

QSPR MODELING OF METAL HALIDES LATTICE ENTHALPIES

Castro, E. A.

INIFTA, Department of Chemistry, Faculty of Exact Sciences,
La Plata National University, Suc. 4, C.C. 16, La Plata (B1906ZAA), Argentina

Fax: 54-221-4254642, E-mail: castro@quimica.unlp.edu.ar

Toropova, A. P.; Toropov, A. A.; Mukhamedjanova, D. V.

Uzbekistan Republic, Academy of Sciences Geology & Geophysics Institute, 700125,
Abdullaev Street 41, Tashkent, Uzbekistan

*Received September 14th, 2005. In final form March 31st, 2006
Dedicated to Prof. Imre G. Csizmadia on the occasion of his 75th birthday*

Abstract

By means of optimisation correlation of local invariant weights in graph of atomic orbitals, one-variable quantitative structure-property relationships have been obtained. Statistical characteristics of the best of such a model are the following: $n = 13$, $r = 0.9999$, $s = 1.21$ kJ/mol, $F = 55060$ (training set); $n = 7$, $r = 0.9990$, $s = 7.68$ kJ/mol, $F = 2485$ (test set)

Resumen

Se ha obtenido una relación estructura-propiedad cuantitativa unidimensional utilizando la optimización de la correlación local invariante de pesos en gráficos de orbitales atómicos. Las características estadísticas de dicho modelo son las siguientes: $n = 13$, $r = 0,9999$, $s = 1,21$ kJ/mol, $F = 55060$ (conjunto de preparación); $n = 7$, $r = 0,9990$, $s = 7,68$ kJ/mol, $F = 2485$ (conjunto de prueba)

Introduction

With the exception of the inert gases, all the elements in the periodic table form halides, often in several oxidation states, and halides generally are quite important and common compounds. There are almost as many ways of classifying halides as there are types of halides, and there are many. There are not only binary halides which can range from simple molecules, ionic or molecular lattices to complicated polymers but also oxyhalides, hydroxy halides, and other complex halides of various structural types.

Metal halides are substances of predominantly ionic character, although partial covalence is important in some of them. On the whole there is a uniform gradation from halides which are for all practical purposes purely ionic to those of intermediate character to those essentially covalent.

As a rough guide we can consider those halides in which the lattice consists of discrete ions rather than definite molecular units to be basically ionic, although there may still be considerable covalence in the metal-halogen interaction. The size and polarizability of the halide ion is also important in determining the character of the halide. Thus we have the rather classic case of the aluminium halides, where AlF_3 is basically ionic, whereas AlCl_3 , AlBr_3 , and AlI_3 exist as covalent dimers.

Metal fluorides have a higher heat of formation than the remaining alkaline metal halides and this feature is due to the reduced heat of formation of the F_2 molecule and the high lattice energy of the compounds themselves. Heat of formation of fluorides decreases when the cation volumes increase, but in the remaining halides this behaviour is just the opposite. This is due to the fact that the magnitude $1/(r_{\text{anion}} + r_{\text{cation}})$ decreases rapidly in the fluoride series due to the smaller anion size. This change has more influence than the decrease of the ionisation potentials and sublimation energies of the corresponding metals.

Heat of formation of chlorides, bromides and iodides increase descending along I, II, IIIA, IVA and VA groups. In the remaining groups of the periodic table in general the inverse behaviour is observed, although there are some interesting enough irregularities, specially in the first two periods. The heat of formation of hypothetical halides can be calculated by means of standard treatments. Numerical models of physical and chemical properties of metal halides in general and their enthalpies in particular have many interesting applications in so diverse fields such as inorganic chemistry, physical chemistry and geochemistry [1].

Study of structure-property and structure-activity relationship continues to attract considerable attention in chemical literature. Various statistical methods have been found useful in such studies, including the Principal Component Analysis [2], the Pattern Recognition [3], the Partial Square method [4], the Artificial Neural Networks [5]. The oldest data reduction method, the Multiple Regression Analysis (MRA) [6], is currently widely used. Most molecular descriptors used in MRA are characterized by fixed numerical values, which are independent of the property/activity under consideration [7]. To better describe the property/activity studies, Randić proposed the employment of some flexible descriptors [8-11] to improve regression results, with the idea that a variable parameter undergoes change during the regression analysis.

