J. Org. Chem. **1998,** *63,* 7445−7448

Molecular Structure Parameters and Predictions of Enthalpies of Formation for Catacondensed and Pericondensed Polycyclic Aromatic Hydrocarbons

William C. Herndon

Department of Chemistry, University of Texas at El Paso, El Paso, Texas 79968

P. Ulrich Biedermann and Israel Agranat*

Department of Organic Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received June 30, 1998

Simple procedures that combine calculated ab initio theoretical energies with empirical structural parameters to correlate experimental enthalpies of formation for polycyclic aromatic hydrocarbons are evaluated for predictive potential. The analyzed data set consists of every benzenoid PAH with an experimentally determined $\Delta H_f°(g)$, i.e., nine catacondensed and three pericondensed aromatic compounds. The tested levels of theory use optimized STO-3G, 3-21G, and 6-31G* calculated HF electronic energies, and energies determined at the correlated, optimized DFT B3LYP/6-31G* and single point MP2/6-31G*//HF/6-31G* levels. The highest precision correlations of the $\Delta H_f°(g)$ data combine computed electronic energies with three types of parametrized carbon structure descriptors and a CH parameter. The predictive accuracy of this protocol is assessed using a statistical cross-validation procedure.

Introduction

Group additivity procedures for estimating the enthalpic properties of organic compounds have been reviewed recently by Cohen and Benson.[1] The average error in calculated $\Delta H_f°(g)$'s for benzenoid aromatics (with and without aliphatic and olefinic substituents) is of the order of 2.5 kcal/mol with several much larger errors for individual outliers. In extending these procedures, several studies of group additivity methods limited to catacondensed and pericondensed unsubstituted benzenoid polycyclic aromatic hydrocarbons (PAHs) have demonstrated that empirical or theoretical estimates of the resonance energy of each compound must be included to yield higher-quality regressions of $\Delta H_f°(g)$ data.[2−6] There also exist many examples of the modeling of $\Delta H_f°(g)$ values for various classes of organic compounds using calculated ab initio or semiempirical energies combined with element type or atom-hybridization incremental additivity parameters.[4,6−23] The rationale and practical justifications for such procedures have previously been discussed.[7,24]

The present paper summarizes further studies of these types of procedures for correlating the $\Delta H_f°(g)$'s of PAHs. We use calculated ab initio electronic energies to model the nonadditive electronic effects, combined with additive molecular structure parameters for four different types of aromatic carbon atoms (dependent upon nearest neighbors). The approach is similar to that used in the PAH studies of Schulman, Peck, and Disch which use a single carbon parameter[10,11] and/or isodesmic reactions.[12,13] However, the objective of the present work is not only to suggest new models or procedures but to evaluate the true capabilities of these types of methods for predicting PAH $\Delta H_f°(g)$'s using different accessible levels of ab initio theory.

The results will be presented as follows. First, experimental $\Delta H_f°(g)$ data and correlations of the data with the theoretical energy terms and the structural parameters will be summarized. Then a general type of cross-validation procedure will be outlined and applied to assess the predictive accuracy of the model equations. Finally, the detailed results of cross-validations will be tabulated, and actual predictions of $\Delta H_f°(g)$'s of PAHs will be discussed.

Correlations of PAH $\Delta H_f°(g)$ Data

Experimental enthalpies of formation and calculated ab initio electronic energies are given in Table 1 for every benzenoid PAH with a reported experimental value of $\Delta H_f°(g)$. The following levels were used in the ab initio calculations:[25] optimized HF/STO-3G, HF/3-21G, HF/6-31G*, and DFT B3LYP/6-31G* levels and single point MP2/6-31G*//HF/6-31G* calculations. Polyhex drawings of the compounds under discussion are given in Figure 1.

---

(1) Cohen, N.; Benson, S. W. *Chem. Rev.* **1993**, *93*, 2419.
(2) Herndon, W. C. *J. Am. Chem. Soc.* **1973**, *95*, 2404.
(3) Herndon, W. C. *Thermochim. Acta* **1974**, *8*, 225.
(4) Herndon, W. C.; Nowak, P. C.; Connor, D. A.; Lin, P. *J. Am. Chem. Soc.* **1992**, *114*, 41.
(5) Herndon, W. C.; Nowak, P. C. *Adv. Theor. Interesting Mol.* **1992**, *2*, 113.
(6) Herndon, W. C.; Knott, S. L. *Polycyclic Aromat. Compd.* **1996**, *11*, 229.
(7) Wiberg, K. B. *J. Comput. Chem.* **1984**, *5*, 197.
(8) Wiberg, K. B. *J. Org. Chem.* **1985**, *50*, 5285.
(9) Ibrahim, M. R.; Schleyer, P. v. R. *J. Comput. Chem.* **1985**, *6*, 157.
(10) Schulman, J. M.; Peck, R. C.; Disch, R. L. *J. Am. Chem. Soc.* **1989**, *111*, 5675.
(11) Peck, R. C.; Schulman, J. M.; Disch, R. L. *J. Am. Chem. Soc.* **1990**, *112*, 6637.
(12) Disch, R. L.; Schulman, J. M.; Peck, R. C. *J. Phys. Chem.* **1992**, *96*, 3998.
(13) Schulman, J. M.; Disch, R. L. *J. Phys. Chem. A* **1997**, *101*, 9176.

