Research Article

Journal of Physical Organic Chemistry

Received: 22 November 2007, Revised: 13 March 2008, Accepted: 21 March 2008, Published online in Wiley InterScience: 2 June 2008

(www.interscience.wiley.com) DOI 10.1002/poc.1376

A statistical-mechanical analysis of group additivity. Calculation of thermochemical values from frequency distributions

Gustav Bojesen[a]*

Molecular vibrational frequencies of homologous series plotted as cumulative frequency distributions are very similar, and the fine structure of the distributions exhibit identical features. An obvious explanation is that the changes in the molecular frequency distributions (MFDs) from one homologue to the next is independent of the chain length and of the functional groups in the molecule. Since group additivity is valid for the chosen homologous series there is here an explanation for the linearity of thermochemical values expressed by group additivity. For these properties the following hypothesis is proposed: group additivity is observed when the MFD is a sum of group frequency distributions (GFDs). This leads to additivity for the zero-point vibrational energy which is confirmed by analysis of the frequencies of 126 organic molecules from 11 homologous series. The frequency distribution for a methylene group is estimated from that of octadecane. From this GFD combined with 11 different MFD it is possible to calculate model frequencies for the homologous series which are in good agreement with frequencies from *ab initio* calculations. For three thermochemical parameters (the logarithm of the vibrational partition function, the vibrational excitation energy and the vibrational heat capacity), the combination of the estimated methylene GFD with 11 different MFDs lead to group additivity values for a methylene group which are identical over a wide temperature range. The derivation of Benson additivity for thermochemical functions from frequency distributions is at step towards a better understanding of Benson additivity.

Supplementary electronic material for this paper is available in Wiley InterScience at http://www.mrw.interscience.wiley.com/suppmat/0894-3230/suppmat/

Keywords: frequency distribution; group additivity

INTRODUCTION

It is the purpose of this work to investigate the statistical-mechanical foundation of additivity methods within the harmonic oscillator stiff rotor approximation. That thermochemical properties of well-behaved organic compounds can be calculated by group additivity is firmly based on experimental evidence.[1–4] However, there appear to have been few attempts to explain the success of group additivity based on the statistical-mechanical description of organic molecules. Some mathematical similarity between *ab initio* calculations and calculations of group additivity values has been noted[5] but generally the justification of the validity of group additivity methods has been based solely on the internal consistency of the results and the agreement between additivity values and experimental or *ab initio* results.

Group additivity is not only used to obtain thermochemical values but does also provide the basis for understanding linear energy relationships, and it will clearly be valuable to have a more fundamental understanding of the method.[6–12]

In the harmonic oscillator stiff rotor approximation, the average energy of a molecule at a given temperature in its electronic ground state can be written as a sum of terms (Eqn (1)) where ε_{ee} is the electronic energy, ε_{zpe} the zero-point energy, ε_{vib} the vibrational excitation energy, ε_{rot} the rotational energy and ε_{tr} is the translational energy

$$\varepsilon = \varepsilon_{ee} + \varepsilon_{zpe} + \varepsilon_{vib} + \varepsilon_{rot} + \varepsilon_{tr} \quad (1)$$

That some thermodynamic property X is additive means that it can be written as a sum of group contributions:

$$X(T) = \sum_i n_i g_i^X(T) \quad (2)$$

where n_i is the number of groups with the contribution $g_i^X(T)$. This leads to the following expressions for the heat of formation

* *Correspondence to: G. Bojesen, Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark. E-mail: bojesen@kiku.dk*

a *G. Bojesen Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark*

ΔH_f, the heat capacity and the entropy:

$$\Delta H_f(T) = \sum_i n_i g_i^{\Delta H}(T) \quad (3a)$$

$$C(T) = \sum_i n_i g_i^{C}(T) \quad (3b)$$

$$S(T) = \sum_i n_i g_i^{S}(T) \quad (3c)$$

where $g_i^{\Delta H}(T)$, $g_i^{C}(T)$ and $g_i^{S}(T)$ are the group contributions to the heat of formation, the heat capacity and the entropy.

In Eqn (1) the energy of the molecule is calculated from the nuclei and electrons whereas the heat of formation is calculated from the elements in their standard states. For molecules in the gas phase, the difference between Eqns (1) and (3a) is essentially that of the standard states apart from the *RT* difference between the internal energy and the enthalpy. Since group additivity is valid at a range of temperatures as long as it is not too low, the most straightforward way for both Eqns (1) and (3a) to be valid is that the individual terms in Eqn (1) are a sum of group contributions or only dependent on the temperature.

Group additivity of the electronic energy has been the subject of many investigations and will not be discussed further.[13–15] The additivity of the electronic energy is a very important element in understanding additivity but even for the heat of formation it can only provide a part of the explanation, and for the heat capacity and the entropy it is unsatisfactory.

Within the approximation chosen here, the translational and rotational energies are only determined by the temperature. This leaves the zero-point energy and the vibrational excitation energy and additivity must be valid for both. Atomic additivity of the zero-point energy has been noticed before and its group additivity (Eqn (4)) will be explored below:[16–21]

$$E_{zpe} = \frac{hc}{2k_B}\sum_i \tilde{\nu}_i \quad (4a)$$

$$= \sum_i n_i g_i^{zpe} \quad (4b)$$

The last term in Eqn (1) to be considered is the vibrational excitation energy which can be derived from the following equation:

$$E^*_{vib}(T) = N_A k_B T^2 \frac{\partial}{\partial T} \ln q_{vib}(T) \quad (5)$$

where q_{vib} is the vibrational partition function calculated from the lowest energy level. Since the differential operator $\partial/\partial T$ follows the distributive rule, it follows from Eqn (5) that the vibrational partition function q_{vib} must be a product of group contributions:

$$q_{vib}(T) = \prod_i \frac{1}{1 - \exp[-hc\tilde{\nu}_i/k_B T]} \quad (6a)$$

$$= \prod_i [g_i^{q_v}(T)]^{n_i} \quad (6b)$$

or that the logarithm of the vibrational partition function is a sum of group contributions:

$$\ln q_{vib}(T) = \sum_i n_i \ln[g_i^{q_v}(T)] \quad (7a)$$

$$= \sum_i n_i g_i^{\ln q_v}(T) \quad (7b)$$

The multiplicativity of the partition function or the additivity of the logarithm of the partition function also follows from the expression for the entropy which in itself is additive:

$$S = \frac{H + RT}{T} + k_B \ln Q \quad (8)$$

The validity of Eqn (6) does suggest that there must be some similarity or systematic variations in the frequencies of compounds for which group additivity is valid, and in order to explore this, cumulative frequency distribution for a variety of molecules have been investigated.

