# QSPR EVALUATION OF THERMODYNAMIC PROPERTIES OF ACYCLIC AND AROMATIC COMPOUNDS

Pablo R. Duchowicz[1], Eduardo A. Castro[1]*, Alexei N. Pankratov[2]

[1]INIFTA, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Diagonal 113 y 64, Sucursal 4, C.C. 16, La Plata 1900, Argentina

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

**PCS: physchem/0207006**

**Abstract:** A QSPR study of three fundamental thermodynamic parameters for a representative set of acyclic and aromatic compounds is made on the basis of rather simple topological descriptors. Results are quite satisfactory and they demonstrate the convenience to take into account fundamental concepts on molecular structure, such as atoms and chemical bonds. Numerical comparisons are performed with respect to previous results derived from semiempirical all-valence MNDO, AM1 and PM3 molecular orbital methods. Some possible further extensions of the present procedure are pointed out.

**Key words:** Thermodynamics – Acyclic and aromatic compounds – QSPR theory – Molecular descriptors – Gibbs free energy – Enthalpy of formation – Entropy.

* Corresponding author (e-mail: castro@quimica.unlp.edu.ar / jubert@arnet.com.ar)

## 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 the macroscopic system, such as the pressure P, the volume V, the temperature T, and the magnetic field B. They are defined experimentally. Thermodynamics is concerned with the mathematical relationships between thermodynamic parameters.

Macroscopic systems (like gases, liquids, or solids) began first to be systematically investigated from a macroscopic phenomenological point of view in the last century. The laws thus discovered formed the subject of “thermodynamics”. The strength of this discipline is its great generality, which allows one to make 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]**.

Experimental measurements of some thermodynamic parameters involve experimental difficulties and they are not always feasible, and the corresponding methods possess real drawbacks **[4-7]**. Consequently, it is necessary to resort to a theoretical calculation of these parameters. This option 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 **[8,9]**. This kind of study is based on the paradigm that physical-chemistry properties and biological activities are dependent on molecular structure. As a consequence, one of the most important points in such research is the selection of adequate descriptors containing the information stored in the molecular structure **[10]**.

In a somewhat recent paper one of the present authors (A. N. P.) have computed standard values of entropy, heat of formation and Gibbs free energy for a representative set of acyclic and aromatic compounds by means of the semiempirical all-valence MNDO, AM1 and PM3 molecular orbital methods **[11]**. It could be demonstrate the existence of quantitative relations between experimental data and theoretical results, although some unsuccessful reproductions were noted. The aim of this paper is to present the results of the calculations of these three fundamental thermodynamic parameters via a different theoretical framework for the same molecular set in order to surmount some previous uncertainties.

This paper is organized as follows: next section deals with the presentation of the theoretical method and its antecedents for this sort of numerical estimations. Then we display results for the predictions of thermodynamic functions and discuss them with respect to previous calculations. Finally, we state our main conclusions and some possible further extensions to other molecular sets and additional physical-chemistry properties.

## THEORETICAL METHOD

The graph-theoretical approach to quantitative structure-property relationships (QSPR) is based on a well-defined mathematical representation of the molecular structure and the basic mathematical relationship is

$$P = f(D) \tag{1}$$

where P is the property, D is a set of molecular descriptors and f stands for any arbitrary function, although usually it represents a polynomial relationship.

The molecular descriptors therefrom are commonly named *"topological indices"* **[12,13]**. These indices are derived from a well-defined mathematical representation of the molecular structure **[14,15]** and they contain relevant information about it. Owing to the complexity of the molecular structure, 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 an active research field within the realm of QSPR theory. However, this search should not be at random and it should follow some regular procedure based on the desired attributes that a molecular structure descriptor needs to possess **[10]**.

