QSPR Correlation of the Melting Point for Pyridinium Bromides, Potential Ionic Liquids

Alan R. Katritzky,*,[#] Andre Lomaka,[#,‡] Ruslan Petrukhin,[#,‡] Ritu Jain,[#] Mati Karelson,[‡]
Ann E. Visser,[§] and Robin D. Rogers[§]

Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P.O. Box 11720,
Gainesville, Florida 32611-7200, Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu, Estonia,
and Department of Chemistry and Center for Green Manufacturing, The University of Alabama,
P.O. Box 870336, Tuscaloosa, Alabama 35487-0336

In an attempt to develop predictive tools for the determination of new ionic liquid solvents, QSPR models for the melting points of 126 structurally diverse pyridinium bromides in the temperature range 30−200 °C were developed with the CODESSA program. Six- and two-descriptor equations with squared correlation coefficients (R^2) of 0.788 and 0.713, respectively, are reported for the melting temperatures. The models illustrate the importance of information content indices, total entropy, and the average nucleophilic reactivity index for an N atom.

INTRODUCTION

Ionic liquids (IL) are a class of novel compounds that are composed entirely of ions, melt at relatively low temperatures, and may be thought to resemble molten metallic ionic melts such as NaCl at 800 °C. In contrast to the molten salts, certain IL are moisture stable, have miniscule volatility, and are water immiscible, all characteristics of a unique class of potentially environmentally benign solvents.[1−3] When considered as a new type of solvent, many IL are highly solvating and noncoordinating which also broadens their scope of applications to include catalysis and synthesis[4] and hence the designation "alternative reaction media".[4−6]

As solvents, IL are capable of dissolving a myriad of materials, ranging from rocks and coal components[1] to spent nuclear fuel cladding.[7,8] As the extracting phase in liquid/liquid separations, the underlying chemical and physical properties of RTIL result in solute partitioning behavior that is similar to that observed utilizing traditional organic solvents.[9,10] Choosing the "best" IL for a particular application is currently problematic; the cation identity (imidazolium, pyridinium, quaternary ammonium, phosphonium), number of alkyl groups attached to the ring, type of substituents (methyl, methoxy, ethyl, propyl, isopropyl), and anion identity all contribute to the observed chemical and physical properties (e.g., melting point, viscosity, hydrophobicity, and water content).[10,11] Given the large number of possible IL, QSPR can be a valuable tool toward modeling and understanding IL systems and anticipating their physical properties.

Ionic liquids (ILs) may be defined as materials which incorporate at least one organic ion in an ion pair and are liquid at temperatures of ∼150 °C or lower.[12] The ILs are used in electrochemistry,[13] biochemistry,[14] and as solvents for synthesis and catalysis.[4,15−17] ILs have many useful properties[4,18] with benefits described as follows: (i) They are good solvents for a wide range of both inorganic and organic materials and allow unusual combinations of reagents to be brought into the same phase.[8,19] (ii) They are often composed of poorly coordinating ions, so they have the potential to be highly polar, yet noncoordinating solvents exhibiting solvatochromic properties rendering them similar to short chain alcohols.[10] (iii) They are immiscible with many organic solvents and many cations combined with anions such as PF_6^- and $N(SO_2CF_3)_2^-$ can sustain a biphasic system where the IL is the extracting phase instead of a traditional organic solvent.[9,20,21] (iv) They have relatively low viscosity, high thermal stability, and a wide temperature range for the liquid phase which encompasses that for water and ammonia.[1] (v) Many ionic liquids are nonvolatile and hence may be used in high-vacuum systems.

Combined, these properties are significant for the consideration of IL as a new class potentially "green" solvents. In particular, the nonvolatile nature of many IL[22,23] could eliminate problems associated with the use of traditional volatile organic solvents which may pose health, fire, and environmental hazards.[24] It must be understood, however, that all of the risks associated with utilizing IL as solvents are not yet known. Toxicity, economic, and environmental impact data are still needed for classes of IL, and research in these areas is currently in progress worldwide. Nonetheless, if it were possible to predict melting points and physical properties of new IL, detailed and expensive analyses of the safety, health, and environmental hazards could be concentrated on only a few of the most promising IL.