Hydrogen-suppressed and hydrogen-filled molecular graphs, which are a particular sort of valuable resource to define variable descriptors, are used as a particular representation of molecular structure in quantitative structure-property relationships (QSPR) of organic compounds, but they are not good tools to encode the main features of molecular structure of inorganic substances such as different crystalline arrangements. However, these substances are object of research works in mineralogy, geochemistry, and geology itself. Under such circumstances, it is preferable to use an alternative set of variable descriptors, recently described as graph of atomic orbitals (GAO) [1, 12-15].

The present study aims to estimate the capabilities of using the GAO approach in QSPR modelling of enthalpies of metal halides and to compare the results with available experimental data.

Method

Two types of molecular graphs are under consideration in this study. The first type calculated with classical molecular graphs (MG) of two vertices (metal and halogen) and one edge. Descriptors for such graphs are calculated as

$$D(\text{MG}) = \sum_{k=1}^N (\text{CW}(A_k) \times \text{CW}(\delta_k)) \quad (1)$$

where A_k denotes a metal or a halogen, N is the number of vertices and CW stands for Correlation Weights. The δ_k is the vertex degree in MG . In the case of the metal halides under consideration, $\delta_k \equiv 1$.

The second type of molecular graph is the GAO. GAO can be obtained from 'classical' graphs by means of a change of atoms by groups of atomic orbitals and the subsequent reconstruction of the adjacency matrix. The procedure was described in detail in Ref. [14], so that we deem it is not necessary to repeat them here. Descriptors for such graphs are obtained by means of the following formula

$$D(\text{GAO}) = \sum_{k=1}^N \text{CW}(\text{AO}_k) + \sum_{k=1}^N \text{CW}(\delta_k) \quad (2)$$

where AO_k is an atomic orbital (i. e., $1s^1, \dots, 2p^3, \dots, 3d^{10}, \dots$), δ_k is the vertex degree of the k -th vertex in the GAO, and N is total number of vertices (atomic orbitals) in the GAO.

Results and discussion

Atomic orbital groups on each atom under consideration are listed in Table 1.

Table 1. Atoms and atomic orbital groups for alkaline and halogen atoms.

Atom	Atomic orbital group
Li	$1s^2, 2s^1$
F	$1s^2, 2s^2, 2p^5$
Na	$1s^2, 2s^2, 2p^6, 3s^1$
Cl	$1s^2, 2s^2, 2p^6, 3s^2, 3p^5$
K	$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 4s^1$
Br	$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^5$
Rb	$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^6, 5s^1$
I	$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^6, 4d^{10}, 5s^2, 5p^5$
Cs	$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^6, 4d^{10}, 5s^2, 5p^6, 6s^1$

Numerical values of the CWs of Eq.(1) or Eq. (2) can be obtained by the method of Monte Carlo optimization [15], which produces the largest as possible correlation coefficient between the descriptors and enthalpies. After the CWs are determined one can calculate $D(\text{MG})$ or $D(\text{GAO})$ for metal halides of the training set, and by means of the Least Squares method one calculates C_0 and C_1 for the linear model

$$H = C_0 + C_1 \times D(G) \quad (3)$$

where G stands for MG or GAO.

One can estimate predictive potential of the Eq. (3) with substances of the test set. Results of three probes for such optimization procedure with MG and GAO are presented in Table 2. Similar final results are obtained for each probe, so that we present data for just one of them. It can be seen from Table 2 that GAO based models of enthalpies are better than MG based models. The GAO based model of enthalpy calculation obtained in the first probe of the optimization procedure is the following

$$\Delta H_L^\circ (\text{kJ/mol}) = 533.43 + 7.323 \times D(\text{GAO}) \quad (4)$$

It can be seen that in spite of the large number of optimised parameters (29), the most important attribute of the model is the lack of any information on metal halides belonging to the test set in the process of obtaining CWs (Table 3) as well as the model to get Eq. (4). Calculation of the enthalpy with Eq.(4) is displayed in Table 4. Experimental data have been taken from Ref. [12]. The percent deviations are rather small, so that the predictions are very accurate (see sixth column in Table 4).

Table 2. Statistical characteristics of models of lattice enthalpies of metal halides obtained in three probes of optimization of the correlation weights in the MG and results of three such probes on correlation weights in the GAO. Here *r*, *s*, *F*, and *n* are correlation coefficient, standard error, Fischer *F*-ratio, and number of compounds in the set, respectively.