Figure 1 shows a graphical representations of the frequencies for the linear alkanes propane, octane, tridecane and octadecane plotted as cumulative molecular frequency distributions (MFDs). The frequencies are calculated at the MP2(FC)/6-31G* level. The plots are constructed in the same way as cumulative distributions known from statistical analyses, such that the ordinate of each point is the fraction of frequencies in the molecule with a wavenumber smaller than or equal to the abscissa. The frequencies fall in three groups within the approximate intervals: $\tilde{\nu} \leq 530\,cm^{-1}$, $740\,cm^{-1} \leq \tilde{\nu} \leq 1580\,cm^{-1}$ and $3070\,cm^{-1} \leq \tilde{\nu}$. Each methylene group contributes with two frequencies in the low region below $530\,cm^{-1}$, five in the middle region between 740 and $1580\,cm^{-1}$ and two in the high region above $3070\,cm^{-1}$. The five frequencies in the region between 740 and $1580\,cm^{-1}$, fall with four between 740 and $1475\,cm^{-1}$ and one in a narrow region around $1570\,cm^{-1}$. For the three longest alkanes in Fig. 1, even the fine structures in the cumulative frequency distributions are clearly similar, which is particularly noticeable in the middle region between 740 and $1580\,cm^{-1}$. Methyl groups have a different number of frequencies in the three regions and consequently the four distributions, although very similar, will not be identical. For the shortest alkanes with highly discontinous distributions, the fine structure of the distributions are difficult to discern but in the whole series from octane (six methylene groups) to octadecane (16 methylene groups) there appear to be no changes in the features of the fine structure. The cumulative frequency distributions of the homologous series listed in Table S1 show the same similarities. Frequency distributions are the basis for analysing the heat capacity of solids and have been used in calculations of heat capacities for liquid and crystalline carbon polymers.[22–26]

It follows from the analysis above, and most clearly from Eqn (6a), (6b) that if Benson groups could be associated with constant group frequencies independently of the number and kind of other groups in the molecule, additivity would follow directly. Many spectroscopic group frequencies are indeed very insensitive to the molecular structure but some, in particular those at low wavenumbers, can occur within wide intervals. However, the most serious shortcoming of spectroscopic group frequencies in any explanation for additivity is that they in general occur at too high frequencies to provide anything but a minor contribution to those thermochemical values which depends on frequencies. This is shown in Fig. 2 where the cumulated vibrational heat capacity for octane is plotted against the wavenumber. For this molecule the normal modes with frequencies below $1200\,cm^{-1}$ contribute more than 90% of the vibrational heat capacity at 300 K. The frequency distribution of this molecule is typical and it shows that spectroscopic group frequencies cannot provide an explanation for the validity of group additivity.

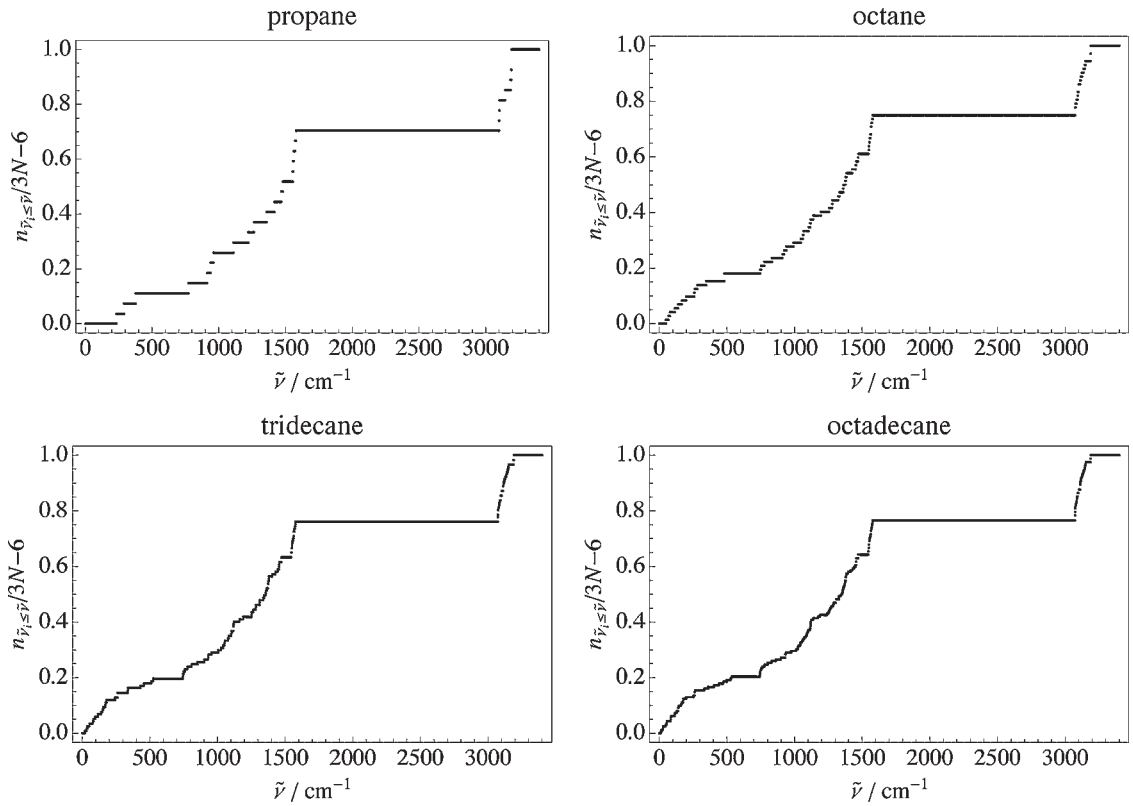

Figure 1. Cumulative molecular frequency distributions (MFDs) for propane, octane, tridecane and octadecane

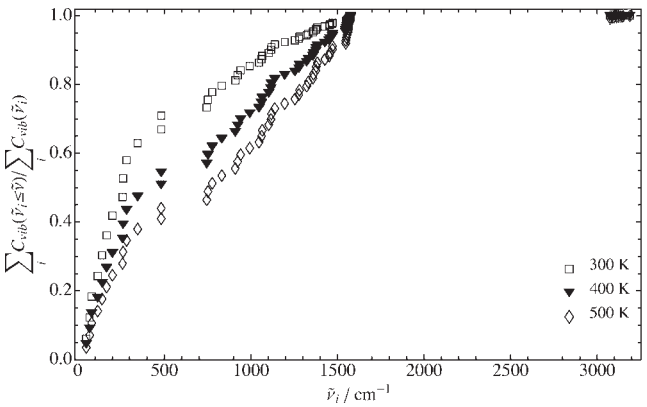

Figure 2. Cumulative plot of the vibrational contribution to the heat capacity of octane

The very similar MFDs of the linear alkanes and in particular the constant fine structures over the series immediately suggest that they can be formed by linear combinations such that the MFD of each homologue is derived from the MFD of the lower homologue combined with a frequency distribution that is characteristic for the methylene group. Such a distribution will be called a group frequency distribution (GFD). The validity of additivity expresses that thermochemical values are linear functions of the groups, and this leads to the central hypothesis of this paper which can briefly be formulated as follows: for properties which depend on the vibrational frequencies, *group additivity is observed when the MFD is a sum of GFDs*. This hypothesis will be tested by calculations of molecular frequencies and by analysis of the group contribution from a methylene group on thermochemical properties, in particular the zero-point energy, the vibrational partition function, the vibrational excitation energy and the vibrational contribution to the heat capacity.