Melting point models have been previously developed for some organic compound data sets.[25] In our group, a QSPR model was developed for the melting points of a large set of mono- and disubstituted benzenes (443 compounds).[26] The most significant parameter in the model was the hydrogen

* Corresponding author phone: (352)392-0554; fax: (352)392-9199; e-mail: katritzky@chem.ufl.edu.
[#] University of Florida.
[‡] University of Tartu.
[§] The University of Alabama.

bonding descriptor *(HDSA2*, atom-charge-related descriptor based on the Zefirov's electronegativity equalization scheme), but quantum chemical descriptors were also found to be of significant utility. A four-parameter model has been developed by P. Gramatica et al. to predict the melting points (range: 16.5−310.0 °C) of polychlorinated biphenyls using WHIM descriptors.[27] The model derived includes size variables (nondirectional A_m and T_u, directional L_{2v} and symmetry variable G_{1p}).

The current work aims to develop QSPR models for the explanation and prediction of the melting points of IL. Successful development of such models will lead to more rapid development and testing of many new potentially environmentally benign designer solvents.

DATA AND COMPUTATIONAL METHODS

The training set for the study was created from the data available from the Beilstein database.[28] For many compounds, the Beilstein database contains several entries for the melting temperature from different literature sources. The agreement between those data entries was used to estimate the reliability of the data.

The following methodology was used to prepare the training set. The temperature range from 0 to 200 °C was divided into 10 degree subintervals, and a search for pyridinium bromides in the Beilstein database[28] was performed for each of those subintervals. A maximum of 10 compounds was selected in each of the subintervals. The compounds with a lower molecular weight were preferred in the selection to minimize the computation time necessary for semiempirical calculations. After excluding duplicate structures from the whole collected data set, the final set of 126 compounds (Table S1) was produced. Within this data set, 32 compounds had several data entries for the melting temperature. In those cases the most recent data entry was used. In most cases, the entries originating from different literature sources were reasonably close showing differences less than 10 °C. In only three cases the melting temperatures differed by more than 50 °C.

The bromide ions were removed in all cases and then the cationic part was optimized using MM+ method of Hyperchem.[29] Final optimizations were performed with MOPAC using the AM1 semiempirical method. AM1 force calculations were also performed to produce thermodynamic parameters. These calculation results gave us the ability to perform QSPR analysis using CODESSA software package.[30] The CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) program computes the descriptor pool of about 500 descriptors including various constitutional, topological, geometrical, electrostatic, quantum chemical, and thermodynamic descriptors. CODESSA has been successfully employed in QSPR studies of many physical properties including boiling points, critical temperatures, vapor pressures, and densities of structurally diverse organic compounds and melting points of some narrow classes of organic compounds.[31] The CODESSA software package includes a tool for effective descriptor generation and several statistical methods for the development of QSPR equations.

Descriptors were calculated for all compounds and also for fragments obtained by removing the pyridinium ring from the cation. The fragment descriptors are further denoted as
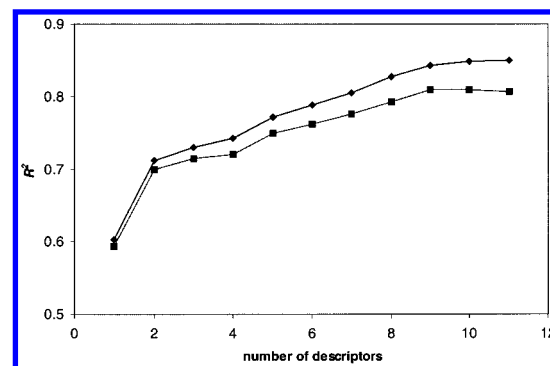
Figure 1. Plot of R^2 and cross-validated R^2 against the number of descriptors.

