Estimation of enthalpies of formation of organometallic compounds from their molecular structures

Jesús Jover [a], Ramón Bosque [a], José A. Martinho Simões [b], Joaquim Sales [a,*]

[a] *Departament de Química Inorgànica, Universitat de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain*
[b] *Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal*

Abstract

A quantitative structure–property relationship (QSPR) was developed, aiming to estimate the gas-phase enthalpies of formation ($\Delta_f H^0$) of a set of 132 organometallic compounds of general formula $MR_n X_{n-m}$, where M is a metal or a semimetal from groups 12 to 16, R is an alkyl, aryl, alkenyl, or alkynyl group, and X is Cl, Br, I, or H. The proposed model, derived from multilinear regression, contains nine descriptors that can be readily calculated from molecular structures. Correlations with R^2 and RMSE of 0.988 (29.1) and 0.990 (30.2) for the training and prediction sets, respectively, are obtained. The ability of QSPR methods to estimate reliable values of enthalpies of formation has been confirmed by the results obtained with a set of 168 organic compounds, which contain the same type of groups of the organometallic compounds. The nine descriptors-derived model, containing only descriptors of the constitutional, topological, and geometrical types, predicts $\Delta_f H^0$ with accuracies comparable to well established additive methods.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Enthalpy of formation; Organometallic compounds; QSPR

1. Introduction

The standard enthalpy of formation, $\Delta_f H^0$, is a basic property of any chemical compound, giving information about its thermodynamic stability and allowing the calculation of enthalpies of reaction. For organic compounds, the experimental enthalpy of formation database is fairly large. In addition, several empirical methods to estimate $\Delta_f H^0$ with good accuracy have been proposed. The most used of these empirical procedures, known as *additive methods*, are based on the structural similarity of molecules and the use of fixed values for given atomic groups or chemical bonds. The most frequently applied, known as the Benson group method, was proposed in 1958 by Benson and Buss [1]. The original parameters have been subsequently extended and refined [2]. Another approach is the Laidler method [3]. In this case, the parameters are assigned to the

chemical bonds, and therefore reflect the strengths of those bonds. This method has been refined for several families of hydrocarbons [4]. Finally, a method developed by Pedley [5] is also based on the additivity of bond enthalpies but includes terms that account for "molecular fragment enthalpies", "bond interactions", and "ring interactions".

The size of the experimental enthalpy of formation database for organometallic compounds is significantly smaller than that for organic compounds. This stems from the fact that the traditional experimental technique used to investigate the thermochemistry of organic compounds (static bomb combustion calorimetry) is usually not suitable to probe organometallic compounds, as the final products of their combustion reactions are not, in general, well characterized. For example, the combustion of an organometallic compound may lead to non-stoichiometric metal oxides, requiring a detailed analysis of the reaction. As a result of a smaller database and also of a much larger diversity of groups and bonds, the application of estimation methods is more restricted for organometallic

* Corresponding author.
  E-mail address: joaquim.sales@ub.edu (J. Sales).

compounds. Indeed, additive methods have been scarcely applied to these substances [6]. Other estimation procedures have been used but their application range is not as broad as the additive methods. For instance, Benson et al. have proposed the "covalent potential method" to estimate $\Delta_f H^0$ of organic and organometallic compounds [7]. It involves relationships of the differences between the enthalpies of formation of a set of compounds RY and RH (where R is an organic or organometallic moiety, and Y is a group bonded to R through an X atom) vs. the covalent potential of X, V_X, which is defined as the ratio between the number of valence electrons and the covalent radius of X. This method has been applied to alkylsilanes [8,9], but the linear correlations with the covalent potential are not observed with a more complete set of organometallic derivatives of main-group elements [10]. Another simple procedure is the so-called Tina's Method [11]. It consists of linear correlations between the enthalpies of formation of a given family of compounds, ML_n, where M is a metal atom or an organometallic moiety and L is a ligand, and the enthalpies of formation of LH. As these latter quantities are usually available with good accuracies, the correlations can be used to derive unknown values of $\Delta_f H^0$, or even to evaluate experimental data.

Quantum chemistry, which is playing an increasingly important role in predicting thermochemical data with chemical accuracy, is also a widely used procedure for estimating enthalpies of formation of organic compounds. However, high-level computational methods are usually too demanding in terms of computer power for calculating enthalpies of formation of organometallic compounds. Therefore, the development of reliable and simple to apply empirical estimation methods is still a rather useful goal.

The quantitative structure–property relationship approach (QSPR) is a widely used method to predict different physical and chemical properties of chemical compounds from numerical descriptors derived from molecular structures and it has been successfully applied to many properties [12]. Some QSPR models using topological descriptors to describe the standard Gibbs energy of formation of alkanes have been reported [13,14]. More recently, the $\Delta_f G^0$ values of a set of 177 organic compounds have been analyzed with six descriptors, by linear and nonlinear methods, giving root-mean-square deviations of 75 and 65 kJ/mol, respectively [15]. In the present paper, we apply the QSPR methodology to predict gas-phase enthalpies of formation of organometallic compounds. The set is composed of 132 derivatives of the general formula MR_nX_{n-m}, where M is a metal or a semimetal from groups 12 to 16, R is an alkyl, aryl, alkenyl, or alkynyl group, and X is Cl, Br, I, or H.

