J. Phys. Chem. A 2009, 113, 2165–2175

Efficient Calculation of Heats of Formation[†]

W. S. Ohlinger, P. E. Klunzinger, B. J. Deppmeier, and W. J. Hehre*

Wavefunction, Inc., 18401 von Karman Avenue, Suite 370, Irvine, California 92617

Received: November 18, 2008; Revised Manuscript Received: December 29, 2008

An efficient procedure has been devised for calculating heats of formation of uncharged, closed-shell molecules comprising H, C, N, O, F, S, Cl, and Br. Known as T1, it follows the G3(MP2) recipe, by substituting an HF/6-31G* for the MP2/6-31G* geometry, eliminating both the HF/6-31G* frequency and QCISD(T)/6-31G* energy and approximating the MP2/G3MP2large energy using dual basis set RI-MP2 techniques. Taken together, these changes reduce computation time by 2−3 orders of magnitude. Atom counts, Mulliken bond orders, and HF/6-31G* and RI-MP2 energies are introduced as variables in a linear regression fit to a set of 1126 G3(MP2) heats of formation. The T1 procedure reproduces these values with mean absolute and rms errors of 1.8 and 2.5 kJ/mol, respectively. It reproduces experimental heats of formation for a set of 1805 diverse organic molecules from the NIST thermochemical database with mean absolute and rms errors of 8.5 and 11.5 kJ/mol, respectively. Heats of formation of flexible molecules have been approximated by the heats of formation of their lowest-energy conformer as given by the T1 recipe. This has been identified by examining all conformers for molecules with fewer than 100 conformers and by examining a random sample of 100 conformers for molecules with more than 100 conformers. While this approximation necessarily yields heats of formation that are too negative, the error for typical organic molecules with less than 10 degrees of conformational freedom (several thousand conformers) is <2−3 kJ/mol. T1 heats of formation have been used to calculate energy differences for a variety of structural, positional, and stereoisomers, as well as energy differences between conformers in a variety of simple acyclic and cyclic molecules for which reliable experimental data are available. In terms of both overall error and errors for individual systems, T1 provides a better account of the experimental thermochemistry than any practical quantum chemical method that we have previously examined. A database of ∼40 000 T1 calculations for both rigid and flexible organic molecules has been produced and is available as part of the Spartan Molecular Database (SMD) in the current version of the Spartan electronic structure program (Spartan'08). (A subset of ∼5000 molecules is provided as part of the standard release, and the full T1 database can be licensed.). This collection differs from the other components of SMD in that the lowest-energy conformation for each molecule has been assigned using a high-level quantum chemical method and not molecular mechanics. Thus, it is not only a source of "high-quality" calculated heats of formation for organic molecules but also a source of conformational preferences.

Introduction

One of the most fundamental properties of a molecule is its internal energy. Internal energy is most commonly reported as a heat of formation, defined as the enthalpy at 298.15 K of a hypothetical chemical reaction in which the molecule is transformed into a set of products that correspond to the most stable forms of its constituent "pure" elements at room temperature. In practice, heat of formation is not directly measured but rather is typically obtained from a measurement of the heat given off in a combustion reaction.

Differences in heats of formation between the products and reactants (reaction enthalpies) indicate the extent to which the reaction will be favorable (exothermic) or unfavorable (endothermic). Together with entropy, heats of formation allow calculation of Gibbs energies and equilibrium constants.

Heats of formation have been determined for several thousand compounds, a large fraction of which are hydrocarbons and oxycarbons (see below). Among the most extensive compilations is the NIST[1] database, freely available online (http://webbook.nist.gov). While care has been taken to ensure the integrity of this collection, because the data derive from a variety of experimental techniques and have been assembled over many decades, heats of formation for individual entries vary widely in quality. The most egregious source of error is that the structure is incorrect, meaning that the reported heat does not correspond to the reported structure. Because most of the compounds in the NIST database are fairly simple and readily available commercially, it is likely that very few structures are incorrectly assigned. More likely sources of error include incomplete combustion and poorly characterized combustion products. Hydrocarbons and oxycarbons present few problems as they lead only to carbon dioxide and water, the amounts of which may easily be determined. However, combustion of molecules with other elements may give rise to a complex mixture of products and greater uncertainty. Among common organic molecules, nitrogen compounds may be particularly problematic in this regard. Stewart recently surveyed the NIST database and found several dozen entries that appear to be far outside the stated error bounds.[2] He suggested alternative heats

---

[†] Part of the "Max Wolfsberg Festschrift".

* To whom correspondence should be addressed.

10.1021/jp810144q CCC: $40.75

for several compounds based on semiempirical molecular orbital calculations, and these have recently been included in the NIST database in addition to the previously reported (experimental) values.

Despite their importance, heats of formation are not routinely determined (or at least reported) for new compounds. While in principle combustion experiments are straightforward and the results easily interpreted, measurements typically need to be performed several times to establish useful error limits. This means that accurate determinations may require (and destroy) significant quantities of compound. Very few synthetic chemists are willing to part with hundreds of milligrams (a huge amount by modern standards) of a compound that they have just spent weeks or months preparing.

