From a Network of Computed Reaction Enthalpies to Atom-Based Thermochemistry (NEAT)

Attila G. Császár* and Tibor Furtenbacher[a]

Abstract: A simple and fast, weighted, linear least-squares refinement protocol and code is presented for inverting the information contained in a network of quantum chemically computed 0 K reaction enthalpies. This inversion yields internally consistent 0 K enthalpies of formation for the species of the network. The refinement takes advantage of the fact that the accuracy of computed enthalpies depends strongly on the quantum-chemical protocol employed for their determination. Different protocols suffer from different sources of error; thus, the reaction enthalpies computed by them have "random" residual errors. Since it is much more natural for quantum-chemical energy and enthalpy results, including reaction enthalpies, to be based on the electronic ground states of the atoms and not on the historically preferred elemental states, and since these two possible protocols can be converted into each other straightforwardly, it is proposed that first-principles thermochemistry should employ the ground electronic states of atoms. In this scheme, called atom-based thermochemistry (AT), the enthalpy of formation of a gaseous compound corresponds simply to the total atomization energy of the species; it is always positive, and it reflects the bonding strength within the molecule. The inversion protocol developed and based on AT is termed NEAT, which represents the fact that the protocol proceeds from a *n*etwork of computed *re*action *e*nthalpies toward *at*om-based *t*hermochemistry, most directly to atom-based enthalpies of formation. After assembling a database that consisted of 361 ab initio reactions and reaction enthalpies involving 188 species, collected from 31 literature sources, the following dependable 0 K atom-based enthalpies of formation, $\Delta_f H_0^{AT}$, all in kJ mol^{-1}, have been obtained by means of NEAT: $H_2 = 432.07(0)$, $CH = 334.61(15)$, $NH = 327.69(25)$, $OH = 425.93(21)$, $HF = 566.13(31)$, $CO = 1072.08(28)$, $O_2 = 493.51(34)$, $CH_2 = 752.40(21)$, $H_2O = 918.05(20)$, $HO_2 = 694.53(32)$, $CO_2 = 1597.77(40)$, $CH_3 = 1209.64(29)$, $NH_3 = 1157.44(33)$, $C_2H_2 = 1625.78(40)$, and $CH_4 = 1641.68(40)$, in which the uncertainty values given in parentheses represent 95% confidence intervals. The average deviation of these values from the well-established active thermochemical tables (ATcT) values is a mere 0.25 kJ mol^{-1}, with a maximum deviation of 0.7 kJ mol^{-1}. This shows that the use of a large number of ab initio reaction enthalpies within a NEAT-type protocol has considerable advantages over the sequential utilization of the ab initio information.

Introduction

There are several databases[1–9]—unfortunately often with rather different internal and overall accuracies—that contain thermochemical information, most importantly temperature-dependent enthalpies of formation and heat capacities. The deposited thermochemical data are important in such diverse scientific and engineering areas as understanding the chemical and physical basis of star formation, the chemistry of the atmospheres of planets and moons, feasibility of chemical reactions,[10] combustion research and the development of combustion engines, and process safety analysis in the chemical industry.[11] It is therefore highly useful that there be a straightforward correspondence between the enthalpies of formation of species deposited in databases and the reaction enthalpies involving these species needed for most specific applications. Even if just the gas-phase enthalpies of formation of a limited number of compounds are

[a] Prof. Dr. A. G. Császár, Dr. T. Furtenbacher
Institute of Chemistry, Laboratory of Molecular Spectroscopy
Loránd Eötvös University, 1518 Budapest 112, P.O. Box 32 (Hungary)
Fax: (+36)1-372-2592
E-mail: csaszar@chem.elte.hu

available over a limited temperature range, based on a number of reasonable assumptions, like the simplification of the reactants (using, for example, the analogue reaction hypotheses[12]), the at least partial neglect of certain solvent, pressure, and temperature effects, and the use of certain bond-energy and/or Benson protocols,[13,14] one can arrive at thermochemical quantities of sufficient accuracy for several of the complex applications mentioned. Nevertheless, improving the accuracy and extending the range of thermochemical data deposited in databases should be an ongoing effort from as many directions as possible.

Over the last decade or so there has been a surge of interest in the determination of accurate thermochemical data, which can be achieved either experimentally or by means of what can be called 'estimation techniques' of different sophistication. The modern blossom of thermochemistry is due to several factors besides the genuine interest in its quantities. First, there has been considerable improvement in the experimental techniques needed to produce thermochemical quantities of interest, ranging from the availability of new and reliable calorimeters to sophisticated measurements using advanced laser spectroscopy (e.g., refs. [15] and [16]). Since our own efforts reported herein constitute a basically computational contribution to the field, these experimental developments are mentioned but are not treated further. Second, there has been a phenomenal progress in the development of new estimation techniques.[17] Perhaps the most promising among these for smaller molecules is first-principles (ab initio) thermochemistry, which utilizes the considerable recent developments in electronic structure and nuclear motion techniques and the hardware on which they are running. The accuracy of the computed first-principles enthalpies of formation is helped by the use of properly chosen reaction schemes. The proper construction and the utility of such schemes has been clarified recently,[18] thus offering improved capabilities for computational thermochemistry of larger systems. The related theoretical developments are reviewed below in more detail. Third, the traditional, historical view, namely, that thermochemistry should be independent from atomic information and that the standard enthalpy of formation of a molecule is best defined if it corresponds to the enthalpy change upon formation from the elements in their standard states has been questioned.[19–23] According to the traditional convention, the standard enthalpy of formation of all elements is zero, independent of the temperature. This convention became a considerable hindrance in modern times since, for example, the elements in their standard states may form gases, liquids, or solids of different constitution. Furthermore, as argued by Sanderson,[19,20] the explanation of chemistry and especially thermochemistry on the basis of atomization enthalpies has obvious and considerable advantages. Another argument favoring what one might call atom-based thermochemistry (AT) over an element-based one comes from von Szentpály,[22,23] who recently showed several examples in which the change to AT revealed highly useful linear relationships, which exhibit predictive power, between the atomization enthalpies of

solids and molecules. Note that, in this respect, quantum-chemical techniques are bound to have an ever greater role in thermochemistry. This provides a further argument in favor of AT: it is much more natural to reference the results of ab initio energy and enthalpy computations to the ground electronic states of atoms rather than to standard states of the elements. For example, the C atom can be subjected to straightforward, high-accuracy electronic structure computations, but graphite, upon which $\Delta_f H°(C_{gas})$ is historically based, cannot be.