Table 1. Correlation Coefficients of Descriptors with the Remaining Descriptors in the Six-Descriptor Model

descriptor	R^2
f-average complementary information content (order 0)	0.399
average nucleophilic reactivity index for a N atom	0.062
average information content (order 2)	0.320
lowest normal mode vibrational frequency	0.277
f-relative number of rings	0.191
min. Coulombic interaction for a C−H bond	0.180

having the prefix f. Next, the resulted descriptor matrices were analyzed using the CODESSA internal statistical engine.

RESULTS AND DISCUSSION

The CODESSA heuristic method[32] of selection of the "best" regression model was applied to the complete set of 126 structures. The heuristic method for descriptor selection proceeds with a preselection of descriptors by eliminating sequentially descriptors which do not match any of the following criteria: (i) Fisher *F*-criterion greater than one unit; (ii) R^2 value less than a value defined at the start (by default 0.01); (iii) Student's *t*-criterion less than that defined (by default 0.1); and (iv) duplicate descriptors having a higher squared intercorrelation coefficient than a predetermined level (usually 0.8), retaining the descriptor with higher R^2 with reference to the property.

The remaining descriptors are then listed in decreasing order of the correlation coefficients when used in a global search for two-parameter correlations. Each significant two-parameter correlation by *F*-criterion is recursively expanded to an *n*-parameter correlation until the normalized *F*-criterion remains greater than the startup value. The top *N* correlations by R^2 as well as *F*-criterion are saved.

The heuristic correlations performed for the whole set provide the optimal equations for different numbers of descriptors in the range of 1−11. The descriptors involved in these correlations are listed in Table S2 in order of their statistical significance according to *t*-test. Plots of R^2 and R^2_{cv} values against the number of descriptors, given in Figure 1, provide guidance regarding the number of descriptors to retain in the model. Figure 1 reveals that most of the increase in the R^2 value was obtained in the first two models. The best two-parameter model has $R^2 = 0.713$, $s = 26.4$ K and contains four outliers according to 95% confidence limit (structures 93, 120, 121, and 123 in Table S1). The next step was studying intercorrelations between descriptors starting from the nine-descriptor model, where R^2 levels off

Table 2. Two- and Six-Parameter Correlations of Melting Points of Pyridinium Bromides by the Heuristic Method

| #P	R^2	R^2_{cv}	F		$X + \Delta X$	t-test	name of the descriptor
2	0.7125	0.7002	151.16	0	$-(1.91 \pm 0.20)\cdot10^2$	-9.6301	intercept
				1	$(6.74 \pm 0.45)\cdot10$	15.0260	total entropy (300 K)/# of atoms
				2	$(2.84 \pm 0.42)\cdot10^3$	6.8270	average nucleophilic reactivity index for a N atom
6	0.7883	0.7624	73.24	0	$(8.25 \pm 1.50)\cdot10^2$	5.5036	intercept
				1	-4.79 ± 0.29	-16.6574	f-average complementary information content (order 0)
				2	$(2.21 \pm 0.37)\cdot10^2$	5.9478	average nucleophilic reactivity index for a N atom
				3	-4.48 ± 0.81	-5.5184	average information content (order 2)
				4	$-(3.69 \pm 0.87)\cdot10^2$	-4.2365	lowest normal mode vibrational frequency
				5	$-(1.60 \pm 0.40)\cdot10^1$	-4.0163	f-relative number of rings
				6	$-(1.15 \pm 0.39)\cdot10^1$	-2.9450	min. Coulombic interaction for a C−H bond
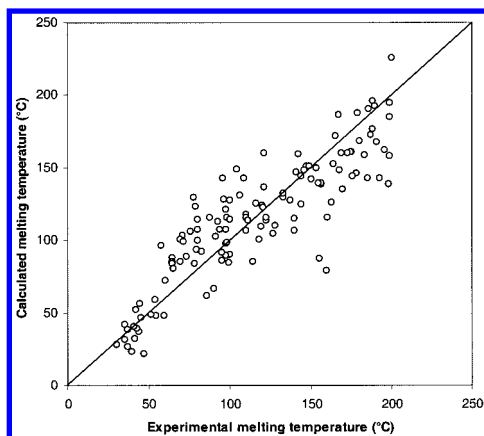
Figure 2. Calculated vs experimental melting temperatures for the set of 126 pyridinium bromides in the six-descriptor model.