2. Data and computational methods

2.1. Organometallic compounds

The values of the gas-phase enthalpies of formation of the organometallic compounds studied have been taken

mainly from the NIST organometallic thermochemistry database [16] and the Cox and Pilcher book [17] (see Table 1). The data set contains 132 organometallic compounds of the following 14 semimetals and non-transition metals: Al, As, B, Bi, Ga, Ge, Hg, P, Pb, Sb, Se, Si, Te, and Sn. As stated above, the compounds have the general formula MR_nX_{n-m}, where R is an alkyl, aryl, alkenyl, alkynyl, or five- and six-membered cycloalkanes; X is Cl, Br, I, or H. The $\Delta_f H^0$ values range from −514.6 kJ/mol for B(octyl)$_3$ to 736.9 kJ/mol for SnPh$_3$(C≡CPh), with a mean value of −23.21 kJ/mol.

The full dataset was split randomly into a 104 (79%) members training set and an independent prediction set of 28 compounds (21%). The random selection has been done to ensure that the prediction set contains compounds with all the metals and types of groups. The training set was used exclusively to derive the model. The prediction set, formed by compounds that were not included in the model development, was used to prove the predictive capacity of the model. Table 1, contains all the experimental and calculated $\Delta_f H^0$ values.

2.2. Organic compounds

The dataset of organic compounds contains 168 values corresponding to alkanes, haloalkanes, alkylbenzenes, halobenzenes, alkenes, alkynes, cycloalkanes, and cycloalkenes. The experimental values have been taken from the Pedley's compilation [5]. The groups present in these compounds are of the same type as those in the set of organometallic compounds studied. The $\Delta_f H^0$ values ranged from −238.9 kJ/mol for 1-chlorooctane to 251.7 kJ/mol for 1,2-diiodobenzene, with a mean value of −80.4 kJ/mol. Analogously to the organometallic compounds, the full set of organic compounds was split randomly into a 133 (79%) member training set and an independent prediction set of 35 compounds (21%). The random selection has been done to ensure that the prediction set contains compounds of all the families. Again, the training set was used exclusively to derive the model and the prediction set, formed by compounds that were not included in the model development, was used to prove the predictive capacity of the model.

2.3. Molecular descriptors

The generation of the descriptors was performed with the CODESSA program [18]. The structures of the compounds were drawn with HyperChemLite and exported in a file format suitable for MOPAC. The geometry optimization was performed with the semi-empirical AM1 or PM3 methods [19,20] using MOPAC 6.0 program [21]. All the geometries have been fully optimized without symmetry restrictions. Frequency calculations have been performed in all cases, to ensure that the calculated geometries correspond to true minima. The MOPAC output files were used by CODESSA to calculate several hundreds of molecular descriptors, which can be classified in constitutional, topological, geometrical,

Table 1
Experimental and calculated gas-phase $\Delta_f H^0$ (kJ/mol) for the organometallic set

