

## 개선된 QSPR 방법에 의한 알켄의 생성열

P. Duchowicz and E. A. Castro\*

CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata  
(2000, 7, 14 접수)

### Improved QSPR Prediction of Heats of Formation of Alkenes

P. Duchowicz and E. A. Castro\*

CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata,  
CC 962, La Plata 1900, Argentina  
(Received July 14, 2000)

**요 약.** 탄화수소의 생성엔탈피를 예측하는 이전의 선형방정식들을 일반화하였다. QSPR 분석에서 사용된 분자를 표현하는 기본적인 인자는 원자와 화학결합이다. 이러한 선택은 이 방법을 매우 간단하게 하며 비용을 줄일 수 있게 한다. 19개의 알켄에 대한 예측치는 실험오차 정도의 편차를 준다. 이 방법의 몇몇 가능한 확장에 대해 지적하였다.

**ABSTRACT.** Some previous linear equations to predict hydrocarbon heats of formation are generalized. The basic molecular descriptors used for the QSPR analysis are atoms and chemical bonds. This particular choice makes the method extremely simple and quite inexpensive. The predictions for a set of 19 alkenes yield deviations which are similar to experimental uncertainties. Some possible extensions of the method are pointed out.

### INTRODUCTION

The quite satisfactory results of applying Quantitative Structure Property Relationships (QSPR) to calculate heats of formation<sup>1-11</sup> seems to indicate this way is a suitable one to compute the enthalpy content of molecules. However, the usual numerical results do not represent true predicted quantities since the same set of molecules chosen to determine the fitting equations is employed to report the heats of formation. Since results are good enough and errors are nearly the same as experimental uncertainties, the QSPR equations show to be a suitable method to systematize data and to derive certain rules regarding the structural elements and group contributions to the molecular enthalpy of formation.

There are a wide variety of molecular descriptors to be used as independent variables in QSPR analysis and this large number of possibilities allows one to make quite different choices to perform the calculations and to interpret in a meaningful way the results. However, on spite

of that, the most natural and directly available descriptors are not employed currently. In fact, the simplest way to describe topologically a molecule is to take into account the atoms and the classical chemical bonds. This particular election has led us to obtain quite satisfactory results in some previous QSPR studies<sup>3-6</sup> for several physical chemistry properties. Regarding hydrocarbon enthalpies of formation from *ab initio* calculations improved through bond parameters we have gotten very good correlations and predictions for a large enough set of hydrocarbons via linear relationships.<sup>3</sup>

Notwithstanding these significant features, one might hope to improve the extent of this sort of approximation by being able to make higher order calculations in order to arrive to more faithful correlations and predictions. Thus, we have deemed very interesting to perform such a sort of calculation in order to test the predictive capabilities of QSPR theory when employing this sort of "natural" molecular descriptors.

Accordingly, the aim of this paper is to report the heats

of formation of a set of representative hydrocarbons and a test set of alkenes on the basis of new second and third order polynomial regression equations, testing the results with the available experimental data and first order theoretical predictions.

The paper is organized as follows: next section deals with the basic equations and then we present the results for the chosen set, comparing the values obtained with other arising from different theoretical standpoints and discussing the relatives merits of each approximation. Finally, we state the main conclusions derived from this study and some possible further extensions.

## METHOD

Five years ago, Herndon presented a simple enough protocol to convert Hartree-Fock *ab initio* total electronic energies for hydrocarbons to accurate heats of formation.<sup>12</sup> The optimum procedures use the number of carbon and hydrogen atoms ( $n_C$  and  $n_H$ , respectively) and 6-31 G\* nonrelativistic total electronic energies ( $E_{\text{elec}}$ ) as independent variables. Thus, the molecular enthalpy of formation at 298.15 K is given by

$$\Delta H_f^0(\text{g}) = A E_{\text{elec}} + B n_C + C n_H + D \quad (1)$$

where A, B, C and D are fitting parameters. The multilinear regression equations were calculated using the MATHEMATICA\* standard software.<sup>13</sup>

Later on, the numerical relationships were improved through the determination of higher-order equations<sup>6</sup> and better agreements between experimental and theoretical data were obtained.

