9196 J. Phys. Chem. A **2001,** *105,* 9196−9202

Heats of Formation from DFT Calculations: An Examination of Several Parameterizations

J. Peter Guthrie*

Department of Chemistry, University of Western Ontario, London, Ontario, Canada N6A 5B7

Received: January 26, 2001; In Final Form: August 2, 2001

DFT calculations of the total energy of molecules (B3LYP/6-31G**//rhf/6-31G**) can be converted to enthalpies of formation by various new parametrization schemes. Three schemes follow the hierarchy proposed by Benson and Buss, of atom, bond, and group contributions. In addition, a modified atomic scheme was found to give better results than the atomic scheme with fewer parameters than the bond scheme. The lower level parametrizations could fit more compounds but had lower precision. To compare the schemes a standard set of 180 compounds was treated by all methods. The results are: Method, number of compounds, number of parameters, standard deviation (kcal/mol): Atom, full: 250, 18, 2.96; std: 180, 12, 2.64; Modified atom, full: 248, 23, 2.53; std: 180, 18, 1.97; Bond, full: 229, 40, 1.90; std: 180, 32, 1.72; Group, full: 183, 79, 1.39; std: 180, 79, 1.26.

Introduction

The calculation of heats of formation, with thermochemical accuracy, i.e., errors of ca. 1 kcal/mol, by ab initio methods is a long standing goal of quantum chemistry. Pople's G1,[1] G2,[2] and G3[3] theories provide a well-established route to this level of accuracy at the expense of very computationally demanding calculations. There has been an ongoing interest in looking for semiempirical methods for achieving useful accuracy with less demanding calculations. In this connection, a method proposed by Schleyer and Ibrahim,[4] following on the path blazed by Wiberg,[5] Baird,[6] and Dewar,[7] for converting ab initio energies into heats of formation was very attractive. Trial calculations quickly showed that B3LYP/6-31G** [8,9] was much better than rhf/6-31G**, particularly for conjugated compounds (as Schleyer and Ibrahim had observed, rhf calculations were not good for these compounds[4]) and for compounds with anomeric effects. B3LYP/6-31G**//RHF/6-31G** was chosen as the calculational method. Four parametrization schemes will be presented. As one might expect, the more detailed the parametrization the better the fit, but what is noteworthy is how well the calculations can do with rather simple parametrizations.

Benson type additivity schemes are very convenient, and work well for simple compounds, provided that appropriate parameters are available, and provided that it is possible to ignore or deal adequately with (i.e., have appropriate additional parameters for) strain and non next neighbor interactions. This means that the Benson approach often has difficulties with polyfunctional, cyclic or polycyclic, or crowded molecules. The approach described here uses DFT calculations to deal with these additional effects and a set of parameters analogous to the basic Benson parameters to convert the DFT energy into a heat of formation at 298 K. This paper deals with enthalpy of formation; in separate work a procedure for calculating the standard entropy of gaseous molecules has been reported,[10] It will be demonstrated that the group contribution scheme is equivalent to a homodesmotic approach[11] to converting DFT energies into heats of formation.

Results

The literature values for the heats of formation used are documented in Table S1, which also includes the calculated total energies for the optimized molecules. To convert the MO theory total energies to enthalpies of formation one must account for the total energies of the component atoms as elements in their standard states, for the zero point energies, for conversion from 0 K to 298 K, and for the presence of conformers,[12,4] All of these can be absorbed into the contributions needed to convert MO energies into enthalpies of formation, leading to a very simple calculation

$$\Delta H_f^\circ = E_{DFT} - \Sigma n_i x_i$$

where x_i = contribution for a structural feature, and n_i = number of times this structural feature appears in the molecule.

Four schemes were considered for conversion of these total energies to enthalpies of formation. The first three followed the hierarchy of contribution schemes recommended by Benson and Buss,[13] i.e., atom, bond, and group contributions. The fourth scheme, which admittedly is a bit of a kludge, is a modified atomic scheme, which gets improved performance for a small increase in number of parameters. Contributions were defined following Benson and Buss[13] with a few exceptions which will be described.

The first scheme used atom equivalents, with different parameters for each coordination number. Thus, C_4, C_3, and C_2, representing tetracoordinate, tricoordinate, and dicoordinate carbon, were treated as different atomic contributions, and similar distinctions were made for other elements. No distinction was made between N_3 with pyramidal or planar geometry. An additional parameter was introduced for C_b, the contribution for a carbon in an aromatic ring. With 250 compounds and 18 parameters, the standard deviation was 2.96 kcal/mol. The parameter values found by least squares are given in Table 1.

The second scheme used bond contributions, with 40 parameters for which two or more instances could be found, and 229 compounds, and a standard deviation of 1.90 kcal/mol. The simplest version of a bond scheme would have a parameter for each possible linkage of two atom types, so that 18 atom

---

* To whom correspondence should be addressed. Fax: 519 661 3022. Phone: 519 661 2111 x86352. E-mail: peter.guthrie@uwo.ca.

10.1021/jp010355k CCC: $20.00

TABLE 1: Atomic Contribution Parameter Values[a]

	full data set		standard data set	
atom	value	error	value	error
[C_4]	−38.120502	0.000603	−38.120 483	0.000586
[C_3]	−38.120693	0.000329	−38.120937	0.000325
[C_2]	−38.118679	0.000763	−38.118282	0.000848
[H]	−0.593561	0.000269	−0.593510	0.000263
[C_b]	−38.120934	0.000218	−38.121098	0.000212
[O_2]	−75.155579	0.000364	−75.156334	0.000368
[O_1]	−75.157204	0.000649	−75.156761	0.000714
[N_3]	−54.768272	0.000912	−54.769489	0.000970
[N_1]	−54.762520	0.001594	−54.763309	0.001563
[Cl]	−460.174377	0.000718		
[F]	−99.749397	0.000481		
[S_6]	−398.139771	0.001879		
[S_4]	−398.173981	0.002279	−398.144348	0.002478
[S_3]	−398.202271	0.000920	−398.176819	0.002614
[S_2]	−341.312408	0.003664	−398.202789	0.001130
[P_5]	−341.323761	0.002546		
[P_4]	−341.358215	0.002822		
[P_3]	−341.358215	0.002828		

[a] At 25 °C, in atomic units; the error is the square root of the variance of the parameter estimated by the least squares procedure.