The focal-point analysis (FPA) approach[24,25] to electronic structure theory was developed in the very early 1990s to meet the stringent accuracy requirements of high-resolution molecular spectroscopy. The most distinguishing design features of the FPA approach are its composite nature and the fact that it facilitates the determination of reasonable uncertainties for the computed quantities (mostly relative energies). Of course, the FPA approach was immediately used for the purposes of conformational analysis[26] and also for computational thermochemistry.[27] For the latter, the accuracy requirements were usually considerably more relaxed, especially at the time of those early publications. The composite nature of FPA means that it considers not only valence electronic energies and their differences but also effects that are usually considered to be 'small': core correlation,[28] a variety of relativistic effects to move beyond non-relativistic electronic structure theory,[29–31] and diagonal Born–Oppenheimer corrections (DBOC)[32] to move beyond the BO approximation and thus allow one to gauge the accuracy of the traditional (non)relativistic BO results. The FPA approach requires the layout of a two-dimensional grid that separates one-electron and many-electron approximations inherent in most electronic structure theories that can be applied to a wide range of molecules. After extrapolating to the appropriate one- and many-particle limits, and assuming additivity for the different 'small' corrections, one arrives at the final (relative) energies and, just as importantly, their uncertainties. Over the years, the FPA approach found a large number of applications.[33–37] In spectroscopy, for example, it allowed the determination of an adiabatic potential energy surface for the water molecule, which yielded, for the first time, first-principles, variationally determined rovibrational energy levels accurate to better than 1 cm^{-1} in an average sense.[37] In thermochemistry, it allowed, for example, the anchoring of the proton affinity scale of molecules in the gas phase[34] and it also proved to be important in an IUPAC effort to determine accurate enthalpies of formation of small radicals.[2] The apparent success of the FPA approach resulted in the belief that "[ab initio] theory achieved a status whereby it can surpass or at least match the accuracy of (most) experiments".[35]

One of the distinctive design features of the FPA approach is that it does not specify particular levels of electronic structure theory; it just calls for the extension of the study to the highest levels of theory affordable. Consequently, the success of the composite FPA approach has been exploited by several groups who came up with restrictive but

"black-box"-type schemes that did specify the levels of theory to be used during the computations. Many of the high-quality ab initio thermochemical approaches in use today, including the Gaussian n ($n=1$–4),[38–41] the Weizmann n ($n=1$–4),[42–45] the HEAT,[46–48] and the correlation-consistent composite (ccCA) approaches,[49,50] and, to some extent, the complete basis set (CBS)[51,52] and the multicoefficient correlation (MCCM)[53–55] methods, can be considered approaches with fixed prescriptions about the levels of theory to be used within the generic composite FPA scheme. These "black-box" approaches are highly useful, though usually they do not provide a straightforward way to estimate the residual uncertainties of the computed quantities.

However accurate the sophisticated first-principles thermochemical computations have become, it should be admitted that over the last few years "experiment" took back its leading role over "theory". The reason is that Ruscic and co-workers could substantially improve the overall accuracy of the "experimental" enthalpies of formation by means of the active thermochemical tables (ATcT) approach.[56–59] Ruscic et al. demonstrated convincingly that by moving away from a sequential utilization of thermochemical information and by using all available experimental information in a grand weighted linear least-squares approach, one can come up with more accurate, more reliable, and internally more consistent thermochemical values, most importantly enthalpies of formation. The linear inversion procedure that underlies the ATcT approach is basically the same as that which has been used by spectroscopists[60–62] for a long time to invert the experimentally measured transition information and determine rovibrational energy levels with improved uncertainties. A particular realization of the spectroscopic weighted linear least-squares refinement procedure by us was termed MARVEL,[61,63,64] meaning, *m*easured *a*ctive *r*otational–*v*ibrational *e*nergy *l*evels. It has been used, for example, to obtain highly accurate rotational–vibrational energy levels of water isotopologues[65] and should help to come up with accurate molecular line lists for larger species as well.

We had several purposes with the present study. First, we wanted to extend the capabilities of the MARVEL approach and code so that it can be used for the benefit of first-principles thermochemistry. What is involved and how this was achieved is described in the Theoretical Background section below. The thermochemical protocol and code developed was termed NEAT, in reference to the fact that we are moving from a *n*etwork of computed reaction *e*nthalpies toward *a*tom-based *t*hermochemistry. Second, we wanted to collect accurate reaction enthalpy information computed ab initio. By utilizing this collected information, we were able to investigate the true accuracy of the different FPA and FPA-based protocols. Our findings are discussed in the Results and Discussion section. Third, we wanted to investigate and demonstrate what one could gain by shifting the reference states in thermochemistry from elements to atoms. It is proposed that ab initio thermochemistry should be further developed based on a scale determined by the lowest electronic states of the atoms rather than the "historical" standard states of the elements. The two scales can be related to each other in a straightforward manner, but for the purposes of error analysis of the quantum-chemical results, the proposed switch in the scale seems preferable. Related results are discussed also in the Results and Discussion section. The paper ends with a Conclusion.

Theoretical Background

Hereby we present a short summary of the simple and fast weighted linear least-squares theory behind the NEAT protocol. The applied method and the code, written in C++, are built upon the MARVEL theory,[61] which has been applied successfully in the area of high-resolution molecular spectroscopy. Similarly to MARVEL, the NEAT protocol took advantage of some of the previously reported results of Flaud et al.,[60] Ruscic et al.,[56] and Watson.[66]

Determination of the enthalpy of formation of a molecule from a network of (computed) reaction enthalpies is based principally upon the following steps: 1) Collect ab initio reaction enthalpy data (values and uncertainties) into a single database that employs a well-defined and universal format. 2) Select the so-called unambiguous reactions. We call a reaction unambiguous if it contains only one unknown species (reactant or product) that is not defined by an earlier unambiguous reaction within the database. 3) Use the selected reactions for setting up a vector that contains all the chosen reaction enthalpies, another one that comprises the requested enthalpies of formation, and a sparse matrix, usually called the design matrix, which basically describes the stoichiometry within the given reaction ensemble. 4) Solve the resulting set of overdetermined linear equations that correspond to the chosen pair of vectors and the inversion (design) matrix. During solution of the set of linear equations, uncertainties in the calculated reaction enthalpies can be incorporated, which results in uncertainties for the determined enthalpies of formation.

Database: The principal input for the NEAT code is a list of N_r-calculated reaction enthalpies. The data to be deposited include values of the (computed) reaction enthalpies, $\Delta_r H_0(i)$, together with their associated uncertainties, δ_i ($i=1, ..., N_r$), coupling N_f enthalpies of formation with values $\Delta_r H_0(j)$ ($j=1, ..., N_f$). The relation between reaction enthalpy and enthalpies of formation for a sample gaseous reaction [Eq. (1)] is shown in Equation (2):

$$x R_1(g) + y R_2(g) \rightarrow a P_1(g) + b P_2(g) \qquad (1)$$

$$\Delta_r H_0 = [\{a\,\Delta_f H_0(P_1) + b\,\Delta_f H_0(P_2)\} - \{x\,\Delta_f H_0(R_1) + y\,\Delta_f H_0(R_2)\}] \qquad (2)$$

In Equation (1), R_1 and R_2 are the reactants, P_1 and P_2 are the products, and x, y, a, and b are stoichiometric factors. Figure 1 shows a pictorial representation of Equation (1). For the sake of simplicity, during this study the temperature was chosen as $T=0$ K. Other temperature choices are, of course, possible; most common among these is $T=298.15$ K, but would lead to further complications in the assessment of the accuracy of the computed reaction enthalpies.