However, on spite of its quite satisfactory degree of accuracy to predict molecular enthalpies of formation, the method was improved by means of the inclusion of the number and nature of chemical bonds. Consequently, it should take into account the existing differences among isomers, which under the previous approach they are considered equivalents.

According to Cioslowski,<sup>11</sup> a correction term must account for the electron correlation effects associated with bond formation and within this approximate scheme we have resorted to a more general equation than Eq. (1), i.e.

$$\Delta H_f^0(\text{g}) = A E_{\text{elec}} + B n_C + C n_H + \sum_{i,j}^{\text{bonds}} b_{ij} n_{ij} + D \quad (2)$$

where  $n_{ij}$  is the number of i-j chemical bonds and  $b_{ij}$  is an empirical parameter obtained by multilinear regression analysis. The final linear fitting equations were<sup>3</sup>

$$\begin{aligned} \Delta H_f^0(\text{g}) = & 593.858373 E_{\text{elec}} - 22498 n_C - 338.850811 n_H \\ & - 0.620987 n_{C-C} - 1.181333 n_{C=C} - 0.397152 n_{C=C} \\ & - 0.701003 n_{C-C(\text{aromatic})} \end{aligned} \quad (3)$$

$$\begin{aligned} \text{Standard Error} = & 1.4437, R^2 = 0.9987, \text{Average Deviation} \\ & = 1.08 \end{aligned}$$

$$\begin{aligned} \Delta H_f^0(\text{g}) = & 589.899403 E_{\text{elec}} + 22349.17 n_C - 336.410813 n_H \\ & - 0.626674 n_{C-C} - 1.330811 n_{C=C} - 0.768674 n_{C=C} \\ & - 0.787758 n_{C-C(\text{aromatic})} + 0.787197 \end{aligned} \quad (4)$$

$$\begin{aligned} \text{Standard Error} = & 1.4504, R^2 = 0.9987, \text{Average Deviation} \\ & = 1.08 \end{aligned}$$

It is interesting to point out that regression equations can be interpreted as being mainly concerned with total electronic energy and the number of constituent atoms, with correction terms related to the different chemical bonds (i.e. compare coefficients A, B, C with  $b_{ij}$  ones).

We have concentrated in a rather specialized molecular set, which is the current approach for this sort of analysis, although this option does not necessarily implies a lack of molecular variations within such restricted choice. In fact, Herndon's choice of 65 hydrocarbons<sup>12</sup> comprises examples of planar, non-planar, alternant and non-alternant aromatic hydrocarbons, alkyl- and alkenyl-substituted benzene derivatives, acyclic and polycyclic alkanes, strained and unstrained olefines and alkynes. The fitting equations were then applied to a set of 19 selected alkenes<sup>15</sup> and quite satisfactory results were obtained.

But in order to take a step further to attain even better concordances between experimental and theoretical values, we have resorted to the determination of second and third-order fitting equations within the realm of the present approximation scheme. Some previous studies have shown that one can arrive to more meaningful quantitative agreements when using higher-order polynomial equations.<sup>16</sup> Then, final equations to correlate and predict molecular enthalpies of formation have the general form

$$\begin{aligned} \Delta H_f^0(\text{g}) = & A E_{\text{elec}} + A' E_{\text{elec}}^2 + B n_C + B' n_C^2 + C n_H + C' n_H^2 \\ & + \sum_{i,j}^{\text{bond}} b_{ij} n_{ij} + b'_{ij} n_{ij}^2 + D \end{aligned} \quad (5)$$

$$\begin{aligned} \Delta H_f^0(g) = & A E_{\text{elect}} + A' E_{\text{elect}}^2 + A'' E_{\text{elect}}^3 + B n_C + B' n_C^2 \\ & - B'' n_C^3 - C n_H + C' n_H^2 + C'' n_H^3 + \sum_{\text{bond}}^{i,j} [b_{ij} n_{ij} \\ & - b'_{ij} n_{ij}^2 + b''_{ij} n_{ij}^3] - D \end{aligned} \quad (6)$$

where A, A', A'', B, B', B'', C, C', C'', b<sub>ij</sub>, b'<sub>ij</sub>, b''<sub>ij</sub> and D are empirical parameters determined with the aid of the set comprising 65 hydrocarbons (*i.e.* training set).