Compound	Set[*]	Experimental	Calculated
Al(i-Bu)$_3$	t	−231.5[a]	−222.5
AlBu$_3$	t	−216.0[b]	−207.3
AlEt$_3$	p	−114.1[a]	−96.6
AlMe$_3$	t	−86.5[a]	−41.3
AlPr$_3$	t	−153.3[b]	−151.9
AsMe$_3$	t	12.5[c]	−44.4
AsPh$_3$	p	408.4[c]	365.3
B(3-Me-Bu)$_3$	t	−381.8[c]	−350.6
B(i-Bu)$_3$	t	−280.3[c]	−299.1
B(i-Pr)$_3$	t	−251.4[c]	−295.2
B(p-tol)Cl$_2$	p	−294.4[c]	−239.9
B(s-Oct)$_3$	p	−507.9[c]	−516.5
BBu$_2$Br	t	−300.4[c]	−327.2
BBu$_2$Cl	t	−365.7[c]	−372.9
BBu$_2$I	t	−225.5[c]	−269.3
BBu$_3$	t	−287.0[c]	−280.0
BCy$_3$	p	−397.9[c]	−341.9
BEt$_3$	t	−152.7[c]	−169.4
BHex$_3$	t	−395.0[c]	−390.6
BHpt$_3$	t	−454.8[c]	−446.0
BMe$_3$	t	−122.6[c]	−114.1
BOct$_3$	t	−514.6[c]	−501.3
BPhCl$_2$	t	−266.1[c]	−274.0
BPr$_3$	p	−236.0[c]	−224.7
BiBu$_3$	t	71.2[b]	3.4
BiMe$_3$	t	194.4[a]	169.3
BiPh$_3$	p	600.6[a]	579.0
BiPr$_3$	t	133.9[b]	58.7
Ga(i-Bu)$_3$	p	−239.5[a]	−184.5
GaBu$_3$	t	−226.9[a]	−169.3
GaEt$_3$	t	−62.3[a]	−58.6
GaMe$_3$	t	−35.9[a]	−3.3
GaPr$_3$	t	−125.5[b]	−113.9
GeBu$_4$	p	−310.0[b]	−304.8
GeBz$_4$	t	388.2[a]	389.0
GeEt$_3$H	t	−111.9[a]	−123.8
GeEt$_4$	t	−164.9[a]	−157.3
GeMe$_3$Br	t	−222.2[a]	−186.0
GeMe$_3$Cl	p	−266.1[a]	−231.8
GeMe$_4$	p	−107.5[a]	−83.5
GePh$_2$(CH$_2$)$_4$	t	138.9[a]	174.1
GePh$_3$(C≡CPh)	t	578.9[a]	633.9
GePh$_3$(CH=CH$_2$)	t	362.6[a]	408.9
GePh$_4$	t	438.6[a]	462.8
GePr$_4$	t	−229.7[c]	−231.0
Hg(C≡CPh)$_2$	t	720.5[a]	707.8
Hg(CH2-i-Bu)$_2$	t	−82.7[a]	−78.6
Hg(i-Bu)$_2$	t	−38.3[a]	−28.3
Hg(i-Pr)$_2$	t	37.0[a]	8.6
HgBr(i-Pr)	t	−53.8[d]	−52.0
HgBrEt	t	−30.3[a]	−28.5
HgBrMe	p	−18.6[a]	−10.0
HgBrPr	t	−51.4[d]	−46.9
HgBu$_2$	p	−32.4[c]	−18.2
HgBz$_2$	p	278.0[a]	328.7
HgCl(i-Pr)	t	−86.9[d]	−97.7
HgClEt	t	−67.7[a]	−74.2
HgClMe	t	−55.1[a]	−55.8
HgClPr	p	−88.1[d]	−92.7
HgEt$_2$	t	75.0[a]	55.6
HgI(i-Pr)	t	−5.8[d]	5.9
HgIEt	p	14.3[a]	29.4
HgIMe	t	21.6[a]	47.9
HgIPr	t	−3.7[d]	11.0
HgMe$_2$	t	92.4[a]	92.5
HgPh$_2$	t	394.2[a]	365.6
HgPr$_2$	t	30.3[a]	18.7
PbEt$_4$	p	109.6[a]	45.1
PbMe$_4$	t	136.1[a]	118.8
PbPh$_2$Br$_2$	t	177.7[a]	186.9
PbPh$_2$I$_2$	t	288.2[a]	302.8
PbPh$_3$Br	t	406.3[a]	426.0
PbPh$_3$I	t	455.4[a]	483.9
PbPh$_4$	p	674.1[a]	665.1
PMe$_3$	p	−94.1[c]	−111.2
PPh$_3$	t	328.4[c]	298.5
SbBu$_3$	p	−100.5[b]	−133.2
SbEt$_3$	t	48.7[a]	−22.6
SbMe$_3$	p	32.2[a]	32.7
SbPh$_3$	t	435.4[a]	442.4
SbPr$_3$	t	−37.7[b]	−77.9
Se(i-Pr)$_2$	t	−108.0[e]	−91.3
SeBu$_2$	p	−131.8[e]	−118.0
SeEt$_2$	t	−57.3[f]	−44.7
SeMe$_2$	t	17.8[g,h]	−7.4
SePent$_2$	t	−172.7[e]	−154.9
SePh$_2$	t	289.7[i]	265.8
SiEt$_4$	t	−265.7[b]	−253.7
SiMe$_2$Cl$_2$	p	−461.1[c]	−476.5
SiMe$_3$Br	t	−297.5[c]	−282.5
SiMe$_3$Cl	t	−354.0[c]	−328.2
SiMeCl$_2$H	t	−415.0[c]	−456.1
SiPh$_2$Cl$_2$	t	−208.8[c]	−203.3
Sn(i-Bu)$_4$	t	−272.2[j,k]	−251.7
Sn(i-Pr)$_4$	t	−119.4[l]	−177.9
SnBu$_3$Br	t	−270.6[a]	−278.5
SnBu$_4$	t	−217.4[a]	−231.3
SnEt$_3$Cl	t	−193.3[b]	−213.6
SnEt$_3$H	t	−54.7[j,m]	−54.0
SnEt$_4$	t	−42.0[a]	−83.8
SnEtCl$_3$	t	−429.3[b]	−473.3
SnEtPh$_3$	t	380.0[n]	381.2
SnMe$_2$Cl$_2$	t	−337.0[o]	−306.6
SnMe$_2$I$_2$	t	−150.0[o]	−99.3
SnMe$_3$(CH=CH$_2$)	p	91.5[a]	72.6
SnMe$_3$(i-Pr)	p	−43.7[a]	−52.0
SnMe$_3$(t-Bu)	t	−67.0[b]	−77.2
SnMe$_3$Br	t	−138.1[a]	−112.6
SnMe$_3$Bz	t	90.4[c]	108.1
SnMe$_3$Cl	t	−174.9[b]	−158.3
SnMe$_3$Et	t	−26.3[a]	−28.5
SnMe$_3$H	t	24.9[j,p]	1.3
SnMe$_3$I	t	−82.4[a]	−54.7
SnMe$_3$Ph	p	113.6[c]	126.5
SnMe$_3$Pr	t	−46.8[b]	−47.0
SnMe$_4$	t	−17.6[b]	−10.1
SnMeCl$_3$	t	−417.1[b]	−454.8
SnMePh$_3$	t	406.0[n]	399.6
SnPh$_2$(CH$_2$)$_4$	t	301.8[a]	247.5
SnPh$_2$(CH$_2$)$_5$	p	289.0[a]	229.1
SnPh$_3$(C≡CPh)	t	736.9[a]	707.3
SnPh$_3$(CH=CH$_2$)	t	528.6[a]	482.3
SnPh$_4$	t	575.4[a]	536.2
SnPr$_4$	t	−142.9[a]	−157.6
Te(i-Pent)$_2$	t	−148.0[f]	−88.7
Te(i-Pr)$_2$	t	−46.0[f]	−15.0
TeBu$_2$	t	−56.0[e]	−41.7

Table 1 (continued)

Compound	Set[*]	Experimental	Calculated
TeEt$_2$	t	−4.4[e,f]	32.1
TeMe$_2$	t	28.6[e,q]	68.9
TePent$_2$	t	−137.0[f]	−78.6
TePh$_2$	p	306.8[f,r]	342.1
TePr$_2$	t	−42.5[e,f]	−4.8

[a] Ref. [16].