After some (admittedly incomplete) search of the literature, 31 publications[9,18,27,33–36,44,46,48,67–87] were identified that directly reported 0 K ab initio reaction enthalpies of interest to us. For some tests, the database formed was augmented with enthalpies of formation computed with the help of the Gaussian 03 program package[88] and its built-in, one-click facility for determining 0 K enthalpies for species of interest according to the G1,[38] G2,[39] G2MP2,[89] G3,[40] and W1U[90] protocols.

To help critical evaluation and tracking of the computed reaction enthalpies as well as the NEAT enthalpies of formation even after a prolonged time, the dataset compiled must also contain information about the sources of the computed results. Therefore, besides the obvious data (i.e., the actual reactions and the computed reaction enthalpies and their uncer-

$$.... \quad R_1 \quad ... \quad R_2 \quad ... \quad P_1 \quad ... \quad P_2 \quad ...$$

$$.... \quad -x \quad ... \quad -y \quad ... \quad a \quad ... \quad b \quad ... \qquad \begin{matrix} \vdots \\ \Delta_f H_0(R_1) \\ \vdots \\ \Delta_f H_0(R_2) \\ \vdots \\ \Delta_f H_0(P_1) \\ \vdots \\ \Delta_f H_0(P_2) \\ \vdots \end{matrix} \quad = \quad \begin{matrix} \cdot \\ \cdot \\ \cdot \\ \Delta_r H_0 \\ \cdot \\ \cdot \\ \cdot \end{matrix}$$
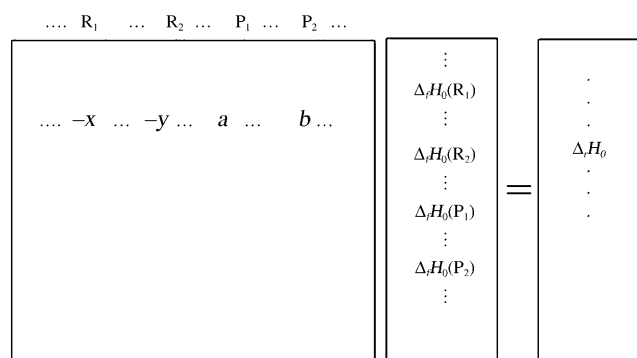
Figure 1. Pictorial representation of the NEAT protocol, which, after an inversion, leads to 0 K enthalpies of formation ($\Delta_f H_0$) from computed reaction enthalpies ($\Delta_r H_0$).

tainties), the same dataset gives each data source a tag based on the year of publication and the names of the authors. This tag is used throughout the NEAT procedure to provide a unique identifier of the primary data sources. Each computed reaction enthalpy has an individual label that involves this tag and a sequence number of the reaction enthalpy in the original publication. The complete dataset built, including the tags, is given in Table S1 of the Supporting Information.

Enlargement of the database, should further or revised computed data become available, involves simply adding extra lines to the reaction enthalpy file presented as the Supporting Information to this work. Thus, we plan to use the dataset in an active way,[56] which means that rerunning the NEAT procedure is envisioned on a regular basis. This will ensure an instant and correct propagation of the consequences of the new computed data throughout the established network.

Determination of enthalpies of formation: Since the relation between reaction enthalpies and enthalpies of formation can be written in a similar way as in Equation (2) for each reaction, the reactions in the database can be summarized in a matrix equation [Eq. (3)]:

$$aX = Y \qquad (3)$$

in which the vector Y (with dimension N_r) contains the N_r-computed reaction enthalpies and the vector X (with dimension $N_f - NB$, in which NB is the number of the selected base elements) contains the $\Delta_f H$ values as results. The elements of the input matrix a, with dimension $N_r \times (N_f - NB)$, are the signed stoichiometric factors, and the sign depends on whether the species is a reactant, in which case its stoichiometric factor is negative, or a product, when its stoichiometric factor is positive.

The weighted least-squares solution of Equation (3) is obtained by solving[56] Equation (4):

$$AX = B \qquad (4)$$

in which $A = a^T g a$, $B = a^T g Y$, and $g_i = 1/\delta_i^2$, in which δ_i is the uncertainty attached to the ith reaction enthalpy. The dimension of the extremely sparse matrix A is $(N_f - NB) \times (N_f - NB)$, and Equation (4) is a simple system of linear equations that can be solved in a straightforward way.

Uncertainties of computed reaction enthalpies: Attachment of dependable uncertainties to computed reaction enthalpies is more or less straightforward within the composite FPA approach. The results of the present study prove that in the studies performed using this approach the attached uncertainties appear to be conservative in all cases studied. However, it is rather unclear how to attach an uncertainty to a reaction enthalpy computed by means of a model chemistry scheme. One possible solution was suggested by Klopper et al.,[77] who, at a particular level of electronic structure theory, came up with the formula $\sigma = n \times 0.13$ kJ mol^{-1} for the total statistical uncertainty of their computed atomization ener-

gies, in which n is the number of electrons in the valence shell. These are the uncertainties that have been entered into our database along with the computed atomization energies from this source (09KlRuTeBi).[77] In most other cases, perhaps the only way to come up with a reasonable uncertainty is to assign a mean uncertainty to the computed reaction enthalpies based on the performance of the model chemistry for a number of investigated systems. Of course, this way one could grossly fail for many of the reactions that deviate from the average accuracy of the model chemistry either in a positive or a negative sense. Thus, to achieve an internally consistent dataset, adjustment of the uncertainties for many of the computed reaction enthalpy entries is mandatory. The 2σ uncertainties attached originally by us to the computed reaction enthalpies can be checked in Table S1 of the Supporting Information.

In our previous spectroscopic studies that employed MARVEL,[61,64,65] it turned out that the robust reweighting technique[66] is especially well suited for the adjustment of the measured uncertainties of the spectroscopic transitions and self-consistency can be achieved. Based on this experience, we preferred to use the same technique for adjusting the uncertainties of the computed reaction enthalpies. In the robust reweighting technique, the adjustment formula is very simple and one uses the expression [Eq. (5)]:

$$\tilde{g}_i = \frac{1}{\delta_i^2 + \alpha \Delta_i^2} \qquad (5)$$

to increase the values of the uncertainties. In Equation (5), Δ_i is the difference between the original database entry and its prediction and α is a positive number, $\alpha \leq {}^1/_3$,[66] chosen for the given dataset. Smaller values of α mean that the adjustment process proceeds somewhat slower but smoother. Adjustment of the uncertainties in the NEAT protocol is done iteratively, based on Equation (5), and the iteration can be stopped when the quantity [Eq. (6)]:

$$\sum_i \frac{\tilde{g}_i \Delta_i^2}{N_r - N_f + NB} \qquad (6)$$

becomes as close to 1 as desired.

The input uncertainties present in the dataset were chosen to correspond to 2σ. Thus, before using them in the iterative reweighting scheme as δ_i, they were divided by a factor of 2. The recomputed, robustly reweighted uncertainties of the reaction enthalpies are given in Table S2 of the Supporting Information.

