

QSPR EVALUATION OF THERMODYNAMIC PROPERTIES OF ACYCLIC AND AROMATIC COMPOUNDS

Duchowicz, P.R.¹; Castro, E.A.¹; Fernández, F.M.¹; Pankratov, A.N.²

¹INIFTA, C.C. 16, Suc. 4, Department of Chemistry, La Plata National University, La Plata (B1906ZAA), Argentina

FAX: +54 221 4254642, E-mail: prduchowicz@yahoo.com.ar

²Department of Chemistry, N. G. Chernyshevskii Saratov State University, 83 Astrakhanskaya Street, Saratov 410012, Russia

Received June 15th, 2006. In final form October 30th, 2006

Abstract

Although parametrized Semiempirical Molecular Orbital Methods were especially designed to obtain enthalpies of formation and other related thermodynamical parameters, in some situations the predictions cannot be compared to the experimental results simply because they present serious drawbacks. Therefore, there is room to apply a wide variety of predictive methods for estimating thermodynamical properties, such as the well-known Group Contribution Methods and the Quantitative Structure-Property Relationships. A QSPR study of 163 enthalpies of formation, 37 Gibbs free energy changes, and 40 standard entropies of elements is established for a representative set of acyclic and aromatic compounds, on the basis of fundamental concepts on molecular structure such as the count of atoms and types of chemical bonds. A recent method discovered in our group called the Replacement Method is employed here to find the best models in a pool containing 33 descriptors. An 8 parameters-model was able to correlate the heats of formation with atoms and bond types ($R=0.9553$, $R_{1-25\%-o}=0.8200$) but with great dispersion ($S=19.173$ Kcal/mol). For the case of standard entropies ($R=0.9813$, $R_{1-20\%-o}=0.8659$) and for free energies ($R=0.9869$, $R_{1-20\%-o}=0.9000$) the models obtained perform poorer than those calculated with Semiempirical Methods.

Resumen

A pesar de que los métodos parametrizados de Orbitales Moleculares Semiempíricos se diseñaron especialmente para obtener entalpías de formación u otros parámetros termodinámicos, en algunas situaciones las predicciones no pueden compararse con los resultados experimentales simplemente porque presentan serias anomalías. Por tanto, hay lugar para aplicar una amplia variedad de métodos predictivos para estimar propiedades termodinámicas, tales como los bien conocidos Métodos de Contribución de Grupos y las Relaciones Cuantitativas Estructura-Propiedad. Se establece un estudio QSPR para 163 entalpías de formación, 37 cambios de energía libre de Gibbs, y 40 entropías estándar a partir de elementos para un conjunto representativo de compuestos acíclicos y aromáticos, sobre la base de conceptos fundamentales de estructura molecular como lo son la cuenta de átomos y tipos de enlace químico. Se emplea un método reciente descubierto en nuestro

grupo llamado Método de Reemplazo para encontrar los mejores modelos a partir de un conjunto total de 33 descriptores. Un modelo de 8 parámetros fue capaz de correlacionar los calores de formación con los átomos y tipo de enlaces ($R=0,9553$, $R_{1-25\%-0}=0,8200$) pero con gran dispersión ($S=19,173$ Kcal/mol). Para el caso de entropías estándar ($R=0,9813$, $R_{1-20\%-0}=0,8659$) y para energías libres ($R=0,9869$, $R_{1-20\%-0}=0,9000$) los resultados que se obtienen resultan más pobres a los alcanzados con los Métodos Semiempíricos.

Introduction

Thermodynamics is a phenomenological theory of matter. As such, it draws its concepts directly from experiments [1]. Thermodynamic parameters are measurable macroscopic quantities associated with macroscopic systems, such as pressure P , volume V , temperature T , and magnetic field B , which are defined experimentally. Macroscopic systems (like gases, liquids, or solids) began first to be systematically investigated from a macroscopic phenomenological point of view in the last century, and the laws thus discovered formed the subject of Thermodynamics. The strength of this discipline is its great generality, which allows making valid statements based on a minimum number of postulates without requiring any detailed assumptions about the microscopic (*i.e.* molecular) properties of the system [2].

Thermodynamics, which makes up a logical subject of great elegance, is a powerful method for studying chemical phenomena and can be developed quite independently of the atomic and molecular theory. It has a permanence which might, for example, be compared with that of Euclid's geometric theorems in plane geometry, which is not shared by our ever-changing views on the nature of atoms and molecules [3].

The prediction of thermodynamic and physical properties for organic compounds in different conditions (*i.e.* temperature, pressure) is vital for the design of chemical and petrochemical plants. Also, experimental measurements of some thermodynamic parameters involve experimental difficulties and they are not always feasible, and the corresponding methods possess real drawbacks [4,5]. Thus, it is necessary to resort to a theoretical calculation of these parameters, which is now accessible because an important, fruitful and current field of research in contemporary chemistry is the model and prediction of physical-chemistry properties of molecules [6,7]. This kind of study is based on the paradigm that physical-chemistry and biological properties are dependent on molecular structure. As a consequence, one of the most important points in such a research is the selection of adequate descriptors containing the information stored in the molecular structure [8].

The most common software packages used in chemical engineering design incorporate efficient algorithms for the prediction of thermodynamic and physical properties of interest, by means of the Group Contribution Methods (*GCM*) [9-11]. These techniques are easy to apply, relying solely on the sum of contributions of each molecular structure fragment to a given thermodynamic property. The basic assumption of these methods is the transferability concept for a group; if this hypothesis does not hold, then *GCM* can be corrected with experimental data, when available, to achieve better predictions. This converts *GCM* into a computational intensive technique.