In view of the value and scarcity of experimental thermochemical data, it is not surprising that considerable attention has been directed at the use of quantum chemical calculations to obtain heats of formation.[3] This eliminates the need for having to actually prepare (and later destroy) the compound and does away with uncertainties in the data due to incomplete combustion and the identity of the combustion products (not to mention the identity of the compound itself). However, the goal of calculating heats of formation for moderate-size organic molecules (molecular weight <400−500 amu) within "experimental accuracy" (<4−8 kJ/mol) has thus far proven elusive. Computational methods that are practical (fast) are not sufficiently accurate, and methods that are sufficiently accurate are very time-consuming and only applicable to very small molecules (molecular weight <150−200 amu). The underlying problem is that accurate energy (heat of formation) calculations require high-order correlation treatments and large atomic basis sets, a combination that is very costly in terms of overall computation. In our view, the best "compromise" yet proposed is G3(MP2),[4] one of a series of "recipes" that also includes G2,[5] G3,[6] and G4.[7] (The term "recipe" is perhaps more appropriate than "model" or "method" as all of these involve combinations of different quantum chemical models.) While heats of formation from G3(MP2) are generally very close to experimental values, the recipe can only be applied to molecules with 10−15 heavy (non-hydrogen) atoms. G3(MP2) calculations on molecules near the top of this range present significant demands, both in terms of computer time and memory and disk usage. The remaining GX recipes are all more costly than G3(MP2) in terms of computation and therefore are even more limited in their application.

G3(MP2) (and more generally all GX recipes) are in practice applicable only to rigid molecules or molecules with only limited conformational freedom such as *n*-butane. This is a major shortcoming, as most organic molecules are flexible and may exist in a multitude of different conformers. At the very least, determining the heat of formation of a flexible molecule requires identifying the lowest-energy conformer which in most cases means having to examine all conformers.

Although comparisons between experimental conformational preferences and those assigned from G3(MP2) are limited, they suggest that the recipe performs reasonably well both in assigning the lowest-energy conformer and in accounting for conformational energy differences. The absence of experimental data for molecules with multiple degrees of conformational freedom complicates more thorough assessment.

Criteria for Practical Thermochemical Recipes. Thermochemical recipes like the theoretical models they comprise should satisfy a small number of criteria. Most important, the recipe should be well-defined and provide unique results given only the number of each kind of nucleus, the number of electrons, the number of unpaired electrons, and a conformation (as a starting point for geometry optimization to the closest local minimum). To the maximum extent possible, the recipe should be unbiased. Any parameters that are incorporated should be based on a large and diverse set of molecules. If possible, the recipe should be size-consistent, meaning that the error in the heat of formation should scale roughly in proportion to molecular size. Finally, the recipe needs to be practical for the molecules of interest given the computational resources that are available.

The GX recipes are all well-defined and unbiased. Because they incorporate a term based on a limited configuration interaction procedure (QCISD(T) in the case of the G3(MP2) recipe), none are size-consistent. However, the QCISD(T) correction is small, and it is likely that errors due to lack of size consistency will also be small. The primary drawback of the GX recipes is that they are practical only for very simple molecules with few conformers.

T1 from G3(MP2). A new recipe, designated T1, has been formulated to reproduce G3(MP2) heats of formation for uncharged closed-shell organic molecules comprising the elements H, C, N, O, F, S, Cl, and Br only.[8] While this already encompasses the vast majority of organic molecules, there is no reason that T1 could not be extended to include other elements or molecules that are charged or contain unpaired electrons. The objective has been to reduce overall computational effort of G3(MP2) by 2 orders of magnitude or more without seriously impacting its accuracy, thereby enabling reliable heat of formation calculations to be routinely carried out on organic molecules with molecular weights up to 400−500 amu. The approach taken has been to simplify (or eliminate altogether) the four steps in the G3(MP2) recipe that are the most computationally demanding:

1. MP2/6-31G* geometry calculation
2. HF/6-31G* frequency calculation
3. MP2/G3MP2large energy calculation
4. QCISD(T)/6-31G* energy calculation

1. G3(MP2) uses an MP2/6-31G* equilibrium geometry for both large basis set MP2 and small basis set QCISD(T) correlation energy calculations. (It also requires an HF/6-31G* geometry for an HF/6-31G* frequency calculation.) T1 uses an HF/6-31G* geometry.

Both HF/6-31G* and MP2/6-31G* models are routinely employed for equilibrium geometry determination. For example, the Spartan Molecular Database (SMD)[9] contains ~150 000 HF/6-31G* structures and ~40 000 MP2/6-31G* structures. Both models have been critically assessed and extensively documented.[10] Several hundred comparisons have been made with gas-phase experimental geometries and show that MP2/6-31G* bond lengths are usually (but not always) closer to experimental bond lengths and typically longer than HF/6-31G* bond lengths. They also show that differences in bond lengths between the two models are usually very small (<0.02 Å), except where both elements involved in the bond are highly electronegative. The worst case of significance is the OO bond in hydrogen peroxide (as a model for OO bonds in organic peroxides), where the HF/6-31G* model gives a bond that is too short by 0.06 Å, while the MP2/6-31G* model gives a bond that is too long by 0.03 Å. HF/6-31G* and MP2/6-31G* NN bond lengths in hydrazine (as a model for organic hydrazines and related compounds) are closer and both smaller than the experimental distance (by 0.04 and 0.01 Å, respectively). The two models yield similar (and acceptable) lengths for bonds involving only a single electronegative element, for example, CO bonds. Finally, bond angles obtained from the two models are almost always very similar (and within experimental error limits), and it is usually not possible to say with confidence which model actually provides the better account.