### RESULTS

Here we present some representative results for the sake of conciseness, and complete data are available upon request to one of us (JAC). The second and third order multilinear regression equations are

$$\begin{aligned} \Delta H_f^0(g) = & 586.0365 E_{\text{elect}} - 0.00031 E_{\text{elect}}^2 - 22201.9580 n_C \\ & - 0.4468 n_C^2 - 335.5647 n_H - 0.0745 n_H^2 - 0.5169 n_{\text{C(prim)}} \\ & - 0.0247 n_{\text{C(prim)}}^2 - 0.0417 n_{\text{C(sec)}} - 0.2907 n_{\text{C(sec)}}^2 \\ & - 0.6105 n_{\text{C(tert)}} + 1.0098 n_{\text{C(tert)}}^2 - 0.2511 n_{\text{C(aron)}} \\ & - 0.0108 n_{\text{C(aron)}}^2 - 3.2420 \end{aligned} \quad (7)$$

$$\begin{aligned} \Delta H_f^0(g) = & 581.1576 E_{\text{elect}} - 0.0142 E_{\text{elect}}^2 - 0 E_{\text{elect}}^3 \\ & + 22018.2954 n_C - 21.1270 n_C^2 + 0.8187 n_C^3 \\ & - 333.3537 n_H + 0.3796 n_H^2 + 0.012 n_H^3 - 0.1135 n_{\text{C(prim)}} \\ & - 0.0683 n_{\text{C(prim)}}^2 + 0.0082 n_{\text{C(prim)}}^3 + 0.1575 n_{\text{C(sec)}} \\ & - 0.5062 n_{\text{C(sec)}}^2 - 0.1618 n_{\text{C(sec)}}^3 - 0.3806 n_{\text{C(tert)}} \\ & - 0.8042 n_{\text{C(tert)}}^2 - 0.7049 n_{\text{C(tert)}}^3 - 0.7150 n_{\text{C(aron)}} \\ & - 0.2452 n_{\text{C(aron)}}^2 - 0.0119 n_{\text{C(aron)}}^3 - 5.6179 \end{aligned} \quad (8)$$

Experimental  $\Delta H_f^0(g)$  and the negative of calculated Hartree-Fock Self-Consistent Field (HF-SCF) electronic energies, optimized at the 6-31 G\* basis set level are listed in columns 3 and 2, respectively, in Table 1. The heats of formation modelled by first-, second-, and third-order polynomial equations are displayed in columns 4, 5, and 6, respectively. Predicted heats of formation of alkenes (test set) are given in Table 2.

We see that, in general, there is a quite satisfactory agreement between theoretical and experimental results. Average deviations are even minor than experimental uncertainties (2-3 kcal/mol) and they diminish when the polynomial order increases. Furthermore, the true predictions (*i.e.* alkenes set) are very encouraging and practically there are not pathological cases. In this sense, present results makes up an improvement regarding former values derived on the basis of linear equations.

The largest deviation in the training set appears for the neopentane molecule (-4 kcal/mol). This fact is hardly surprising since the existence of two branches in neopentane with respect to the regular alkanes makes the variation of physical properties does not follow the normal pattern. In fact, it is well known the effect of branching on physical chemistry properties within all families of organic compounds. With branching, the shape of the molecule tends to approach that of a sphere and as this happens the surface area decreases, with the result that the intermolecular forces become weaker and are overcome with a minor energy expenditure.<sup>16</sup> Regarding the test set, only four molecules present somewhat large deviations when using the linear equation (3,3-dimethyl-1-butene; 2,3-dimethyl-2-butene; *trans*-1,2-di-*t*-butylethylene and *trans*-4,4-dimethyl-2-pentene), but all predictions fall within the same range of experimental uncertainties when resorting to second and third order formulae. Perhaps a more suitable way to organize calculations should be to consider branched molecules as a separate set and it would yield better numerical results. However, it should duplicate the number of fitting parameters. Since our main aim has been to provide a better alternative way to predict heats of formation instead of making numerical gymnastic, we have chosen the present way to make the calculations.