A drawback of *GCM* is that in its basic form (without corrections) it cannot model isomeric structures; this is not a problem for small organic compounds, although the situation gets worse for bigger size compounds with increasing number of conformers. Another associated problem is

that there are not always measured data available to extend these methods to less common compounds such as molecules containing fused aromatic rings or to organometallic compounds. This is an inconvenience also present in semiempirical methods if they are not properly parametrized.

All commented limitations for *GCM* point to the need of employing a different theoretical framework, such as that encompassed in the realms of the Quantitative Structure-Property Relationships (*QSPR*). A fundamental difference between both methodologies is that in *QSPR* the user can consider only theoretically defined molecular descriptors to represent the molecular structure, not relying on empirical parameters for improving the model's limitations. The graph-theoretical approach to *QSPR* is based on a well-defined mathematical representation of the chemical structure, relating a property (*P*) with a set of molecular descriptors (*d*) through an arbitrary function (*f*), which usually represents a polynomial relationship. These molecular descriptors are commonly named "topological indices" [12,13], they contain relevant information about the structure. Owing to the complexity of molecular structures, it seems to be nearly impossible to expect that a single set of descriptors would contain all the relevant structural information. Therefore, the search for novel molecular structure descriptors continues and it is a field of active research within the realm of *QSPR* theory. However, this search should not be carried out at random. Instead it should follow some regular procedure based on the desired attributes that a molecular structure descriptor needs to possess [8].

The most simple and obvious sort of graph theoretical indices, having a quite direct chemical interpretation, are atoms and chemical bond types. Although they have been considered as suitable molecular descriptors, they have not been widely employed. Several applications made by two of us (P. R. D. and E. A. C.) have demonstrated their usefulness to predict physical-chemistry properties and biological activities [14-18]. They can be computed rather easily and have the advantage that they may be applied to a quite diverse sets of structures.

In a recent paper one of the authors (A. N. P.) has computed standard values of ΔH_f° , S° and ΔG_f° for a representative set of acyclic and aromatic compounds by means of the semiempirical all-valence MNDO, AM1 and PM3 molecular orbital methods [19]. It was demonstrated the existence of quantitative relationships between experimental data and theoretical results, although some inconsistencies were noted.

The aim of this study is to present the results of the estimations of these three fundamental thermodynamic parameters by exploring the performance of the elemental atom and bond-type molecular descriptors, comparing the resulting predictions with those reported for the three semiempirical methods. In contrast to *GCM*, where each structure is dissected in all its constituent fragments (being a different number for each molecule) and using all of them to predict the property value, in *QSPR* one has to select among all the fragments resulting from the entire set of compounds, the most representative and common contributing the most to the property being modeled. For this purpose we resort to the recently proposed *RM* technique, which is described in next section. Then, we report results for the predictions of thermodynamic functions and discuss them with respect to previous calculations. Finally, we state our main conclusions derived from this study.

Materials and Methods

Chemical Data

Using a set of 33 atom and bond types we represented the molecular structure for the three thermodynamic properties under study. Although many *QSPR* studies have been proposed to predict thermodynamical properties, most of them rely on homogeneous type of calibration sets, since in this way the predictions are better. In the present report, we use a diverse data set of compounds. The 163 ΔH_f shown in Table 1 were extracted from the MOPAC manual [20], in order to compare the model predictions with the reported MNDO, AM1 and PM3 calculations. We chose the molecules appropriately in order to create a balanced calibration set, that is, including in the set approximately the same number of molecules possessing a given functional group in its structure. The experimental values of 40 S° and 37 ΔG_f were both obtained from Pankratov's paper [19] since the MOPAC manual does not report experimental values for them.

Computer Software

We computed the best calibration models with the Replacement Method [21], which constitutes a good approximation to the combinatorial search (*FS*) of variables and allow studying a pool D containing thousands of them. This new procedure simply consists in replacing a chosen variable of the regression by another that minimizes S . The method is as follows: choose d descriptors $\{X_1, X_2, \dots, X_d\}$ at random and do a linear regression. Choose one of the descriptors of this set, say X_i , and replace it for each of the D descriptors of the pool (except itself) keeping the best resulting set. Since one can start replacing any of the d descriptors in the initial model, then a regression equation with d variables has d possible paths to achieve the final result; for example, the choice above will develop into path i . Next, choose the variable with greatest relative error in its coefficient (except the one replaced in the previous step) and replace it with all the D descriptors (except itself) keeping again the best set. Replace all the remaining variables in the same way bypassing those replaced in previous steps. When finishing, start again with the variable having greatest relative error in the coefficient and repeat the whole process. Repeat this process as many times as necessary until the set of descriptors remains unchanged. At the end, we have the best model for the path i . Proceed in exactly the same way for all possible paths $i=1, 2, \dots, d$, compare the resulting models, and keep the best one. Our numerical experiments show that in this way one obtains a model almost as good as the best achieved with *FS*.

Results and Discussion

It is very common in the *QSPR* theory to built models not containing an excessive number of descriptors, in order to make possible the interpretation of the established relationship in terms of interaction mechanisms [22].

Table 1. Details of the descriptors appearing in the present study.

Label	Description	label	description
<i>C</i>	Number of C atoms	<i>Br</i>	Number of Br atoms
<i>H</i>	Number of H atoms	<i>N-H</i>	Number of bonds between N and H
<i>Cl</i>	Number of Cl atoms	<i>I</i>	Number of I atoms
<i>O</i>	Number of O atoms	<i>S</i>	Number of S atoms
<i>F-N</i>	Number of bonds between F and N	<i>N</i>	Number of N atoms
<i>C-C_{arom}</i>	Number of aromatic bonds between two C atoms	<i>C-O</i>	Number of single bonds between C and O
<i>F</i>	Number of F atoms	<i>C=O</i>	Number of double bonds between C and O
<i>NO₂</i>	Number of nitro groups	<i>C-C_{triple bond}</i>	Number of triple bonds between two C atoms
<i>C-C</i>	Number of single bonds between two C atoms		

A simple procedure to control the model expansion is by plotting the correlation coefficients of calibration (R) and leave-one-out cross-validation (R_{loo}), as a function of the number of variables present in the model. By analyzing the plot, it is found that the statistical parameters of the model improve up to a certain point called “break point”, beyond that, the improvement can be considered negligible because the relative change in the parameters is less important. Consequently, the model corresponding to the break point is supposed to be the best model in the set of variables analyzed.