In summary, HF/6-31G* and MP2/6-31G* models yield broadly similar geometries. While there are exceptions, peroxides, for example, these are relatively few and may be easily anticipated. Finally, note that because the potential energy surface in the vicinity of an equilibrium structure is "flat", any error in energy resulting from use of a slightly different geometry is likely to be small.

Replacement of the MP2/6-31G* geometry by an HF/6-31G* geometry leads to significant reduction in computation. MP2 is an O(N^5) process (N is the total number of basis functions), compared to an effective O(N^3) or smaller scaling for Hartree−Fock models. The HF/6-31G* geometry calculation is typically the most costly step in the T1 recipe for molecules with molecular weights <150 amu.

2. G3(MP2) requires vibrational frequencies both to provide the zero-point energy and to correct the energy (enthalpy) for finite temperature. The zero-point energy is dominated by high-frequency vibrations and may be accurately estimated from quantum chemical calculations, while the temperature correction is dominated by low-frequency vibrations which are believed to be more difficult to accurately calculate, and for flexible molecules are likely to depend on conformation. (Assessment is complicated by a lack of experimental data.) G3(MP2) uses vibrational frequencies from the HF/6-31G* model (in lieu of MP2/6-31G* frequencies), requiring an HF/6-31G* equilibrium geometry (in addition to the MP2/6-31G* geometry used for the other steps in the recipe). HF/6-31G* (stretching) frequencies are known to be uniformly larger than experimental values by ~12%, consistent with the fact that bond distances are shorter than experimental values. To account for this, all frequencies are scaled by 0.88 prior to their use in calculation of both the zero-point energy and the temperature correction.

The T1 recipe does not include a frequency calculation. This means that neither the zero-point energy nor the temperature dependence of the energy is taken into account explicitely. We do not view this omission as significant. The zero-point-energy correction (dominated by high-frequency vibrations) should be taken into account by the bond-order terms in the T1 parameter set (see discussion following), while the temperature correction (dominated by low-frequency vibrations) is not likely to be well-described in G3(MP2) for all but rigid molecules.

The absence of an explicit frequency calculation means that T1 unlike G3(MP2) cannot be used to calculate the entropy (and therefore the Gibbs energy). (Frequency calculation can be performed following the T1 calculation, making this information available.) It should be noted, however, that entropy is dominated by low-frequency vibrations and, like the temperature correction, is likely to be problematic for any but very small (rigid) molecules.

Frequency calculation typically requires computer time equal to or greater than that required for geometry calculation. Significant time savings results from its elimination in the T1 recipe.

3. G3(MP2) requires an MP2 energy calculation with a very large basis set, specifically, the so-called G3MP2large basis set.[11] This involves multiple valence shells, polarization through f orbitals, and diffuse functions. This step is maintained as is the choice of basis set, but two approximations are introduced to significantly reduce computation. The first approximation is to perform the underlying Hartree−Fock calculation in two steps using a so-called dual basis set.[12] SCF convergence is first reached using the much smaller 6-311G* basis set (a subset of the G3MP2large basis set), and, following this, corrections for the additional basis functions present in the G3MP2large basis set are introduced perturbatively. In practice, the dual basis set calculation is only slightly more costly than a single SCF cycle with the G3MP2large basis set, reducing computation time (for the Hartree−Fock part) by an order of magnitude or more.

The second approximation is the use of the RI-MP2 procedure[13] in lieu of MP2. This replaces an O(N^5) step with an O(M^3) step (where M is typically 2−3 times larger than N). In practice, an RI-MP2 energy calculation requires an order of magnitude less computation cost than the comparable MP2 calculation.

While total energy changes as a result of these two approximations, neither has a significant effect on calculated *relative* energies.[14]

4.G3(MP2) requires a QCISD(T) calculation[15] using the 6-31G* basis set. This step is O(N^7) and completely dominates the overall recipe for all but the smallest molecules. It is also the most demanding part of the G3(MP2) recipe in terms of memory and disk usage and is the major reason for practical size limitations. This step has been eliminated in T1, based on the observation that the MP2 and QCISD(T) correlation corrections in the G3(MP2) recipe are related provided that differences in the numbers of each element and each type of formal chemical bond are taken into account (see discussion following). The RI approximation to MP2, together with the use of a dual basis set to obtain the underlying Hartree−Fock wave function, strongly favors elimination of the QCISD(T) correction over the MP2 correction.