RESULTS AND DISCUSSION

Zero-point energy and additivity

It is easily seen that a MFD determined as described above will lead to zero-point energy additivity. When the MFD is a sum of GFDs, it follows that the mean of all the frequencies in the molecule: $< \tilde{\nu}_M >$ can be written as the weighted sum of group frequency means $< \tilde{\nu}_i >$:

$$< \tilde{\nu}_M > = \frac{1}{(3N-6)} \sum_i n_i w_i \ < \tilde{\nu}_i > \qquad (9)$$

where n_i is the number of ith groups in the molecule and w_i is the number of frequencies in the group. From this follows:

$$\frac{hc}{2k_B}(3N-6) < \tilde{\nu}_M > = \frac{hc}{2k_B} \sum_i n_i w_i \ < \tilde{\nu}_i > \qquad (10)$$

and comparison with Eqn (4b) shows that the zero-point energy group additivity value g_i^{zpe} can be written as

$$g_i^{zpe} = \frac{hc}{2k_B} w_i < \tilde{\nu}_i > \qquad (11)$$

As mentioned above, several studies have shown the approximate linear dependency of the zero-point energy on the molecular composition whereas group additivity has

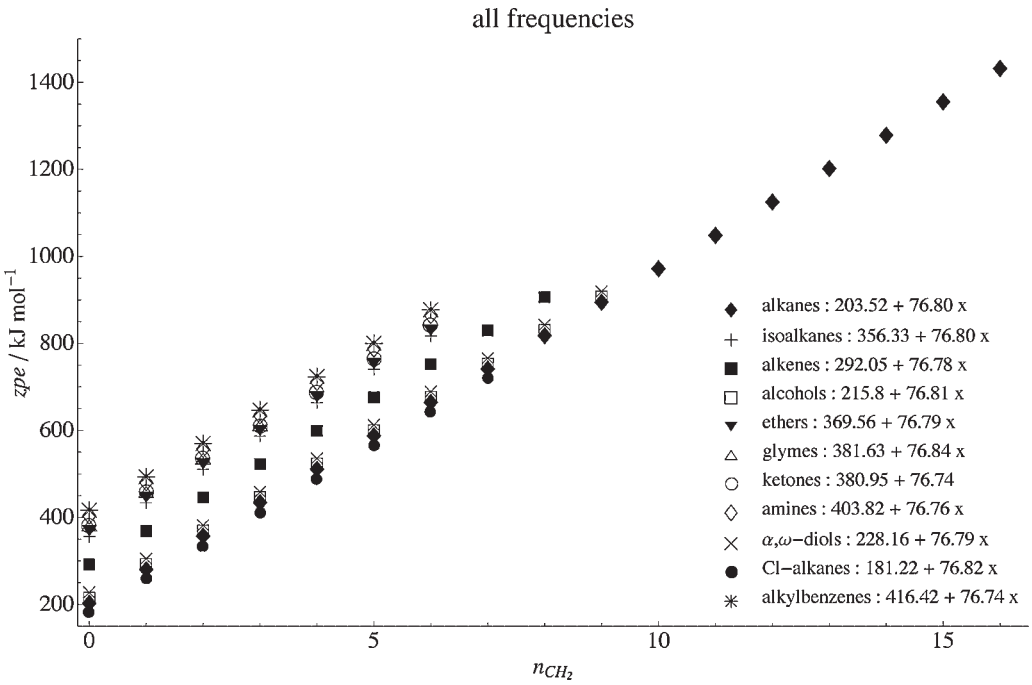

Figure 3. Zero-point energy dependency on number of CH_2 groups. The values are from *ab initio* calculations. The equations in the legend are least-square fits to the points for each homologous series

attracted much less attention. Group additivity of the zero-point energy must necessarily provide more variables than any atomic additivity scheme and thus be more accurate.

For the 126 compounds in Table S1, the zero-point energy values from *ab initio* calculations are summarized in Fig. 3. The average total zero-point additivity value for a methylene group as calculated from the slopes of the fits in Fig. 3(a) is equal to 76.79 kJ mol^{-1}. According to the central hypothesis the methylene group additivity value should be independent of the other groups in the molecule or on the number of methylene groups. This is confirmed by the very small variations in the slopes and by the quality of the linear fits (all standard errors on the slopes are smaller than 0.03).

When the number of frequencies in a selected region changes with a constant number from one homologue to the next, the zero-point energy for the frequencies in that region should, according to the hypothesis, be a linear function of the number of methylene groups. For the three regions mentioned above, this criteria is fulfilled for the homologous series in Table S1, except for the alkylbenzenes for which there is no empty region between 530 and 740 cm^{-1}. As shown in Figs S1 and S2 the linearity of the fits are excellent. The lowest point in the alkane series from ethane show the only deviation from the predicted linearity. The linearity of the zero-point energy in the selected regions also proves that the overall linearity of the zero-point energy is not caused by the two additional high C—H stretch frequencies. The observed behaviour of the zero-point energy is in agreement with the hypothesis and directly explains the group additivity of the zero-point energy required by the concurrent validity of Eqns (1) and (3a).

Comparisons of molecular frequency distributions

One obvious consequence of the central hypothesis is that isomers with the same Benson groups should have identical frequencies. An indication that this is indeed the case can be seen in Figs 3, S1 and S2. For the ethers, glymes, ketones and secondary amines several of the points arise from more than one isomer. For the total zero-point energy and the zero-point energy calculated from the frequency in the middle regions these points coincide. Only for the zero-point energy calculated from the normal modes in the lowest region small deviations are seen.

The results of direct comparisons of the frequencies is shown in Table 1. In this table the frequencies from isomeric dodecenes are also included. When the frequencies of the isomers are plotted against each other a straight line with a slope of one is obtained, and the temperature dependency of the vibrational heat capacities are very close to identical. Notice that the linear regressions are based only on frequencies below 1200 cm^{-1} which means that the spectroscopic group frequencies are largely excluded. The observation that isomers with the same Benson groups have nearly identical frequencies confirms the hypothesis.