[b] S.W. Benson, J.T. Francis, T.T. Tsotsis, J. Phys. Chem. 92 (1988) 4515.

[c] Ref. [17].

[d] Value recalculated from data in C.T. Mortimer, H.O. Pritchard, H.A. Skinner, Trans. Faraday Soc. 48 (1952) 220.

[e] Value recalculated from data in M.G. Voronkov, V.A. Klyuchnikov, S.N. Kolabin, G.N. Shvets, P.I. Varushin, É. N. Deryagina, N.A. Korchevin, S.I. Tsvetnitskaya, Dokl. Akad. Nauk. SSSR 307 (1989) 1139.

[f] Value recalculated from data in V.I. Tel'noi, I.B. Rabinovich, Russ. Chem. Rev. 49 (1980) 603.

[g] Value recalculated from data in V.I. Tel'noi, V.N. Larina, E.N. Karataev, V.K. Stankevich Metalloorg. Khim. 1 (1988) 1102.

[h] Value recalculated from data in V.I. Tel'noi, V.N. Larina, E.N. Karataev, S.V. Amosova, Zh. Obshch. Khim. 59 (1989) 2012.

[i] Value recalculated from data in D.S. Barnes, C.T. Mortimer, J. Chem. Thermodyn. 5 (1973) 371.

[j] Value recalculated from data in W.F. Lautsch, A. Tröber, W. Zimmer, L. Mehner, W. Linck, H.-M. Lehmann, H. Brandenburger, H. Körner, H.-J. Metzchker, K. Wagner, R. Kaden, Z. Chem. 3 (1963) 415.

[k] Value recalculated from data in B.I. Kozyrkin, L.N. Larionova, A.E. Sokolovskii, A.K. Baev, N.A. Golovanov, Zh. Prikl. Khim. (Leningrad) 63 (1990) 1978.

[l] Value recalculated from data in D.J. Coleman, H.A. Skinner, Trans. Faraday Soc. 62 (1966) 1721.

[m] Value recalculated from data in C.R. Dillard, E.H. McNeill, D.E. Simmons, J.B. Yeldell, J. Am. Chem. Soc. 80 (1958) 3607.

[n] D.B. Chambers, F. Glocking, Inorg. Chim. Acta 4 (1970) 150.

[o] P.H. De Ryck, L. Verdock, G.P. van der Kelen, Bull. Soc. Chim. Belg. 94 (1985) 621.

[p] Value recalculated from data in A.E. Finholt, A.C. Bond Jr., K.E. Wilzbach, H.I. Schlessinger, J. Am. Chem. Soc. 69 (1947) 2692.

[q] Value recalculated from data in V.I. Tel'noi, V.N. Larina, E.N. Karataev, E.N. Dergyagina, Russ. J. Phys. Chem. 62 (1988) 1623.

[r] Value recalculated from data in V.I. Tel'noi, M.S. Sheiman, V.N. Larina, G.P. Kamelova, Russ. J. Phys. Chem. 70 (1996) 1261.

[*] t: training set; p: prediction set.

electrostatic, quantum-chemical, and charge partial surface area types. As metal descriptors we have used several physicochemical properties such as atomic number, atomic weight, ionization energy, electron affinity, different electronegativity scales (Pauling, Mulliken, Sanderson, Allred-Rochow), radii and ratios charge/radius, hardness and softness, etc. Their values have been taken mainly from the Emsley's handbook [22]. A total of 33 metal descriptors were considered for each organometallic compound. These descriptors were imported to the CODESSA program to make the first reduction of the descriptors pool. The heuristic multilinear regression procedures available in the framework of the CODESSA program were used to derive the models.

The goodness of the correlation was tested by the determination coefficient (R^2), the F-test, the root-mean square error (RMSE), and the mean absolute error (MAE). The stability of the correlation was tested against the cross-val-

idated coefficient, R_{cv}^2, which describes the stability of a regression model by focusing on the sensitivity of the model to the elimination of any single data point. The t-test of each coefficient, as well as the standardized regression beta coefficients is also reported. As indicated before, the validation of the model was performed on the external prediction set of compounds.

3. Results and discussion

Although the molecular descriptors were derived from the molecular structure of the organometallic compounds and so they encode information on the nature of the metal, we have observed that the inclusion of descriptors associated exclusively with the metal improves significantly the correlation. Thus, we have imported 33 descriptors of each metal to the CODESSA program before starting the analysis. In the initial attempts, the heuristic procedure of CODESSA gave models which contained only constitutional molecular descriptors. These descriptors reflect the chemical composition of the compounds, without any reference to the geometric or electronic structure of the amolecule. The models derived by CODESSA had constitutional descriptors of several types (number of simple, double, or triple bonds, number of rings or aromatic rings, etc.) and contained also a number of descriptors of the metal, such as electronegativity, ionic radius, number of valence electrons, etc. These models presented good statistical parameters, $R^2 \approx 0.99$ and RMSE ≈ 30. Yet, they did not distinguish between isomers like propyl or isopropyl derivatives of the same metal. For this reason, after a detailed analysis of the descriptors involved in the models derived by the heuristic and also by the best multilinear regression routines of the CODESSA program, the topological descriptor Kier–Hall index of order zero was incorporated in the final model. Besides this topological descriptor, the proposed model (Table 2) contains six constitutional molecular descriptors: molecular weight, number of single, double, triple and aromatic bonds, number of halogen atoms, and two descriptors derived for the metal, viz. Mulliken electronegativity and the charge/radius ratio. The Mulliken electronegativity has been calculated as the average of the ionization energy and the