Uncertainties of the enthalpies of formation: After executing the robust reweighting procedure, one obtains an internally consistent dataset, in which each reaction enthalpy can be reproduced by the calculated enthalpies of formation within the given uncertainty limits. In a self-consistent database the uncertainties of the enthalpies of formation can be computed as [Eq. (7)]:

$$\varepsilon_j = t \sqrt{A_{jj}^{-1}} \qquad (7)$$

For $t = 2$, ε_j represents an approximately 95% confidence limit for the uncertainty of the enthalpies of formation. These are the uncertainties that are reported in Table 1 as the primary results of the present investigation.

Reference values: Since in NEAT one needs to solve an overdetermined system of linear equations, we need references that fix the zero point(s) of the enthalpies of formation. There are two possible choices: either the chemical elements (H_2, N_2, O_2, and so on) are the zero points, as historically has been chosen, or the atoms (H, N, O, and so on) in their ground electronic states. Our preferred choice is the atom-based referencing that leads to atom-based (AT) thermochemistry.

Atom-based thermochemistry appears to be a much better choice when the database contains ab initio reaction enthalpies as, for example, the reaction enthalpy of the $C_{graphite} \rightarrow C_{gas}$ reaction cannot be computed straightforwardly, thus leading to considerable uncertainties in the enthal-

Table 1. Comparison of $\Delta_f H_0^{AT}$ enthalpies of formation, the results of the NEAT protocol, with those of the ATcT approach. The enthalpy values are in kJ mol^{-1}.

Species	No. of reactions	No. of sources	NEAT	ATcT[a]
H$_2$O	17	11	918.05(20)	917.83(3)
C$_2$H$_2$	13	9	1625.78(40)	1626.16(24)
CO	13	9	1072.08(28)	1072.13(9)
CH	13	6	334.61(15)	334.66(23)
NH$_3$	12	8	1157.44(33)	1157.25(4)
OH	12	6	425.93(21)	425.62(3)
CH$_3$	11	8	1209.64(29)	1209.63(13)
H$_2$	11	9	432.07(0)	432.07(0)
CO$_2$	11	8	1597.77(40)	1598.27(9)
CH$_4$	10	7	1641.68(40)	1642.24(12)
CH$_2$	10	4	752.40(21)	752.70(26)
NH	9	4	327.69(25)	328.43(29)
O$_2$	7	6	493.51(34)	493.69(0)
HF	7	6	566.13(31)	565.97(1)
HO$_2$	7	5	694.53(32)	694.46(22)

[a] All ATcT values are taken from ref. [48], except CH$_4$, taken from ref. [74].

Table 2. Comparison of $\Delta_f H_0^o$ enthalpies of formation, the converted results of the NEAT protocol, with those of ATcT. All values are in kJ mol^{-1}.

Species	NEAT[a]	ATcT[b]
H$_2$O	−239.18(23)	−238.92(4)
C$_2$H$_2$	228.67(35)	228.20(64)
CO	−114.09(32)	−113.81(17)
CH	592.62(24)	593.19(36)
NH$_3$	−38.52(26)	−38.91(40)
OH	36.90(16)	37.09(5)
CH$_3$	149.66(34)	149.94(11)
CO$_2$	−392.98(40)	−393.11(1)
CH$_4$	−66.35(46)	−66.58(6)
CH$_2$	390.86(29)	390.65(54)
NH	359.17(17)	−
HF	−272.82(35)	−272.73(24)
HO$_2$	15.09(23)	14.96(64)

[a] The conversion factors applied, all in kJ mol^{-1}, to the entries of Table 1 are as follows: $\Delta_f H_0^o$(H) = 126.035, $\Delta_f H_0^o$(C) = 711.194, $\Delta_f H_0^o$(N) = 470.818, $\Delta_f H_0^o$(O) = 246.795, and $\Delta_f H_0^o$(F) = 77.27.[6] [b] All ATcT values are taken from ref. [46], except CH$_4$, taken from ref. [56].

py of formation of the carbon atom.[91] Furthermore, using the atom-based referencing all of the known atomization reactions results in unambiguous reactions. Therefore, it is best to start building the database of computed reaction enthalpies with atomization reactions. Then, one can add reactions to the database in which a maximum of one species is not defined by one of the atomization reactions. The present database contains 361 reactions that correspond to 188 species. Note that the electron was also given a zero reference value to allow the inclusion of cations and anions in the database. In the AT approach, the reaction enthalpy of a gaseous reaction [e.g., that of Eq. (1)], is simply the difference between the total atomization energies of the reactants and the products.

To make the conversion between atom-based and the traditional element-based enthalpies of formation, one needs the nonzero enthalpies of formation of the atoms when the elements form the zero reference points. All the necessary reference atomic values, for example, $\Delta_f H_0$(C$_{gas}$), were chosen to correspond to the most recent ATcT values available to us, compiled in ref. [46].

Results and Discussion

In what follows, the principal thermochemical results of the NEAT protocol are discussed along with comments related to the protocol itself. Selected enthalpies of formation obtained within the atom-based and the element-based conventions are presented in Tables 1 and 2, respectively. Table 1 presents $\Delta_f H_0^{AT}$ enthalpies of formation determined in this study along with ATcT total atomization energies taken from the literature. Table 2 provides a similar assessment of the NEAT results with respect to the standard states of the elements; it also contains some further literature values.

As mentioned before, the complete database assembled during this work contains 361 reactions and 188 species. The species for which results are given in Tables 1 and 2 were selected as follows. A NEAT enthalpy of formation, $\Delta_f H_0^{AT}$, is called dependable if the chosen species has participated in at least 7 reactions and the reaction enthalpies come from at least 4 different sources. Only dependable enthalpies of for-

mation are given in the main body of the paper and our discussion is focused on these species. Further NEAT enthalpies of formation are given in the Supporting Information for all species in the present database, but these numbers should not be taken literally due to the insufficiency of the information determining these enthalpies of formation.

Reaction networks: When a database of computed reaction enthalpies is built, it can happen that it contains so-called floating networks. A network of reactions is called floating if none of its members are connected, through other reactions in the database, to the reference species. The present database, as reported in Table S1 in the Supporting Information, contains no floating networks.

One particular significance of the reaction network is that it can increase the reliability, or decrease the uncertainty, of the computed enthalpies of formation. When a species is involved only in reactions that do not form circles, the uncertainty of $\Delta_f H_0$ is larger than in cases when the reactions form at least one circle. To show this, we note that if one calculates the enthalpies of formation of CH, CH$_2$, and CH$_3$ using only their atomization reactions, with uncertainties of 1.0, the uncertainties of $\Delta_f H_0$ would be 1.0. However, if one adds to the database combination reactions (e.g., CH$_2$= CH + H and CH$_3$= CH$_2$ + H, with the same uncertainties), the uncertainties become 0.71.

Overall cost of NEAT: The true utility of the NEAT protocol depends on the cost associated with the weighted least-squares refinement and the adjustment of the initial uncertainties by means of robust reweighting. It is reassuring that the overall cost is minuscule due basically to the sparsity of the design matrix. The present database contains 361 entries, so one needs to invert a matrix of this size. The whole protocol was executed in just a few seconds on a modern personal computer.