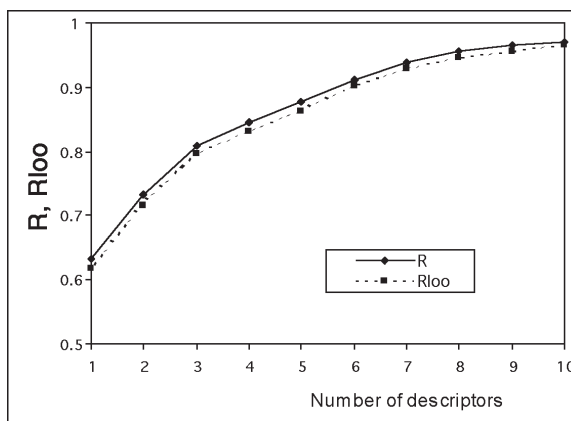
**Figure 1.** Model selection for ΔH_f

Figure 1 shows this type of plot for ΔH_f with a break point appearing at 8 variables included in the model. The statistical parameters of the equation are the following:

$$\Delta H_f = 24 (\pm 3) + 19 (\pm 2) C - 12.0 (\pm 0.8) H - 20 (\pm 2) Cl - 48 (\pm 2) O + 44 (\pm 4) F - N - 7 (\pm 1) C-C_{arom} - 60 (\pm 2) F + 88 (\pm 5) NO_2 \quad (1)$$

$$N=163, R=0.9553, S=19.173 \text{ Kcal.mol}^{-1}, F=200.924$$

$$R_{loo}=0.9468, S_{loo}=20.307 \text{ Kcal.mol}^{-1}$$

$$R_{l-25\%-o}=0.8200, S_{l-25\%-o}=23.240 \text{ Kcal.mol}^{-1}$$

where the leave-many-out technique was studied for 100000 cases of exclusion of compounds generated at random (for any of the three thermodynamic properties considered). Details of the descriptors appearing in all the equations are given in Table 1. Figure 2 plots the dispersions (difference between the experimental and predicted values of the property) as a function of the experimental property, revealing that the deviations are randomly distributed and not following any kind of pattern, also suggesting that data clustering is absent.

From Figure 3 it is clear that is possible to correlate the 163 ΔH_f with atoms and bonds but with great dispersion in the predicted values (S). This can be better seen in Table 2, where the absolute errors in most of the cases are much greater than those achieved with any of the three semiempirical methods. It is well known that experimental uncertainties for ΔH_f are around 2-3 kcal.mol⁻¹ [20,23-26].

The second property to consider is S^o ; the set comprising 40 compounds is given in Table 3, and the corresponding plot for R and R_{loo} in Figure 4. The best regression found is for 8 variables not manifesting clustering of data (see Figure 5).

$$S^o = 46 (\pm 1) + 3.5 (\pm 0.2) C + 5.3 (\pm 0.4) C-C + 11 (\pm 2) I + 5.8 (\pm 0.8) O + 12 (\pm 1) S + 4 (\pm 1) N-H + 7 (\pm 2) Cl + 10 (\pm 2) Br \quad (2)$$

$$N=40, R=0.9813, S=2.946 \text{ cal.mol}^{-1} \cdot \text{K}^{-1}, F=101.037$$

$$R_{loo}=0.9722, S_{loo}=3.162 \text{ cal.mol}^{-1} \cdot \text{K}^{-1}$$

$$R_{l-20\%-o}=0.8659, S_{l-20\%-o}=6.925 \text{ cal.mol}^{-1} \cdot \text{K}^{-1}$$

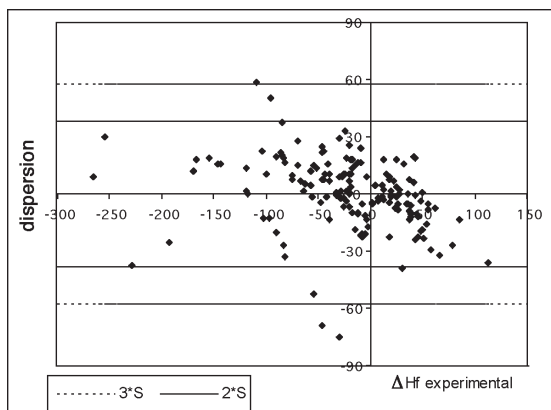


Figure 2. Dispersion plot for (2).

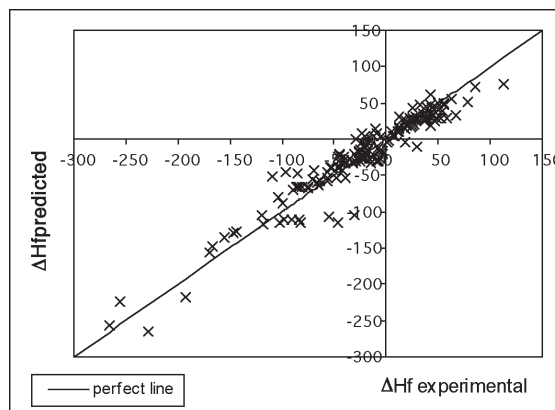


Figure 3. Linearity of model (2).