For more than a century most chemists have used constitutional formulae without realizing that by representing "connectedness" of atoms such formulae are graphs or multigraphs **[16]**. As a matter of fact, the structural (or constitutional) formula of a chemical compound may be regarded as molecular graph where the vertices represent atoms while the edges stand for valence bonds **[17]**. Evidently, the most simple and obvious sort of graph theoretical indices are bonds and chemical bonds. 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 demonstrate their usefulness to predict physical-chemistry properties and biological activities **[18-22]**. These parameters may be calculated solely from consideration of molecular structure and their chemical interpretation is quite direct. They can be computed very readily and have the advantage that they may be applied to quite diverse sets of structures.

The mathematical approach consists on defining a polynomial relationship such as

$$P = aA + bB + \ldots + \alpha A^2 + \beta B^2 + \ldots + \aleph A^n + \Im B^n + \ldots. + t \tag{2}$$

where a, b, …, $\alpha$, $\beta$, …, $\aleph$, $\Im$, …, t are numerical coefficients determined from the fitting procedure and A, B, … are the molecular descriptors. In this case, for example, A is the number of carbon atoms, B is the number of hydrogen atoms, C is the number of single carbon-carbon bonds, etc. for each molecule. In this work n = 1, 2, 3, respectively (i.e. linear, quadratic and cubic polynomials.

The fitting equations were determined through the well known procedures of polynomial regressions employing the MATHEMATICA® software **[23-25]**. Besides, we have performed the fitting procedure at first, second and third polynomial-orders. In fact, many correlations, particularly when involving molecules of different size, need not be linear. But even if we have molecules of the same or similar size, a higher-order regression function may result in a better description of the relationship than a simple linear model. In general, it is advisable to test polynomial regressions for quadratic dependence and, if warranted, for higher order polynomial fitting polynomials or other suitable functional dependence **[26]**.

The molecular set was chosen identical to that chosen by Pankratov **[11]** in order to make a simple and direct comparison with previous results. Molecular set is displayed in Table 1 together with experimental thermodynamic properties.

**TABLE 1**

This molecular set is quite representative of relative small organic molecules since it contains, for example, alkanes, benzenoic derivatives, nitro derivatives, halogen derivatives, alkyl benzene compounds, thiol derivatives, alkyne and alkene hydrocarbons, acids, etc. among others, so that there is a wide variety of molecules. The empirical data used in this work are experimental results reported in a previous publication **[11]**.

## RESULTS AND DISCUSSION

We have made linear, quadratic and cubic polynomial fittings and statistical estimations are given in Table 1 (columns 3, 5, and 7) only for linear regression equations, since results do not improve markedly when computing second- and third-order formulae. Statistical results are displayed in Table 2.

**TABLE 2**

The comparison between experimental and theoretical data show that a satisfactory agreement exists between both quantities and deviations are within the experimental errors. Thus, for example, it is well known that experimental uncertainties for heats of formation is around 2-3 kcal/mol **[18, 27-30]**.

In order to make a direct comparison with previous Pankratov's estimations **[11]** we have calculated the relationship

$$P(exper.) = q\,P(theor.) \tag{3}$$

where P is any of the chosen thermodymanic properties. Results are given in Table 3.

## TABLE 3

Figures 1-3 present the relationships between experimental and theoretical thermodynamic properties

