* **2008**, *469* (1−2), 8−11.

(18) Gharagheizi, F.; Sattari, M.; Tirandazi, B. Prediction of crystal lattice energy using enthalpy of sublimation: a group contribution-based model. *Ind. Eng. Chem. Res.* **2011**, *50* (4), 2482−2486.

(19) Zhokhova, N. I.; Baskin, I. I.; Palyulin, V. A.; Zefirov, A. N.; Zefirov, N. S. Calculation of the enthalpy of sublimation by the QSPR method with the use of a fragment approach. *Russ. J. Appl. Chem.* **2003**, *76* (12), 1914−1919.

(20) Mathieu, D. Simple alternative to neural networks for predicting sublimation enthalpies from fragment contributions. *Ind. Eng. Chem. Res.* **2012**, *51* (6), 2814−2819.

(21) Tabernero, A.; Martin del Valle, E. M.; Galán, M. A. Estimation of sublimation enthalpies of solids constituted by aromatic and/or polycyclic aliphatic rings by using a group contribution method. *AIChE J.* **2012**, *58* (9), 2875−2884.

(22) Dearden, J. C. Quantitative structure-property relationships for prediction of boiling point, vapor pressure, and melting point. *Environ. Toxicol. Chem.* **2003**, *22* (8), 1696−1709.

(23) Preiss, U. P.; Beichel, W.; Erle, A. M. T.; Paulechka, Y. U.; Krossing, I. Is universal, simple melting point prediction possible? *ChemPhysChem* **2011**, *12* (16), 2959−2972.

(24) Ouvrard, C.; Mitchell, J. B. O. Can we predict lattice energy from molecular structure? *Acta Crystallogr., Sect. B: Struct. Sci.* **2003**, *59* (5), 676−685.

(25) DIPPR Project 801, Evaluated Process Design Data, Public Release Documentation, Design Institute for Physical Properties (DIPPR).

(26) *PHYSPROP (Physical/Chemical Property database)*; Syracuse Research Corporation: New York, 2011.

(27) Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J. Pharm. Sci.* **2001**, *90* (6), 749−784.

(28) Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *11* (Supplement2), S29−S39.

(29) Stanton, D. T.; Dimitrov, S.; Grancharov, V.; Mekenyan, O. G. Charged partial surface area (CPSA) descriptors QSAR applications. *SAR QSAR Environ. Res.* **2002**, *13* (2), 341−351.

(30) Stanton, D. T.; Jurs, P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. *Anal. Chem.* **1990**, *62* (21), 2323−2329.

(31) Burden, F.; Winkler, D. Bayesian regularization of neural networks. *Methods Mol. Biol.* **2008**, *458*, 25−44.

(32) Tarasova, A.; Burden, F.; Gasteiger, J.; Winkler, D. A. Robust modelling of solubility in supercritical carbon dioxide using Bayesian methods. *J. Mol. Graphics Modell.* **2010**, *28* (7), 593−597.

(33) Burden, F. R.; Winkler, D. A. Optimal sparse descriptor selection for QSAR using Bayesian methods. *QSAR Comb. Sci.* **2009**, *28* (6−7), 645−653.

(34) Burden, F. R.; Winkler, D. A. Robust QSAR models using Bayesian regularized neural networks. *J. Med. Chem.* **1999**, *42* (16), 3183−3187.

(35) Burden, F. R.; Winkler, D. A. An optimal self-pruning neural network and nonlinear descriptor selection in QSAR. *QSAR Comb. Sci.* **2009**, *28* (10), 1092−1097.

(36) Polley, M. J.; Winkler, D. A.; Burden, F. R. Broad-based quantitative structure-activity relationship modeling of potency and selectivity of farnesyltransferase inhibitors using a Bayesian regularized neural network. *J. Med. Chem.* **2004**, *47* (25), 6230−6238.