Vibrational excitation energy and heat capacity

The vibrational excitation energy $E^*(T)$ is equal to the integrated heat capacity:

$$E^*_{\mathrm{vib}}(T) = \int_0^T C_{\mathrm{vib}}(T)\mathrm{d}T \tag{12}$$

Assuming additivity for the heat capacity, Eqn (12) can be written as

$$E^*_{\mathrm{vib}}(T) = \int_0^T \sum_i n_i g_i^C(T)\mathrm{d}T \tag{13a}$$

$$= \sum_i n_i \int_0^T g_i^C(T)\mathrm{d}T \tag{13b}$$

Table 1. Comparison of frequencies and vibrational heat capacities from isomers with the same Benson groups

	y		
x	Heptylpropylether	Butylhexylether	Dipentylether
Ethyloctylether	0.999259, 0.00431[a] 87, 20, −16, −35[b]	0.999863, 0.00293 −2, −75, −105, −114	0.999054, 0.00500 31, −35, −66, −78
2-Pentoxyethanol	3-Butoxypropanol 0.998359, 0.00578 −393, −324, −271, −216	4-Propoxybutanol 0.99767, 0.00639 −337, −248, −191, −147	5-Ethoxypentanol 0.998692, 0.00441 −415, −265, −175, −112
3-Undecanone	4-Undecanone 0.99605, 0.00548 −45, −17, −6, −2	5-Undecanone 0.994159, 0.00797 −44, −24, −15, −11	6-Undecanone 0.994202, 0.00750 −42, −22, −15, −11
Ethyloctylamine	Heptylpropylamine 1.00087, 0.00336 82, 9, −24, −39	Butylhexylamine 1.00108, 0.00233 34, −35, −64, −72	Dipentylamine 1.00082, 0.00399 43, −26, −54, −64
Trans-3-dodecene	*Trans*-4-dodecene 0.998876, 0.00481 7, 12, 6, 1	*Trans*-5-dodecene 0.998137, 0.00488 13, 29, 26, 20	*Trans*-6-dodecene 0.998301, 0.00555 3, 20, 17, 12

[a] These numbers are the slope and its standard error from the regression analysis of the points: $(\tilde{\nu}_i^x, \tilde{\nu}_i^y)$ for $\tilde{\nu}_i < 1200\,\mathrm{cm}^{-1}$.
[b] These numbers are the difference in the vibrational heat capacity calculated as $C^y - C^x$ ($10^{-3}\,\mathrm{J\,K^{-1}\,mol^{-1}}$) for T = 300, 400, 500, 600 K.

Since the heat capacity is an increasing function of the temperature any deviation from additivity at low temperatures will give a proportionally small contribution to the vibrational energy and thus additivity of the vibrational excitation energy will follow from the additivity of the heat capacity.

Model calculations

Whereas the additivity of the zero-point energy follows directly from its definition and the way of calculating MFDs from GFDs, this is neither the case for the vibrational excitation energy nor for the heat capacity or the logarithm of the partition function. A more detailed analysis for what is required is presented in the Appendix. Without an analytical proof that additivity of the three mentioned thermochemical functions follows from the construction of MFDs from GFDs, confirmation of the hypothesis must come from indirect arguments. The approach chosen here is the model calculations described below.

The model calculations include the following steps: (1) the GFD of a methylene group is estimated from the frequencies of a long-chained alkane. (2) Calculation of model frequencies from combinations of the methylene GFD with the *ab initio* based MFDs of the lowest homologues in Table S1 and comparison of the resultant model frequencies of the higher homologues with *ab initio* frequencies. (3) Comparison of heat capacities at 300 K calculated from model frequencies and from *ab initio* frequencies. (4) Proof that the thermochemical values calculated from the model frequencies lead to additivity, that is that Eqn (A7) is satisfied.

In steps 2 and 3 the outcome of model calculations is compared with *ab initio* results. The approximations of the harmonic oscillator stiff rotor description are well understood, and *ab initio* frequencies are successfully used in accurate calculations of thermochemical values in which corrections for the approximations are included. When the frequencies are sufficiently similar the corrections for the approximations will be the same independently of the source of the frequencies, and consequently it is sufficient to compare the model frequencies with *ab initio* frequencies The most significant part of the model calculations is the last step where it is tested whether the linear combinations of frequency distributions do in fact leads to additivity of those thermochemical functions which depend on the molecular frequencies. Thus this step tests whether the required mathematical relationships are obtained.

It is relatively easy to find a GFD for the methylene group since a long-chained linear alkane will have a MFD which is nearly determined by the methylene groups alone. In octadecane there are a total of 162 normal modes: 33 frequencies with $10.2\,\mathrm{cm}^{-1} \le \tilde{\nu} \le 534.4\,\mathrm{cm}^{-1}$, 91 with $743.7\,\mathrm{cm}^{-1} \le \tilde{\nu} \le 1576.0\,\mathrm{cm}^{-1}$ and 38 with $3072.4\,\mathrm{cm}^{-1} \le \tilde{\nu}$. For the methylene group the ratio of frequencies in the same regions are 2, 5 and 2. The methylene GFD was generated by keeping the value at 0 for $\tilde{\nu} \le 10.1\,\mathrm{cm}^{-1}$, 2/9 for $534.5\,\mathrm{cm}^{-1} \le \tilde{\nu} \le 743.6\,\mathrm{cm}^{-1}$ and 7/9 for $1576.1\,\mathrm{cm}^{-1} \le \tilde{\nu} \le 3072.4\,\mathrm{cm}^{-1}$. The points in the octadecane MFD were subsequently linearly fitted to the endpoints and a linear interpolation was used to calculate the function for every $0.1\,\mathrm{cm}^{-1}$.

The combination of the methylene GFD with the MFDs of the lowest homologues is illustrated in Fig. 4. Figure 4 (panel A) shows the low region ($\tilde{\nu} \le 1200\,\mathrm{cm}^{-1}$) of the MFD from ethanol together with the distributions obtained for three higher homologues by addition of the methylene GFD one, three and five times. The resultant normalized functions are models for the MFDs of propanol, pentanol and heptanol. When the number of frequencies in a molecule is known, the frequencies can be determined from the MFD as described in the Appendix and as indicated by the dashed lines in Fig. 4A. Ethanol has eight frequencies below $1200\,\mathrm{cm}^{-1}$ and the three higher homologues have 11, 19 and 27. The figure illustrates how as expected, the repeated addition of the methylene GFD necessarily leads to