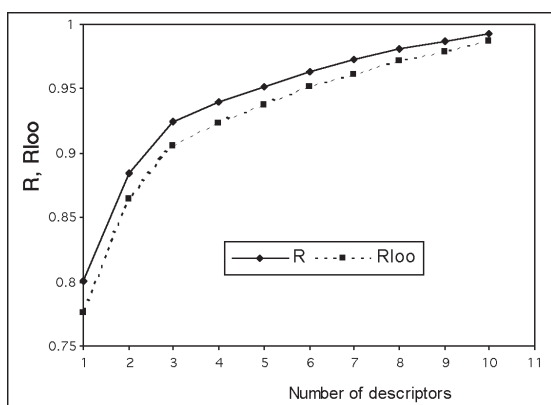


Figure 4. Model selection for S° .

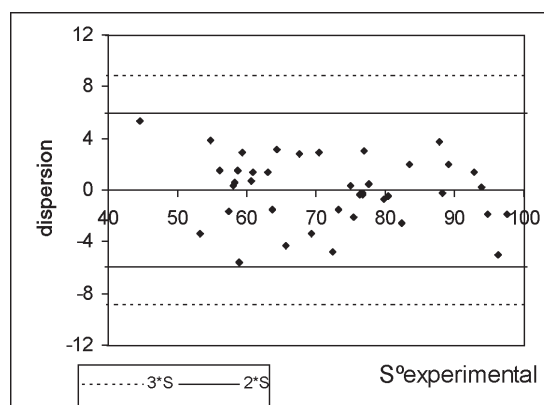


Figure 5. Dispersion plot for (3).

Other thermodynamical quantities under study are 37 ΔG_f . According to Figure 7, the best model has again 8 descriptors, with predictions randomly distributed in Figure 8 and indicated in Table 3.

$$\Delta G_f = 4 (\pm 3) + 9.1 (\pm 0.5) C - 4.1 (\pm 0.5) H + 14 (\pm 3) N - 41 (\pm 3) C-O - 45 (\pm 4) C=O + 32 (\pm 5) C-C_{\text{triple bond}} - 14 (\pm 4) Cl \quad (3)$$

$$N=37, R=0.9869, S=6.754 \text{ Kcal.mol}^{-1}, F=130.681$$

$$R_{\text{loo}}=0.9795, S_{\text{loo}}=7.330 \text{ Kcal.mol}^{-1}$$

$$R_{\text{l-20%-o}}=0.9000, S_{\text{l-20%-o}}=8.538 \text{ Kcal.mol}^{-1}$$

Tables 2 and 3 reveal that, as was the case for ΔH_f , the constitutional descriptors are not enough to obtain better predictions for S° and ΔG_f than those achieved with the more sophisticated molecular orbital theory methods. Additional structural information is needed to describe better

the behavior of the thermodynamical quantities, like that present in more elaborated topological indices, geometrical or electronic variables. Despite these results, the naive descriptors showed to be quantitative and able to establish relationships with the properties, although not as good as desired. In the three cases considered a linear equation was found suitable as the mathematical function of the model, according to the proximity of R to unity.

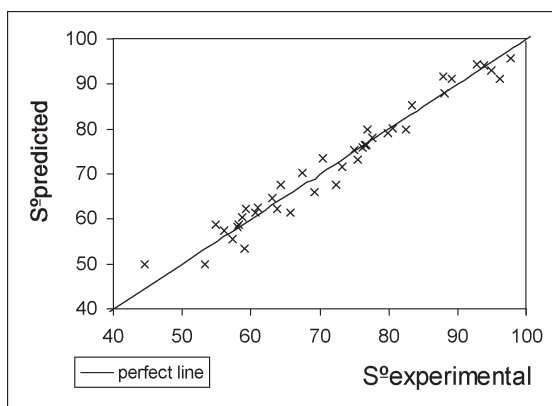


Figure 6. Linearity of model (3).

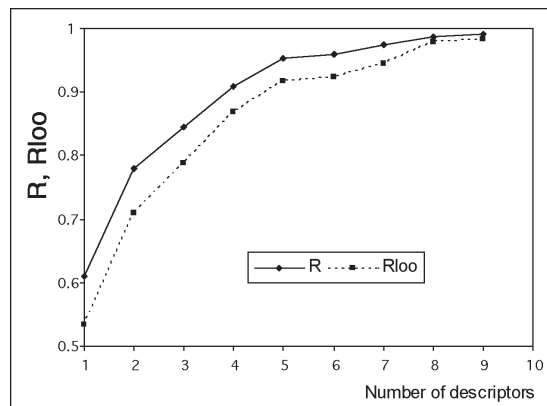


Figure 7. Model selection for “ G_f ”.

Table 2. Experimental and predicted ΔH_f (Kcal.mol⁻¹) with associated errors when using semiempirical methods.

N°	Molecule	exp.	predicted eq. (1)	error MNDO	error AM1	error PM3	error eq. (1)
1	Methane	-17.89	-4.638	5.93	9.1	4.86	13.25
2	Acetylene	54.34	38.61	3.53	0.44	3.65	15.72
3	Ethylene	12.45	14.53	2.93	4	4.16	2.08
4	Ethane	-20.24	-9.556	0.49	2.8	2.08	10.68
5	Allene	45.63	33.7	1.74	0.48	1.4	11.92
6	Cyclopropene	66.2	33.7	2.07	8.58	1.94	32.49
7	Cyclopropane	12.73	9.612	1.55	5.01	3.5	3.12
8	Propane	-24.83	-14.47	0.15	0.53	1.17	10.35
9	1,2-Butadiene	38.8	28.78	5.3	1.73	0.85	10.01
10	1-Butyne	39.5	28.78	3.4	2.05	3.85	10.71
11	Cyclobutene	37.45	28.78	6.48	8.26	0.17	8.67
12	Methyl cyclopropene	58.2	28.78	4.48	6.56	0.63	29.41
13	Cyclobutane	6.78	4.694	18.72	7.82	10.62	2.09
14	Isobutene	-4.3	4.694	2.25	3.1	0.93	8.99
15	Cyclopentadiene	32.1	47.95	0.08	4.9	0.4	15.85