Without further modification, these changes to G3(MP2) result in heats of formation that are not sufficiently accurate to be useful in thermochemical calculations. Significant improvement results from using the calculated HF/6-31G* and MP2/G3MP2large energies, together with atom counts and Mulliken bond orders, as variables in a linear regression. The regression involves a total of 67 terms, although it can be argued that only 48 of these are involved in the fit. Eight terms involve atom counts and merely serve to convert between total energies and heats of formation,[16] and 11 terms refer to unique molecules (H_2, HF, HCl, HBr, N_2, F_2, FCl, FBr, Cl_2, ClBr, and Br_2) and, with the possible exception of the term for H_2, could easily be eliminated.

One thousand one hundred and twenty-six uncharged, closed-shell organic molecules containing only the elements H, C, N, O, F, S, Cl, and Br, and that are small enough for G3(MP2) calculations to be performed, have been used to determine linear coefficients in the regression.[17] This results in mean absolute and rms errors of 1.8 and 2.5 kJ/mol, respectively. A plot of T1 vs G3(MP2) heats of formation for this set is provided in Figure 1. The regression coefficients are provided in Appendix A.

Heats of Formation for Conformationally-Flexible Molecules. By assuming that a molecule adopts a particular shape, chemists are often able to draw parallels between molecular properties and structure. In practice, however, very few organic molecules are confined to one shape. Instead, they exist as a mixture of different shapes or conformers. This means that molecular properties correspond to averages over these conformers. In fact, except for very simple molecules such as

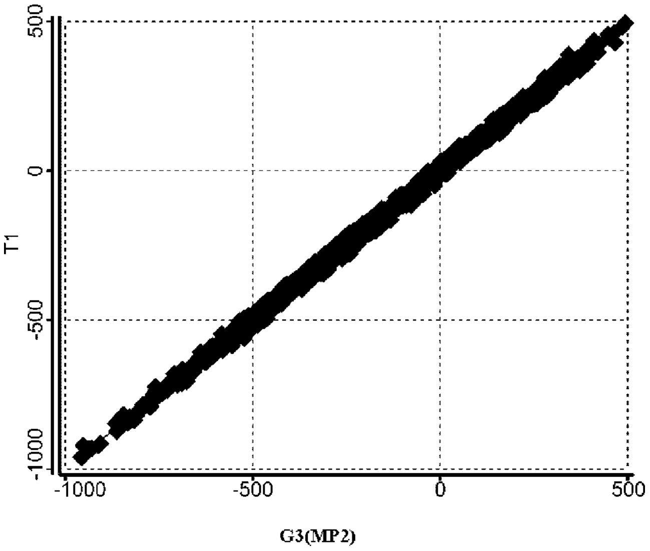


Figure 1. Comparison of heats of formation obtained from the T1 and G3(MP2) recipes (kJ/mol).

n-butane, it is generally not possible to experimentally establish the properties of individual conformers.

The average value of a property **A** may be calculated by summing over all available conformers. The value of the property for each conformer, a_i is multiplied by the product of the number of times that conformer appears, n_i, and its Boltzmann weight, W_i.

$$\mathbf{A} = \Sigma_i a_i n_i W_i$$

The Boltzmann weight depends on the energy of the conformer relative to the energy of the lowest-energy conformer, ΔE_i, and on the temperature, T. k is the Boltzmann constant.

$$W_i = \exp(-\Delta E_i/kT)/\sum_j n_j[\exp(-\Delta E_j/kT)]$$

In practice, calculation of average property values is difficult for two reasons. First, it is necessary to explicitly consider all possible conformers. Second, calculation of accurate Boltzmann weights (that is, accurate conformational energy differences) requires high-order correlation schemes and large basis sets. Accurate calculations of average property values are not likely to be practical for all but the simplest molecules.

In the case of interest here where the property is the energy, simplification follows by assuming that the energy of the lowest-energy conformer is the same as the Boltzmann average.[18] Identifying the lowest-energy conformer is easier than accounting for energy differences among conformers. Note that the energy of the lowest-energy conformer is necessarily lower than (or equal to) the average energy. In practice, for molecules with less than 10 degrees of conformational freedom (yielding hundreds to thousands of different conformers), the two differ at most by only 1−3 kJ/mol. The difference can become larger for molecules with many more degrees of freedom (polymers). The single (lowest-energy) conformer approximation will be used in the T1 recipe.

The only way to guarantee that the lowest-energy conformer will actually be found is to systematically examine all conformers. However, this approach rapidly becomes impractical with increasing number of degrees of conformational freedom. A common practical alternative to systematic searching is to sample conformations according to the so-called Monte-Carlo method.[19] This involves random changes to single-bond and ring conformations, with the decision to keep or discard a conformer (as a starting point for the next random move) depending on its energy relative to the lowest-energy conformer yet found. Practical experience suggests that the Monte-Carlo method nearly always locates the lowest-energy conformer (or a conformer that is very close in energy), even though it examines only a small fraction of the possible number of conformers.