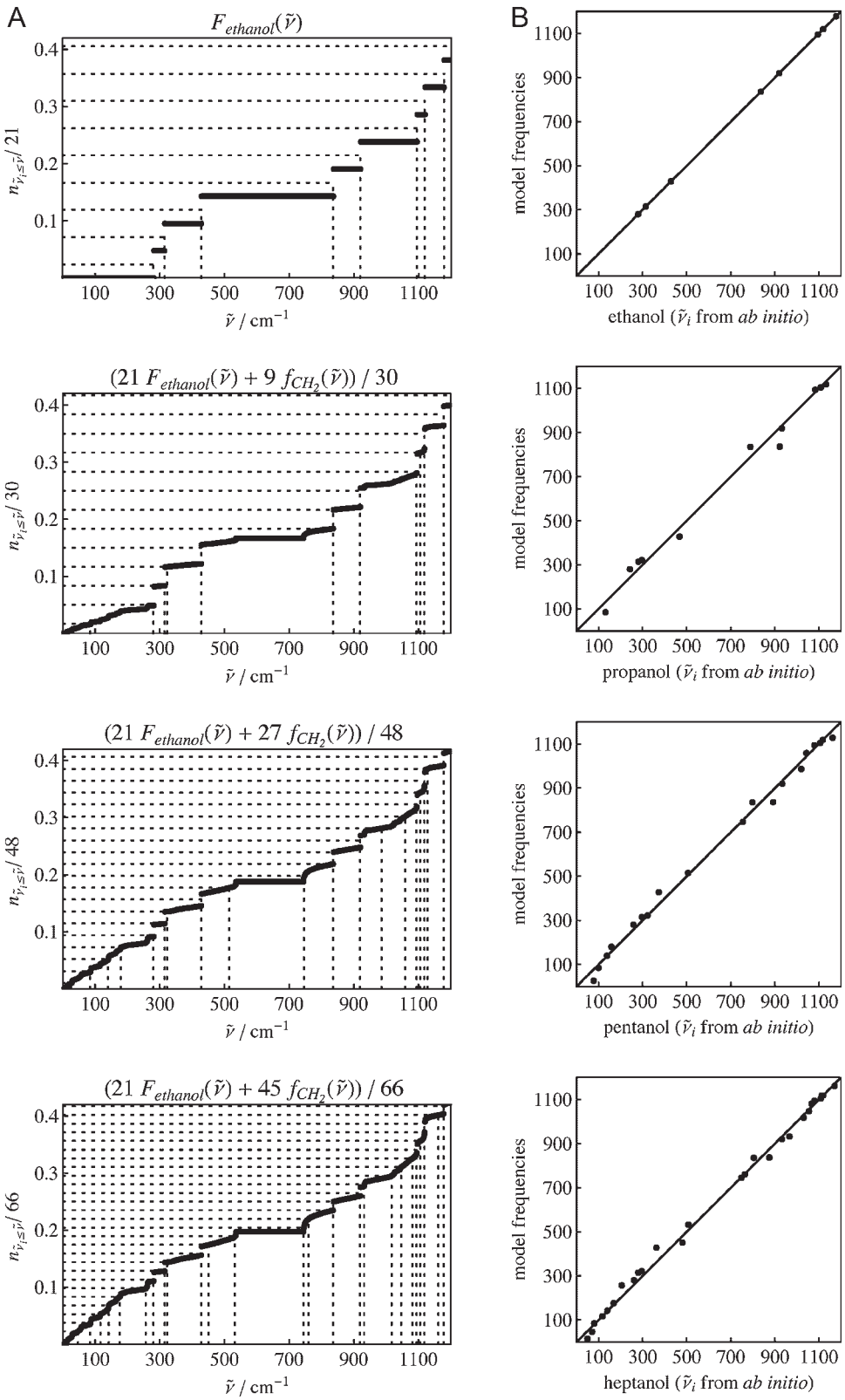

Figure 4. (A) Cumulative molecular frequency distribution for ethanol and model distributions for propanol, pentanol and heptanol. The dashed lines illustrate how the frequencies are determined from the distributions. (B) Comparison of the frequencies found from the distributions in A and frequencies from *ab initio* calculation. The full lines are drawn with a slope of 1

frequencies at progressively lower wavenumbers. Figure 4 (panel B) shows a comparison between the model frequencies found from the MFDs and the frequencies from *ab initio* calculations of propanol, pentanol and heptanol. For ethanol itself the chosen procedure for finding the frequencies from the MFD ensures that the original *ab initio* frequencies are recovered except for the rounding error to the nearest 0.1 cm^{-1}. For the higher homologues the agreement between the *ab initio* and the model frequencies is very satisfactory. The points on the figures in panel B are spread around the line drawn with a slope of one. Similar calculations were done for all compounds in the 11 homologous series in Table S1, and the model and *ab initio* frequencies for the 126 molecules are available in the Supplementary Material.

In estimating the difference between the *ab initio* results and the results of the model calculation, the 11 molecules in the first row are excluded. For the remaining 115 molecules the absolute value of the relative error on the frequencies is 3.2%. This satisfactory calculation of molecular frequencies for the homologous series provides further confirmation of the central hypothesis. However, it is also important to show that this leads to consistent and correct values of the thermochemical functions. That this is indeed the case is shown below.

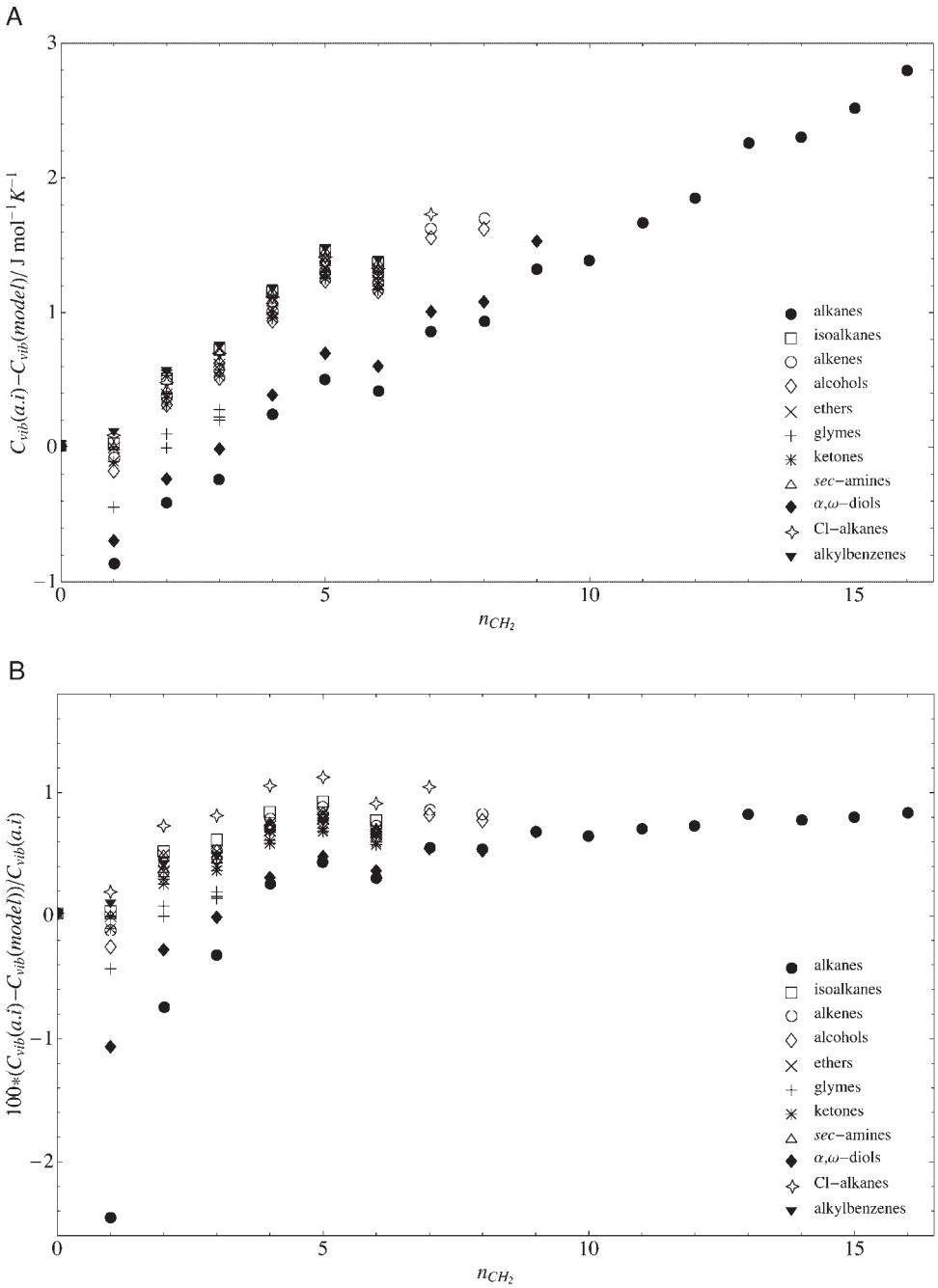

Figure 5. (a) The difference in the vibrational heat capacity ($T = 300$ K) calculated from *ab initio* frequencies, $C_{vib}(a.i.)$ or from model frequencies C_{vib}(model). (b) The relative error on the vibrational heat capacity values