16	Methyl-silane	-7.8	-28.72	5.86	3	4.16	20.92
17	Propene	4.88	9.612	0.07	1.66	1.49	4.73
18	Toluene	11.99	16.46	1.52	2.35	2.04	4.47
19	Propyne	44.39	33.7	3.03	1.02	4.2	10.68
20	1.3-Cyclohexadiene	25.4	43.03	10.98	7.93	5.06	17.63
21	Bicyclobutane	51.9	28.78	12.11	26.15	17.29	23.11
22	Hexane	-39.92	-29.22	0.71	4.98	0.06	10.69
23	Cycloheptatriene	43.2	62.2	9.47	4.98	0.77	19.00
24	2.2.3.3-Tetramethylbutane	-53.83	-39.06	44.93	17.41	9.38	14.76
25	Naphthalene	36.05	35.85	2.15	4.42	4.51	0.20
26	1.2.6.7-Cyclodecatetraene	85.11	71.53	17.25	13.62	5.45	13.57
27	Acetaldehyde	-39.73	-33.26	2.59	1.87	4.5	6.46
28	Acetone	-51.99	-38.18	2.54	2.75	1.37	13.80
29	Ethanol	-56.2	-57.35	6.83	6.5	0.69	1.15
30	Propanal	-45.5	-38.18	2.17	2.86	3.83	7.32
31	Cyclopentanone	-46.03	-23.93	11.04	9.35	9.19	22.09
32	Acetic acid	-103.3	-81.06	2.14	0.27	1.26	22.23
33	Ethylamine	-11.4	-21.6	1.88	3.78	1.16	10.20
34	Acetonitrile	17.7	26.57	1.5	1.55	5.56	8.87
35	Methylamine	-5.5	-16.68	2.07	1.91	0.29	11.18
36	Trimethylamine	-5.67	-26.51	2.83	3.92	5.24	20.84
37	Iodoform	50.4	31.49	18.41	12.6	10.21	18.90
38	Methanol	-48.1	-52.43	9.28	8.95	3.8	4.34
39	2-Methoxyethanol	-90.07	-110	9.72	13.46	1.16	19.99
40	1.2.4-Trichlorobenzene	1.16	-4.351	0.98	1.68	3.56	5.51
41	Bromoethylene	18.71	26.57	2.92	0.78	5.05	7.86
42	1.3-Diiodopropane	10.8	9.612	3.98	1.64	8.04	1.19
43	Acetyl-iodine	-30.2	-21.22	3.21	9.42	0.23	8.98
44	2.4.6-Tribromoaniline	38	45.46	2.96	2.46	10.96	7.47
45	1.4-Dibromobutane	-20.55	4.694	2.89	1.49	5.77	25.24
46	Isobutane	-32.41	-31.43	5.58	2.99	2.83	0.97
47	Nitromethane	-17.9	0.09314	21.17	7.91	1.91	17.99
48	Hexachlorobenzene	-8.6	-30.08	5.92	0.66	0.6	21.48
49	Fluoromethane	-56.8	-52.53	4.13	4.24	2.98	4.27
50	Chloromethane	-20	-13.21	2.52	1.03	5.3	6.78
51	1.2-Dichlorobenzene	7.89	4.225	0.58	1.25	3.17	3.66
52	1.4-Pentadiene	25.3	23.86	0.7	0.53	1.25	1.44
53	1-Chloronaphthalene	27.5	27.27	4.57	7.03	7.04	0.23

54	2-Bromonaphthalene	41.97	47.89	1.11	3.35	6.12	5.92
55	Bromomethane	-9.1	7.405	1.27	2.9	7.1	16.50
56	1.2.4.5-Tetrachlorobenzene	-7.8	-12.92	6.23	5.8	7.29	5.13
57	Trifluoroacetic acid	-255	-224.7	17.52	12.24	10.93	30.24
58	Chloroform	-7.8	-30.36	6.23	5.8	7.29	22.56
59	Dichloromethylsilane	-96	-45.87	1.45	2.48	7.08	50.12
60	Dichloromethane	-22.96	-21.79	5.09	2.91	5.8	1.17
61	Chlorotrimethylsilane	-84.6	-47.13	2.85	0.25	2.98	37.46
62	Trifluorochloromethane	-169.2	-156.9	9.56	6.51	0.08	12.29
63	Benzene	19.81	21.37	1.44	2.14	3.58	1.57
64	Cyclopentane	-18.3	-0.2233	12.23	10.56	5.64	18.07
65	Biphenyl	43.53	42.47	2.46	3.94	4.41	1.05
66	Fluorene	41.83	61.64	2.98	12.38	7.04	19.81
67	Phenanthrene	49.5	50.32	6	7.78	5.36	0.82
68	Anthracene	55.2	50.32	3.47	7.56	6.3	4.88
69	Octadecane	-99.08	-88.24	3.18	28.15	6.11	10.83
70	Pyrrole	25.9	17.95	6.48	13.91	1.15	7.94
71	Pyridine	34.6	29.5	5.86	2.63	4.3	5.10
72	Neopentane	-40.3	-24.31	15.63	7.46	4.4	15.98
73	Quinoline	47.92	43.97	3.3	4.16	0.49	3.95
74	Bicyclo(2.2.2)-octane	-24.1	9.11	2.36	12.05	3.79	33.21
75	Acetamide	-56.96	-45.31	9.79	6.22	7.56	11.64
76	3-Chloro-4-methylaniline	18	-4.159	10.54	11.63	12.36	22.15
77	Chloroethane	-26.8	-18.13	2	0.61	4.71	8.67
78	Cyclopropyl benzene	36.02	30.71	1.52	8.15	6.24	5.31
79	Benzoic acid	-70.1	-55.04	2.46	2.05	3.83	15.05
80	Fluoroacetylene	30	-9.277	14.37	14.81	11.91	39.27
81	Methylethyl sulfone	-97.6	-110	144.6	21.75	18.94	12.46
82	Hexamethyl benzene	-18.5	-8.129	21.36	5.83	3.42	10.37
83	Phenol	-23.04	-26.41	3.72	0.71	1.29	3.38
84	Methyl thiocyanate	38.33	26.57	14.52	15.58	9.41	11.75
85	Methyl isothiocyanate	27.1	26.57	9.82	0.55	9.04	0.53
86	Thiophenol	26.9	21.37	3.55	1.24	0.7	5.52
87	Butanethiol	-21.1	-19.39	1.85	3.16	1.59	1.71
88	Thiophene	27.6	29.99	1.18	0.21	3.07	2.40
89	2.3-Dihydrothiophene	21.8	28.78	18.03	11.67	8.43	6.98
90	5.6-Dimethyl chrysene	62.7	54.95	13.54	8.44	0.13	7.74
91	Iodobenzene	39.4	33.42	7	1.35	5.27	5.98