and R^2_{cv} is maximized. To this end, a multilinear least-squares analysis was performed for each descriptor with respect to the rest of the descriptors in nine-, eight-, seven- and six-parameter models. The analysis showed that the six-parameter equation was the first model where all R^2 values remained below 0.5. These R^2 values are given in Table 1. The correlation coefficients for the descriptors in two- and six-parameter models are given in Table 2. The best correlation with six descriptors has $R^2 = 0.788$, $F = 73.24$, and $s = 23.0$ K. The melting temperatures predicted by the six-parameter equation are included in Table S1, and the corresponding correlation chart is given in Figure 2. Six data points in this model lie outside the range of ±2 s (95% confidence limit) from the predicted value (structures 31, 47, 59, 93, 96, 120, and 122).

The most significant descriptor in most of the models of Table S2 is the zeroth order average complementary information content ($^0\overline{CIC}$). This descriptor and the second-order average information content ^2IC describe the connectivity and branching in a molecule and can be related to molecular shape and symmetry. The relative number of rings in the fragments can also be related to molecular shape. The negative regression coefficients for $^0\overline{CIC}$ and $^2\overline{IC}$ reflect the fact that cations with lower symmetry have weaker coordination ability that leads to lower melting temperatures. The second most significant descriptor in both two- and six-descriptor models is the average nucleophilic reactivity index for an N atom. This descriptor is likely connected to electrostatic intermolecular interactions including multipole interactions involving nitrogen atom of the pyridinium ring. The lowest normal mode vibrational frequency can be associated with the flexibility of the molecule. The most

Table 3. Validation of the Six-Descriptor Melting Temperature Correlation Equation

subset	training R^2	standard training error	prediction R^2	standard prediction error
A+B	0.8002	23.71	0.7570	24.05
A+C	0.7857	23.29	0.7785	25.53
B+C	0.7905	22.55	0.7691	27.10
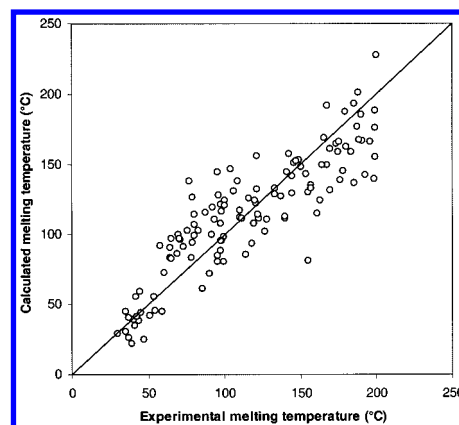
Figure 3. Cross-validation plot for the six-descriptor model.

significant descriptor in the two-parameter model, the total entropy per atom, can be related to an increase in the conformational and rotational degrees of freedom for the molecules in the liquid phase.

VALIDATION

To validate the six-descriptor equation developed for the prediction of melting temperatures, the whole set of compounds was divided into three subsets A, B, and C. Structures 1, 4, 7, etc. in Table S1 formed subset A, structures 2, 5, 8, etc. formed subset B, and structures 3, 6, 9, etc. formed subset C. Correlation models were obtained for the subsets A+B, A+C, and B+C. The resulting new correlation models were used in turn to predict the melting temperatures for subsets C, B, and A, respectively. Correlation coefficients and standard errors of fit and prediction are presented in Table 3. The correlation chart of the validation showing the summary of all three predictions is given in Figure 3.