Figure 5(a) shows the difference between the heat capacities of the 126 compounds in Table 1 calculated from the *ab initio* frequencies and from the model frequencies. The temperature chosen is 300 K at which a relatively small number of frequencies at low wavenumbers contribute very significantly to the heat capacity, and which consequently is a stricter test of the model frequencies than heat capacities calculated at a higher temperature. The error on the heat capacity values is between -1 and 3 $J\,K^{-1}\,mol^{-1}$ and a small systematic increase with the chain length is observed. The relative error on the heat capacities at 300 K is shown in Fig. 5(b) and for most of the molecules the relative error of less than 1%. Comparison with the data in the NIST Computational Chemistry Comparison and Benchmark database[27] indicate that the differences for the frequencies obtained here between model calculations and *ab initio* results are within the differences between *ab initio* calculations with different methods or basis sets.

The points from the various homologous series in Fig. 5(a), (b) show a systematic increase with the chain length in the difference between the *ab initio* frequencies and the model frequencies. This difference appears to depend very little on the homologous series and can consequently be eliminated by adjusting the GFD for the methylene group.

As illustrated in Fig. 4 the discrete levels in the MFD of the lowest homologues are partially repeated in the MFDs of the higher homologues. Better frequencies of the higher homologues could undoubtedly be generated with a continuous function for the MFDs of the lowest homologues but how this continuous function is best generated will require further model calculations. The satisfactory agreement between calculations of heat capacities from the model frequencies and from *ab initio* frequencies is a further argument in support of the central hypothesis.

The central part of the hypothesis is formulated in Eqns (A1) and (A2). Ideally a test should be based on several GFDs and so far it has only been possible to estimate a GFD for the methylene group. However, the correct mathematical relationship between frequency distributions and additivity of the thermochemical values can nonetheless be tested as described in the Appendix and expressed in Eqn (A7). The test of this equation is the last step in the model calculations and was done as follows: the values for $\ln[q_{vib}(T)]$, $E_{vib}(T)$ and $C_{vib}(T)$, were calculated for the 11 lowest homologues that is the compounds in the first row of Table 1 from their *ab initio* frequencies. Subsequently the values of the same functions were calculated from the model frequencies of the 11 homologues with one C—$(H)_2(C)_2$ group. The $\ln[q_{vib}(T)]$, $E_{vib}(T)$ and $C_{vib}(T)$ functions were determined in steps of 1 K in the temperature interval 1–1500 K. In the last step, the methylene group additivity values $g^{\ln q_v}_{CH_2}(T)$, $g^{E}_{CH_2}(T)$ and $g^{C}_{CH_2}(T)$ values were determined according to Eqn (A7) by subtraction of the values for the 11 different lower homologues from those of the 11 higher homologues.

This gives 11 different estimates of the group additivity values $g^{\ln q_v}_{CH_2}(T)$, $g^{E}_{CH_2}(T)$ and $g^{C}_{CH_2}(T)$ in the chosen temperature interval. These results are shown in Figs S3, S4 and 6. It is remarkable that the 11 different estimates of all three functions in the whole temperature interval are indistinguishable. For the heat capacity group additivity values, it is possible to compare with the values used for the C—$(H)_2(C)_2$ group in the NIST database.[28] These are shown as the point in Fig. 6. The agreement with the results based on the model calculations is satisfactory, in particular considering that the methylene GFD ($f_{CH_2}(\tilde{\nu})$) used in the model calculations is based on unscaled frequencies. It is a very strong argument in favour of the central hypothesis that the addition of the methylene GFD to any MFD is proven to give consistent values for $g^{\ln q_v}_{CH_2}(T)$, $g^{E}_{CH_2}(T)$ and $g^{C}_{CH_2}(T)$ in a wide temperature interval as well as a constant contribution to the zero-point energy.

In the model calculations the additivity of the heat capacity appears to extend to much lower temperatures than what is found when Benson schemes are constructed from experimental values. Several things may contribute to this difference, in particular the MFDs which have been used in the model calculations are based on a single conformer. Since the relative concentrations of the different conformers will depend on the

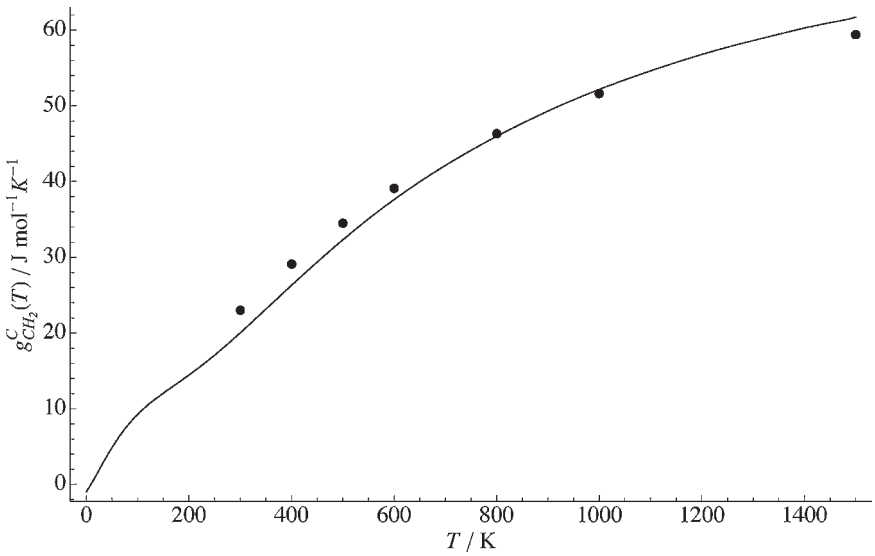

Figure 6. Methylene group additivity values: $g^{C}_{i}(T)$ from 11 different model calculations. The points are the group additivity values from Reference [18] for C–$(H)_2(C)_2$

temperature the additivity observed at low temperatures in the model calculations cannot be expected for experimental heat capacities. The harmonic oscillator stiff rotor approximation will also give the largest deviations when the thermochemical value is most dependent on the lowest frequencies.

CONCLUSIONS

The compounds chosen for this study include such a variety of functional groups that their frequency distributions are representative for wider selections of organic compounds. Thus the effect of an additional methylene group on their frequency distribution must be a general effect of that group.