Table 2
Descriptors of the model for the organometallic set

Descriptor	Coefficient	SD	t-Test	β
Intercept	217.37	32.08	6.78	
Molecular weight	1.61	0.11	14.82	0.64
Number of single bonds	−6.36	0.76	−8.31	−0.32
Number of double bonds	72.17	23.13	3.12	0.04
Number of triple bonds	169.81	14.28	11.89	0.15
Number of aromatic bonds	20.57	0.96	21.43	0.56
Number of halogen atoms	−196.06	7.02	−27.91	−0.54
Kier–Hall index(order 0)	−31.04	4.68	−6.63	−0.42
Electronegativity (Mulliken)	−34.92	5.17	−6.75	−0.09
Charge/radius	−6.53	1.29	−5.06	−0.10

electron affinity of the element. The charge is the formal oxidation number of the metal in the compound, and the radius refers to the corresponding cation in its oxidation state.

The topological descriptors give information on the atomic connectivity of a molecule, which is a measure of the degree of branching of the structure. In particular, the descriptor Kier–Hall index of order zero belongs to the well-known valence connectivity indices (χ) defined by Kier and Hall [23] to account for the presence of heteroatoms and the hybridization of atoms in the molecule. The specific descriptor of order zero, $^0\chi$, is calculated as follows:

$$^0\chi = \sum_i^n \left(\frac{1}{\delta_i^v} \right)^{1/2}$$

where n is the total number of atoms (without counting hydrogen) and δ_i^v is the 0th order atomic connectivity of each atom. This atomic connectivity is computed as

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_i^v - 1}$$

where Z_i is the total number of electrons in the ith atom, Z_i^v is the number of valence electrons of this atom, and H_i is the number of hydrogen atoms directly attached to the ith non-hydrogen atom. As a example, the calculation of this descriptor is detailed here for two chain isomers: BBu$_3$ and B(i-Bu)$_3$. The calculation of this descriptor for each molecule is

$$^0\chi_{BBu_3} = \left(\frac{1}{\delta_B^v} \right)^{\frac{1}{2}} + 3 \left[3 \left(\frac{1}{\delta_{CH_2}^v} \right)^{\frac{1}{2}} + \left(\frac{1}{\delta_{CH_3}^v} \right)^{\frac{1}{2}} \right]$$

$$^0\chi_{B(i-Bu)_3} = \left(\frac{1}{\delta_B^v} \right)^{\frac{1}{2}} + 3 \left[\left(\frac{1}{\delta_{CH}^v} \right)^{\frac{1}{2}} + \left(\frac{1}{\delta_{CH_2}^v} \right)^{\frac{1}{2}} + 2 \left(\frac{1}{\delta_{CH_3}^v} \right)^{\frac{1}{2}} \right]$$

respectively. The boron atom contribution, which remains constant, is calculated as

$$\delta_B^v = \frac{(3-0)}{(5-3-1)} = 3$$

The main difference between the molecules is due to the substituent terms, which are calculated as follows:

$$\delta_{CH}^v = \frac{(4-1)}{(6-4-1)} = 3, \quad \delta_{CH_2}^v = \frac{(4-2)}{(6-4-1)} = 2$$

$$\text{and} \quad \delta_{CH_3}^v = \frac{(4-3)}{(6-4-1)} = 1$$

Since all atomic connectivities are known, the descriptor $^0\chi$, for these two molecules, BBu$_3$ and B(i-Bu)$_3$, is 9.9413 and 10.4307, respectively.

The training set composed of 104 compounds, give a good correlation: $R^2 = 0.988$, $F = 842.1$, $R_{cv}^2 = 0.983$, RMSE = 29.1, and MAE = 22.7. According to the β coefficients, the three more significant descriptors are the

molecular weight, the number of aromatic bonds, and the number of halogen atoms. Similar results are obtained with the prediction set of 28 compounds, thus confirming the high prediction capacity of the model. The statistical parameters are $R^2 = 0.990$, RMSE = 30.2, and MAE = 23.9. Fig. 1, shows a plot of the calculated vs. observed $\Delta_f H^0$ values for the training and prediction sets. The results are also very similar for the two subsets of compounds with or without halogen atoms. Thus, the subset of 92 compounds without halogen atoms bonded to the metal, MR$_n$, yields $R^2 = 0.987$ and RMSE = 31.2; and the subset with the other 40 compounds of the type MR$_n$X$_m$, has a determination coefficient of 0.991 and a RMSE of 23.8. On the other hand, the accuracy of the estimated values for isomeric pairs of compounds, such as normal- and iso-derivatives, follows the general trend.

Our linear model allows the estimation of $\Delta_f H^0$ by addition of the values of the nine descriptors for the corresponding compound. The constitutional descriptors, which reflect the chemical composition of the molecules, have positive sign, except the number of single bonds and the number of halogen atoms; the topological descriptor, encoding information about the molecular structure, and the two descriptors of the metal have negative sign. The two metal descriptors are clearly related to the chemical nature of the M–C or M–X bonds and consequently to their strength since the difference of electronegativities between the two atoms of a bond is reflected by the bond dissociation energy. On the other hand, the charge/radius ratio of an atom is a measure of its polarizing power, and consequently of the covalent-ionic degree of the involved bond and therefore of its bond dissociation energy.