It is important to point out that present approximation is extremely simple and direct to apply (just to count atoms and chemical bonds !!) and these results reveals once again the significative relevance of the primary building blocks of the molecular structure.

### CONCLUSIONS

We have verified the need to resort to higher-order polynomial equations when modeling molecular heats of formation. Certainly, whenever converting HF total molecular electronic energies to standard enthalpies of formation, it is convenient to employ higher-order multilinear regression equations in order to achieve chemical accuracy. In order to judge the relative merits of the present approximation scheme, we can compare the average absolute deviation obtained from a rather restricted molecular set (just four molecules) presented in Ref.<sup>16</sup> (1.81 kcal/mol) with present values (*i.e.* and for

Table 1. SCF-HF total electronic energy (atomic units) calculated at the 6-31G\* basis set level and molecular enthalpies of formation (kcal/mol) for the training set.

Molecule	-Energy	$\Delta H_f^\circ(\text{g})$ (exp.)	$\Delta H_f^\circ(\text{g})$ (linear equation)	$\Delta H_f^\circ(\text{g})$ (quadratic equation)	$\Delta H_f^\circ(\text{g})$ (cubic equation)
Methane	40.19517	-17.79	-15.51	-15.98	-17.28
Acetylene	76.81783	54.55	56.38	55.18	54.48
Ethylene	78.03172	12.56	12.58	12.14	12.03
Ethane	79.22876	-20.04	-20.04	-19.84	-20.36
Propyne	115.86432	44.41	44.25	44.40	44.71
Allene	115.86110	45.31	44.88	43.91	44.22
Propene	117.07147	4.79	4.41	4.74	5.07
Propane	118.26365	-24.93	-25.96	-25.29	-25.30
1,3-butadiene	154.91965	26.01	25.62	25.55	26.19
2-butyne	154.90926	34.69	33.02	34.07	34.46
1-butene	156.10608	0.07	-1.43	-0.57	-0.18
(Z)-2-butene	156.10786	-1.77	-1.77	-1.15	-1.01
(E)-2-butene	156.11041	-2.84	-3.28	-2.64	-2.49
Isobutene	156.11067	-4.27	-3.43	-2.79	-2.64
Cyclobutane	156.09720	6.78	4.59	4.70	5.20
n-butane	157.29840	-30.33	-30.55	-30.34	-30.69
Isobutane	157.29897	-32.24	-30.88	-30.67	-31.02
Cyclopentadiene	192.79172	32.12	32.83	32.47	32.81
1,3-pentadiene	193.95916	18.29	17.60	17.89	18.08
1,4-pentadiene	193.94721	25.27	24.65	24.89	24.98
Cyclopentene	193.97719	8.44	7.05	7.50	7.77
Cyclopentane	195.16358	-18.26	-19.28	-19.13	-18.84
n-pentane	196.33302	-35.60	-35.68	-36.17	-36.40
Cyclohexane	234.20800	-29.49	-30.20	-30.55	-30.44
Cyclopropene	115.82305	66.22	67.40	66.19	66.47
Cyclopropane	117.05887	12.73	11.92	11.62	12.43
Cyclobutane	154.89961	37.45	37.52	37.37	37.89
Neopentane	196.33383	-40.14	-36.16	-36.64	-36.87
Cubane	307.39391	148.69	146.41	147.25	147.54
Bicyclo(1.1.0)butane	154.87177	51.90	54.03	53.38	54.09
Bicyclo(2.1.0)pentane	193.92697	37.70	36.75	36.72	37.05
Bicyclo(2.2.0)hexane	232.96556	29.90	29.27	29.33	29.06
Bicyclo(2.2.1)heptane	272.06116	-12.40	-11.84	-11.91	-12.00
Bicyclo(2.2.2)octane	311.10358	-23.67	-21.58	-22.47	-22.21
Spiropentane	193.91753	44.25	42.32	42.25	42.50
Bicyclo(2.1.0)pentene	192.71022	79.70	80.98	80.50	80.02
Bicyclo(2.2.0)hexene	231.76849	62.50	61.90	62.10	61.10
n-hexane	235.36779	-39.94	-40.91	-42.55	-42.18
Cycloheptatriene	269.68233	43.56	44.30	43.59	43.56
Norbomadiene	269.65251	59.18	61.97	62.02	60.16
Quadricyclane	269.61822	81.04	82.35	82.87	82.83
Cyclooctatetraene	307.52422	70.30	70.76	70.85	70.65
Benzene	230.70310	19.80	17.91	17.90	18.74
Naphthalene	383.35500	36.00	34.20	34.54	35.22
Anthracene	535.99880	55.20	55.27	55.46	55.45
Phenanthrene	536.00980	49.70	48.78	49.02	49.09