The strength of the Monte-Carlo procedure, rapidly “zeroing in” on low-energy regions of conformation space, is also potentially its weak point. The reason for this apparent contradiction is that the energy function used for conformational searching will likely not be T1 (or another method that can realistically be expected to reliably account for conformer energy differences) but will instead be a less costly method based on a molecular mechanics or semiempirical molecular orbital procedure. It is not reasonable to expect that such simplified energy functions will always be able to distinguish low-energy conformers from higher-energy alternatives, in particular where significant nonbonded van der Waals or Coulombic interactions are possible. When they do not, a straightforward Monte-Carlo search may completely bypass “important regions” of conformational space.

A hybrid search procedure, referred to as *randomized systematic search*, together with the PM3 semiempirical model[20] in lieu of molecular mechanics as an energy function,[21] will be used in this work to provide an initial selection of conformers for later heat of formation calculations with T1. Conformers are “stepped through” as in a systematic search, but some are randomly discarded in order to maintain a fixed maximum total number of conformers. The remaining conformers are optimized with the PM3 method, duplicate conformers are eliminated, and T1 heats of formation (including reoptimization of geometry using the HF/6-31G* model as part of the T1 recipe) are calculated. Finally, the conformer with the lowest heat of formation is selected. In practice, the search has been limited to ~100 conformers (although this number is arbitrary and can be changed). This means that the procedure is identical to a full systematic search for molecules with up to 100 conformers.[22] While such a procedure requiring up to 100 T1 calculations is inefficient, it is still practical for molecules with molecular weights up to 400−500 amu.

For molecules with more than 100 conformers, a randomized systematic search does not guarantee that the conformer with the lowest T1 heat will actually be found. However, the fact that diverse regions of conformation space are sampled suggests that there is a high probability that a “good conformer” will be located.

Properties of the T1 Recipe. The T1 recipe as specified is well-defined both for rigid molecules as well as for molecules with fewer than 100 conformers (or whatever limit is placed on the number of conformers considered). The nature of the randomized systematic search introduces ambiguity for molecules with more conformers, as their selection depends both on the starting conformer as well as on the random number generator. It seems unlikely that the final heat of formation (based on the best conformer found) will differ significantly from one investigation to another, provided that the number of conformers examined is a sizable fraction of the actual number of conformers.

While the T1 recipe has the been extensively parametrized, given the size and diversity of the set of molecules used in the parametrization, it seems unlikely that any bias favoring or disfavoring particular types of molecules will be significant.

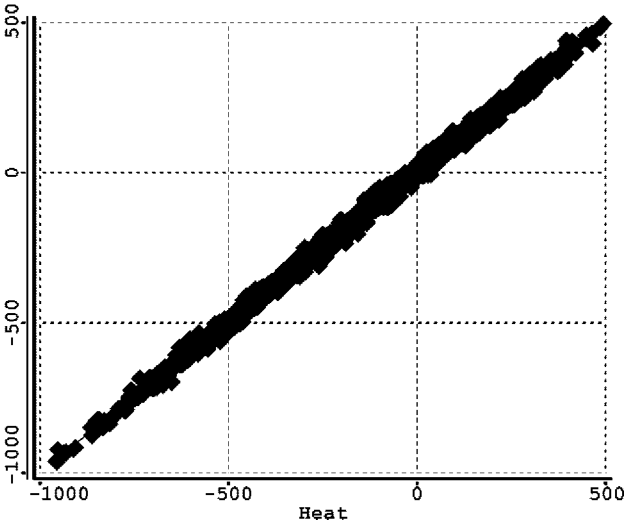


Figure 2. Comparison of T1 and experimental heats of formation for 1805 molecules from the NIST thermochemical database (kJ/mol).

The T1 recipe is not size-consistent. While is does not include the QCISD(T) term that destroys size consistency in G3(MP2), the bond parameters used in the regression depend on the local atomic environment. It is unclear whether the lack of size consistency will lead to significant errors.

Assessment of T1 Heats of Formation. T1 was developed to provide heats of formation that are sufficiently accurate to be used to supplement (or replace) experimental data in thermochemical comparisons. One thousand eight hundred and five experimental heats of formation found in the NIST thermochemical database (drawn from the full set of ~2000 uncharged, closed-shell organic molecules containing only H, C, N, O, F, S, Cl, and Br) have been employed to assess its overall performance and provide error bounds. Compounds with (experimental) heats outside the range of −1000 to +500 kJ/mol, as well as those for which the stated experimental errors are 10 kJ/mol or greater, have been excluded as have those based on mass spectrometry and other indirect measurement techniques. Also removed (following T1 calculations) are compounds for which calculated heats differ from experimental values by more that 35 kJ/mol. G3(MP2) calculations (where they have been practical) also show large deviations for these same compounds, leading us to conclude that the errors are due to the experimental data and not the calculations. Note that several of the excluded compounds have previously been flagged by Stewart as likely in error.[2]

Mean absolute and rms errors for T1 heats of formation vs experimental values for this set of molecules are 8.5 and 11.5 kJ/mol, respectively. Plots of T1 vs experimental heats of formation and of signed error vs number of molecules with this error are provided in Figures 2 and 3, respectively.