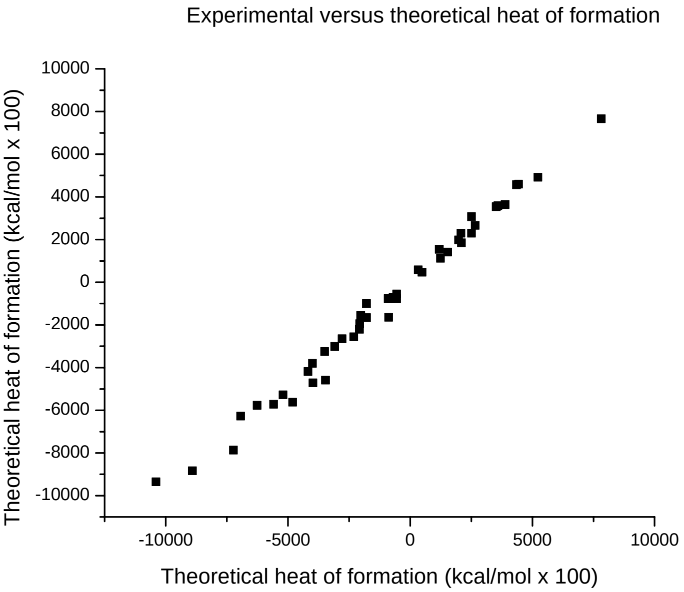

Figure 1.

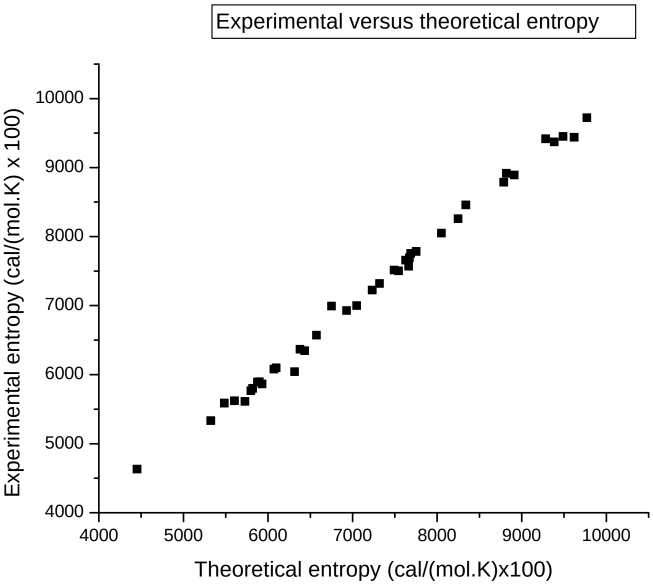

Figure 2.

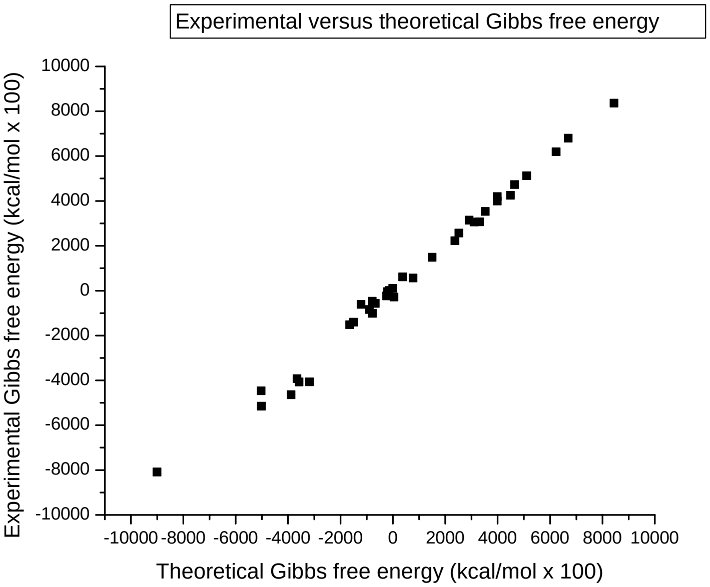

Figure 3.

The analysis of numerical and graphical results shows that present approach gives quite satisfactory estimations. In fact, when comparing present results for standard heats of formation with those derived from semiempirical methods (Table I in Ref.11), we note that maximum deviations from MNDO, AM1 and PM3 methods are 148.68, 22.62 and 21.99 kcal/mol, respectively, while present calculation yields a maximum deviation equal to 10.43. Besides, average absolute deviations derived from MNDO, AM1 and PM3 semiempirical methods are 19.52, 4.34 and 4.08 kcal/mol, respectively, and present procedure gives an average absolute deviation equal to 2.82 kcal/mol. Furthermore, semiempirical estimations yields some rather pathological results, such as those given for dimethylsulphoxide, dimethylsulphone, methylisocyanide, nitromethane, methylphenylsulphone, and nitrobenzene molecules with deviations larger than 20 kcal/mol from experimental values according to the MNDO method, for example. Present calculations do not yield such large deviations from experimental data. In order to assess the quality of present calculations, it is interesting to point out that experimental uncertainties lies within the range 2-3 kcal/mol.