The hypothesis which has been tested states that MFDs are linearly dependent on group distributions, and this leads directly to the required group additivity of the zero-point energy. However, it must also be shown to lead to group additivity of the vibrational excitation energy and the heat capacity. The model calculations show that this is indeed the case. Thus the hypothesis is confirmed by a number of observations: (1) it is shown that a methylene group contributes linearly to the zero-point energy and that this linearity is not a consequence of the two high C—H stretch frequencies associated with each methylene group. (2) In agreement with the hypothesis, isomers with the same groups have nearly identical frequency including frequencies at low wavenumbers where spectroscopic group frequencies are absent. (3) It is shown for the methylene group that the way of calculating frequencies from linear combinations of frequency distributions, leads to frequencies which compares well with *ab initio* frequencies. (4) It is proven that frequencies derived from the 11 combinations of 11 different MFDs and an estimated methylene GFD leads to additivity for the logarithm of the partition function, the vibrational excitation energy and the heat capacity of the methylene group in a wide temperature range, and that the result for the heat capacity are in good agreement with experimental values. Thus the mathematical relationship expressed by Eqn (A7) is proven. This provide at strong indication for the validity of Eqns (A1) and (A2). It must also be mentioned that the observation of striking similarity of the MFDs of homologous series which lead to this work in itself supports the hypothesis.

Hitherto the foundation of Benson additivity on fundamental molecular properties has been limited to the electronic energy as expressed by the atoms-in-molecules theory. The validity of Benson additivity for the heat capacity and the entropy has been based on the consistent agreement between experimental (or *ab initio*) results and the values calculated by additivity. The analyses and hypothesis presented here shows that spectroscopic group frequencies are an insufficient explanation and states that additivity is not the result of cancellation-of-errors or some such phenomenon. The confirmation of the hypothesis shows that the additivity of functions which depends on molecular vibrations depends on a fundamental property of the organic molecules. Thus it provides important support for the validity of Benson additivity.

Since Benson additivity has also been shown to be valid for transition states, the MFDs of transition states should also be linear functions of GFDs.[6–12] This is the subject of a forthcoming publication.

COMPUTATIONAL METHODS

Frequencies were calculated with the Gaussian 03 programmes at the MP2(FC)/6-31G* level.[29] For all the compounds an all anti-conformation of the aliphatic chain was used. All other calculations were done with Mathematica version 5.2 or 6.0

SUPPLEMENTARY MATERIAL

In Table S1 the 126 compounds from 11 homologous series included in the study are listed. Figures S1 and S2 are Plots of the zero-point energy of the 126 compounds in the regions $600\,cm^{-1} < \tilde{\nu} \leq 2000\,cm^{-1}$ and $\tilde{\nu} < 600\,cm^{-1}$ shown as a function of the number of CH_2-groups. Figures S3 and S4 show the methylene group additivity values $g^{\ln q_v}_{CH_2}(T)$ and $g^{E}_{CH_2}(T)$ from 11 different model calculations. *Ab initio* and model frequencies for the 126 compounds listed in Table S1 are provided as are the *ab initio* frequencies for the four isomeric dodecenes in Table 2.

Acknowledgements

I am grateful to Steen Hammerum and Theis Sølling for assistance with the Gaussian calculations and for valuable comments to earlier versions of the manuscript. Jørn Børling Olsson is thanked for his patience when answering questions in elementary mathematics.

REFERENCES

[1] S. W. Benson, J. H. Buss, *J. Chem. Phys.* **1958**, *29*, 546–572.
[2] S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, R. Walsh, *Chem. Rev.* **1969**, *69*, 279–324.
[3] S. W. Benson, *Thermochemical Kinetics*, John Wiley, New York, 1976.
[4] N. Cohen, S. W. Benson, *Chem. Rev.* **1993**, *93*, 2419–2438.
[5] I. Fishtik, R. Datta, *J. Phys. Chem. A* **2003**, *107*, 6698–6707.
[6] T. N. Truong, *J. Chem. Phys.* **2000**, *113*, 4958–4964.
[7] S. Zhang, T. N. Truong, *J. Phys. Chem. A* **2003**, *107*, 1138–1147.
[8] R. Sumathi, H.-H. Carstensen, W. H. Green, *J. Phys. Chem. A* **2001**, *105*, 6910–6925.
[9] R. Sumathi, H.-H. Carstensen, W. H. Green, *J. Phys. Chem. A* **2001**, *105*, 8969–8984.
[10] R. Sumathi, H.-H. Carstensen, W. H. Green, *J. Phys. Chem. A* **2002**, *106*, 5474–5489.
[11] R. Sumathi, W. H. Green, *Theor. Chem. Acc.* **2002**, *108*, 187–213.
[12] M. Saeys, M. F. Reyniers, V. van Speybroeck, M. Waroquier, G. B. Marin, *Chemphyschem* **2006**, *7*, 188–199.
[13] R. F. W. Bader, *Atoms in Molecules*, Clarendon, Oxford, 1990.
[14] R. F. W. Bader, P. L. A. Popelier, T. A. Keith, *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 620–631.
[15] R. F. W. Bader, D. Bayles, *J. Phys. Chem.* **2000**, *104*, 5579–5589.
[16] H. J. Bernstein, A. D. E. Pullin, *J. Chem. Phys.* **1953**, *21*, 2188–2195.
[17] S. Arimoto, *Phys. Lett. A* **1985**, *113*, 126–132.
[18] J. M. Schulman, R. L. Disch, *Chem. Phys. Lett.* **1985**, *113*, 291–293.
[19] M. Rahal, M. Hilali, A. El Hammadi, M. El Mouhtadi, A. El Hajbi, *J. Mol. Struct. (THEOCHEM)* **2001**, *572*, 73–83.
[20] P. Winget, T. Clark, *J. Comput. Chem.* **2003**, *25*, 725–733.
[21] D. Baric, Z. B. Maksic, R. Vianello, *J. Mol. Struct. (THEOCHEM)* **2004**, *672*, 201–207.
[22] P. Debye, *Ann. Phys.* **1912**, *39*, 789–839.
[23] T. L. Hill, *Statistical Thermodynamics*, Addison-Wesley, Reading, 1960.
[24] N. W. Ashcroft, N. D. Mermin, *Solid State Physics*, Holt, Rinehart and Winston: New York, 1976.
[25] H. S. Bu, S. Z. D. Cheng, B. Wunderlich, *J. Phys. Chem.* **1987**, *91*, 4179–4188.
[26] K. Loufakis, B. Wunderlich, *J. Phys. Chem.* **1988**, *92*, 4205–4209.

[27] *NIST Computational Chemistry Comparison and Benchmark Database*, NIST Standard Reference Database Number 101 (Ed.: R. D. Johnson, III), 2005 (http://srdata.nist.gov/cccbdb) National Institute of Standards and Technology: Gaithersburg, MD.

[28] S. E. Stein, R. L. Brown, Structures and properties group additivity model, in *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69 (Eds: P. J. Linstrom, W. G. Mallard) National Institute of Standards and Technology, Gaithersburg, MD, 2005, p. 20899 (http://webbook.nist.gov).

[29] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.