While G3(MP2) calculations are not practical for the majority of the molecules in the NIST database, they can be and have been performed on a set of 218 of the smaller molecules. Flexible molecules in this subset are assumed to be in the conformation found by T1. Compared to experimental values, G3(MP2) heats show mean absolute and rms errors of 6.2 and 8.3 kJ/mol, respectively, which are nearly identical to similar to mean absolute and rms errors of 6.6 and 9.0 kJ/mol, respectively, from the T1 recipe for this same subset of molecules.

Assessment of the T1 Model for Thermochemical Calculations. Heats of formation may be employed to establish reaction thermochemistry, for example, deciding which of several possible isomers is most stable. Because T1 is restricted to uncharged closed-shell (electron-paired) molecules, it may not be applied to several important classes of reactions, most notably reactions that lead to bond dissociation and reactions that compare acid and base strengths. Comparisons here are limited to the relative energies of structural, positional, and stereoisomers where ample gas-phase experimental thermochemical data exist with which to assess T1.

Calculated heats of formation may also be used to establish the preferred shape (conformation) of isolated molecules and determine conformational energy differences. Although a great deal is known from X-ray diffraction about the shapes of molecules packed into crystalline solids, it is not clear to what extent their shapes are influenced by environment. Also, crystal structures do not provide the relative energies of alternative (higher-energy) conformers. The preferred shapes of several hundred very simple molecules (with one or two degrees of conformational freedom) have been established in the gas phase by microwave spectroscopy, but information about the shapes of alternative conformers and conformational energy differences is far more limited. Therefore, assessment of T1 is limited.

We have compared T1 isomer and conformer energies results not only with experimental data but also with the results of G3(MP2) calculations. The lack of experimental data and uncertainties in much of the data that do exist make the latter of considerable practical importance.

Structural Isomers. We use the term *structural isomer* to associate molecules that have the same total number of single-bond equivalents but differ in the number of each kind of chemical bond (as defined by element and hybridization) and/or in the number of each kind of ring. Thus, we refer to 1-butyne (one single bond involving two sp^3 hybrids, one single bond involving sp and sp^3 hybrids, and one triple bond) and 2-butyne (two single bonds involving sp and sp^3 hybrids and one triple bond) as structural isomers.

Table 1 compares structural isomer energies obtained from T1 heats of formation for a variety of simple systems with those obtained from experimental heats. The mean absolute error is 5 kJ/mol, significantly lower than achieved previously with practical Hartree−Fock, density functional, and MP2 models for similar sets of structural energy comparisons.[9] Equally important, the largest individual errors are on the order of 10 kJ/mol, also less that previously achieved.

Energy differences from G3(MP2) heats are also provided. These yield a similar mean absolute error (4 kJ/mol) and show similar individual errors.

Positional Isomers. We use the term *positional isomer* to associate molecules that have the same number of each kind of bond but with different bond arrangements. Ortho-, meta-, and para-xylene are positional isomers as are 1,3-dioxane and 1,4-dioxane.

Table 2 compares positional isomer energies obtained from T1 heats of formation for a variety of simple systems with those obtained from experimental heats. The mean absolute error is only 2 kJ/mol or about half of that found for comparisons involving structural isomers. This is consistent with the fact that the comparisons here are more subtle and that energy differences among positional isomers are typically much smaller than those between structural isomers. There are some significant errors in isomer assignments. For example, T1 incorrectly assigns

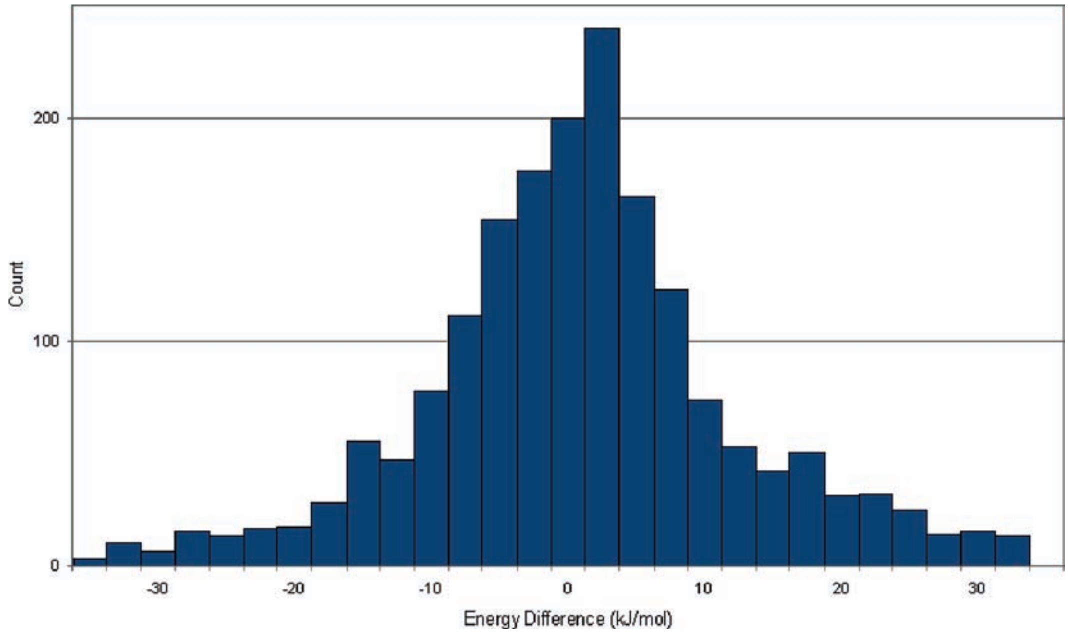


Figure 3. Signed errors in T1 heats of formation for 1805 molecules from the NIST thermochemical database.

ortho-xylene as 2 kJ/mol more stable that meta-xylene, in contrast to the experimental result where ortho-xylene is less stable by 2 kJ/mol.