Table 1. (continued)

Molecule	-Energy	$\Delta H_f^\circ(g)$ (exp.)	$\Delta H_f^\circ(g)$ (linear equation)	$\Delta H_f^\circ(g)$ (quadratic equation)	$\Delta H_f^\circ(g)$ (cubic equation)
Bicyclo(2.1.0)pentane	193.92697	37.70	36.75	36.72	37.05
Bez(a)anthracene	688.65688	68.10	67.91	67.56	67.37
Chrysene	688.66090	66.00	65.54	65.21	65.03
Triphenylene	688.66030	66.50	65.89	65.56	65.38
Benz(c)phenanthrene	688.64950	69.60	72.27	71.88	71.66
Pyrene	611.76800	54.00	55.03	55.11	55.13
Perylene	764.40650	78.40	79.23	78.63	78.89
Acenaphthene	460.26060	37.23	36.43	37.04	37.07
Bhiphenylene	459.01460	99.80	98.62	98.42	98.13
Acenaphthylene	459.07380	62.20	63.70	63.75	63.96
Azulene	383.28260	73.53	76.91	76.95	76.96
Fluoranthene	611.74562	69.20	68.24	68.21	68.10
Cis-stilbene	537.13326	60.30	59.44	59.81	60.45
Trans-stilbene	537.13943	56.40	56.00	56.20	56.88
Biphenyl	460.25394	43.30	41.30	41.51	41.78
o-xylene	308.77622	4.56	5.35	5.70	5.09
m-xylene	308.77724	4.14	4.75	5.10	4.50
p-xylene	308.77704	4.31	4.87	5.21	4.62
Styrene	307.58540	35.40	34.29	34.99	33.35
Toluene	269.74016	11.95	11.34	11.74	12.06
Average absolute error*	-	-	1.08	0.97	0.87

\*Average absolute error =  $\sum_{i=1}^n |\Delta H_f^\circ(g)(exp.)_i - \Delta H_f^\circ(g)(theor.)_i|$ ; n: n-number of molecules

Table 2. Alkenes 6-31G\* total electronic energies (atomic units) and enthalpies of formation (kcal/mol) for the test set.

Molecule	-Energy	$\Delta H_f^\circ(f)$ (exp.)	$\Delta H_f^\circ(f)$ (linear)	$\Delta H_f^\circ(f)$ (quadratic)	$\Delta H_f^\circ(f)$ (cubic)
Cis-2-pentene	195.14229	-6.60	-6.80	-7.17	-7.00
Trans-2-pentene	195.14504	-7.62	-8.42	-8.78	-8.60
2-methyl-2-butene	195.14577	-9.99	-8.85	-9.21	-9.48
2-methyl-1-butene	234.18020	-15.99	-13.87	-15.23	-15.67
2-methyl-1-butene	195.14370	-8.44	-7.65	-8.02	-8.30
3,3-dimethyl-1-butene	234.17395	-14.46	-10.19	-11.57	-12.79
3-methyl-1-butene	195.14116	-6.60	-6.13	-6.50	-6.59
2,3-dimethyl-2-butene	234.17701	-16.30	-11.99	-13.36	-13.90
2,3-dimethyl-1-butene	234.17700	-14.96	-11.99	-13.35	-13.64
1-methylcyclopentene	233.01793	-0.91	-1.70	-2.40	-2.25
3-methylcyclopentene	233.01300	1.77	1.21	0.48	1.16
Cycloheptene	272.04603	-2.20	-2.99	-4.63	-3.50
1-methylcyclohexene	272.05863	-10.35	-10.42	-12.08	-11.57
Norbomene	270.86184	21.51	22.12	21.11	21.30
Cis-4,4-dimethyl-2-pentene	273.20350	-17.35	-12.33	-15.26	-16.27
Trans-4,4-dimethyl-2-pentene	273.21254	-21.22	-17.66	-20.55	-20.89
2-bicyclo(2.2.2)octene	309.91244	4.90	7.55	5.80	5.00
Cyclohexene	233.01965	-1.20	-2.71	-3.41	-2.78
Trans-1,2-di- <i>t</i> -butylethylene	390.31453	-39.56	-31.97	-42.49	-41.56
Average absolute error	-	-	2.25	1.44	0.92