Regarding estimations of standard entropies, semiempirical methods show maximum deviations for MNDO, AM1 and PM3 procedures equal to 9.89, 8.21, and 7.63 cal/(mol K), respectively, while present calculation yields a maximum deviation of 9.33 cal/(mol K). The average absolute deviations derived from the semiempirical methods are 2.18, 2.17, and 2.31 cal/(mol K) for the MNDO, AM1 and PM3 techniques, respectively, while this quantity is equal to 0.55 cal/(mol K) for present calculation. Once again, we note the best predictive capability of the QSPR method.

The maximum deviations estimated for free Gibbs energies of formation are equal to 23.94, 14.11 and 18.16 kcal/mol when computed via MNDO, AM1 and PM3 semiempirical methods, respectively, while our calculations present a maximum deviation of 9.13 kcal/mol. The average absolute deviations derived from semiempirical calculations are equal to 3.31, 2.82 and 3.21 kcal/mol according to MNDO, AM1 and PM3 methods, respectively, and our estimation gives an average absolute deviation equal to 2.22 kcal/mol.

Thus, present results show the existence of a quite good agreement between theoretical and experimental results when calculated through a QSPR scheme resorting to the most naïve topological descriptors: atoms and chemical bonds. In order to appreciate the value and usefulness of these findings one must take into account the extremely simple computational procedure to derive molecular descriptors: just to count and to identify atoms and chemical bonds. Then, the advantages of this method are quite evident: a) extreme simplicity to perform the calculations; b) direct possibility to extend the approach to other classes of compounds; c) better results than those provided by other more complex computational schemes; and d) capability to make predictions when experimental results are not available or/and the experimental determination is rather expensive.

## 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. The choice of topological descriptors is based on the most intuitive chemical concepts: atoms and chemical bonds. The comparison between theoretical estimations and experimental data reveals a very good agreement degree. It is particularly noticeable the lack of “pathological” predictions (*i.e.* outliers) within the chosen molecular set for any of the three thermodynamic properties. The comparison of the present results with those derived from MNDO, AM1 and PM3 semiempirical methods shows the QSPR theory estimations are better than those derived from molecular orbital theory. This finding is really significant since this last procedure resort to rather elementary concepts as the foundation for computing the basic variables (*i.e.* topological descriptors), while molecular orbital theory have to make some involved calculation procedures to build the Fock matrix.

These results are in line with other previous ones **[18-22]** and they give evidence of the theoretical soundness and practical usefulness of those primitive chemical concepts, such as atoms and chemical bonds, when suitably employed within the context of the QSPR theory. Perhaps it should be necessary to explore other similar applications before to deduce more definitive conclusions. That is to say, it could be interesting to calculate other physical-chemistry properties for the same molecular set and other sort of molecules. At present, research along this line is under development in our laboratories and results will be published elsewhere in the forthcoming future.