92	Diethyl sulfone	-102.5	-114.9	142.9	21.16	21.56	12.48
93	Isopropylamine	-20	-26.51	3.61	0.74	1.22	6.52
94	Thiolacetic acid	-43.5	-33.26	2.17	4.28	4.6	10.23
95	n-Propylamine	-16.8	-26.51	1.47	5.33	1.12	9.72
96	Toluenethiol	23.6	16.46	3.42	1.29	1.35	7.14
97	Methylpropylthioether	-19.5	-19.39	8.34	2.88	0.04	0.11
98	Tris(diethylamino) phosphine	-54.7	-106.9	11.84	23.93	14.45	52.21
99	Ethylphosphine	-12	-21.6	9.12	8.11	0.3	9.60
100	Tetraethylsilane	-64.4	-63.15	17.73	7.52	13.61	1.25
101	Hexafluorobenzene	-228.5	-265.9	15.06	2.78	0.89	37.49
102	Trifluoroacetonitrile	-118.4	-117.1	5.15	1.15	3.3	1.29
103	Trifluoromethylbenzene	-143.2	-127.2	15.57	8.71	8.17	15.97
104	Trifluoromethane	-166.3	-148.3	2.45	6.22	4.3	17.97
105	N,N'-Difluorobenzylamine	1.79	-2.571	9.5	9.3	12.32	4.36
106	Fluorobenzene	-27.76	-26.51	2.43	4.37	7.45	1.24
107	Tetranitromethane	18.5	14.28	76.32	34.4	12.22	4.21
108	1-Iodonaphtalene	55.9	47.89	3.99	2.5	9.93	8.01
109	Methyl nitrate	-29.11	0.09314	16.61	2.29	3.37	29.20
110	N-Nitrodimethylamine	-3.2	-16.86	25.45	24.83	4.39	13.66
111	Nitrobenzene	15.4	26.11	22.08	9.78	0.97	10.71
112	Nitroethane	-23.5	-4.824	21.35	6.62	2.59	18.67
113	Ethyl decanoate	-146.3	-130.2	9.85	10.05	5.07	16.05
114	Methyl 4-methylbenzoate	-74.71	-64.88	8.78	5.53	7.11	9.83
115	4-Ethylbenzoic acid	-84.99	-64.88	7.6	8.76	6.08	20.10
116	3,4-Dimethylbenzoic acid	-86.62	-64.88	8.45	4.27	2.82	21.73
117	Formaldehyde	-25.95	-28.34	6.95	5.56	8.15	2.40
118	1,2-Dimethoxyethane	-81.93	-114.9	11.51	17.78	7.81	33.05
119	t-Butanol	-74.7	-67.19	10.36	3.03	3.37	7.51
120	Dimethylperoxide	-30.1	-105.1	1.75	4.32	4.05	75.05
121	1,3-Propanediol	-97.61	-110	12.63	16.6	2.83	12.45
122	Dimethoxymethane	-83.21	-110	9.38	16.7	5.59	26.85
123	Methyl benzoate	-66.84	-59.96	8.71	5.53	8.82	6.88
124	2-Ethylbenzoic acid	-81.43	-64.88	6.92	7.17	4.62	16.54
125	2,5-Dimethylbenzoic acid	-83.91	-64.88	8.09	1.88	1.8	19.02
126	2,4,6-Trimethylbenzoic acid	-89.44	-69.8	3.77	2.22	1.72	19.63
127	Methylenecyclopropane	47.92	28.78	10.07	0.31	3.45	19.13
128	2,4,6-Trinitrotoluene	12.9	30.65	61.59	28.1	9.74	17.75
129	Methylcyclopropene	58.2	28.78	4.48	6.56	0.63	29.41