TABLE 2: Bond Contribution Parameter Values[a]

	full data set		standard data set	
bond	value	error	value	error
[C_4-C_4]	−19.058563	0.000267	−19.058397	0.000264
[C_4-H]	−10.124343	0.000113	−10.124368	0.000113
[C_3-H]	−19.653498	0.000321	−19.653559	0.000330
[C_3-C_4]	−28.589790	0.000518	−28.589743	0.000523
[C_3-C_3]	−38.118217	0.001935	−38.117893	0.001864
[C_b-H]	−13.300263	0.000427	−13.300577	0.000429
[$C_b = C_b$]	−25.414240	0.000265	−25.414030	0.000262
[C_b-C_4]	−22.235491	0.000651	−22.235477	0.000626
[C_b-C_3]	−31.766869	0.001238		
[C_b-C_b]	−25.413666	0.001753	−25.414103	0.001688
[C_2-H]	−38.708191	0.001205	−38.708221	0.001105
[C_2-C_4]	−47.648468	0.001018	−47.648438	0.000945
[C_a-C_3]	−19.061277	0.000974	−19.061184	0.000917
[C_4-O_2]	−47.107574	0.000211	−47.107780	0.000215
[O_2-H]	−38.165695	0.000706	−38.165932	0.000672
[C_b-O_2]	−50.288097	0.000907	−50.287788	0.000848
[O_2-O_2]	−75.156677	0.000794	−75.156166	0.000746
[$CO-H$]	−57.231541	0.000933	−57.231522	0.001115
[$CO-O_2$]	−94.218903	0.000737	−94.218773	0.000684
[$CO-C_4$]	−66.166351	0.000550	−66.166245	0.000522
[$CO-C_b$]	−69.349648	0.001955	−69.348389	0.001850
[N_3-C_4]	−27.785290	0.000373	−27.785660	0.000388
[N_3-C_b]	−30.963179	0.002677	−30.961954	0.002493
[N_3-H]	−18.848175	0.000591	−18.848759	0.000583
[N_3-CO]	−74.893883	0.001320	−74.894485	0.001438
[$NC-C_4$]	−102.411911	0.000983	−102.412697	0.000969
[$NC-C_b$]	−105.592575	0.002173	−105.592499	0.002006
[NO_2-C_b]	−217.792191	0.002021	−217.791992	0.001860
[$F-C_4$]	−109.279305	0.000348		
[$Cl-C_4$]	−469.705017	0.000540		
[$Cl-CO$]	−516.814819	0.001342		
[S_2-C_4]	−208.631226	0.000511	−208.631271	0.000497
[S_2-H]	−199.692810	0.001143	−199.692535	0.001234
[S_2-C_b]	−211.805908	0.001927	−211.805542	0.002152
[S_2-C_3]	−218.151245	0.001584		
[S_2-CO]	−255.740158	0.001186		
[$SO-C_4$]	−246.195023	0.000931	−246.195068	0.000867
[SO_2-C_4]	−283.756226	0.000661	−283.757385	0.000716
[SO_2-C_3]	−293.286835	0.001507		
[SO_2-C_b]	−286.930939	0.001428		

[a] At 25 °C, in atomic units; the error is the square root of the variance of the parameter estimated by the least squares procedure.

types would lead to $18*17/2 + 18 = 171$ bond parameters. Not all, or even most, of these are exemplified in the data set, so the number is decreased but we have also increased the number of atom types treated as bond termini by adding CO, SO, and SO_2 as divalent "atoms", and CN and NO_2 as univalent "atoms". Trigonal carbon is defined[13] to include one-quarter of the contribution of the double bond, so there is no explicit contribution for the C═C in an isolated alkene. In a conjugated alkene, there is an explicit contribution for the bond joining the two alkenes, symbolized [C_3-C_3]. To handle allenes, a contribution for [C_a-C_3], the bond from an allenic (sp) to an olefinic (sp^2) carbon was introduced. An allene would have two [C_a-C_3] contributions; this approach allows for latter extension to ketenes or ketenimines when sufficient data become available. For aromatic carbons the symbol [$C_b = C_b$] symbolizes a bond joining two aromatic carbons within an aromatic ring, whether isolated or fused, whereas [C_b-C_b] symbolizes a bond joining two arenes, as in biphenyl. The parameter values found by least squares are given in Table 2.

The third scheme used group contributions, with 79 parameters for 183 compounds, and a standard deviation of 1.39 kcal/mol. Groups were defined in a way which generally followed Benson[13,14] with the exception that Hine's revised notation,[15] in which atoms accounted for in the group contribution are not in parentheses, whereas atoms whose identity affects the value of the group contribution, but will themselves be accounted for in a separate contribution, are enclosed in parentheses. The various forms of carbon are as follows: C − tetrahedral carbon; C_d − doubly bonded carbon; C_a − allenic carbon; C_t − triply bonded carbon; C_b − aromatic carbon (not involved in a fused ring junction); C_f − aromatic carbon at a fused ring junction. The univalent functions which were treated as "atoms" attached to a central atom and included in the group contribution are −CN, and −NO_2. Carbonyl, sulfinyl, and sulfonyl were treated as divalent structural units. The parameter values found by least squares are given in Table 3.

Finally, a modified atomic scheme, where the number of attached hydrogens, and the presence of attached oxygen (as in carbonyl, sulfinyl, sulfonyl, and phosphoryl) were treated as significant, as was being part of an aromatic ring. This scheme, with 23 parameters for 248 compounds, and a standard deviation of 2.53 kcal/mol was distinctly better than the simple atomic contributions scheme and needed many fewer parameters than the bond contributions scheme for not much worse performance. Variations with triply bonded carbon treated as a unique contribution were not significantly better. In the modified atomic contributions scheme, three changes were made from the atomic contributions scheme. First, attached hydrogens were considered part of the central atom, so that CH_3, CH_2, CH, and C were different "atoms". Second, attached unicoordinate oxygens were treated as part of the central atom so that CO, CHO, NO_2, SO, SO_2, and PO were treated as "atoms". Third, no distinction was made between sp^3, sp^2, or sp carbon, except that aromatic carbon was treated as distinct. Thus, $CH(X)_3$ could be sp^3 with three singly bonded substituents, sp^2 with one doubly and one singly bonded substituent, or sp with one triply bonded substituent. The goal in devising this scheme was to have as few parameters as possible while still giving satisfactory results. With a small increase in the number of parameters the scheme gives distinctly better results than the strict atomic scheme. The parameter values found by least squares are given in Table 4.