## REFERENCES

**[1]** K. Huang, **Statistical Mechanics**, Wiley, New York, 1963, p.3.
**[2]** F. Reif, **Fundamentals of Statistical and Thermal Analysis**, McGraw-Hill, New York, 1965, pp.2-4.
**[3]** G. M. Barrow, **Physical Chemistry**, McGraw-Hill, New York, 1961, p.86.
**[4]** L. V. Vilkov, Yu. A. Pentin, **Physical Methods of Investigation in Chemistry. Structural Methods and Optical Spectroscopy**, Vysshaya Shkola, Moscow, 1987, 367
**[5]** D. R. Stull, E. F. Westrum, G. C. Sinke, **The Chemical Thermodynamics of Organic Compounds**, Wiley, New York, 1969.
**[6]** L. V. Gurvich, G. V. Karachevtsev, V. N. Kondrat'ev, Yu. A. Lebedev, V. A. Medvedev, V. K. Potapov, Yu. S. Khodeev, **Energies of Chemical Bonds Splitting. Ionization Potentials and Electron Affinity**, V. N. Kondrat'ev, Editor-in-Chief, Nauka, Moscow, 1974, p. 351.
**[7]** O. A. Osipov, V. I. Minkin, A. D. Gamovsky, **Reference Book on Dipole Moments**, Vysshaya Shkola, Moscow, 1971, p. 416.
**[8]** M. Randic, N. Trinajstic, *New J. Chem.* **18** (1994) 179.
**[9]** M. Randic, *New J. Chem.* **20** (1996) 10019.
**[10]** E. Estrada, O. Ivanciuc, I. Gutman, A. Gutiérrez, L. Rodríguez, *New J. Chem.* **22** (1998) 819.
**[11]** A. N. Pankratov, *Afinidad* **LVI** (1999) 257.
**[12]** A. Sablijc, N. Trinajstic, *Acta Pharm. Jugosl.* **31** (1981) 189.
**[13]** P. J. Hansen, P. C. Jurs, *J. Chem. Educ.* **65** (1988) 574.
**[14]** P. G. Seybold, M. May, V. A. Bagal, *J. Chem. Educ.* **64** (1987) 189.
**[15]** Z. Mihalic, N. Trinajstic, *J. Chem. Educ.* **69** (1992) 701.
**[16]** A. T. Balaban, Ed., **Chemical Applications of Graph Theory**, Academic Press, New York, 1976.
**[17]** D. Bonchev, **Information Theoretic Indices for Characterization of Chemical Structures**, UMI, Bell & Howell Company, Ann Arbor, 1998, p. 63.
**[18]** P. R. Duchowicz, E. A. Castro, *Arkivoc* **2**(6) (2001) 227.
**[19]** P. R. Duchowicz, E. A. Castro, *J. Indian Chem. Soc.* **78** (2001) 192.
**[20]** P. R. Duchowicz, E. A. Castro, *J. Korean Chem. Soc.* **44**(6) (2000) 501.
**[21]** P. R. Duchowicz, E. A. Castro, *Acta Chim. Slov.* **47** (2000) 281.

**[22]** P. R. Duchowicz, E. A. Castro, *J. Korean Chem. Soc.* **43**(6) (1999) 621.
**[23]** S. Wolfram, **The MATHEMATICA® Book**, Fourth edition, Cambridge University Press, Cambridge, 1999.
**[24]** P. T. Tam, **A Physicist's Guide to MATHEMATICA®**, Academic Press, San Diego, 1997.
**[25]** W. H. Cropper, **MATHEMATICA® Computer Programs for Physical Chemistry**, Springer, New York, 1998.
**[26]** M. Randic, S. C. Basak, ***Variable Molecular Descriptors***, in **Some Aspects of Mathematical Chemistry**, S. C. Basak, R. K. Mohany, I. N. Busa Mallick, Eds., Visva Bharati University Press, Santiniketan, India, 1999, p. 24.
**[27]** A. Mercader, E. A. Castro, A. A. Toropov, *Int. J. Mol. Sci.* **2** (2001) 121.
**[28]** A. Mercader, E. A. Castro, A. A. Toropov, *Chem. Phys. Lett.* **330** (2000) 612.
**[29]** L. Gavernet, M. Firpo, E. A. Castro, *Rev. Roum. Chim.* **43**(11) (1998) 1079.
**[30]** C. Vericat, E. A. Castro, *Egyp. J. Chem.* **41**(1-6) (1998) 109.