second order and third order equations, respectively). Besides, the extreme simplicity of the method makes it very easy to apply, because it does not demand any sort of extra computational effort, and calculations are elementary and quite inexpensive.

Since this approximation has been employed just for one set of organic molecules, before stating more definitive conclusions about its relative qualifications, it should be necessary to extend the procedure to different molecular sets (*i.e.* organic molecules including heteroatoms, inorganic compounds, radicals, ions, etc.). At present, work along this line is being carried out in our laboratory and results will be published elsewhere in the forthcoming future.

A final remark is in order to point out the existence of other important approximation schemes to predict heats of formation. Application of the group additivity method within the readily applicable version suggested by Benson *et al.*<sup>16-19</sup> has provided a powerful means of studying chemical on the basis of thermochemistry, alleviated planning of experimental studies, evaluation of experimental results, estimation of the equilibrium and rate constants needed by simulation studies in olefin chemistry, oxidation, air pollution, etc. The transferable groups values relating to molecules are based on well-established experimental data on many compounds. At present we are analysing the application of similar equations as those employed in this work in order to extend this sort of approximation. Results will be given at due time elsewhere.

**Acknowledgments.** The authors thank the valuable comments of two referees which helped to improve the

presentation of this paper.

## REFERENCES

1. Castro, E. A. *Rev. COPAQUA* **1999**, *19(1,2)*, 32.
2. Vericat, C.; Castro, E. A. *Egypt. J. Chem.* **1998**, *41(1-6)*, 109.
3. Duchowicz, P.; Castro, E. A. *J. Korean Chem. Soc.* **1999**, *43(6)*, 621.
4. Gavernet, L.; Firpo, M.; Castro, E. A. *Rev. Roum. Chim.* **1998**, *43(11)*, 1079.
5. Castro, E. A.; Gavernet, L.; Firpo, M. *Acta Chim. Slov.* **1997**, *44(3)*, 327.
6. Castro, E. A. *Comp. & Chem.* **1887**, *21(5)*, 305.
7. Castro, E. A. *Ing. Cienc. Quím.* **1996**, *16(2)*, 75.
8. Vericat, C.; Castro, E. A. *Commun. Math. Comp. Chem. MATCH* **1996**, *34*, 167.
9. Castro, E. A. *J. Chem. Soc. Pak.* **1995**, *17(3)*, 156.
10. Castro, E. A. *J. Molec. Struct. THEOCHEM* **1995**, *339*, 239.
11. Castro, E. A. *J. Molec. Struct. THEOCHEM* **1994**, *304*, 93.
12. Hemdon, W. C. *Chem. Phys. Lett.* **1995**, *234*, 82.
13. Cropper, W. H. *Mathematica® Computer Programs for Physical Chemistry*; Springer-Verlag: New York, 1998.
14. Cioslowski, J.; Liu, G.; Piskorz, P. *J. Phys. Chem. A* **1998**, *102*, 9890.
15. Schmitz, L. R.; Chen, Y. R. *J. Comp. Chem.* **1994**, *15(12)*, 1437.
16. Morrison, R. T.; Boyd, R. N. *Organic Chemistry*, 3<sup>rd</sup> ed.; Allyn and Bacon, Inc.: Boston, 1975; pp. 85-87.
17. Benson, S. W.; Buss, J. H. *J. Chem. Kinet.* **1958**, *29*, 546.
18. O'Neal, H.; Benson, S. W. *Int. J. Chem. Kinet.* **1969**, *1*, 221.
19. Benson, S. W. *Thermochemical Kinetics* 2<sup>nd</sup> ed.; John Wiley & Sons: New York, 1976.