The least squares procedure used to evaluate the parameters[16] provides a covariance matrix, leading to estimated standard deviations for the parameters; these are reported as "errors" in Tables 1−4. These values should give an idea of how well the

TABLE 3: Group Contribution Parameter Values[a]

group	value	error	group	value	error
$[CH_3(X)]$	−39.903004	0.000280	$[O(O)_{2]}$	−75.158669	0.001594
$[CH_2(C)_2]$	−39.308296	0.000215	$[OH(C)]$	−75.744118	0.001441
$[CH(C)_3]$	−38.710243	0.000909	$[O(C)_2]$	−75.156776	0.001050
$[C(C)_4]$	−38.111214	0.001293	$[O(C_b)(C)]$	−75.158958	0.001991
$[C_dH_2]$	−39.307198	0.000588	$[OH(O)]$	−75.743034	0.000714
$[C_dH(C)]$	−38.713299	0.000704	$[O(O)(C)]$	−75.154793	0.000720
$[C_d(C)_2]$	−38.120335	0.001694	$[CH_2(C)(O)]$	−39.307659	0.000445
$[C_dH(C_d)]$	−38.713497	0.000907	$[CH(C)_2(O)]$	−38.712936	0.001033
$[C_d(C)(C_d)]$	−38.116943	0.001236	$[C(C)_3(O)]$	−38.114769	0.001213
$[CH_2(C)(C_d)]$	−39.308083	0.000957	$[CH_2(O)_2]$	−39.305977	0.001885
$[CH_2(C_d)2]$	−39.307835	0.001998	$[CH(C)(O)_2]$	−38.709557	0.002283
$[CH(C)_2(C_d)]$	−38.710258	0.001799	$[CH(O)_3]$	−38.712807	0.003093
$[C(C)_3(C_d)]$	−38.112732	0.002111	$[C(C)_2(O)_2]$	−38.115177	0.002286
$[Ca]$	−38.122021	0.001523	$[C(C)(O)_3]$	−38.115005	0.003059
$[C_tH]$	−38.707764	0.000931	$[C(O)_4]$	−38.112278	0.003674
$[C_t(C)]$	−38.117760	0.000971	$[C_b(O)]$	−38.122669	0.001667
$[CH_2(C_t)(C)]$	−39.309322	0.001864	$[CHO(X)]$	−113.868858	0.001755
$[C_bH]$	−38.714615	0.000193	$[CO(C)_2]$	−113.273308	0.001484
$[C_f(C_f)(C_b)_{2]}$	−38.120934	0.000817	$[CO(C)(O)]$	−113.274101	0.002788
$[C_f(C_f)_2(C_b)]$	−38.121075	0.000454	$[CO(N)(C)]$	−113.273392	0.002866
$[C_f(C_f)_3]$	−38.120926	0.001435	$[C(C)_3(CO)]$	−38.113083	0.002091
$[C_b(C)]$	−38.119995	0.000795	$[CH(C)_2(CO)]$	−38.711132	0.001517
$[C_b(C_b)]$	−38.120899	0.001082	$[CH_2(C)(CO)]$	−39.309364	0.002041
$[CH_2(C)(C_b)]$	−39.307568	0.001692	$[C_b(CO)]$	−38.124466	0.002991
$[CH_2(C_b)_2]$	−39.306625	0.003101	$[CO(N)(C_b)]$	−113.273499	0.004806
$[CH(C)_2(C_b)]$	−38.710773	0.001989	$[CH_2(CO)_2]$	−39.307201	0.002835
$[C(C)_3(C_b)]$	−38.113110	0.001879	$[CO(O)_2]$	−113.273933	0.005405
$[NH_2(X)]$	−55.953499	0.001391	$[O(C)(CO)]$	−75.159195	0.002542
$[NH(C)_2]$	−55.361214	0.002139	$[OH(CO)]$	−75.746208	0.002455
$[N(C)_3]$	−54.766075	0.002064	$[SH(C)]$	−398.793060	0.001417
$[NH_2(CO)]$	−55.953323	0.002599	$[S(C)_2]$	−398.201141	0.001793
$[N(CO)(C)_2]$	−54.770016	0.002599	$[CH_2(S)(C)]$	−39.308685	0.001097
$[CH_2(C)(N)]$	−39.307713	0.000743	$[CH(S)(C)_2]$	−38.713154	0.001096
$[CH(C)_2(N)]$	−38.709976	0.001332	$[C(S)(C)_3]$	−38.113113	0.002124
$[CH_2(C)CN]$	−132.191040	0.001399	$[C_b(S)]$	−38.119392	0.002340
$[CH(C)_2CN]$	−131.592819	0.001306	$[SO(C)_2]$	−473.329193	0.001866
$[C(C)_3CN]$	−130.994659	0.001297	$[CH_2(C)(SO)]$	−39.307411	0.000936
$[C_bCN]$	−131.006134	0.001891	$[SO_2(C)_2]$	−548.457275	0.001787
$[C_bNO_{2]}$	−243.205994	0.001305	$[C(C)_3(SO_2)]$	−38.110420	0.001526

[a] At 25 °C, in atomic units; the error is the square root of the variance of the parameter estimated by the least squares procedure.

individual parameters are defined. There are also significant covariance terms reported by the least squares procedure, indicating that there is some correlation of parameters, so that results can be better (or worse) than the standard deviations alone would indicate.

There is a problem in comparing the schemes because quite different numbers of compounds were treated. This is a consequence of the larger numbers of parameters in the bond and group schemes, which would require inclusion of data for many more compounds to specify all of the parameters needed for the full 250 compounds considered in the atomic contributions scheme. The necessary heats of formation were not available in the literature. In addition, some compounds would require unique parameters for single compounds in the higher levels of parametrization. For instance, CH_4, CF_4, and CCl_4 all require unique group contributions as do all CH_3X, CH_2XY, and CHXYZ where X, Y, and Z are univalent species (which may be the same). To have a direct comparison, a reduced set of 180 compounds was treated by all four systems, This required somewhat fewer parameters for some of the schemes since some classes of compounds were excluded. With this standard set, the atomic contribution scheme with 12 parameters had a standard deviation of 2.64 kcal/mol. The bond contribution scheme, with 32 parameters, had a standard deviation of 1.72 kcal/mol, and the modified atomic contribution scheme with 18 parameters had a standard deviation of 1.97 kcal/mol. The group contribution scheme had a standard deviation of 1.26 kcal/mol. The parameters evaluated by least-squares fitting of the four parametrization schemes to the standard data set are also found in Tables 1−4.