There seem to be no practical limits in using the NEAT protocol for the purposes of computational thermochemistry. For a test database that contained 10000 reactions, the memory requirement was less than 50 MB and the refinement took less than 10 min on a personal computer. The expected cost when the database of computed reaction enthalpies grows to a size on the order of 100000 can be judged from the performance of the related code MARVEL. Preliminary MARVEL studies performed for the water isotopologues HD^{16}O and H$_2$^{16}O show that such large databases can easily be handled with the protocol and the code written, though the CPU time might increase to several hours.

Accuracy: As mentioned in the Introduction, considerable advantages of the ATcT[56,57] and MARVEL[61] protocols include that, by a grand least-squares refinement of all the available related experimental information given in a network, one can judge the accuracy of most of the individual measurements and, at the same time, reduce the uncertainties of the derived enthalpies of formation or energy levels. Since NEAT is a protocol similar to both ATcT and MARVEL, it is anticipated that the uncertainties in a given $\Delta_f H_0$ value can be reduced if reaction enthalpies from a large number of computations are available.

This indeed turned out to be the case. For example, the best ab initio determination of the atomization energy of the CH radical,[33] obtained within the FPA scheme, is $592.48^{+0.47}_{-0.56}$ kJ mol^{-1}. The corresponding NEAT value is (592.53 ± 0.26) kJ mol^{-1}. Clearly, the least-squares refinement lowered the uncertainty substantially but left the recommended $\Delta_f H_0^{AT}$ value basically unchanged.

Model chemistries: As part of this project, for a large number of relatively simple species in the NEAT database G1, G2, G2MP2, G3, and W1U model chemistry computations have been performed (Figure 2). The best utilization of the resulting energies and enthalpies was investigated in some detail, and NEAT produced some useful results that faci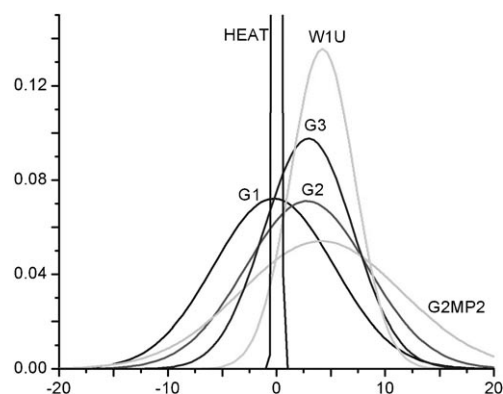litated judgment of the overall accuracies of these model chemistries. This is important, as these simple model chemistries can be applied to considerably larger species than are present in our database at the moment.

The simplest way to utilize the computed enthalpies through model chemistries is to define a total atomization reaction for each new species and enter the resulting reaction and the corresponding reaction enthalpy into the database. This way one can include as many new species as possible and all these reactions are unambiguous by definition. It is clear, however, that the uncertainties that should be attached to these reactions are much larger than if a chemically motivated reaction was used in which the new species participates. The reactions that lead to the smallest possible uncertainties are those in which a minimum number of bonds is broken, involve the maximum number of closed-shell species both as reactants and products, and the products are as close in their bonding pattern to the reactants as possible. These requirements help to take advantage of the often employed "cancellation of error principle" of electronic-structure theory. It is often not possible to fulfill all these criteria simultaneously and thus some care should be exercised when selecting the best reaction. For example, when trying to determine the enthalpy of formation of formaldehyde it is advantageous[80] to use the reaction H$_2$CO = H$_2$ + CO over either H$_2$CO = C + H$_2$O or H$_2$CO = 2H + C + O.

The error analysis of the G1, G2, G2MP2, G3, and W1U model chemistries was based on the 25 species treated by the HEAT protocol[46] and on the NEAT enthalpies of formation as reference values. This error analysis, which results in the distribution of the errors, was based on the following formula [Eq. (8)]:

$$\rho(P) = \frac{1}{\Delta_{std}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{P - \Delta}{\Delta_{std}} \right)^2 \right] \tag{8}$$

in which Δ_{std} is the standard deviation and Δ is the mean error characterizing each model chemistry (in kJ mol^{-1}). The Gaussian distribution curves corresponding to the G1, G2, G2MP2, G3, and W1U atom-based enthalpies of formation with respect to NEAT are given in Figure 2. Clearly, the worst performer is the G2MP2 model chemistry. Somewhat surprising is the good performance of the G1 model chemistry. The other three model chemistries show the expected behavior. Note that the high-accuracy HEAT model chemistry.

Conversion between the atom-based and the element-based approaches: Conversion of the enthalpies of formation between the atom-based (AT) and the standard element-based approaches is rather simple. Suppose that the standard enthalpies of formation of the atoms, $\Delta_f H_0^\circ(A)$, are known from some source. Then, the standard (element-based) enthalpy of formation of an A_aB_b species, $\Delta_f H_0^\circ$, can be computed from the NEAT result, $\Delta_f H_0^{AT}$, as in Equation (9):

Figure 2. The overall accuracy of the model chemistries G1, G2, G2MP2, G3, and W1U based on the set of 25 species treated in the HEAT database,[46] presented as a probability distribution function (abscissa in kJ mol^{-1}). Performance of the model chemistries is referenced to the NEAT values obtained in this study.

$$\Delta_f H_0^\circ(A_aB_b) = \Delta_f H_0^{AT}(A_aB_b) + a \cdot \Delta_f H_0^\circ(A) + b \cdot \Delta_f H_0^\circ(B) \tag{9}$$

Clearly, this conversion process has only one considerable drawback: one needs to know the $\Delta_f H_0^\circ(A)$ values a priori. Fortunately, by now these entities are available from several sources and their uncertainty is reasonably small, though with some important exceptions (for example, the atoms C and B).

Effect of the inclusion of highly accurate entries into the database: One question one might ask concerning NEAT is what happens if one includes in the database, besides the computational results, those experimental entries, usually obtained by means of sophisticated spectroscopic measurements, that have uncertainties much smaller than those of modern quantum-chemical techniques. Of course, ultimately one arrives this way at the ATcT database, which principally contains measured quantities (and also some computed results). Since it was not the purpose of the present work to duplicate the ATcT database, we performed only a few "experiments".