(37) Manallack, D. T.; Burden, F. R.; Winkler, D. A. Modelling inhalational anaesthetics using Bayesian feature selection and QSAR modelling methods. *ChemMedChem* **2010**, *5* (8), 1318−1323.

(38) Thomas, V. H.; Bhattachar, S.; Hitchingham, L.; Zocharski, P.; Naath, M.; Surendran, N.; Stoner, C.; El-Kattan, A. The road map to oral bioavailability: an industrial perspective. *Expert Opin. Drug Metab. Toxicol.* **2006**, *2* (4), 591−608.

(39) Yalkowsky, S. H. *Solubility and Solubilization in Aqueous Media*; Oxford University Press: New York, 1999; p 480.

(40) Ando, H. Y.; Radebaugh, G. W. *Remington: The Science and Practice of Pharmacy*; UotSi, Ed.; Lippincott, Williams and Wilkins: Philadelphia, USA, 2005.

(41) Charton, M.; Charton, B. I. The nature of topological parameters. II. The composition of topological parameters. *J. Comput.-Aided Mol. Des.* **2003**, *17* (2), 211−221.

(42) Gopal, R. Relation between lattice energy and melting points of some crystalline substances. 2. Alkali metals. *Z. Anorg. Allg. Chem.* **1955**, *278* (1−2), 42−45.

(43) Liu, Y.; Holder, A. J. A quantum mechanical quantitative structure-property relationship study of the melting point of a variety of organosilicons. *J. Mol. Graphics Modell.* **2011**, *31*, 57−64.

(44) Dearden, J. C. The QSAR prediction of melting point, a property of environmental relevance. *Sci. Total Environ.* **1991**, *109−110*, 59−68.

(45) Bhhatarai, B.; Teetz, W.; Liu, T.; Öberg, T.; Jeliazkova, N.; Kochev, N.; Pukalov, O.; Tetko, I. V.; Kovarich, S.; Papa, E.; Gramatica, P. CADASTER QSPR Models for Predictions of Melting and Boiling Points of Perfluorinated Chemicals. *Molecular Informatics* **2011**, *30* (2−3), 189−204.

(46) Karthikeyan, M.; Glen, R. C.; Bender, A. General Melting Point Prediction Based on a Diverse Compound Data Set and Artificial Neural Networks. *J. Chem. Inf. Model.* **2005**, *45* (3), 581−590.

(47) Guendouzia, A.; Mekelleche, S. M. Prediction of the melting points of fatty acids from computed molecular descriptors: A quantitative structure−property relationship study. *Chem. Phys. Lipids* **2012**, *165* (1), 1−6.

(48) Modarressi, H.; Dearden, J. C.; Modarressi, I. QSPR Correlation of Melting Point for Drug Compounds Based on Different Sources of Molecular Descriptors. *J. Chem. Inf. Model.* **2006**, *46*, 930−936.

(49) Nigsch, F.; Bender, A.; van Buuren, B.; Tissen, J.; Nigsch, E.; Mitchell, J. B. O. Melting Point Prediction Employing k-Nearest

Neighbor Algorithms and Genetic Parameter Optimization. *J. Chem. Inf. Model.* **2006**, *46* (6), 2412−2422.

(50) Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. Why Are Some Properties More Difficult To Predict than Others? A Study of QSPR Models of Solubility, Melting Point, and Log P. *J. Chem. Inf. Model.* **2008**, *48* (1), 220−232.

(51) Habibi-Yangjeh, A.; Pourbasheer, E.; Danandeh-Jenagharad, M. Prediction of melting point for drug-like compounds using Principal Component−Genetic Algorithm−Artificial Neural Network. *Bull. Korean Chem. Soc.* **2008**, *29*, 833−841.

(52) Lazzús, J. A. Neural Network Based on Quantum Chemistry for Predicting Melting Point of Organic Compounds. *Chin. J. Chem. Phys.* **2009**, *22* (1), 19−26.