The errors in these parametrized fits appear to be mostly normally distributed, with a few outliers which represent systematic errors. To show this, we have calculated the number of errors falling in 1 kcal bins, i.e., between −5 and −4, between −4 and −3, etc. These are compared with the expectation for a normal distribution with the observed standard deviation. These expected values were calculated using Table C2 in Bevington,[16] The results are shown in Figure 1. The observed errors are in approximate accord with the theoretical expectation, with two differences which can be noted. The observed errors actually give steeper distributions, corresponding to smaller standard deviations, except that there are more outliers with large errors than predicted by the normal distribution. It is thus of interest to look for patterns in the outliers to see if there are characteristic structural features which lead to poor performance by these procedures for estimating heats of formation. Outliers were defined as compounds with errors for a particular parametrization greater than 3 standard deviations. At the numbers of compounds in these fits, there should be less than one error this large on either wing of the distribution so that any such error is clearly outside the normal distribution. At this level, the full data set for the atom level parametrization had

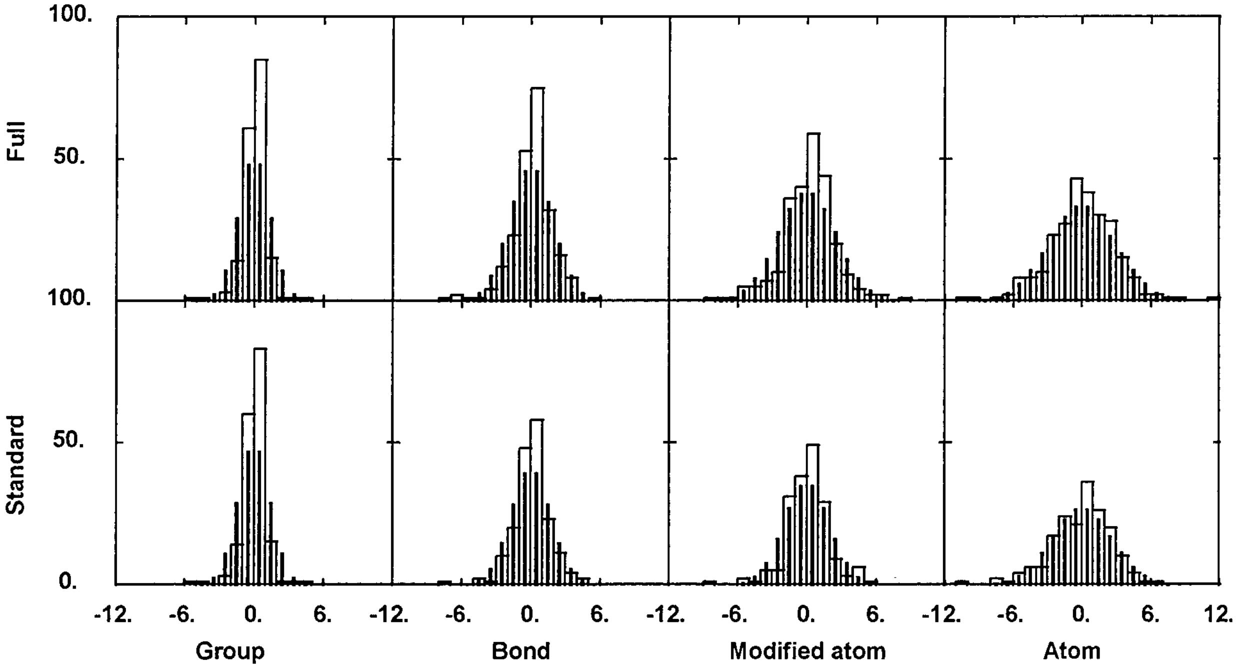

Figure 1. Error distributions for the least-squares fits to the various parametrization schemes. Open bars: observed number of errors with magnitude between x-1 and *x* kcal/mol; solid lines: expected number of errors for a normal distribution with the observed standard deviation.

TABLE 4: Modified Atomic Contribution Parameter Values[4]

"atom"	full data set		standard data set	
	value	error	value	error
$[CH_3(X)]$	−39.902378	0.000225	−39.902256	0.000211
$[CH_2(X)_2]$	−39.307804	0.000195	−39.308048	0.000171
$[CH(X)_3]$	−38.712749	0.000357	−38.712566	0.000343
$[C(X)_4]$	−38.116291	0.000541	−38.115711	0.000504
$[C_bH]$	−38.714256	0.000173	−38.714527	0.000175
$[C_b(X)]$	−38.121784	0.000481	−38.121265	0.000428
$[CO(X)_2]$	−113.276718	0.000719	−113.276688	0.000703
$[CHO(X)]$	−113.871162	0.001548	−113.870934	0.001213
$[O(X)_2]$	−75.156059	0.000345	−75.156876	0.000311
$[OH(X)]$	−75.744827	0.000691	−75.744789	0.000579
$[N(X)_3]$	−54.766453	0.000862	−54.767059	0.000756
$[NH(X)_2]$	−55.360847	0.002132	−55.360523	0.001689
$[NH_2(X)]$	−55.951916	0.001149	−55.952232	0.001155
$[NO_2(X)]$	−205.086288	0.001891	−205.086777	0.001899
$[SO_2(X)_2]$	−548.454346	0.001193	−548.456604	0.001526
$[Cl(X)]$	−460.174988	0.000613		
$[F(X)]$	−99.750572	0.000414		
$[S(X)_2]$	−398.202850	0.000903	−398.202667	0.001046
$[SO(X)_2]$	−473.328857	0.001885	−473.330017	0.001914
$[P(X)_3]$	−341.354004	0.002425		
$[P(X)_5]$	−341.304871	0.003195		
$[PO(X)_3]$	−416.476624	0.002124		
$[SH(X)]$	−398.793945	0.001450	−398.793732	0.001306