CONCLUSIONS

A six-descriptor equation with a reasonably good correlation has been developed for the prediction of the melting temperatures of pyridinium bromides based on the correlation analysis of the melting point data for a diverse set of 126 compounds. The following descriptors appeared to be the most significant contributors to the model: (i) two informa-

tion content indices reflecting the coordination ability of a cation, (ii) the average nucleophilic reactivity index for an N atom related to electrostatic intermolecular interactions, and (iii) total entropy per atom related to the difference in conformational and rotational degrees of freedom in solid and liquid phases.

Because of the dominating role of the cation in determining the properties of ionic liquids, the developed models can help to screen candidate lead compounds in search for new potential ILs or to suggest directions for new lead molecules. The technique used in this work can be extended by developing models for different sufficiently narrow classes of cations commonly occurring in known ionic liquids, such as pyridinium, phosphonium, and ammonium cations.

ACKNOWLEDGMENT

Supporting Information Available: Table S1, experimental and predicted melting temperatures for the set of 126 pyridinium bromides, and Table S2, descriptor names, correlation coefficients, and statistical parameters for the optimal equations derived for different numbers of descriptors in the range of 1−11. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Freemantle, M. Designer Solvents. *Chem. Eng. News* **1998**, *76*, 13, 32−37.

(2) Holbrey, J. D.; Seddon, K. R. The Phase Behaviour of 1-alkyl-3-methylimidazolium Tetrafluoroborates; Ionic Liquids and Ionic Liquid Crystals. *J. Chem. Soc., Dalton Trans.* **1999**, 2133−2139.

(3) Suarez, P. A. Z.; Einloft, S.; Dullius, J. E. L.; de Souza, R. F.; Dupont, J. Synthesis and Physical − Chemical Properties of Ionic Liquids Based on 1-*n*-butyl-3-methylimidazolium Cation. *J. Chim. Phys.* **1998,** *95*, 1626−1639.

(4) Welton, T. Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis. *Chem. Rev.* **1999**, *99*, 2071−2083.

(5) Hussey, C. L. Room-Temperature Haloaluminate Ionic Liquids. Novel Solvents for Transition Metal Solution Chemistry. *Pure Appl. Chem.* **1988,** *60*, 1763−1772.

(6) Wilkes, J. S.; Levisky, J. A.; Wilson, R. A.; Hussey, C. L. Dialkylimidazolium Chloroaluminate Melts: A New Class of Room-Temperature Ionic Liquids for Electrochemistry, Spectroscopy, and Synthesis. *Inorg. Chem.* **1982**, *21*, 1263−1264.

(7) Fields, M.; Hutson, G. V.; Seddon, K. R.; Gordon, C. M. Ionic Liquids as Solvents. *World Pat.* WO 98/06106, **1998**, CAN 128: 185727.

(8) Thied, R. C.; Seddon, K. R.; Pitner, W. R.; Rooney, D. W. Nuclear Fuel Reprocessing. *World Pat.* WO 99/41752, **1999**, CAN 131: 150588.

(9) Huddleston, J. G.; Willauer, H. D.; Swatloski, R. P.; Visser, A. E.; Rogers, R. D. Room-Temperature Ionic Liquids as Novel Media for 'Clean' Liquid−Liquid Extraction. *Chem. Commun.* **1998**, 1765−1766.

(10) Huddleston, J. G.; Visser, A. E.; Reichert, W. M.; Willauer, H. D.; Broker, G. A.; Rogers, R. D. Characterization and Comparison of

(11) Bonhôte, P.; Dias, A.-P.; Papageourgiou, N.; Kalyanasundaram, K.; Grätzel, M. Hydrophobic, Highly Conductive Ambient-Temperature Molten Salts. *Inorg. Chem.* **1996**, *35*, 1168−1178.