S0022-3263(98)01280-8 CCC: $15.00

Table 1. Experimental $\Delta H_f°$(g) (kcal/mol) and ab Initio Electronic Energies $E_{\text{ab initio}}$ (Negative Atomic Units) for PAHs

compound	$\Delta H_f°$(g)	HF/STO-3G	HF/3-21G	HF/6-31G*	B3LYP/6-31G*	MP2/6-31G* [a]
benzene	20.00	227.8913602	229.4194454	230.7031370	232.2486574	231.4565052
naphthalene	36.00	378.6868486	381.2158091	383.3550455	385.8927269	384.6118842
phenanthrene	49.70	529.4874508	533.0157232	536.0097644	539.5386549	537.7720098
anthracene	55.20	529.4724848	533.0034726	535.9987691	539.5305197	537.7609308
pyrene	54.00	604.3253247	608.3457870	611.7679568	615.7731340	613.7827553
chrysene	66.00	680.2838027	684.8117924	688.6609246	693.1820034	690.9291395
triphenylene	66.50	680.2860601	684.8122687	688.6602533	693.1810750	690.9307944
benz[*a*]anthracene	68.10	680.2772156	684.8069457	688.6568818	693.1789632	690.9242889
benzo[*c*]phenanthrene	69.60	680.2702233	684.8005824	688.6495210	693.1724881	690.9227530
naphthacene	72.30[b]	680.2534797	684.7870715	688.6385477	693.1658074	690.9070166
perylene	78.40	755.1125211	760.1298830	764.4064824	769.4061174	766.9297428
coronene	68.93	904.8227039	910.8210024	915.9525511	921.8978931	918.9797713

[a] The MP2 values are single point calculations, carried out at the HF 6-31G* optimized geometries. See text. [b] Questionable experimental value.[26,27]

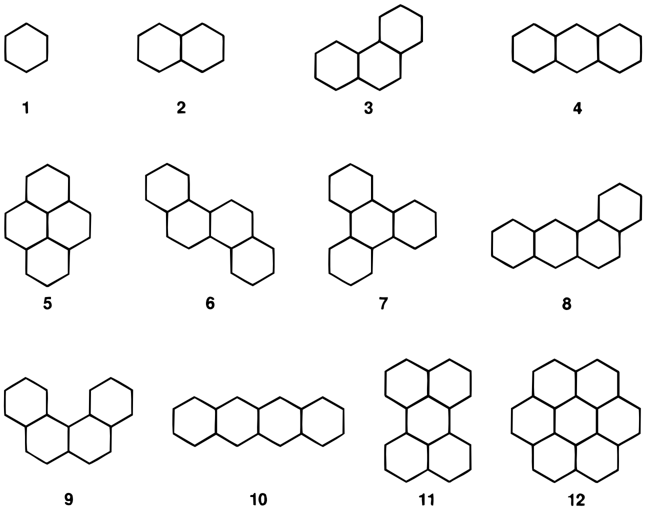

Figure 1. Polyhex drawings for PAHs with reported experimental values of ΔH_f (g).

The original references for values of $\Delta H_f°$(g) were checked using the standard compilations to obtain primary literature citations. The small differences from values recommended by Pedley, Naylor, and Kirby[26,27] for benzene, anthracene, phenanthrene, and perylene are based on recently obtained or on previously overlooked data.[28] The value of $\Delta H_f°$(g) for coronene listed in Table 1 is an important new addition to previous lists of experimental $\Delta H_f°$(g) data for PAHs. The recently determined precise heat of combustion data (Kiyobayashi and Sorai) give $\Delta H_f°(\text{solid}) = 36.45$ kcal/mol.[29,31] This is combined with the last published value[30] of several measured heats of sublimation,[26,27] which determines the $\Delta H_f°$(g) value for coronene given in Table 1.