APPENDIX

The hypothesis suggested here for the additivity of the vibrational excitation energy, the vibrational heat capacity and the logarithm of the vibrational partition function can be formulated as follows:

The cumulative *group* frequency distributions (GFDs): $f_1(\tilde{\nu})$, $f_2(\tilde{\nu})$, ..., $f_i(\tilde{\nu})$ are continuous functions for $0 < \tilde{\nu} \leq \tilde{\nu}_{max}$ with $f_i(0) = 0$, all with a maximum at $f_i(\tilde{\nu}_{max}) = 1$ and with $\tilde{\nu}_{max} \leq 4000\,cm^{-1}$.

From these functions the cumulative *molecular* frequency distributions (MFDs) can be constructed by linear combinations:

$$\begin{aligned} F_1(\tilde{\nu}) &= \frac{1}{k_1}\sum_i n_{1,i} w_i f_i(\tilde{\nu}) \\ F_2(\tilde{\nu}) &= \frac{1}{k_2}\sum_i n_{2,i} w_i f_i(\tilde{\nu}) \\ &\cdot \\ &\cdot \\ F_j(\tilde{\nu}) &= \frac{1}{k_j}\sum_i n_{j,i} w_i f_i(\tilde{\nu}) \end{aligned} \tag{A1}$$

$n_{j,i}$ refer to the number of ith groups in the jth molecule and w_i is the number of frequencies in the ith group. $1/k_j$ are normalization factors which are given by

$$\begin{aligned} k_1 &= n_{1,1}w_1 + n_{1,2}w_2 + \ldots + n_{1,i}w_i \\ k_2 &= n_{2,1}w_1 + n_{2,2}w_2 + \ldots + n_{2,i}w_i \\ &\cdot \\ &\cdot \\ k_j &= n_{j,1}w_1 + n_{j,2}w_2 + \ldots + n_{j,i}w_i \end{aligned}$$

The frequencies of the molecules are sets A_1, A_2, ..., A_j of lengths given by k_j which is equal to $3N_j - 6$, where N_j is the number of atoms in the jth molecule. In the model calculation it has been used that when the lengths of the sets are known, the elements (i.e. the frequencies) can be calculated from the cumulative MFDs $F_1(\tilde{\nu})$, $F_2(\tilde{\nu})$, ..., $F_j(\tilde{\nu})$. The procedure is equivalent to a quantile analysis where the inverse of the cumulative distribution function (the probability distribution) is used. k_j fractions between $1/(2k_j)$ and $1 - 1/(2k_j)$ are found on the ordinate of the MFD and the frequencies are then determined from the function. The procedure is illustrated by the dashed lines in Fig. 4.

As outlined in the main text, the sum z_j of the elements in the set A_j can then be calculated from:

$$z_j = n_{j,1}w_1 < \tilde{\nu}_1 > + n_{j,2}w_2 < \tilde{\nu}_2 > + \ldots + n_{j,i}w_i < \tilde{\nu}_i >$$

where $<\tilde{\nu}_i>$ is the mean of the GFD functions: $< \partial[f_i(\tilde{\nu})]/\partial\tilde{\nu} >$ in the interval $0 < \tilde{\nu} \leq \tilde{\nu}_{max}$. The zero-point energy of jth molecule is equal to z_j multiplied by the constant: $hc/2k_B$.

Additivity of a thermochemical property requires that when the associated function $X(\tilde{\nu},T)$ is mapped on the sets A_1, A_2, ..., A_j the sums for all the sets can be written as

$$\begin{aligned} \sum_{l=1}^{l=k_1} X(\tilde{\nu}_l \in A_1, T) &= n_{1,1}g_1^X(T) + n_{1,2}g_2^X(T) + \ldots + n_{1,i}g_i^X(T) \\ \sum_{l=1}^{l=k_2} X(\tilde{\nu}_l \in A_2, T) &= n_{2,1}g_1^X(T) + n_{2,2}g_2^X(T) + \ldots + n_{2,i}g_i^X(T) \\ &\cdot \\ &\cdot \\ \sum_{l=1}^{l=k_j} X(\tilde{\nu}_l \in A_j, T) &= n_{j,1}g_1^X(T) + n_{j,2}g_2^X(T) + \ldots + n_{j,i}g_i^X(T) \end{aligned} \tag{A2}$$

where $\tilde{\nu}_l$ refer to the elements in the sets A_j and $g_i^X(T)$ is the group additivity value for the X-property of the ith group. The coefficients $n_{j,i}$ in Eqns (A1) and (A2) are identical. In the present work $X(\tilde{\nu},T)$ is either the logarithm of the vibrational partition function (Eqn (A3)), the vibrational excitation energy function (Eqn (A4)) or the vibrational heat capacity function (Eqn (A5)), and the coefficients $g_i^X(T)$ are the group additivity values for the logarithm of the vibrational partition function $g_{CH_2}^{\ln q_v}(T)$, the vibrational excitation energy $g_i^E(T)$ or the vibrational heat capacity $g_i^C(T)$.

$$\ln[q_{vib}(\tilde{\nu},T)] = \ln\{1/[1 - \exp(-hc\tilde{\nu}/k_BT)]\} \tag{A3}$$

$$E^*_{vib}(\tilde{\nu},T) = N_A k_B hc\tilde{\nu}/[\exp(hc\tilde{\nu}/k_BT) - 1] \tag{A4}$$

$$C_{vib}(\tilde{\nu},T) = N_A k_B (hc\tilde{\nu}/k_BT)^2 \exp(hc\tilde{\nu}/k_BT) / [\exp(hc\tilde{\nu}/k_BT) - 1]^2 \tag{A5}$$

The procedure described under 'Model calculations' for testing the consistency of this way of finding thermochemical parameters is based on Eqn (A2) which for two homologues reduces to Eqn 6, where X_M and X_{M+CH_2} are the values of a thermochemical property for a molecule and its higher homologue:

$$X_M(T) = \sum_{l=1}^{k_M} X(\tilde{\nu}_l \in A_M, T) \tag{A6a}$$

$$X_{M+CH_2}(T) = \sum_{l=1}^{k_M+9} X(\tilde{\nu}_l \in A_{M+CH_2}, T) \tag{A6b}$$

X_M was found from the set of frequencies A_M obtained from the *ab initio* calculations. For the higher homologue the set of frequencies A_{M+CH_2} were found from the weighted sum of frequency distributions $[k_M F_M(\tilde{\nu}) + 9f_{CH_2}(\tilde{\nu})]/(k_M + 9)$, where k_M is the number of normal modes for the lower homologue and $f_{CH_2}(\tilde{\nu})$ is the GFD of a methylene group derived from the octadecane frequencies. The group additivity value $g^X_{CH_2}(T)$ for a methylene group was subsequently calculated from the following equation:

$$g^X_{CH_2}(T) = X_{M+CH_2}(T) - X_M(T) \quad \text{(A7a)}$$

$$= \sum_{I=1}^{k_M+9} X(\tilde{\nu}_I \in A_{M+CH_2}, T) - \sum_{I=1}^{k_M} X(\tilde{\nu}_I \in A_M, T) \quad \text{(A7b)}$$