(12) Seddon, K. R. Ionic Liquids for Clean Technology. *J. Chem. Technol. Biotech.* **1997**, *68*, 351−356.

(13) McFarlane, D. R.; Sun, J.; Golding, J.; Meakin, P.; Forsyth, M. High Conductivity Molten Salts Based on the Imide Ion. *Electrochim. Acta* **2000**, *45*, 1271−1278.

(14) Cull, S. G.; Holbrey, J. D.; Vargas-Mora, V.; Seddon, K. R.; Lye, G. J. Room-Temperature Ionic Liquids as Replacements for Organic Solvents in Multiphase Bioprocess Operations. *Biotech. Bioeng.* **2000**, *69*, 228−233.

(15) Olivier, H. Recent Developments in the Use of Non-Aqueous Ionic Liquids for Two-Phase Catalysis. *J. Mol. Catalysis A: Chemical* **1999**, *146*, 285−289.

(16) Fischer, T.; Sethi, A.; Welton, T.; Woolf, J. Diels−Alder Reactions in Room-Temperature Ionic Liquids. *Tetrahedron Lett.* **1999**, *40*, 793−796.

(17) Lee, C. W. Diels−Alder Reactions in Chloroaluminate Ionic Liquids: Acceleration and Selectivity Enhancement. *Tetrahedron Lett.* **1999**, *40*, 2461−2464.

(18) Wasserscheid, P.; Keim, W. Ionic Liquids − New "Solutions" for Transition Metal Catalysis. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*, 3773−3789.

(19) Blanchard, L. A.; Brennecke, J. F. Recovery of Organic Products from Ionic Liquids Using Supercritical Carbon Dioxide. *Ind. Eng. Chem. Res.* **2001**, *40*, 287−292.

(20) Visser, A. E.; Swatloski, R. P.; Reichert, W. M.; Griffin, S. T.; Rogers, R. D. Traditional Extractants in Nontraditional Solvents: Group 1 and 2 Extraction with Crown Ethers in Room-Temperature Ionic Liquids. *Ind. Eng. Chem. Res.* **2000**, *39*, 3596−3604.

(21) Visser, A. E.; Swatloski, R. P.; Griffin, S. T.; Hartman, D. H.; Rogers, R. D. Liquid/Liquid Extraction of Metal Ions in Room-Temperature Ionic Liquids. *Sep. Sci. Technol.* **2001**, *36*, 785−804.

(22) Anastas, P. T. The United States Green Chemistry Program. *Critical Rev. Anal. Chem.* **1999**, *29*, 267−268.

(23) United States Environmental Protection Agency. World Wide Web, http://www.epa.gov/opptintr/greenchemistry/whatis.htm.

(24) United States Environmental Protection Agency. World Wide Web, http://www.epa.gov/epahome/laws.htm.

(25) Katritzky A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. Structurally Diverse Quantitative Structure − Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1−18.

(26) Katritzky A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 913−919.

(27) Gramatica, P.; Navas, N.; Todeschini, R. 3D-modelling and Prediction by WHIM Descriptors. Part 9. Chromatographic Relative Retention Time and Physico-Chemical Properties of Polychlorinated Biphenyls (PCBs). *Chemometrics Intelligent Lab. Syst.* **1998**, *40,* 53−63.

(28) The Beilstein Database. MDL Information Systems GmbH. World Wide Web, http://www.Beilstein.com.

(29) Hypercube, Inc. World Wide Web, http://www.hyper.com.

(30) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA Reference Manual. Version 2.0*; University of Florida, 1996.

(31) Karelson, M.; Maran, U.; Wang, Y. QSPR and QSAR Models Derived Using Large Molecular Descriptor Spaces. A Review of CODESSA Applications. *Collect. Czech. Chem. Commun.* **1999**, *64*, 1551−1571.

(32) Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley: New York, 2000.