As an example, let us take the first dissociation energy of water, which is known very accurately as $D_0 = (492.215 \pm 0.002)$ kJ mol^{-1}.[15] Inclusion of this entry in the database fixes, of course, the difference between the AT enthalpies of formation of OH and H_2O, as none of the computed results come even close to the precision of this particular experiment. This, in turn, changes slightly the two enthalpies of formation as well. Without this entry, the values are $\Delta_f H_0^{AT}(OH) = (425.91 \pm 0.21)$ and $\Delta_f H_0^{AT}(H_2O) = (918.05 \pm 0.21)$ kJ mol^{-1}, whereas after the inclusion of this entry into the database they become $\Delta_f H_0^{AT}(OH) = (425.87 \pm 0.17)$ and $\Delta_f H_0^{AT}(H_2O) = (918.09 \pm 0.17)$ kJ mol^{-1}, now showing the expected D_0 difference correctly, just like the reference ATcT values (Table 1). The new datum makes $\Delta_f H_0^{AT}(OH)$ move closer to the supposedly correct ATcT value but that of H_2O moves away, thus showing the intrinsic imprecision of the computed entries at the level of 0.2 kJ mol^{-1} in this case. Since after robust reweighting the original dataset becomes self-consistent, the inclusion of a new and accurate entry should not change the enthalpies of formation outside their uncertainty limits. As to the values of the uncertainties, these did not change significantly either. The principal reason for this is that the first dissociation energy of water is not connected directly to the reference (atomic) values and thus a single accurate entry is unable to have a substantial effect on the overall uncertainties. Nevertheless, if a highly accurate number was entered for a species directly connected to the references (e.g., the dissociation enthalpy of OH was known with similar accuracy as that of H_2O), it would move the precise bottom value to a subsequent diatomic level, thereby providing increased accuracy and "rigidity" to the whole dataset. Of course, this can be extended upwards, and moving from atoms to diatoms and then to triatomic species seems the best way to increase the overall accuracy and thus the utility of the present NEAT approach and of any similar ones.

Finally, it is important to emphasize that the simple structure of the NEAT protocol enables straightforward hypothe-

sis testing. Assuming tentative values or related uncertainties for certain data entries, one can test for consistency and for the effect of new computed data on the NEAT $\Delta_f H$ values. This should help to identify "weak links" within the existing network and contributes toward the meaningful utilization of available computational resources to improve thermochemical knowledge.

Conclusion

It is common knowledge and requires no further justification that state-of-the-art computational quantum chemistry that employs wavefunction-based methods has become a significant source of accurate thermochemical data, competitive with even some of the best experimental techniques. These often costly computations, best performed within the framework of the composite focal-point analysis (FPA) philosophy, can provide a large number of accurate thermochemical information to the scientific and engineering communities. Since ab initio computations do not suffer from the same constraints as experiments (they do, however, suffer from others), it is usually easier to perform computations than doing experiments on, for example, reactive, unstable, or transient species. Clearly, an accurate quantum-chemical reaction enthalpy, even at a single temperature (say, at 0 K), or a related computed enthalpy of formation can be used directly and have significance on its own. Nevertheless, as this study demonstrated, this is not the optimal use of computational quantum chemistry for thermochemical applications. The seemingly optimal way to extract the full amount of information present in computed reaction enthalpies obtained using several protocols of different sophistication is to assemble them into a network and do a simultaneous analysis of all the available information by means of weighted linear least squares. This way, one achieves several important goals, all at the same time, as the reaction network built provides: 1) the best currently known computed values of the enthalpies of formation, including uncertainties lower than that of even the best related individual computation, thus representing the overall consensus of wavefunction-based ab initio theory; 2) a simple scheme that allows for an efficient update and extension of the database, perhaps over the Web, with full allowance of tracking each entry; 3) a straightforward testing facility of new theoretical data; and 4) strong and, even more importantly, weak links within the database and a possibility of identifying missing links between species of interest or uncertain reaction enthalpies, the updating of which would be most beneficial in enhancing the information content of the complete database. Furthermore, the reaction network provides, through the robust reweighting scheme, a simple way to identify literature sources that contain overly optimistic uncertainties for the entry reaction enthalpies and a straightforward and easily controllable adjustment technique to improve the uncertainties until self-consistency is achieved within the database. Finally, the weighted linear least-squares approach

that has been chosen not only presents a simple route to utilize the interdependencies between the species within the database but also a way to plan further reaction enthalpy computations and forecast their overall effect on the thermochemical knowledge attainable.

To minimize the externally required information, the NEAT protocol developed in this publication, in which NEAT stands for a *ne*twork of computed reaction *e*nthalpies that lead toward *a*tom-based *t*hermochemistry, does not utilize any other information except the zero reference values for the ground electronic states of the atoms, which can be computed with relative ease. To have cations and anions in the database, the reference value of the electron is also set to zero. For the same reason, we plan to extend the NEAT database only at 0 K and propose to do any temperature or other corrections a posteriori. Note that, in this respect, the accuracy of the temperature corrections of the thermodynamic quantities is considerably better than the intrinsic precision of the underlying 0 K enthalpy values, especially if the desired temperature was as low as 298.15 K. The switch to an atom-based reference scheme forms the basis of atom-based thermochemistry (AT), whereby the enthalpy of formation of a gaseous compound would directly reflect the bonding within the molecule, a highly useful feature of AT over element-based thermochemistry. Of course, for liquids and solids, one would need to include additional enthalpy as well as entropy terms if the temperature was not set to 0 K. Atom-based thermochemistry has several advantages well beyond first-principles thermochemistry, though they are not explored here.

As to technicalities, NEAT is fast and with the robustly reweighted linear least-squares protocol described it does not seem to be difficult to treat a database that contains reaction enthalpies on the order of 100 000. Thus, the technical limits of NEAT are rather far, even if the database were extended to lot more elements than treated presently.

The only notable drawback of the protocol described, shared by ATcT[56] and MARVEL,[61] is that introduction of each new reaction enthalpy into the database results in new enthalpies of formation and uncertainties for all species. First, this is just a small, unavoidable disadvantage that needs to be compared to all the related advantages of the proposed protocol. Second, with proper accounting and keeping versions of the database and its refinements after all major changes, this becomes an even less significant issue.

From a reaction network assembled as part of this study that contained 361 entries and 188 species, the following dependable 0 K atom-based enthalpies of formation, $\Delta_f H_0^{AT}$, all in kJ mol^{-1}, have been obtained in this study: H$_2$ = 432.07(0), CH = 334.61(15), NH = 327.69(25), OH = 425.93(21), HF = 566.13(31), CO = 1072.08(28), O$_2$ = 493.51(34), CH$_2$ = 752.40(21), H$_2$O = 918.05(20), HO$_2$ = 694.53(32), CO$_2$ = 1597.77(40), CH$_3$ = 1209.64(29), NH$_3$ = 1157.44(33), C$_2$H$_2$ = 1625.78(40), and CH$_4$ = 1641.68(40), in which the uncertainty values given in parentheses represent 95 % confidence intervals and, due to the NEAT protocol,

are usually lower by about 50 % than those characterizing even the most accurate directly computed ab initio reaction enthalpies (atomization energies). The full potential of the NEAT method will be revealed if the underlying database of high-quality ab initio reaction enthalpies is developed much further.

[1] J. B. Pedley, *Thermochemical Data and Structures of Organic Compounds, Vol. 1*, Thermodynamics Research Center, Texas, **1994**.

[2] B. Ruscic, J. E. Boggs, A. Burcat, A. G. Császár, J. Demaison, R. Janoschek, J. M. L. Martin, M. Morton, M. J. Rossi, J. F. Stanton, P. G. Szalay, P. R. Westmoreland, F. Zabel, T. Bérces, *J. Phys. Chem. Ref. Data* **2005**, *34*, 573–656.

[3] L. V. Gurvich, I. V. Veyts, C. B. Alcock, *Thermodynamic Properties of Individual Substances, Vols. 1–3*, 4th ed., Hemisphere, New York, **1989–1996**.