The proposed model is a useful method to estimate the enthalpy of formation of organometallic compounds of the elements and chemical types analyzed. Using some caution, the model can even be used to assess data quality. For example, experimental values of $\Delta_f H^0$ of some alkyl derivatives MR$_2$ of Zn and Cd have been experimentally determined. The values have low accuracy, especially for the zinc derivatives [16]: ZnMe$_2$ (52.9 ± 1.3), ZnEt$_2$ (55.5 ± 3.9), ZnPr$_2$ (−17.2 ± 23.0), ZnBu$_2$ (−55.2 ± 23.4), CdMe$_2$ (104.8 ± 1.3), and CdEt$_2$ (105.3 ± 2.0). All the attempts to incorporate these data in the training set of our QSPR analysis, produces worse fits and a significant increase of the RMSE up to ca. 40 kJ/mol. Therefore, more precise experimental data and values for more Zn and Cd compounds are needed to derive a better multilinear regression including these metals.

To test the capacity of the QSPR approach for estimating enthalpies of formation, we have analyzed a set of organic compounds for which reliable experimental gas-phase values of $\Delta_f H^0$ are available. We selected a set of organic compounds similar to the organometallic compounds studied, both in the number of compounds and in their chemical nature. Thus, we have chosen 168 organic molecules, including alkanes, haloalkanes, alkylbenzenes,
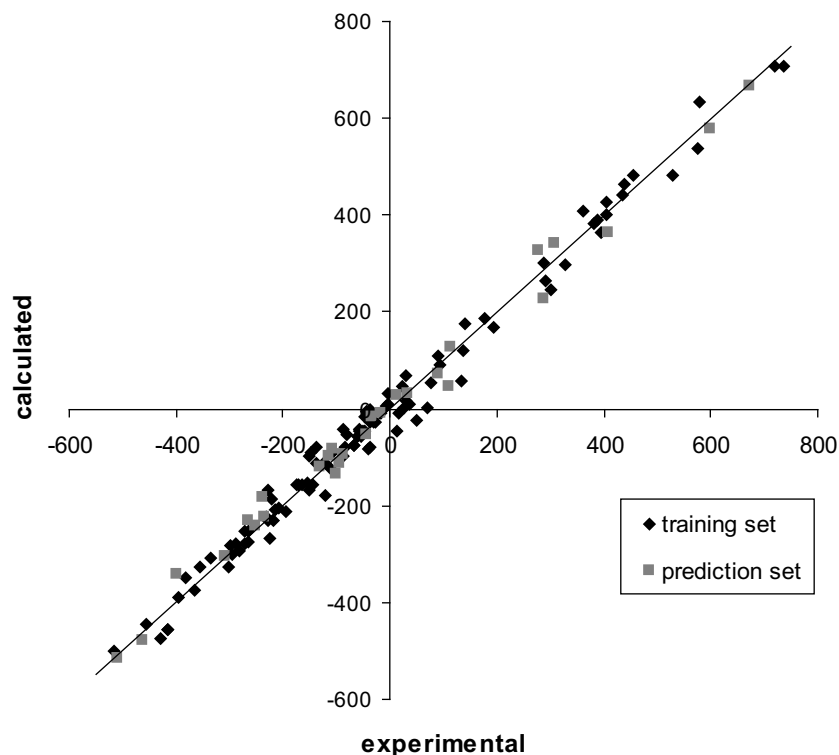
Fig. 1. Plot of calculated vs. experimental gas-phase $\Delta_f H^0$ of the organometallic compounds. Data in kJ/mol.

halobenzenes, alkenes, alkynes, cycloalkanes, and cycloalkenes. These organic compounds have the same type of groups present in the organometallic set. The CODESSA program has been used in the usual way, but the pool of descriptors was restricted to the constitutional, topological, and geometrical ones. In other words, the descriptors of electrostatic, quantum-chemical, and charge partial surface area types were excluded. The derived nine descriptor model is shown in Table 3. It presents six constitutional descriptors, viz. number of single and double bonds, number of rings and benzene rings, number of halogen atoms, and relative number of hydrogen atoms (number of hydrogen atoms divided by the total number of atoms in the molecule). It also contains two topological descriptors (the Kier shape index of order 1, and the Randic index of order 2) and one geometrical descriptor (gravitation index) for all bonds. The Kier shape indexes [24] are topological descriptors that depend on the number of skeletal atoms and the molecular branching. The Randic molecular connectivity indexes [25] also encode information about the size and branching of the structure. The geometrical descriptors represent more advanced structural molecular descriptors, derived from the three-dimensional coordinates of the atomic nuclei and the atomic masses and/or radii in the molecule. The gravitation index for all bonds [26] reflects the molecular shape and the mass distribution in the molecule, and is calculated according to Newton's law.

The correlation obtained for the training set of organic compounds is good: $n = 133$, $R^2 = 0.996$, $F = 3501.6$, $R_{cv}^2 = 0.995$, RMSE = 6.8, and MAE = 4.9. According to