The evaluation studies begin with group additivity analyses of the experimental $\Delta H_f°$(g). Four types of carbon atom substructures were selected for the PAHs based on examination of their molecular graphs. The CH group is one of these descriptors; its abundance, [CH], is equal to the number of hydrogen atoms. The other carbon types are fusion carbons, C(C,C,C), C(CH,C,C), and C(CH,CH,C), where the nearest neighbors are given in parentheses. This four-parameter set of descriptors comprises the set of core parameters that are used in the Benson additivity procedure for calculating enthalpic properties of PAHs.[1] Benson parameters give an average absolute error equal to 2.31 kcal for the $\Delta H_f°$(g) of the 12 PAHs listed in Table 1. Perylene and benzo[*c*]phenanthrene are serious outliers, with calculated errors of +11.58 and +5.39 kcal/mol, respectively.

The stabilizing nonadditive effects of delocalization or resonance, which is not included in the Benson analysis, are essential components of the quantitative descriptions of PAH systems.[3-5] In this paper, the nonadditive contributions to the molecular enthalpy are estimated using theoretical calculations, i.e., we evaluate (in turn) the results of correlations using ab initio HF, DFT, and MP2 methodologies. The values of the structural pa-

---

(14) Wang, H.; Frenklach, M. *J. Phys. Chem.* **1993**, *97*, 3867.

(15) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. *J. Phys. Org. Chem.* **1990**. *3*, 732.

(16) Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K.; Allinger, N. L. *J. Phys. Org. Chem.* **1992**, *5*, 225.

(17) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. *J. Am. Chem. Soc.* **1992**, *114*, 2880.

(18) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C. Labanowski, J. K. *J. Comput. Chem.* **1992**, *13*, 838.

(19) Allinger, N. L.; Yang, I.; Motoc, I.; Bender, C.; Labanowski, J. K. *Heteroat. Chem.* **1992**, *3*, 69.

(20) Liu, R.; Allinger, N. L. *J. Phys. Org. Chem.* **1993**, *6*, 551.

(21) Yala, Z. *J. Mol. Struct. (Theochem)* **1990**, *207*, 217.

(22) Castro, E. A. *J. Mol. Struct. (Theochem)* **1994**, *304*, 93.

(23) Herndon, W. C. *Chem. Phys. Lett.* **1995**, *234*, 82.

(24) Dewar, M. J. S.; Storch, D. M. *J. Am. Chem. Soc.* **1985**, *107*, 3898.

(25) Gaussian 94, Revision E.2, Frisch, M. J., Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Chesseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanyakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A.; Gaussian, Inc., Pittsburgh, PA, 1995.

(26) Pedley, J. B.; Naylor, R. N.; Kirby, S. P. *Thermochemical Data of Organic Compounds*, 2nd ed.; Chapman and Hall; London, 1986.

(27) Pedley, J. B. *Thermochemical Data and Structures of Organic Compounds*, Vol. 1; TRC: Texas A&M Research Center, 1994.

(28) Exceptions or differences from recommended values[26,27] are based on various sources as follows (all values in kcal/mol): $\Delta H_f°$(g, benzene) = 20.0 uses $\Delta H°$(vaporization) = 8.30 (J. Chickos, private communication and independently confirmed in our laboratory (WCH)); $\Delta H_f°$(g, anthracene) = 55.2 uses the average of three recommended values; $\Delta H_f°$(g, phenanthrene) = 49.7 uses an average of recommended $\Delta H°$(sublimation) = 23.95 and $\Delta H°$(sublimation) = 24.15 (Malaspina, L.; Bardi, G.; Gigli, R. *J. Chem. Thermodyn.* **1974**, *6*, 1053); $\Delta H_f°$(g, perylene) = 78.4 uses recommended $\Delta H_f°$(solid) = 43.7 and $\Delta H°$(sublimation) = 34.7 (Gigli, R.; Malaspina, L.; Bardi, G. *Ann. Chim. (Rome)* **1973**, *63*, 627); $\Delta H_f°$(g, coronene) = 68.93 uses $\Delta H_f°$(solid) = 36.45[29] and $\Delta H°$(sublimation) = 32.49.[30] The experimental $\Delta H_f°$(g) value quoted for naphthacene is questionable.[5,6,10,11,23] The average of the literature values (adduced or estimated) of the experimental errors for the remaining compounds is ±0.4 kcal/mol. Considering the ranges of values quoted for individual heats of sublimation,[26,27] a more relastic general error limit for $\Delta H_f°$(g) for the benzenoid PAHs may be ±0.8 kcal/mol.