At 25 °C, in atomic units; the error is the square root of the variance of the parameter estimated by the least squares procedure.

four outliers: bicyclobutane-1-carbonitrile, trimethylphosphine, trimethylphosphine oxide, and 1-phenylcyclohexene. The modified atomic level parametrization had four outliers: bicyclobutane-1-carbonitrile, trimethylphosphine, trimethylphosphine oxide, and triethyl phosphite. (1-Phenylcyclohexene had a large error but less than the 3σ criterion.) The bond level paramerization fit had three outliers: 2,2,3,3-tetramethylbutane, bicyclobutane-1-carbonitrile, and 1-phenylcyclohexene. Phosphorus compounds had been excluded from the bond and group data sets. The group level parametrization fit had two outliers: bicyclobutane-1-carbonitrile, and. [2.1.0]bicyclopentane-1-carbonitrile. Compounds with a phenyl to alkene bond had been excluded from the data set, so 1-phenylcyclohexene was not fitted. For the standard data set of 180 compounds, the atomic parameters gave one outlier, bicyclobutane-1-carbonitrile, as did the bond scheme. The modified atomic scheme had three outliers, bicyclobutane-1-carbonitrile, *N*,*N*-dimethyl-4-methoxybenzamide, and acetylene. The group scheme had only two outliers, bicyclobutane-1-carbonitrile, and. [2.1.0]bicyclopentane-1-carbonitrile

The lessons seem clear. Heats of formation of *highly* strained molecules are likely to have substantial errors for any of the parametrizations discussed here. Note that less extreme cases of strained molecules are fitted satisfactorily: norbornane, methylenecyclopropane, methylcyclopropene.

Phosphorus compounds were left in the atomic and modified atomic contributions data sets because most of the limited body of data were fitted satisfactorily, but compounds with substituents on phosphorus of quite different electronegativity from the oxygen or halogen which dominates the data set led to deviations, as seen for trimethylphosphine and trimethylphosphine oxide. It seems likely that these molecules require a higher level of parameter (bond or group) but there are not enough experimental values at this time to define these parameters satisfactorily. That triethyl phosphite deviates so badly is surprising because PF_3 is fitted satisfactorily. Clearly there is a need for more well established thermochemical values to settle whether PX_3 compounds, with X an electronegative atom, can be adequately described at these lower levels of parametrization. It is already clear that such schemes are not adequate for the full range of PX_3 compounds with no restriction on electronegativity. The few phosphorus compounds for which there are experimental heats of formation are fitted reasonably well by the low level treatments although the large deviation for triethyl phosphite with the modified atomic contributions scheme shows that the treatment does not cope well with large changes in the electronegativity of the attached atoms. This is demonstrated by omitting the point for trimethylphosphine from the fit. The standard deviation falls to 2.42 and the only deviations greater than 3σ are bicyclobutane-1-carbonitrile and trimethylphosphine oxide. There is no justification for omitting trimethyl phos-

TABLE 5: Parameters Needed for Some Sample Calculations

compound	atom	bond	group	modified atom
ethylene oxide	$2[C_4]$	$4[C_4-H]$	$1[O(C)_2]$	$2[CH_2(X)_2]$
	$4[H]$	$1[C_4-C_4]$	$2[CH_2(C)(O)]$	$1[O(X)_2]$
	$1[O_2]$	$2[C_4-O_2]$		
2-methoxy-2-methyl-1,3-dioxolane	$5[C_4]$	$10[C_4-H]$	$2[CH_3(X)]$	$2[CH_3(X)]$
	$10[H]$	$2[C_4-C_4]$	$3[O(C)_2]$	$2[CH_2(X)_2]$
	$3[O_2]$	$6[C_4-O_2]$	$2[CH_2(C)(O)]$	$1[C(X)_4]$
			$[C(O)_3(C)]$	$3[O(X)_2]$
1,2-butadiene	$1[C_4]$	$3[C_4-H]$	$[CH_3(X)]$	$1[CH_3(X)]$
	$2[C_3]$	$3[C_3-H]$	$[C_dH_2]$	$1[CH_2(X)_2]$
	$1[C_2]$	$1[C_3-C_4]$	$[C_dH(C)]$	$1[CH(X)_3]$
	$6[H]$	$2[C_a\text{-}C_3]$	$[C_a]$	$1[C(X)_4]$
phenanthrene	$14[C_b]$	$10[C_b\text{-}H]$	$10[C_bH]$	$10[C_bH]$
	$10[H]$	$16[C_b = C_b]$	$2[C_f(C_f)(C_b)_2]$	$4[C_b(X)]$
			$2[C_f(C_f)_2(C_b)]$	
bicyclo[1.1.0]butane-1-carbonitrile	$4[C_4]$	$5[C_4-H]$	$2[CH_2(C)_2]$	$2[CH_2(X)_2]$
	$1[C_2]$	$5[C_4-C_4]$	$1[CH(C)_3]$	$1[CH(X)_3]$
	$1[N_1]$	$1[NC-C_4]$	$1[C(C)_3CN]$	$2[C(X)_4]$
	$5[H]$			$1[N(X)_3]$

phine: the reported uncertainty in its enthalpy of formation is small, as is that for triethyl phosphite. This simply serves as a warning about the dangers of trying to have a single parameter for tricoordinate phosphorus with a wide range of substituents. Curiously enough there is no analogous problem with carbon bonded to groups with a range of electronegativities: CX_4 with $X = H$, CH_3, F, Cl, and OCH_3 are all treated satisfactorily. For sulfuric or sulfurous acid derivatives, the problem of a large effect of changing electronegativities of substituents is even more serious.[17] There are too few data to allow extensive testing at higher levels of parametrizations, so these compounds will not be discussed further.

Finally 1-phenylcyclohexene is an outlier in both schemes where it is included. On checking the origins of this value we found that it is traced from Pedley[18] to Cox and Pilcher [19] to an experimental heat of combustion dating from 1935.[20] It is possible that the problem here is in the experimental value.

Acetylene is not treated well by the modified atomic scheme; in fact, it also has a large deviation with the atomic scheme, though it does not count as an outlier there. Other alkynes are treated well by both schemes.