Training set, n = 13			Test set, n = 7			All metal halides, n = 20		
r*	s*	F*	r	s	F	r	s	F
<i>QSPR based on classic molecular graph (MG), 10 CWs optimized</i>								
0.9892	12.514	502	0.9674	43.270	73	0.9708	26.271	295
0.9893	12.493	504	0.9736	41.577	91	0.9733	25.386	323
0.9892	12.524	501	0.9724	41.987	87	0.9728	25.608	317
<i>QSPR based on graph of atomic orbitals (GAO), 29 CWs optimized</i>								
0.9999	1.208	55060	0.9990	7.678	2485	0.9993	4.420	12280
0.9999	1.209	54998	0.9989	7.794	2326	0.9992	4.484	11744
0.9999	1.207	55184	0.9991	7.372	2675	0.9993	4.252	13217

* *r* = regression coefficient, *s* = standard deviation, *F* = Fisher coefficient

Table 3. Correlation weights of local invariants in MG and GAO on three probes of the Monte Carlo optimization method.

Invariant	Correlation weight		
	Probe 1	Probe 2	Probe 3
<i>Atoms in classical molecular graph</i>			
Li	1.03870	1.03480	1.01493
F	1.04000	1.03610	1.01750
Na	1.02312	1.01970	1.00500
Cl	1.00357	1.00235	0.99498
K	0.99618	0.99375	0.98753
Br	0.99749	0.99592	0.98992
I	0.97632	0.97755	0.97761
Rb	0.98734	0.98746	0.98245
Cs	0.97510	0.97633	0.97510
<i>Vertex degree in the MG ($\delta_k \equiv 1$)</i>			
0001	0.97877	1.02010	1.00495
<i>Atomic orbitals in the GAO</i>			
1s²	3.43741	3.72242	4.00030
2s¹	9.73563	9.75210	10.01208
2s²	1.76161	1.95397	2.08304
2p⁵	9.25218	9.49227	9.89069
2p⁶	0.44504	0.44943	0.63368
3s¹	6.46161	6.21753	6.19379
3s²	0.25752	0.16496	0.20073
3p⁵	3.62485	3.68319	3.60195
3p⁶	0.62491	0.58201	0.50928
3d¹⁰	0.47608	0.56669	0.53403
4s¹	1.68541	1.43652	1.57951
4s²	0.53967	0.48002	0.50084
4p⁵	1.51444	1.55596	1.50496
4p⁶	0.39830	0.42293	0.30058
4d¹⁰	0.16733	0.13953	0.16601
5s¹	2.28638	2.22994	2.38266
5s²	0.16679	0.14367	0.16006
5p⁵	0.41282	0.41451	0.38563
5p⁶	0.15455	0.14161	0.17758
6s¹	0.16211	0.14918	0.17068

Vertex degrees (δ_k) in the GAOs

0002	2.75304	2.84486	2.94301
0003	2.10578	2.12581	2.18399
0004	1.94673	2.01932	2.08281
0005	1.31370	1.31567	1.34344
0006	1.52068	1.57429	1.57582
0008	1.02412	1.01689	1.03744
0009	0.90996	0.93548	0.95119
0011	0.54134	0.52282	0.53131
0012	0.80873	0.82093	0.82911

Table 4. Calculation of the Enthalpy with Eq. (4).

ID	Metal halides	D(GAO)	$\Delta H_L^\circ(\text{exp})$ (kJ/mol)	$\Delta H_L^\circ(\text{calc})$ (kJ/mol)	ΔH_L° (exp-calc)
	<i>Training Set</i>				
1	NaF	53.6	926.0	926.1	-0.1
2	KF	39.3	821.0	821.0	0.0
3	RbF	34.7	789.0	787.4	1.6
4	CsF	29.8	750.0	751.5	-1.5
5	LiCl	43.5	852.0	852.2	-0.2
6	NaCl	34.4	786.0	785.7	0.3
7	RbCl	22.1	695.0	695.3	-0.3
8	LiBr	38.4	815.0	814.8	0.2
9	RbBr	18.7	668.0	670.5	-2.5
10	CsBr	16.1	654.0	651.5	2.5
11	LiI	31.0	761.0	760.8	0.2
12	NaI	23.5	705.0	705.3	-0.3
13	RbI	13.4	632.0	631.9	0.1
	<i>Test Set</i>				
1	LiF	67.5	1037.0	1027.9	9.1
2	KCl	25.3	717.0	718.5	-1.5
3	CsCl	19.0	678.0	672.5	5.5
4	NaBr	30.0	752.0	753.3	-1.3
5	KBr	22.2	689.0	695.9	-6.9
6	KI	17.7	649.0	662.7	-13.7
7	CsI	11.7	620.0	618.9	1.1