130	Hexanitroethane	42.8	18.83	121.8	42.99	14.84	23.96
131	Fluoroethylene	-32.5	-33.36	2.05	1.58	3.88	0.87
132	Fluorodinitrophenyl methane	-44.2	-21.97	58.64	44.28	22.69	22.22
133	1.1-Difluoroethane	-118.8	-105.3	5.32	0.15	6.84	13.45
134	Trimethylphosphine	-22.5	-26.51	25.71	0.46	7.36	4.02
135	Diethylphosphine	-25	-31.43	20.06	6.62	1.73	6.44
136	Isopropanethiol	-18.1	-14.47	1.88	2.05	3.73	3.63
137	Diethyl thioether	-19.89	-19.39	8.17	2.26	0.66	0.50
138	Thiacyclohexane	-15.12	-0.2233	12.97	8.81	0.67	14.89
139	Bicyclo(2.1.0)-pentane	37.3	23.86	7.17	8.69	0.44	13.43
140	1.2-Dimethyl cyclopropene	46.4	23.86	7.15	8.36	0.64	22.53
141	Benzoyl chloride	-26.1	-15.82	2.61	10.28	7.63	10.27
142	Chloro- pentafluoroethane	-266.3	-257.6	18.58	10.52	1.55	8.69
143	Chlorodimethylsilane	-69.9	-42.22	3.17	0.1	6.14	27.67
144	Tetrachloroethylene	-2.7	-19.77	5.33	9.78	5.42	17.07
145	Diacetylene	113	76.95	9.85	6.96	10.58	36.04
146	Bromotrifluormethane	-155.1	-136.2	8.47	10.49	2.78	18.81
147	1.5-Hexadiene	20.1	18.94	0.47	2.37	0.92	1.15
148	Benzyl iodide	25.1	28.5	3.47	5.7	12.45	3.40
149	Fluoroethane	-62.9	-57.45	2.25	3.44	2.65	5.45
150	1.2-Dimethyl cyclobutene	19.8	18.94	13.46	7.07	3.71	0.85
151	t-Butyldifluoroamine	-46	-38.42	24.84	17.01	14.39	7.57
	N.N. N'.N'-tetrafluoro-						
152	1.1-cyclohexanediamine	-41.6	-43.2	16.84	4.95	9.94	1.61
153	t-Butylamine	-28.9	-31.43	13.38	7.63	3.63	2.54
154	Diethyl peroxide	-46.1	-114.9	7.76	7.22	0.78	68.88
155	Acetyl bromide	-45.6	-21.22	2.33	11.23	2.03	24.37
156	Benzoyl bromide	-11.6	4.795	0.6	12.12	3.03	16.39
157	1.1-Dibromoethane	-9.8	14.53	0.56	4.18	5.15	24.33
158	Difluoro-dichloromethane	-117.5	-117.5	7.3	10.45	1.42	0.08
159	Triphenylphosphine	78.5	51.53	27.11	20.54	11.47	26.96
160	Triethylsilane	-39.5	-53.31	24.94	15.51	15.93	13.81
161	Ethylsilane	-15	-33.64	6.71	1.51	4.8	18.64
162	Dichlorodimethylsilane	-109.5	-50.79	1.76	5.22	1.48	58.70
163	Pentafluorobenzene	-192.5	-218.1	9.28	0.4	3.78	25.60

Error = absolute value of the difference between predicted and experimental value for the property.

Table 3. Experimental and predicted values of S° ($\text{cal.mol}^{-1}.\text{K}^{-1}$) and ΔG_f (Kcal.mol^{-1}).

N ^o (a)	molecule	exp. S°	predicted S° eq. (2)	exp. ΔG_f	predicted ΔG_f eq. (3)
1	Methane	44.52	49.84	-12.15	-3.68
2	Ethane	54.85	58.73	-7.87	-2.85
6	Propene	63.8	62.27	14.99	6.21
7	Toluene	76.64	76.42	29.16	34.21
8	Propyne	59.3	62.27	46.47	46.94
9	Acetaldehyde	63.15	64.51	-31.86	-40.10
10	Acetone	70.49	73.39	-36.58	-39.26
11	Acetic acid	67.52	70.29	-90.03	-81.90
12	Acetonitrile	58.19	58.73	25.24	23.82
13	Methylamine	57.98	58.28	7.71	6.54
16	Methanol	57.29	55.62	-38.84	-45.48
23	Methylisothiocyanate	69.29	65.94	—	—
25	Nitromethane	65.73	61.4	-1.66	14.76
26	Fluoromethane	53.25	49.84	-50.19	-51.87
27	Chloromethane	56.04	57.53	-15.03	-13.55
28	Bromomethane	58.75	60.24	-6.73	0.44
31	n-Hexane	92.83	94.27	-0.06	0.48
38	Benzene	64.34	67.53	30.99	33.37
40	Biphenyl	93.85	94.11	66.94	71.26
44	Benzoic acid	88.19	87.98	-50.29	-44.85
48	Phenol	75.43	73.31	-7.86	-8.43
61	Fluorobenzene	72.33	67.53	-16.5	-14.82
64	Iodobenzene	79.84	79.12	44.88	37.49
17	Methanethiol	60.96	62.4	-2.37	-3.68
18	Dimethylsulphoxide	73.2	71.72	—	—
24	Methylisocyanide	58.98	53.38	39.91	23.82
29	Iodomethane	60.71	61.42	3.74	0.44
30	n-Pentane	83.4	85.38	-2	-0.35

32	1-Pentanol	96.21	91.16	-35.79	-42.16
35	1-Chloropentane	94.89	93.08	-8.94	-10.23
36	1-Bromopentane	97.7	95.79	-1.37	3.76
39	Styrene	82.48	79.96	51.1	43.26
41	Ethynylbenzene	76.88	79.96	84.46	83.99
43	Acetophenone	89.12	91.08	0.44	-2.21
45	Benzonitrile	76.73	76.42	62.35	60.87
46	Aniline	76.28	75.97	39.84	4.38
49	Benzenethiol	80.51	80.09	35.28	43.59
53	Benzenesulphonic acid	87.88	91.65	—	—
62	Chlorobenzene	74.92	75.22	23.7	33.37
63	Bromobenzene	77.53	77.93	33.11	23.50

^(a)Number as in Ref. [19].