On one hand the success of these schemes suggests that for many classes of compounds very good estimates of the heat of formation at 298 K can be obtained with relatively inexpensive calculations, requiring far less computational resources than G2 [2], or G3 [3] level calculations, which would lead more directly to enthalpies of formation of calorimetric accuracy. On the other hand, there are patterns of failure of the approach which indicate that caution must be exercised in using such methods to obtain enthalpies of formation.

To carry out the calculations needed to convert a B3LYP energy to a heat of formation, it is convenient to set up a spreadsheet with the parameter labels and values in rows at the top. Then the DFT energy and the numbers of contributions for each parameter can be entered and the heat of formation calculated. Once the formula has been entered, subsequent calculations are easy. The use of a spreadsheet is advisable because of the need to carry many digits in the DFT energies and parameters. By copying the DFT energy from the output file to the spreadsheet using the Windows clipboard, errors in transcription can be avoided, and the only human data entry needed is the sequence of integers specifying the number of contributions to be used. The process of analyzing the structure for contributions will be illustrated for a few examples in Table 5.

Clearly, one should use the highest level parametrization for which all parameters are available. The quality of the estimates of the enthalpy of formation at 298 K for the methods presented here can be summarized in Table 6.

Discussion

The goal of this investigation was to find a simple scheme which would allow comparatively easy estimation of heats of formation at 298 K, in as general as possible a way for the widest possible range of compounds. In terms of this goal, a group contributions scheme involves too many parameters, and is of too limited a range of applicability. In fact, it has really become a way of using the ab initio calculation to deal with deviations from simple behavior because group contribution schemes lead to very good estimates of heat of formation with no molecular orbital component, provided that the molecules are simple, with no severe strain or nonstandard interactions. However, Benson's tables[14] include numerous correction terms which must be added to the simple group additivity parameters (ring corrections, gauche corrections, cis corrections, ortho corrections, 1,5 H repulsions, ditertiary ether corrections, etc.) and it is all of these nonstandard terms which one expects to be accounted for in the molecular orbital calculations. Thus, when the necessary parameters are well-defined, the group contribution approach, though it may offend one's sense of scientific esthetics because of the large number of parameters, will give the best results. The same criticisms apply to the bond scheme, although to a lesser extent, because bond contributions alone do not lead to particularly good estimates of heats of formation.[13] By contrast the bond contribution scheme for converting DFT energies into enthalpies of formation leads to quite good results with a wider range of applicability than the group schemes. The atomic scheme, even with the inclusion of contributions dependent upon the coordination number, seems to be too simple to give useful accuracy. A modified atomic scheme with CO, SO, SO_2, and PO, as well as aromatic carbon treated as distinct "atoms" allowed better estimations for a small increase in number of parameters. Results of this scheme are good enough for some purposes, and are in fact of comparable accuracy to some experimental values.

The necessary warning about the use of any of these schemes is that they will not account for highly strained molecules. For the heavy elements (S and P were examined) it is not satisfactory to ignore differences in the electronegativity of the atoms attached to this heavy element in the parametrization. This in turn means that at least bond level and probably group level

TABLE 6: Quality of the Estimates of the Enthalpy of Formation at 298 K

	data set					
	full			standard		
method	no. of compds	no. of parameters	std. dev (kcal/mol)	no. of compds	no. of parameters	std. dev (kcal/mol)
atom	250	18	2.96	180	12	2.64
modified atom	248	23	2.53	180	18	1.97
bond	229	40	1.90	180	32	1.72
group	183	79	1.39	180	79	1.26

treatments will be needed to cover the full range of organosulfur or organophosphorus compounds. Unfortunately, there is not yet the body of thermochemical data to permit this in any comprehensive fashion.

Comparison with other Methods. The B3LYP functional[21,22,8,23,24] was developed to give good agreement with thermochemistry. A recent examination[25] of its application to alkanes showed that simple conversion of the DFT energies to enthalpies gave good results for C1−C3 alkanes but increasingly poor results for larger ones, with a 30 kcal/mol error for hexadecane. Use of an isodesmic scheme reduces the error to 16 kcal/mol for hexadecane. A homodesmotic[11] scheme, maintaining equal numbers of CH_3 and CH_2 groups in reactants and products in the hypothetical reactions used to calculate heats of formation reduced the error to 2 kcal/mol for hexadecane. This is looking like a group contribution scheme. In fact, it can be shown that the group contribution scheme presented here is equivalent to a homodesmotic scheme. The fundamental assumption, illustrated for the reactions considered by Redfern et al.,[25] is that for the hypothetical reaction

$$C_nH_{2n+2} + (n-3)C_2H_6 \rightleftharpoons (n-2)\ C_3H_8$$

$\Delta H^\circ \approx \Delta E_{MO}$, where ΔE_{MO} is the difference in calculated energies for this reaction by the molecular orbital method used. Now, these may both be expanded to give

$$\Delta H^\circ = (n-2)\ \Delta H_f^\circ(C_3H_8) - (n-3)\ \Delta H_f^\circ(C_2H_6) - \Delta H_f^\circ(C_nH_{2n+2})$$

$$\Delta E_{MO} = (n-2)\ E_{MO}(C_3H_8) - (n-3)\ E_{MO}(C_2H_6) - E_{MO}(C_nH_{2n+2})$$

Solving these for $\Delta H_f^\circ(C_nH_{2n+2})$, with the assumption that $\Delta H^\circ \approx \Delta E_{MO}$, one obtains

$$\Delta H_f^\circ(C_nH_{2n+2}) = (n-2)\{\ \Delta H_f^\circ(C_3H_8) - E_{MO}(C_3H_8)\} - (n-3)\{\ \Delta H_f^\circ(C_2H_6) - E_{MO}(C_2H_6)\} + E_{MO}(C_nH_{2n+2})$$

Now by the definition of group parameters

$$\Delta H_f^\circ(C_2H_6) = E_{MO}(C_2H_6) - 2^*[CH_3(X)]$$

and

$$\Delta H_f^\circ(C_3H_8) = E_{MO}(C_3H_8) - 2^*[CH_3(X)] - 1^*[CH_2(X)_2]$$

Thus

$$\Delta H_f^\circ(C_nH_{2n+2}) = (n-2)\{-2^*[CH_3(X)] - 1^*[CH_2(X)_2]\} - (n-3)\{-2^*[CH_3(X)]\} + E_{MO}(C_nH_{2n+2}) = E_{MO}(C_nH_{2n+2}) - 2^*[CH_3(X)] - (n-2)^*[CH_2(X)_2]$$

Thus, the group contribution scheme presented here is a systematic application of the homodesmotic approach to converting molecular orbital energies to heats of formation. Similarly, the bond contribution scheme presented here is a somewhat more elaborate version of an isodesmic approach with some account taken of the local environment of the atoms at the ends of the bonds.