(29) Sorai, M.; Kiyobayahsi, T. *Microbomb Combustion Calorimetry of Non-planar Aromatic Hydrocarbon* (Doctoral thesis of T. Kiyobayashi (1995); Microcalorimetry Research Center, Faculty of Science, Osaka University, Japan, Chapter 5).

(30) Murray, J. J.; Pottie, R. F. *Can. J. Chem.* **1974**, *52*, 557.

Table 2. Parameters[a] and Statistical Results for Correlations of PAH $\Delta H_f°$(g) Data

	HF/STO-3G	HF/3-21G	HF/6-31G*	B3LYP/6-31G*	MP2/6-31G* [b]
c_1	23837.29658	23997.12894	24131.37110	24293.03160	24210.18927
c_2	23484.61511	23638.00643	23771.97101	23920.85410	23851.61324
c_3	23484.09857	23637.81759	23771.79644	23921.39564	23851.66447
c_4	23482.96342	23637.29816	23771.54758	23921.26513	23850.92375
			error statistics[c] for heat of formation residuals (kcal/mol)		
max	3.61	2.58	2.92	1.88	0.39
ave	0.91	0.65	0.64	0.63	0.20
rms	1.30	0.92	0.99	0.87	0.23
sd	1.36	0.96	1.04	0.91	0.24
			prediction error (kcal/mol) for naphthacene (not included in regression analyses)[d]		
	−7.49	−6.62	−6.18	−3.51	−4.77

[a] In atomic units, listed with the number of significant figures required to reproduce the tabulated statistical errors. [b] Single point MP2/6-31G* calculations at HF/6-31G* geometry. [c] Maximum, average absolute, root-mean-square, and standard deviation. [d] See discussion.

rameters are defined by the c_i coefficients in a multilinear regression equation as defined in eq 1. In agreement with previous work,[7−22] the dependent variable for the statistical analysis is taken to be a composite value, composed of the experimental heat of formation $\Delta H_f°$(g, expt) minus the ab initio energy term $E_{\text{ab initio}}$.

$$\Delta H_f°(\text{g, expt}) - E_{\text{ab initio}} = c_1[\text{CH}] + c_2[\text{C(C,C,C)}] + c_3[\text{C(CH,C,C)}] + c_4[\text{C(CH,CH,C)}] \quad (1)$$

The statistical results of the correlations of the $\Delta H_f°$(g, expt) and calculated electronic energy data with the four structural parameters are tabulated in Table 2. The significant aspects of these results are as follows. First, in agreement with previous questions regarding the accuracy of the experimental $\Delta H_f°$(g) value for naphthacene,[5,6,10−13,23] the overall quality of these regression studies is considerably improved by its omission. The table gives the correction term which would be required to bring the naphthacene experimental $\Delta H_f°$(g) into line with the other experimental values using these procedures. This correction term is at least three times the standard deviation of the residuals from the multilinear equation for the remaining compounds in every analysis. Second, at the HF/3-21G, HF/6-31G*, and B3LYP/6-31G* level, there are no practical differences in the very good quality of the correlations. Finally, the highly precise correlation of the $\Delta H_f°$(g) data with the MP2 energies should be weighed with some skepticism since the calculated standard deviation is somewhat smaller than the literature value of the estimated average experimental errors.[26−28]

Predicting PAH Enthalpies of Formation

By any statistical criterion, each regression equation (Table 2) is a very good correlation of the experimental data. We now would like to pose a different and more important question. What is the expected accuracy of these procedures when they are employed for *predictions* of $\Delta H_f°$(g) values rather than just correlations? We would particularly like to know if there is a reliable (low) level of ab initio theory that can be used for predictions in investigations of general thermochemical problems involving PAHs.

A well-known procedure for establishing predictive capabilities of model equations is known generally as cross-validation or the leave-*n*-out method.[32] In the present work we use a leave-one-out protocol, which is practical for small data sets, and which we infer will be useful as additional experimental data are made available. The leave-one-out procedure requires N trials for N cases, each case left out in turn. A model rectification of data for each set of $N - 1$ cases is calculated, and a value of the dependent variable for each missing case is computed from the relevant least-squares regression equation. Thus a complete and authentic set of predicted values of the dependent variable is obtained. The calculated and experimental values are then compared to assess the predictive (rather than correlative) power of the model equation. The results of this type of cross-validated investigation are summarized in Table 3. Naphthacene is again excluded from the analysis for the reason given previously.