[4] W. Tsang, R. F. Hampson, *J. Phys. Chem. Ref. Data* **1986**, *15*, 1087.

[5] D. R. Stull, E. F. Westrum, G. C. Sinke, *The Chemical Thermodynamics of Organic Compounds*, Krieger, Malabar, **1987**.

[6] J. D. Cox, D. D. Wagman, V. A. Medvedev, *CODATA Key Values for Thermodynamics*, Hemisphere, New York, **1998**. See also http://www.codata.org/codata/databases/key1.html.

[7] "JANAF Thermochemical Tables", 3rd ed.: M. W. Chase, Jr., C. A. Davies, J. R. Downey, D. J. Frurip, R. A. McDonald, A. N. Syverud, *J. Phys. Chem. Ref. Data* **1985**, 14.

[8] "NIST-JANAF Thermochemical Tables", 4th ed.: *J. Phys. Chem. Ref. Data* **1998**, 9, edited by M. W. Chase, Jr.

[9] One of the most useful free reference sources is the NIST WebBook, available at http://webbook.nist.gov/chemistry.

[10] J. Zádor, I. G. Zsély, T. Turányi, M. Ratto, S. Tarantola, A. Saltelli, *J. Phys. Chem. A* **2005**, *109*, 9795–9807.

[11] G. A. Weisenburger, R. W. Barnhart, J. D. Clark, D. J. Dale, M. Hawksworth, P. D. Higginson, Y. Kang, D. J. Knoechel, B. S. Moon, S. M. Shaw, G. P. Taber, D. L. Tickner, *Org. Process Res. Dev.* **2007**, *11*, 1112–1125.

[12] J. D. Cox, G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds*, Academic Press, London, **1970**.

[13] E. S. Domalski, E. D. Hearing, *J. Phys. Chem. Ref. Data* **1993**, *22*, 805.

[14] N. Cohen, S. W. Benson, *Chem. Rev.* **1993**, *93*, 2419–2438.

[15] P. Maksyutenko, T. R. Rizzo, O. V. Boyarkin, *J. Chem. Phys.* **2006**, *125*, 181101.

[16] M. Grechko, O. V. Boyarkin, T. R. Rizzo, P. Maksyutenko, N. F. Zobov, S. V. Shirin, L. Lodi, J. Tennyson, A. G. Császár, and O. L. Polyansky, *J. Chem. Phys.* **2009**, *131*, 221105.

[17] See, for example, the CHETAH software at http://www.chetah.usouthal.edu.

[18] S. E. Wheeler, K. N. Houk, P. von R. Schleyer, W. D. Allen, *J. Am. Chem. Soc.* **2009**, *131*, 2547–2560.

[19] R. T. Sanderson, *Polar Covalence*, Academic Press, New York, **1983**.

[20] R. T. Sanderson, *Chemical Bonds and Bond Energy*, Academic Press, New York, **1976**.

[21] J. N. Spencer, R. S. Moog, R. J. Gillespie, *J. Chem. Educ.* **1996**, *73*, 631–636.

[22] L. von Szentpály, *J. Am. Chem. Soc.* **2008**, *130*, 5962–5973.

[23] L. von Szentpály, *J. Phys. Chem. A* **2008**, *112*, 12 695–12 701.

[24] W. D. Allen, A. L. L. East, and A. G. Császár, *Structures and Conformations of Non-Rigid Molecules* (Eds.: J. Laane, M. Dakkouri, B. van der Veken, H. Oberhammer), Kluwer, Dordrecht, **1993**, p. 343.

[25] A. G. Császár, W. D. Allen, H. F. Schaefer III, *J. Chem. Phys.* **1998**, *108*, 9751–9764.

[26] A. G. Császár, *J. Am. Chem. Soc.* **1992**, *114*, 9568–9575.

[27] A. L. L. East, W. D. Allen, *J. Chem. Phys.* **1993**, *99*, 4638–4650.

[28] A. G. Császár, W. D. Allen, *J. Chem. Phys.* **1996**, *104*, 2746–2748.

[29] R. D. Cowan, D. C. Griffin, *J. Opt. Soc. Am.* **1976**, *66*, 1010.

[30] G. Tarczay, A. G. Császár, W. Klopper, H. M. Quiney, *Mol. Phys.* **2001**, *99*, 1769–1794.

[31] A. G. Császár, J. S. Kain, O. L. Polyansky, N. F. Zobov, J. Tennyson, *Chem. Phys. Lett.* **1998**, *293*, 317–323; A. G. Császár, J. S. Kain, O. L. Polyansky, N. F. Zobov, J. Tennyson, *Chem. Phys. Lett.* **1999**, *310*, 385–389.

[32] N. C. Handy, Y. Yamaguchi, H. F. Schaefer III, *J. Chem. Phys.* **1986**, *84*, 4481.

[33] A. G. Császár, P. G. Szalay, M. L. Leininger, *Mol. Phys.* **2002**, *100*, 3879–3883.

[34] G. Czakó, E. Mátyus, A. C. Simmonett, A. G. Császár, H. F. Schaefer III, W. D. Allen, *J. Chem. Theory Comput.* **2008**, *4*, 1220–1229.

[35] A. G. Császár, M. L. Leininger, A. Burcat, *J. Phys. Chem. A* **2003**, *107*, 2061–2065.

[36] A. G. Császár, M. L. Leininger, V. Szalay, *J. Chem. Phys.* **2003**, *118*, 10 631–10 642.

[37] O. L. Polyansky, A. G. Császár, S. V. Shirin, N. F. Zobov, P. Barletta, J. Tennyson, D. W. Schwenke, P. J. Knowles, *Science* **2003**, *299*, 539–542.

[38] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, L. A. Curtiss, *J. Chem. Phys.* **1989**, *90*, 5622.

[39] L. A. Curtiss, K. Raghavachari, G. W. Trucks, J. A. Pople, *J. Chem. Phys.* **1991**, *94*, 7221–7230.

[40] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, J. A. Pople, *J. Chem. Phys.* **1998**, *109*, 7764–7776.

[41] L. A. Curtiss, P. C. Redfern, K. Raghavachari, *J. Chem. Phys.* **2007**, *126*, 084108.

[42] J. M. L. Martin, O. De Oliveira, *J. Chem. Phys.* **1999**, *111*, 1843–1856.

[43] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, J. Gauss, *J. Chem. Phys.* **2004**, *120*, 4129–4141.

[44] A. Karton, E. Rabinovich, J. M. L. Martin, B. Ruscic, *J. Chem. Phys.* **2006**, *125*, 144108.

[45] A. Karton, P. R. Taylor, J. M. L. Martin, *J. Chem. Phys.* **2007**, *127*, 064104.

[46] A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, J. F. Stanton, *J. Chem. Phys.* **2004**, *121*, 11599–11613.

[47] Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, J. F. Stanton, *J. Chem. Phys.* **2006**, *125*, 064108.

[48] M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, J. F. Stanton, *J. Chem. Phys.* **2008**, *128*, 114111.

[49] N. J. DeYonker, T. R. Cundari, A. K. Wilson, *J. Chem. Phys.* **2006**, *124*, 114104.