**Table 1.** Molecular set and experimental values (taken from **[11])** and theoretical results of standard heats of formation, entropies and Gibbs free energies of formation. Numbering as in Ref. 11.

| **Compound** | **Δ$H_f$(kcal/mol)** | | **S(cal mol$^{-1}$K$^{-1}$)** | | **Δ$G_f$(kcal/mol)** | |
|---|---|---|---|---|---|---|
| | **Exper.** | **Theor.** | **Exper.** | **Theor.** | **Exper.** | **Theor.** |
| **1.** Methane | -17.89 | -10.02 | 44.52 | 46.30 | -12.15 | -6.11 |
| **2.** Ethane | -20.24 | -15.62 | 54.85 | 55.87 | -7.87 | -4.69 |
| **3.** Methyl-silane | -7.8 | -7.80 | - | - | - | - |
| **6.** Propene | 4.88 | 4.71 | 63.80 | 63.68 | 14.99 | 14.88 |
| **7.** Toluene | 11.95 | 15.46 | 76.64 | 75.69 | 29.16 | 31.44 |
| **8.** Propyne | 44.32 | 45.93 | 59.30 | 68.63 | 46.47 | 47.30 |
| **9.** Acetaldehyde | -39.76 | -47.13 | 63.15 | 60.43 | -31.86 | -40.65 |
| **10.** Acetone | -52.00 | -52.74 | 70.49 | 70.00 | -36.58 | -39.23 |
| **11.** Acetic acid | -103.93 | -93.50 | 67.52 | 69.94 | -90.03 | -80.90 |
| **12.** Acetonitrile | 21.00 | 18.49 | 58.19 | 58.00 | 25.24 | 25.63 |
| **13.** Methylamine | -5.5 | -7.71 | 57.98 | 57.67 | 7.71 | 5.61 |
| **14.** Mehtylphosphine | -7.0 | -7.00 | - | - | - | - |
| **16.** Methanol | -48.08 | -56.25 | 57.29 | 56.11 | -38.84 | -46.42 |
| **17.** Methanethiol | -5.49 | -5.49 | 60.96 | 60.96 | -2.37 | -2.37 |
| **18.** Dimethylsulphoxide | -34.6 | -45.87 | 73.20 | 73.20 | - | - |
| **19.** Dimethylsulphone | -89.1 | -88.37 | - | - | - | - |
| **22.** Methylthiocyanate | 25.1 | 30.74 | - | - | - | - |
| **23.** Methylisothiocyanate | - | - | 69.29 | 69.29 | - | - |
| **24.** Methylisocyanide | 35.9 | 35.90 | 58.98 | 58.98 | 39.91 | 39.91 |
| **25.** Nitromethane | -17.86 | -16.57 | 65.73 | 65.73 | -1.66 | -1.66 |
| **26.** Fluoromethane | -55.90 | -57.22 | 53.25 | 53.33 | -50.19 | -51.51 |
| **27.** Chloromethane | -20.63 | -19.44 | 56.04 | 56.21 | -15.03 | -14.10 |
| **28.** Bromomethane | -9.00 | -7.68 | 58.75 | 58.92 | -6.73 | -5.67 |
| **29.** Iodomethane | 3.34 | 5.76 | 60.71 | 60.81 | 3.74 | 6.15 |
| **30.** n-Pentane | -35.00 | -32.43 | 83.40 | 84.59 | -2.00 | -0.42 |
| **31.** n-Hexane | -39.96 | -38.03 | 92.83 | 94.17 | -0.06 | 1.00 |
| **32.** 1-Pentanol | -72.27 | -78.66 | 96.21 | 94.40 | -35.79 | -40.74 |
| **35.** 1-Chloropentane | -41.80 | -41.84 | 94.89 | 94.50 | -8.94 | -8.41 |
| **36.** 1-Bromopentane | -30.87 | -30.09 | 97.70 | 97.21 | -1.37 | 0.02 |
| **38.** Benzene | 19.82 | 19.84 | 64.34 | 63.46 | 30.99 | 30.61 |
| **39.** Styrene | 35.22 | 35.39 | 82.48 | 82.60 | 51.10 | 51.21 |
| **40.** Biphenyl | 43.52 | 45.73 | 93.85 | 93.70 | 66.94 | 67.97 |
| **41.** Ethynylbenzene | 78.22 | 76.01 | 76.88 | 77.55 | 84.46 | 83.63 |
| **42.** Benzaldehyde | -8.8 | -16.46 | - | - | - | - |
| **43.** Acetophenone | -20.76 | -22.07 | 89.12 | 88.93 | 0.44 | -2.90 |
| **44.** Benzoic acid | -69.36 | -62.69 | 88.19 | 89.16 | -50.29 | -44.64 |
| **45.** Benzonitrile | 52.30 | 49.17 | 76.73 | 76.92 | 62.35 | 61.96 |
| **46.** Aniline | 20.76 | 22.97 | 76.28 | 76.59 | 39.84 | 41.94 |
| **48.** Phenol | -23.03 | -25.57 | 75.43 | 75.03 | -7.86 | -10.10 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **49.** Benzenethiol | 26.66 | 26.66 | 80.51 | 80.51 | 35.28 | 35.28 |
| **52.** Methylphenylsulphone | -62.6 | -57.69 | - | - | - | - |
| **53.** Benzenesulphonic acid | - | - | 87.88 | 87.88 | - | - |
| **60.** Nitrobenzene | 15.4 | 14.11 | - | - | - | - |
| **61.** Fluorobenzene | -27.86 | -26.54 | 72.33 | 72.25 | -16.5 | -15.18 |
| **62.** Chlorobenzene | 12.39 | 11.24 | 74.92 | 75.14 | 23.70 | 22.23 |
| **63.** Bromobenzene | 25.10 | 23.00 | 77.53 | 77.85 | 33.11 | 30.66 |
| **64.** Iodobenzene | 38.85 | 36.43 | 79.84 | 79.74 | 44.88 | 42.47 |
| **Average absolute Deviation** | ***2.82*** | | ***0.55*** | | ***2.22*** | |