Conclusions

The enthalpies of metal halides are defined by the structure features of metal and halide atoms (i.e., by the presence of different types of atomic orbitals), since the statistical characteristics of the models based on information of structure atoms are better than those of models based only on information of the presence of metal and halogen atoms. We have outlined a novel way of

deriving powerful structure-property models to calculate enthalpies of metal halides lattices. Even though the approach has been demonstrated here for one physical-chemical property of metal halides, it is general and can be applied to the analysis of other properties for different sort of molecules, as shown in several previous articles (see, for example, references [16-20]).

The advantage of the outlined approach is that it yields regressions accompanied with considerably smaller standard error than those given by similar studies using standard molecular descriptors (i.e. rigid molecular descriptors). The “flexibility” of the molecular descriptor, such as Graph of Atomic Orbitals employed in this study, makes it possible to describe very accurately metal halides lattice enthalpies. At present we are extending this treatment to other molecular sets in order to model different physical-chemical properties. Results will be presented elsewhere in the forthcoming future.

Acknowledgements

The authors are indebted to a reviewer for suggestions.

References

- [1] Jenkins, H. D. B.; Glasser, L., *Inorg. Chem.* **2002**, *41*, 4378.
- [2] Hotelling, H., *J. Educ. Psychol.* **1993**, *24*, 417.
- [3] Wold, S.; Sjöström, M., in *Chemometrics: Theory and Applications*, ACS Symp. Ser. No. 52, Kowalski, B. R., Ed., American Chemical Society, Washington, D.C., **1977**, p.243.
- [4] Wold, S.; Sjöström, M.; Eriksson, L., Partial least squares projections to latent structures (PLS) in chemistry, in *Encyclopedia of Computational Chemistry*, von Schleyer, R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds., Wiley, Chichester, England, **1998**, 2006-2021.
- [5] Zupan, J., Neural networks in chemistry, in *Encyclopedia of Computational Chemistry*, von Schleyer, R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds., Wiley, Chichester, England, **1998**, 1813-1827.
- [6] Malinowski, E.R.; Howery, D.G., *Factor Analysis in Chemistry*, Wiley-Interscience, New York, **1991**.
- [7] Randić, M., *New J. Chem.* **2000**, *24*, 165-171.
- [8] Randić, M., *Chemom. Intel. Labl. Syst.* **1991**, *10*, 213-223.
- [9] Randić, M., *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 970-980.
- [10] Randić, M.; Pompe, M., *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 631-638.
- [11] Randić, M.; Basak, S. C., *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614-618.
- [12] Pogliani, L., *Croat. Chem. Acta*, **2004**, *77*, 193-201
- [13] Castro, E. A.; Toropova, A. P.; Toropov, A. A.; Mukhamedjanova, D. V., *J. Argent. Chem. Soc. (this Journal)* **2003**, *91*, 85-90.
- [14] Mercader, A.; Castro, E. A.; Toropov, A. A., *Chem. Phys. Lett.* **2000**, *330*, 612-623.
- [15] Toropova, A. P.; Toropov, A. A., *Russ. J. Struct. Chem.* **2001**, *42*, 1230-1232.
- [16] Pérez González, M.; Toropov, A. A.; Duchowicz, P. R.; Castro, E. A., *Molecules* **2004**, *9*, 1019-1033.
- [17] Castro, E. A.; Torrens, F.; Toropov, A. A.; Nesterov, I. V.; Nabiev, O. M., *Molec. Simul.* **2004**, *30*, 691-696.

-
- [18] Castro, E. A.; Toropov, A. A.; Nesterova, A. I.; Nazarov, A. U., , *J. Theor. Comput. Chem.* **2004**, *3*, 31-41.
- [19] Duchowicz, P. R.; Castro, E. A.; Toropov, A. A.; Nesterov, I. V.; Nabiev, O. M., *Molec. Diversity* **2004**, *8*, 325-330.
- [20] Nesterov, I. V.; Toropov, A. A.; Duchowicz, P. R.; Castro, E. A., *TheScientificWorld Journal* **2004**, *4*, 956-964.