Ibrahim and Schleyer[4] reported average errors on the order of 2 kcal/mol for all molecules with their scheme that used atomic contributions which depended on what atoms were attached to the central atom. The results for benzene were unsatisfactory and no other aromatics were considered. Their set of compounds did not include any with anomeric effects except for some polyfluoro methanes and ethanes.

Yala[26] reported an atom equivalents scheme for converting rhf/6-31G** energies to heats of formation, with an rms deviation of 1.80 kcal/mol for a set of 62 compounds with 11 parameters, including benzene as the only aromatic compound (fitted poorly) and very few examples of molecules with O or N containing functional groups (20 compounds representing 10 functional groups). Castro[27] reported a similar scheme, and had an average error of 3.0 kcal/mol for O and N compounds.

Cioslowski et al.[28] reported a general approach to converting B3LYP energies to enthalpies of formation which used atomic equivalents, bond density functions and corrections for molecular charge and spin multiplicity. This approach was able to calculate reasonable enthalpies of formation for a wide range of molecules and ions, but even at B3LYP/6-311G** and the most elaborate system the standard deviation was 5.84 kcal/mol in enthalpy.

For hydrocarbons, atomic schemes can do very well.[29] For polycyclic aromatic hydrocarbons, simple group schemes can give good accuracy either at rhf/6-31G* [30] or MP2/6-31G* [31] levels.

Allinger has shown that for various families of compounds a bond contributions scheme, derived from the one used in MM2 and MM3, can be used to convert rhf/6-31G* energies into heats of formation with accuracy of less than a kcal/mol. This was done for alkanes[32] (somewhat better results were obtained by B3LYP/6-31G* calculations), alkylamines,[33] carboxylic acids and esters,[34] aliphatic aldehydes and ketones,[35] alcohols and ethers,[36] and thiaalkanes.[37]

For limited ranges of structural variation, there is abundant evidence that various kinds of parametrization schemes can give very good enthalpies of formation from RHF or DFT energies. The present investigation looked at a range of molecular sizes and functionality, including molecules with multiple functional groups, to test the applicability of various schemes for converting molecular orbital energies to enthalpies of formation for the kinds of polyfunctional compounds which arise in studies of reaction mechanisms. The conclusion from this work is that the conversion can be done with useful accuracy for a wide range of compounds, though there are limits to the currently inexpensive methods. It should be pointed out that by limiting the compounds considered in this work to subsets with only one functional group and no sources of strain, the standard group contribution parameters derived here give smaller standard deviations: for 10 alkanes the rms error = 0.63; for 10 alkylamines the rms error = 0.67; for 10 alcohols and simple ethers (including ethylene oxide) the rms error = 0.75; for 11 aliphatic carboxylic acids and esters the rms error = 0.69; for 5 aliphatic aldehydes and ketones he rms error = 0.72. Thus, with the standard group contribution parameters and limited ranges of compounds the calculated values are in very good agreement with experiment. With a much wider range of compounds, including polyfunctional molecules, the agreement is less good but still quite useful.

Calculations

Methods. Molecular orbital calculations were carried out using Gaussian 94.[38] Initial structures were optimized by molecular mechanics using PCModel.[39] Then the structure was optimized using Gaussian at the rhf/6-31G** level, and finally a single point calculation at the optimized geometry was carried out using B3LYP/6-31G**,[8,9] as implemented in Gaussian 94.

Acknowledgment. We thank the Natural Sciences and Engineering Research Council of Canada for financial support of this work.

Supporting Information Available: Appendix: Heats of formation for compounds where only the heat of formation of the liquid is available from the literature. Table S1: Enthalpies of formation for compounds discussed in this work. Table S2: Experimental and calculated heats of formation for the various parameterizations, with the full data sets. Table S3: Observed and calculated heats of formation for the standard set of compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622−5629.

(2) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221−7230.

(3) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764−7776.

(4) Ibrahim, M. R.; Schleyer, P. v. R. *J. Comput. Chem.* **1985**, *6*, 157−167.

(5) Wiberg, K. B. *J. Comput. Chem.* **1984**, *5*, 197−199.

(6) Baird, N. C.; Hadley, G. C. *Chem. Phys. Lett.* **1986**, *128*, 31−37.

(7) Dewar, M. J. S.; Storch, D. M. *J. Am. Chem. Soc.* **1985**, *107*, 3898−3902.

(8) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(9) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(10) Guthrie, J. P. *J. Phys. Chem.* **2001**, in press.

(11) George, P.; Trachtman, M.; Bock, C.; Brett, A. *Tetrahedron* **1976**, *32*, 317−323.

(12) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio Molecular Orbital Theory*; John Wiley and Sons: New York, 1986.

(13) Benson, S. W.; Buss, J. H. *J. Chem. Phys.* **1958**, *29*, 546−572.

(14) Benson, S. W. *Thermochemical Kinetics*, Second ed.; Wiley: New York, 1976.

(15) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 292−298.

(16) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.

(17) Calculations carried out in this project but not reported here.

(18) Pedley, J. B.; Naylor, R. D.; Kirby, S. P. *Thermochemical Data of Organic Compounds*, Second ed.; Chapman and Hall: London, 1986.

(19) Cox, J. D.; Pilcher, G. *Thermochemistry of Organic and Organometallic Compounds*; Academic Press: London, New York, 1970.

(20) Brull, L. *Gazzetta Chimica Italiana* **1935**, *65*, 19.

(21) Becke, A. D. *J. Chem. Phys.* **1992**, *96*, 2155−2160.

(22) Becke, A. D. *J. Chem. Phys.* **1992**, *97*, 9173−9177.

(23) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040−1046.

(24) Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554−8560.

(25) Redfern, P. C.; Zapol, P.; Curtiss, L. A.; Raghavachari, K. *J. Phys. Chem. A* **2000**, *104*, 5850−5854.

(26) Yala, Z. *J. Mol. Struct. (THEOCHEM)* **1990**, *207*, 217−223.