As expected, the sets of actual predicted $\Delta H_f°$(g) differ from experimental values to a greater extent than in the correlation studies. Average deviations for each analysis are about twice those observed in the correlative calculations. The predictive mean absolute accuracy through the B3LYP level is on the order of 1.0 kcal/mol, except for the STO-3G calculations (1.56 kcal/mol). It should be noted that errors of 1 to 2 kcal/mol are generally considered acceptable in many previous studies, but this size of discrepancy would cause difficulties in accurate treatments of, for example, chemical equilibria problems.

One observes that the level of accuracy is dramatically improved upon proceeding to the MP2/6-31G*//HF/6-31G* level. Approximately 50% of the total of prediction errors at the four other ab initio levels is due to the same two molecules, coronene and benzo[*c*]phenanthrene. However, note the change in sign for the calculated error for the coronene (B3LYP) case and the large positive prediction error found for perylene. The rather similar magnitudes of the errors across the range of methods would normally indicate the possibility of errors in experiments, systematic mistakes in parametrization, or inadequacies in the quantum theoretical calculations. However, the cross-validated results obtained for the MP2 model equation suggest the possibility that the experimental $\Delta H_f°$(g) values are correct within small limits of experimental error and that the four additive structure-type parameters constitute a good descriptor set for the PAHs.

The larger errors in predicted $\Delta H_f°$(g) using the first four levels of ab initio calculations indicate that caution

---

(31) For a recent discussion of the experimental $\Delta H_f°$ and a comparison with ab initio derived values, see ref 13.

(32) Franke, R. *Theroretical Drug Design Methods*; Elsevier; Amsterdam, 1984; pp 227, 297.

Table 3. Errors in Predicted PAH $\Delta H_f°$(g)'s (kcal/mol)

compound	HF/STO-3G	HF/3-21G	HF/6-31G*	B3LYP/6-31G*	MP2/6-31G* [a]
benzene	0.72	0.35	0.62	0.19	0.09
naphthalene	1.74	1.23	1.17	0.78	0.35
phenanthrene	1.18	0.71	0.61	−0.32	−0.34
anthracene	−0.92	−0.81	−0.56	0.58	−0.43
pyrene	−0.17	0.35	0.47	0.53	−0.34
chrysene	0.91	0.68	0.74	−0.22	−0.20
triphenylene	0.74	0.59	0.32	−1.11	−0.20
benz[*a*]anthracene	1.55	1.03	1.04	0.47	0.64
benzo[*c*]phenanthrene	−4.10	−2.94	−3.32	−2.13	0.27
perylene	−0.56	−0.38	0.12	2.65	0.18
coronene	4.61	2.51	1.94	−2.71	0.02
error statistics for predicted heats of formation					
max dev	4.61	2.94	3.32	2.71	0.64
mean dev	1.56	1.05	0.99	1.06	0.28
rms dev	2.09	1.34	1.32	1.41	0.32
standard dev	2.12	1.37	1.35	1.47	0.34

[a] Single point at HF/6-3G* geometry (MP2/6-31G*//HF/6-31G*).

should be exercised when these types of parametrized methods are used to predict thermochemical properties for PAHs. On the other hand, the results also indicate that MP2/6-31G*//HF6-31G* calculations, coupled with the simple carbon-substructure parameters, may constitute a very accurate approach to modeling the PAH system of compounds. We tentatively conclude that the parametrized MP2 model chemistry can be used to *predict* values of $\Delta H_f°$(g) for benzenoid PAHs, close to the order of the accuracy of experimental thermochemical measurements. As an aside, the general results also illustrate that standard and mean deviations of a correlation equation (without cross-validation) are not sufficient measures of the capability to yield accurate predictions. The magnitudes of the largest errors in a cross-validation procedure will provide a better criterion for this purpose.

The general results are encouraging. The MP2 calculations do require a rather high degree of computer power, substantial memory, and considerable CPU time. However, it seems possible that additional parametrization, which might include separate terms for the internal C(CCC) term in planar and nonplanar systems, could reduce errors in the predicted values of $\Delta H_f°$(g) from the more economical HF or the DFT calculations to even more useful levels. Investigations along these lines, a more extensive testing of basis sets, the effects of thermochemical and statistical mechanical corrections to the calculated electronic energies, and extensions to other types of aromatic systems, are in progress.

Acknowledgment. The financial support of the Welch Foundation of Houston, TX, is gratefully acknowledged. The University of Texas at El Paso Center for Environmental Resource Management (E.P.A. Superfund Research Program) provided funds for student support and purchase of a computer in the early stages of closely related research.

JO981280S