Table 3
Descriptors of the model for the organic set

Descriptor	Coefficient	SD	t-Test	β
Intercept	162.40	14.43	11.25	
Number of single bonds	−67.82	1.50	−45.25	−3.90
Number of double bonds	−51.10	3.06	−16.68	−0.21
Number of benzene rings	−504.55	12.59	−40.07	−1.90
Number of rings	302.93	10.03	30.19	1.33
Relative number of H atoms	−193.38	21.26	−9.10	−0.21
Number of halogen atoms	−280.42	4.76	−58.94	−2.56
Kier shape index (order 1)	177.67	5.59	31.78	2.81
Randic index (order 2)	−12.45	1.56	−7.96	−0.12
Gravitation index (all bonds)	0.11	0.01	8.33	0.21

the β coefficients, the three more significant descriptors are the number of simple bonds, the Kier shape index, and the number of halogen atoms. The model has good prediction capacity, since with the 35 compounds of the prediction set, R^2 is 0.995, RMSE is 7.9, and MAE is 5.8. Fig. 2, shows the plot of the calculated vs. experimental values for these two sets or organic compounds. The obtained results are similar to those derived by the Pedley method, based on the additivity of bond enthalpies. This method gives $R^2 = 0.998$, RMSE = 5.0 and MAE = 2.6 for the overall set of 168 organic compounds. The statistical values for the QSPR results also for the 168 compounds are $R^2 = 0.996$, RMSE = 7.0 and MAE = 5.1. Table 4 gives experimental and predicted values obtained for two subsets of organic compounds: the C_8H_{18} alkanes and several polychlorobenzenes, using the QSPR approach and Pedley's method. These two approaches lead to similar
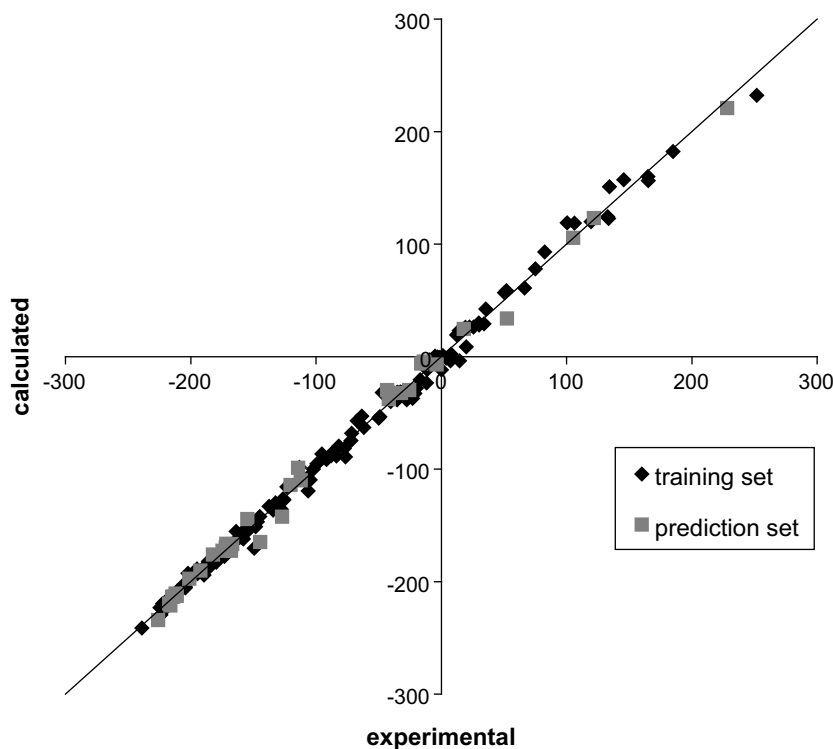
Fig. 2. Plot of calculated vs. experimental gas-phase $\Delta_f H^0$ of the organic compounds. Data in kJ/mol.

predictions of $\Delta_f H^0$ for isomeric molecules. Thus, for the 27 compounds reported in Table 4, the determination coeffi-

Table 4
Gas-phase $\Delta_f H^0$ (kJ/mol) for several organic isomeric compounds

Compound	Formula	Experimental	QSPR[a]	Pedley[b]
1,2,3,4-Tetrachlorobenzene	$C_6H_2Cl_4$	−25.4	−29.9	−39.8
1,2,3,5-Tetrachlorobenzene	$C_6H_2Cl_4$	−34.9	−31.6	−39.8
1,2,4,5-Tetrachlorobenzene	$C_6H_2Cl_4$	−32.6	−31.5	−39.7
1,2,3-Trichlorobenzene	$C_6H_3Cl_3$	3.8	−1.4	−9.1
1,2,4-Trichlorobenzene	$C_6H_3Cl_3$	−8.1	−3.0	−9.2
1,3,5-Trichlorobenzene	$C_6H_3Cl_3$	−13.4	−4.9	−9.2
1,2-Dichlorobenzene	$C_6H_4Cl_2$	30.2	28.0	21.4
1,3-Dichlorobenzene	$C_6H_4Cl_2$	25.7	26.2	21.4
1,4-Dichlorobenzene	$C_6H_4Cl_2$	22.5	26.3	18.2
2,2,3,3-Tetramethylbutane	C_8H_{18}	−226.0	−234.3	−224.1
2,2,3-Trimethylpentane	C_8H_{18}	−219.9	−223.7	−222.5
2,2,4-Trimethylpentane	C_8H_{18}	−223.9	−229.6	−223.7
2,3,3-Trimethylpentane	C_8H_{18}	−216.2	−221.4	−219.5
2,3,4-Trimethylpentane	C_8H_{18}	−217.2	−219.6	−219.6
3-Ethyl-3-methylpentane	C_8H_{18}	−214.8	−213.5	−215.3
2,2-Dimethylhexane	C_8H_{18}	−224.5	−223.3	−225.3
2,3-Dimethylhexane	C_8H_{18}	−213.8	−215.1	−216.9
2,4-Dimethylhexane	C_8H_{18}	−219.2	−216.8	−218.9
2,5-Dimethylhexane	C_8H_{18}	−222.5	−219.4	−221.9
3,3-Dimethylhexane	C_8H_{18}	−219.9	−218.4	−220.3
3,4-Dimethylhexane	C_8H_{18}	−212.8	−212.2	−214.9
3-Ethylhexane	C_8H_{18}	−210.7	−208.3	−211.2
2-Methylheptane	C_8H_{18}	−215.3	−213.4	−215.2
3-Methylheptane	C_8H_{18}	−212.5	−210.5	−213.2
4-Methylheptane	C_8H_{18}	−211.9	−210.8	−213.2
3-Ethyl-2-methylpentane	C_8H_{18}	−211.0	−212.8	−214.9
Octane	C_8H_{18}	−208.5	−207.3	−208.5

[a] This work.
[b] Ref. [5].

cient is the same (0.999) and the RMSE is similar (3.6 for the QSPR and 4.2 for the Pedley's method). The QSPR methodology gives better results for the subset of polychlorobenzenes, while Pedley's method is more accurate for the alkanes.