[50] N. J. DeYonker, T. Grimes, S. Yockel, A. Dinescu, B. Mintz, T. R. Cundari, A. K. Wilson, *J. Chem. Phys.* **2006**, *125*, 104111.

[51] J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **1999**, *110*, 2822–2827.

[52] J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **2000**, *112*, 6532–6542.

[53] P. L. Fast, J. C. Corchado, M. L. Sánchez, D. G. Truhlar, *J. Phys. Chem. A* **1999**, *103*, 3139.

[54] P. L. Fast, M. L. Sánchez, D. G. Truhlar, *Chem. Phys. Lett.* **1999**, *306*, 407.

[55] P. L. Fast, J. C. Corchado, M. L. Sánchez, D. G. Truhlar, *J. Phys. Chem. A* **1999**, *103*, 5129.

[56] B. Ruscic, R. E. Pinzon, M. L. Morton, G. Von Laszewski, S. Bittner, S. G. Nijsure, K. A. Amin, M. Minkoff, A. F. Wagner, *J. Phys. Chem. A* **2004**, *108*, 9979.

[57] B. Ruscic, J. V. Michael, P. C. Redfern, L. A. Curtiss, *J. Phys. Chem. A* **1998**, *102*, 10889.

[58] B. Ruscic, M. Litorja, R. L. Asher, *J. Phys. Chem. A* **1999**, *103*, 8625.

[59] B. Ruscic, R. E. Pinzon, M. L. Morton, N. K. Srinivasan, M.-C. Su, J. W. Sutherland, J. V. Michael, *J. Phys. Chem.* **2006**, *110*, 6592.

[60] J.-M. Flaud, C. Camy-Peyret, J. P. Maillard, *Mol. Phys.* **1976**, *32*, 499–521.

[61] T. Furtenbacher, A. G. Császár, J. Tennyson, *J. Mol. Spectrosc.* **2007**, *245*, 115–125.

[62] S. A. Tashkun, V. I. Perevalov, J.-L. Teffo, A. D. Bykov, N. N. Lavrentieva, *J. Quant. Spectrosc. Radiat. Transfer* **2003**, *82*, 165–196.

[63] A. G. Császár, G. Czakó, T. Furtenbacher, E. Mátyus, *Annu. Rep. Comput. Chem.* **2007**, *3*, 155–176.

[64] T. Furtenbacher, A. G. Császár, *J. Quant. Spectrosc. Radiat. Transfer* **2008**, *109*, 1234–1251.

[65] J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, M. R. Carleer, A. G. Császár, R. R. Gamache, J. T. Hodges, A. Jenouvrier, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, R. A. Toth, A. C. Vandaele, N. F. Zobov, L. Daumont, A. Z. Fazliev, T. Furtenbacher, I. F. Gordon, S. N. Mikhailenko, S. V. Shirin, *J. Quant. Spectrosc. Radiat. Transfer* **2009**, *110*, 573–596.

[66] J. K. G. Watson, *J. Mol. Spectrosc.* **2003**, *219*, 326–328.

[67] N. D. K. Petraco, W. D. Allen, H. F. Schaefer III, *J. Chem. Phys.* **2002**, *116*, 10229–10237.

[68] B. A. Flowers, P. G. Szalay, J. F. Stanton, M. Kállay, J. Gauss, A. G. Császár, *J. Chem. Phys.* **2004**, *120–121*, 3195–3199.

[69] M. S. Schuurman, S. R. Muir, W. D. Allen, H. F. Schaefer III, *J. Chem. Phys.* **2004**, *120*, 11586–11599.

[70] S. E. Wheeler, W. D. Allen, H. F. Schaefer III, *J. Chem. Phys.* **2004**, *121*, 8800–8813.

[71] A. Karton, B. Ruscic, J. M. L. Martin, *J. Mol. Struct.* **2007**, *825–846*, 345–353.

[72] D. Feller, K. A. Peterson, *J. Chem. Phys.* **2007**, *126*, 114105.

[73] S. E. Wheeler, K. A. Robertson, W. D. Allen, H. F. Schaefer III, Y. J. Bomble, J. F. Stanton, *J. Phys. Chem. A* **2007**, *111*, 3819–3830.

[74] J. Aguilera-Iparraguirre, A. D. Boese, W. Klopper, B. Ruscic, *Chem. Phys.* **2008**, *346*, 56–68.

[75] A. Karton, A. Tarnopolsky, J. M. L. Martin, *Mol. Phys.* **2009**, *107*, 977–990.

[76] A. Karton, S. Parthiban, J. M. L. Martin, *J. Phys. Chem. A* **2009**, *113*, 4802–4816.

[77] W. Klopper, B. Ruscic, D. P. Tew, F. A. Bischoff, S. Wolfsegger, *Chem. Phys.* **2009**, *356*, 14–24.

[78] A. C. Simmonett, H. F. Schaefer, W. D. Allen, *J. Chem. Phys.* **2009**, *130*, 044301.

[79] G. Czakó, B. C. Shepler, B. J. Braams, J. M. Bowman, *J. Chem. Phys.* **2009**, *130*, 084301.

[80] G. Czakó, B. Nagy, G. Tasi, A. Somogyi, J. Simunek, J. Noga, B. J. Braams, J. M. Bowman, A. G. Császár, *Int. J. Quantum Chem.* **2009**, *109*, 2393–2409.

[81] D. B. Kinghorn, L. Adamowicz, *J. Phys. Chem. A* **1997**, *101*, 4589.

[82] A. Karton, I. Kaminker, J. M. L. Martin, *J. Phys. Chem. A* **2009**, *113*, 7610—7620.

[83] G. Tarczay, T. A. Miller, G. Czakó, A. G. Császár, *Phys. Chem. Chem. Phys.* **2005**, *7*, 2881–2893.

[84] K. Piszczatowski, G. Lach, M. Przybytek, J. Komasa, K. Pachucki, B. Jeziorski, *J. Chem. Theory Comput.* **2009**, *5*, 3039–3048.

[85] J. Breidung, W. Thiel, *J. Phys. Chem. A* **2006**, *110*, 1575–1585.

[86] D. Feller, K. A. Peterson, *J. Chem. Phys.* **2009**, *131*, 154306.

[87] B. Ruscic, A. F. Wagner, L. B. Harding, R. L. Asher, D. Feller, D. A. Dixon, K. A. Peterson, Y. Song, X. Qian, C.-Y. Ng, J. Liu, W. Chen, D. W- Schwenke, *J. Phys. Chem. A* **2002**, *106*, 2727—2747

[88] 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, V. Bakken, 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**.

[89] L. A. Curtiss, K. Raghavachari, J. A. Pople, *J. Chem. Phys.* **1993**, *98*, 1293.

[90] W1U is a slight modification of the W1 procedure described in reference [42], using UCCSD instead of ROCCSD for open-shell systems.

[91] G. Tasi, R. Izsák, G. Matisz, A. G. Császár, M. Kállay, B. Ruscic, J. F. Stanton, *ChemPhysChem* **2006**, *7*, 1664–1667.