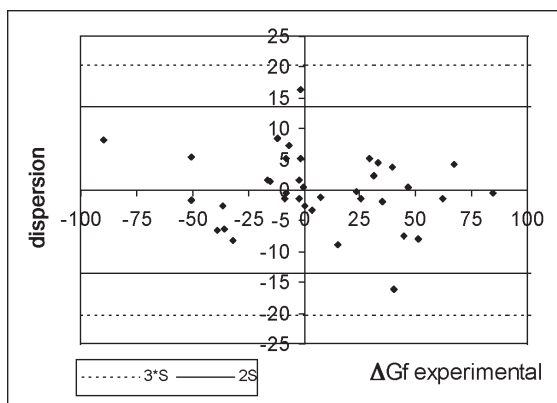


Figure 8. Dispersion plot for (5).

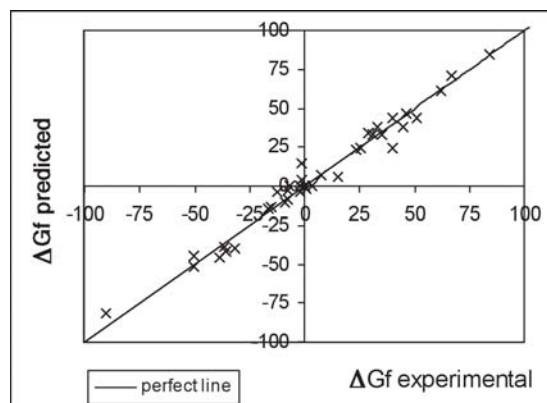


Figure 9. Linearity of model (5).

Conclusions

We have presented a rather simple and direct calculation scheme to derive thermodynamic parameters, such as heat of formation, entropy and Gibbs free energy. This approach is quite different to the molecular orbital theory, which have to make some involved calculation procedures to build the Fock matrix. The choice of topological descriptors is based on the most intuitive chemical concepts: atoms and chemical bonds. Comparisons between the theoretical estimations and experimental data revealed here that this sort of elementary descriptors have to be complemented with others derived in a different way, that is to say, taking into account geometrical and electronic characteristics of the molecular structure. At present, research along this line is under development in our laboratories and results will be published elsewhere in the forthcoming future.

Acknowledgment

P. R. D. acknowledges financial support from CONICET, RA.

References

- [1] Huang, K., *Statistical Mechanics*, Wiley, New York, 1963.
- [2] Reif, F., *Fundamentals of Statistical and Thermal Analysis*, McGraw-Hill, New York, 1965, pp. 2-4.
- [3] Barrow, G.M., *Physical Chemistry*, McGraw-Hill, New York, 1961.
- [4] Stull, D.R.; Westrum, E.F.; Sinke, G.C., *The Chemical Thermodynamics of Organic Compounds*, Wiley, New York, 1969.
- [5] Gurvich, L.V.; Karachevtsev, G.V.; Kondrat'ev, V.N. Lebedev, Yu.A.; Medvedev, V.A.; Potapov, V.K.; Khodeev, Yu.S., *Energies of Chemical Bonds Splitting Ionization Potentials and Electron Affinity*, V. N. Kondrat'ev, Ed. Nauka, Moscow, 1974.
- [6] Randic, M.; Trinajstic, N. *New J. Chem.*, **1994**, 18, 179.
- [7] Randic, M., *New J. Chem.*, **1996**, 20, 10019.
- [8] Estrada, E.; Ivanciuc, O.; Gutman, I.; Gutiérrez, A.; Rodríguez, L., *New J. Chem.*, **1998**, 22, 819.
- [9] Predict, <http://www.mwsoftware.com/dragon/desc.html>
- [10] ChemEng Software Design, <http://www.cesd.com/chempage.htm>
- [11] Artist, http://www.ddbst.de/new/Win_DDBSP/frame_Artist.htm
- [12] Sabljic, A.; Trinajstic, N., *Acta Pharm. Jugosl.*, **1981**, 31, 189.
- [13] Hansen, P.J.; Jurs, P.C. *J. Chem. Educ.*, **1988**, 65, 574.
- [14] Duchowicz, P.R.; Castro, E.A., *Arkivoc*, **2001**, 2, 227.
- [15] Duchowicz, P.R.; Castro, E.A., *J. Indian Chem. Soc.*, **2001**, 78, 192.
- [16] Duchowicz, P.R.; Castro, E.A., *J. Korean Chem. Soc.*, **2000**, 44, 501.
- [17] Duchowicz, P.R.; Castro, E.A., *Acta Chem. Slov.*, **2000**, 47, 281.
- [18] Duchowicz, P.R.; Castro, E.A., *J. Korean Chem. Soc.*, **1999**, 43, 621.
- [19] Pankratov, A.N., *Afinidad*, **1999**, LVI, 257.
- [20] MOPAC, http://www.cachesoftware.com/mopac/Mopac2002manual/table_of_heats.html
- [21] Duchowicz, P.R.; Castro, E.A.; Fernández, F.M., *Commun. Math. Comput. Chem.(MATCH)*, **2006**, 55, 179.
- [22] Katritzky, A.R.; Fara, D.C.; Karelson, M., *Bioorg. Med. Chem.*, **2004**, 12, 3027.
- [23] Mercader, A.; Castro, E.A.; Toropov, A.A., *Int. J. Mol. Sci.*, **2001**, 2, 121.
- [24] Mercader, A.; Castro, E.A.; Toropov, A.A., *Chem. Phys. Lett.*, **2000**, 330, 612.
- [25] Gavernet, L.; Firpo, M.; Castro, E.A., *Rev. Roum. Chim.*, **1998**, 43, 1079.
- [26] Vericat, C.; Castro, E.A., *Egypt. J. Chem.*, **1998**, 41, 109.