4. Conclusions

The ability of the QSPR approach for estimating enthalpies of formation of organometallic compounds has been demonstrated. The method is applicable to a wide variety of organometallic compounds, involving 13 post-transition (block p) elements and mercury. The proposed model, linearly derived, is composed of nine descriptors which can be easily calculated from the chemical formula of the compound. The correlation has a good squared determination coefficient, and its prediction capacity has been tested with an independent prediction set composed of organometallic compounds which were not used in the derivation of the model. The robustness of the model is corroborated by the analogous statistical parameters found for the two subsets of organometallic compounds analyzed, the homoleptic (MR_n) and those with terminal halogen or hydrogen atoms (MR_nX_{n-m}), as well as the fact that the accuracy of the estimated values for isomeric pairs of compounds is not different to those obtained for the whole set. The RMSE for both training and prediction sets (ca. 30 kJ/mol) is rather high, but having in mind that some of the literature $\Delta_f H^0$ values used may have significant errors, this value is acceptable.

The power of QSPR methods to estimate enthalpies of formation has also been clearly shown by the results obtained with the set of organic compounds. The gas-phase $\Delta_f H^0$ values of 168 organic molecules containing the same chemical groups as the analyzed organometallic compounds have been estimated also by a nine-descriptors model. The descriptors are of the constitutional, topological, and geometrical type, and are analogous to those derived for the organometallic set. The predicted values are quite similar to those estimated by Pedley's method.

Acknowledgements

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jorganchem.2008.01.021.

References

[1] S.W. Benson, J.H. Buss, J. Chem. Phys. 29 (1958) 546.
[2] E.S. Domalsky, E.D. Hearing, V.J. Hearing Jr., NIST Estimation of the Chemical Thermodynamic Properties for Organic Compounds at 298.15 K, NIST Standard Reference Database 18, National Institute of Standards and Technology, Gaithersburg, 1994.
[3] K.J. Laidler, Can. J. Chem. 34 (1956) 626.
[4] J.P. Leal, J. Phys. Chem. Ref. Data 35 (2006) 55.
[5] J.B. Pedley, Thermochemical Data and Structures of Organic Compounds, vol. I, TRC, College Station, 1994.
[6] G. Pilcher, H.A. Skinner, Thermochemistry of organometallic compounds, in: F.R. Hartley, S. Patai (Eds.), The Chemistry of the Metal–Metal Bond, Wiley, New York, 1982 (Chapter 2).
[7] Y.-R. Luo, S.W. Benson, Acc. Chem. Res. 25 (1992) 375.
[8] Y.-R. Luo, S.W. Benson, J. Phys. Chem. 93 (1989) 4643.
[9] Y.-R. Luo, S.W. Benson, J. Phys. Chem. 93 (1989) 3791.
[10] R. Becerra, R. Walsh, J. Phys. Chem. A 101 (1997) 8959.
[11] A.R. Dias, J.A. Martinho Simões, C. Teixeira, C. Airoldi, A.P. Chagas, J. Organomet. Chem. 335 (1987) 71.
[12] A.R. Katritzky, V.S. Lobanov, M. Karelson, J. Chem. Inf. Comput. Sci. 40 (2000) 1.
[13] O. Ivanciuc, T. Ivanciuc, D. Cabrol-Brass, A.T. Balaban, J. Chem. Inf. Comput. Sci. 40 (2000) 631.
[14] O. Ivanciuc, T. Ivanciuc, D.J. Klein, W.A. Seitz, A.T. Balaban, J. Chem. Inf. Comput. Sci. 41 (2001) 536.
[15] A. Yan, J. Chem. Inf. Model. 46 (2006) 2299.
[16] J.A. Martinho Simões, Organometallic Thermochemistry Data, in: P.J. Linstrom, W.G. Mallard (Eds.), NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg, June 2005. <http://webbook.nist.gov>.
[17] J.D. Cox, G. Pilcher, Thermodynamic of Organic and Organometallic Compounds, Academic Press, 1970.
[18] A.R. Katritzky, V.S. Lobanov, M. Karelson, CODESSA, Reference Manual V 2.13, Semichem and the University of Florida, 1997.
[19] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.P.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[20] J.P.P. Stewart, J. Comput. Chem. 10 (1989) 209.
[21] J.P.P. Stewart, QCPE, No 455, Indiana University, Bloomington, IN, 1989.
[22] J. Emsley, The Elements, Claredon Press, 1989.
[23] L.B. Kier, L.H. Hall, Eur. J. Med. Chem. 12 (1977) 307.
[24] L.B. Kier, Quant. Struct.-Act. Relat. 2 (1985) 109.
[25] M. Randic, J. Am. Chem. Soc. 97 (1975) 6609.
[26] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson, J. Phys. Chem. 100 (1996) 10400.