Energy differences obtained from G3(MP2) heats are similar to those from the T1 calculations, although some comparisons (the xylenes) are better described by the G3(MP2) recipe than by T1 and vice versa. The overall mean absolute errors are the same for the two recipes.

Stereoisomers. We use the term *stereoisomer* to associate molecules that have exactly the same bonding and differ only in their three-dimensional geometries. Cis and trans disubstituted alkenes and cis and trans disubstituted cycloalkanes are examples.

Table 3 compares stereoisomer energies obtained from T1 heats of formation for a variety of simple systems with those obtained from experimental heats. In general, the agreement between T1 and experimental energy differences is excellent. The worst case is the cis−trans energy difference in cyclooctene, which is smaller than the experimental preference (43 kJ/mol in favor of the cis isomer) by 8 kJ/mol. Another large discrepancy is that T1 finds the trans isomer of 1,2-dichloroethylene to be 3 kJ/mol more stable than the cis isomer. Experimentally, the trans isomer is less stable by 3 kJ/mol. All remaining errors are smaller. As with the previous isomer comparisons, heats obtained from G3(MP2) calculations yield similar energy differences, in terms of both mean absolute error and individual errors.

In summary, diverse energy comparisons show that both T1 and G3(MP2) recipes generally provide heats of formation data that are sufficiently accurate to be useful for thermochemical calculations. While there are significant exceptions, and while some of these could be due to errors in the experimental data, it is likely that other discrepancies are symptomatic of limitations in the G3(MP2) recipe on which T1 is based. That is to say, problems with T1 are also problems with G3(MP2). We see no benefit in performing G3(MP2) calculations where the T1 recipe can be applied.

Performance of the T1 Model for Conformational Energy Differences. As commented previously, the success of T1 to accurately calculate heats of formation for flexible molecules

TABLE 1: Comparison of Energies of Structural Isomers from T1 and G3(MP2) Recipes with Experimental Values (kJ/mol)

formula (reference)	isomer	T1	G3(MP2)	expt.
C_2H_3N (acetonitrile)	methyl isocyanide	103	100	88
C_2H_4O (acetaldehyde)	vinyl alcohol	41	41	43
	oxirane	113	115	118
$C_2H_4O_2$ (acetic acid)	methyl formate	68	70	75
C_2H_6O (ethanol)	dimethyl ether	48	50	51
C_3H_4 (propyne)	allene	−3	1	7
	cyclopropene	92	100	93
C_3H_6 (propene)	cyclopropane	36	38	29
C_4H_6 (1,3-butadiene)	2-butyne	35	37	36
	cyclobutene	49	56	48
	1,2-butadiene	43	51	53
	1-butyne	56	58	56
	methylenecyclopropane	80	85	92
	bicyclo[1.1.0]butane	113	120	108
C_4H_6O (methyl vinyl ketone)	*trans*-2-butenal	7	6	5
	2-methyl-2-propenal	4	3	9
	cyclobutanone	25	23	23
	2-hydroxy-1,3-butadiene	48	38	38
	2,3-dihydrofuran	35	35	43
	divinyl ether	96	97	102
C_4H_8 (2-methylpropene)	*trans*-2-butene	5	6	7
	1-butene	15	16	17
	cyclobutane	46	47	46
C_5H_8 (cyclopentene)	2-methyl-1,3-butadiene	42	37	40
	methylenecyclobutane	89	87	86
	1,4-pentadiene	69	67	70
	1,1-dimethylallene	86	90	93
	1,2-pentadiene	98	101	105
mean absolute error		5	4	−

TABLE 2: Comparison of Energies of Positional Isomers from T1 and G3(MP2) Recipes with Experimental Values (kJ/mol)

molecule	T1	G3(MP2)	expt.
1,1-dimethylallene			
1,3-dimethylallene	4	6	4
2-methyl-1,3-butadiene			
trans-1,3-pentadiene	1	3	0
2-methylpyridine			
3-methylpyridine	7	7	5
4-methylpyridine	5	5	5
2-aminopyridine			
4-aminopyridine	16	15	12
3-aminopyridine	26	27	26
m-xylene			
p-xylene	0	3	1
o-xylene	−2	2	2
1,3,5-trimethylbenzene			
1,2,4-trimethylbenzene	−1	−1	2
1,2,3-trimethylbenzene	3	4	6
2-methylnaphthalene			
1-methylnaphthalene	1	1	1
m-difluorobenzene			
p-difluorobenzene	2	2	3
o-difluorobenzene	15	15	26
p-dichlorobenzene			
m-dichlorobenzene	0	0	4
o-dichlorobenzene	5	6	8
1,4-dicyanobenzene			
1,3-dicyanobenzene	3	1	4
1,2-dicyanobenzene	6	7	9
1,3-dioxane			
1,4-dioxane	23	24	23
mean absolute error	2	2	−