(27) Castro, E. A. *Theochem − J. Mol. Struct.* **1994**, *110*, 93−99.

(28) Cioslowski, J.; Liu, G.; Piskorz, P. *J. Phys. Chem.* **1998**, *A102*, 9890−9900.

(29) Mole, S. J.; Zhou, X. F.; Liu, R. F. *J. Phys. Chem.* **1996**, *100*, 14 665−14 671.

(30) Schulman, J. M.; Peck, R. C.; Disch, R. L. *J. Am. Chem. Soc.* **1989**, *111*, 5675−5680.

(31) Herndon, W. C.; Biedermann, P. U.; Agranat, I. *J. Org. Chem.* **1998**, *63*, 7445−7448.

(32) Labanowski, J.; Schmitz, L.; Chen, K. H.; Allinger, N. L. *J. Comput. Chem.* **1998**, *19*, 1421−1430.

(33) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. *J. Phys. Org. Chem.* **1990**, *3*, 732−736.

(34) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. *J. Comput. Chem.* **1992**, *13*, 838−841.

(35) Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K.; Allinger, N. L. *J. Phys. Org. Chem.* **1992**, *5*, 225−229.

(36) Allinger, N. L.; Schmitz, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. *J. Am. Chem. Soc.* **1992**, *114*, 2880−2883.

(37) Allinger, N. L.; Yang, L. R.; Motoc, I.; Bender, C.; Labanowski, J. K. *Heteroatom Chem.* **1992**, *3*, 69−72.

(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, 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.; Nanayakkara, 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 94*, revision E.1; Gaussian, Inc.: Pittsburgh, PA, 1995.

(39) Serena Software, Box 3076, Bloomington, IN 47402-3076.

(40) Guthrie, J. P. *Can. J. Chem.* **1986**, *64*, 635−640.

(41) Pedley, J. B. *Thermochemical Data and Structures of Organic Compounds*; Thermodynamics Research Center: College Station, TX, 1994.

(42) Verevkin, S. P. *J. Chem. Eng. Data* **2000**, *45*, 953−960.

(43) Roth, W. R.; Adamczak, O.; Breuckmann, R.; Lennartz, H. W.; Boese, R. *Chem. Ber.* **1991**, *124*, 2499−2521.

(44) Verevkin, S. P. *J. Chem. Eng. Data* **1999**, *44*, 175−179.

(45) Chase, M. W. *NIST-JANAF Thermochemical Tables*, 4th ed.; American Chemical Society: Washington, DC, 1998.

(46) Wagman, D. D.; Evans, W. H.; Parker, V. B.; Schumm, R. H.; Halow, I.; Bailey, S. M.; Churney, K. L.; Nuttall, R. L. *J. Phys. Chem. Ref. Data* **1982**, *11, Suppl. 2*.

(47) Giguere, P. A.; Morissette, B. G.; Olmos, A. W.; Knop, O. *Can. J. Chem.* **1955**, *33*, 804−820.

(48) Lay, T. H.; Bozzelli, J. W. *J. Phys. Chem.* **1997**, *A101*, 9505−9510.

(49) Lord, A.; Pritchard, H. O. *J. Chem. Thermodyn.* **1970**, *2*, 187−191.

(50) Kabo, G. Y.; Miroshnichenko, E. A.; Frenkel, M. L.; Kozyro, A. A.; Simirskii, V. V.; Krasulin, A. P.; Vorobeva, V. P.; Lebedev, Y. A. *Bulletin of the Russian Academy of Sciences, Division of Chemical Sciences* **1990**, 662−667.

(51) Guthrie, J. P.; Gallant, R. T. *Can. J. Chem.* **2000**, *78*, 1295−1298.

(52) Choi, J. K.; Joncich, M. J. *J. Chem. Eng. Data* **1971**, *16*, 87−90.

(53) Guthrie, J. P. *J. Am. Chem. Soc.* **1974**, *96*, 3608−3615.

(54) Cruickshank, F. R.; Benson, S. W. *J. Am. Chem. Soc.* **1969**, *91*, 1, 2487−2492.

(55) Wagman, D. D.; Kilpatrick, J. E.; Pitzer, K. S.; Rossini, F. D. *J. Res. Natl. Bur. Stand.* **1945**, *35*, 467496.

(56) Melia, T. P.; Merrifield, R. *J. Appl. Chem.* **1969**, *19*, 79.

(57) Verevkin, S.; Dogan, B.; Beckhaus, H.-D.; Ruchardt, C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 674−675.

(58) Guthrie, J. P. *Can. J. Chem.* **1976**, *54*, 202−209.

(59) Guthrie, J. P. *Can. J. Chem.* **1977**, *55*, 3562−3574.

(60) Wiberg, K. B.; Squires, R. R. *J. Am. Chem. Soc.* **1979**, *101*, 5512−5515.

(61) Sabbah, R.; Gouali, M. *Aust. J. Chem.* **1994**, *47*, 1651−1660.

(62) Acree, W. E.; Tucker, S. A.; Pilcher, G.; Toole, G. *J. Chem. Thermodyn.* **1994**, *26*, 85−90.

(63) Acree, W. E.; Tucker, S. A.; Zvaigzne, A. I.; Yang, M. Y.; Pilcher, G.; Ribeiro da Silva, M. D. M. C. *J. Chem. Thermodyn.* **1991**, *23*, 31−36.

(64) Guthrie, J. P. *Can. J. Chem.* **1978**, *56*, 962−973.

(65) Guthrie, J. P.; Pike, D. C.; Lee, Y. C. *Can. J. Chem.* **1992**, *70*, 1671−1683.

(66) Marsh, K. N.; Mansson, M. *J. Chem. Thermodyn.* **1985**, *17*, 995.

(67) Guthrie, J. P.; Cullimore, P. A. *Can. J. Chem.* **1980**, *58*, 1281−1294.

(68) Guthrie, J. *J. Am. Chem. Soc.* **1977**, *99*, 3391−4001.

(69) Steele, W. V.; Chirico, R. D.; Smith, N. K. *J. Chem. Thermodyn.* **1995**, *27*, 671−678.

(70) Ribeiro da Silva, M. A. V.; Matos, M. A. R.; do Rio, C. M. A.; Morais, V. M. F. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 3061−3065.

(71) Morawetz, E. *J. Chem. Thermodyn.* **1972**, *4*, 455−460.