**Table 2.** Statistical results for first-order fitting polynomials corresponding to the three thermodynamic properties.*

| **Property** | **Regression coefficient** | **Average absolute deviation** | **Standard error of estimate** | **Number of Molecules** |
|---|---|---|---|---|
| **$\Delta H_f$** | 0.9946 | 2.82 | 5.92 | 45 |
| **S** | 0.9979 | 0.55 | 1.37 | 40 |
| **$\Delta G_f$** | 0.9960 | 2.22 | 5.49 | 40 |

* A complete listing of the coefficients corresponding to the three linear regression equations are available upon request to one of the authors (E.A.C.).

**Table 3.** Coefficients q and r (correlation coefficient) in equation (3). Results for MNDO, AM1 and PM3 are taken from Ref. 11.

| **Property** | **Method** | **q** | **r** |
|---|---|---|---|
| **$\Delta H_f$** | MNDO | 0.9924 | 0.9964 |
| **$\Delta H_f$** | AM1 | 0.9635 | 0.9943 |
| **$\Delta H_f$** | PM3 | 1.0053 | 0.9957 |
| **$\Delta H_f$** | present | 0.9899 | 0.9960 |
| **S** | MNDO | 1.0069 | 0.9811 |
| **S** | AM1 | 1.0056 | 0.9843 |
| **S** | PM3 | 0.9975 | 0.9815 |
| **S** | present | 0.9998 | 1.0000 |
| **$\Delta G_f$** | MNDO | 0.9974 | 0.9966 |
| **$\Delta G_f$** | AM1 | 0.9708 | 0.9954 |
| **$\Delta G_f$** | PM3 | 0.9744 | 0.9968 |
| **$\Delta G_f$** | present | 0.9922 | 0.9961 |