TABLE 3: Comparison of Energies of Cis−Trans Stereoisomers from T1 and G3(MP2) Recipes with Experimental Values ($E_{\mathrm{trans}} - E_{\mathrm{cis}}$, kJ/mol)

molecule	T1	G3(MP2)	expt.
1,2-dimethylcyclopentane	6	5	7
1,3-dimethylcyclopentane	−1	−1	2
1,2-dimethylcyclohexane	7	7	8
1,3-dimethylcyclohexane	−8	−7	−8
1,4-dimethylcyclohexane	7	7	8
decalin	11	11	13
2-butene	5	5	3
1,3-pentadiene	5	5	7
1,2-dichloroethylene	−3	−3	3
cyclooctene	−35	−34	−43
bicyclo[4.1.0]hept-3-ene	−110	−112	−112
mean absolute error	2	2	−

TABLE 4: Comparison of Conformer Energy Differences in Acyclic Molecules from T1 and G3(MP2) Recipes with Experimental Values ($E_{\mathrm{low\text{-}energy}} - E_{\mathrm{high\text{-}energy}}$, kJ/mol)

molecule	low-energy/ high-energy conformer	T1	G3(MP2)	expt.
n-butane	trans/gauche	2.7	2.7	2.8
1-butene	skew/cis	0.8	0.3	0.9
1,3-butadiene	trans/gauche	10	13	12
acrolein	trans/cis	7.3	9.0	7.1
N-methylformamide	trans/cis	5.6	5.6	5.9
N-methylacetamide	trans/cis	11	9.5	9.6
formic acid	cis/trans	19	17	16
methyl formate	cis/trans	22	21	20
methyl acetate	cis/trans	32	31	36
propanal	eclipsed/anti	4.4	4.2	2.8
2-methylpropanal	eclipsed/anti	2.4	2.1	3.3
1,2-difluoroethane	gauche/anti	2.6	3.3	2.3
1,2-dichloroethane	anti/gauche	4.9	4.6	4.5
ethanol	anti/gauche	0.0	0.5	0.5
methyl ethyl ether	anti/gauche	6.1	5.8	6.3
methyl vinyl ether	cis/skew	8.7	9.4	7.1
mean absolute error		1.2	1.1	–

TABLE 5: Comparison of Equatorial−Axial Energy Differences in Cyclic Molecules from T1 and G3(MP2) Recipes with Experimental Values ($E_{\mathrm{equatorial}} - E_{\mathrm{axial}}$, kJ/mol)

molecule	T1	G3(MP2)	expt.
methylcyclohexane	7.0	5.9	7.3
tert-butylcyclohexane	21	20	23
cis-1,3-dimethylcyclohexane	22	22	23
fluorocyclohexane	0.1	−0.9	0.7
chlorocyclohexane	1.7	0.8	2.1
piperidine	3.9	2.8	2.2
N-methylpiperidine	16	15	13
2-chlorotetrahydropyran	−12	−12	−7.5
2-methylcyclohexanone	7.2	7.3	8.8
3-methylcyclohexanone	3.6	4.2	5.7, 6.5
4-methylcyclohexanone	5.4	6.1	7.3, 8.8
mean absolute error	1.7	1.8	−

is predicated on its ability to identify the lowest-energy conformer (or a conformer that is among the lowest-energy conformers). Experimental data with which to assess the success or failure of the T1 recipe are limited. The structures of several hundred small molecules for which more than one conformer is possible are known in the gas phase primarily from microwave spectroscopy and electron diffraction.[23] (The crystal structures of upward of 250 000 flexible organic molecules have been established by X-ray diffraction,[24] but the energy of crystal packing is of the same order of magnitude as energy differences among conformers, and it is not unreasonable to expect changes in conformation from those of isolated molecules.) Energy differences between conformers have been established experimentally for only a very few molecules. Most of the experimental data derive from equilibrium measurements in the gas phase or in dilute solution and would be expected to be less and less precise with increasing difference in conformer energy.

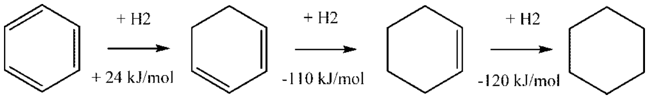
Conformer energy differences for a small selection of acyclic molecules obtained from the T1 recipe are compared with experimental values in Table 4. The level of agreement between T1 and experimental conformational energy differences is excellent. The correct conformer is assigned in all cases, and the mean absolute error is only 1.2 kJ/mol. Individual energy differences obtained from the G3(MP2) recipe are very similar to those from T1, and the two sets show a similar mean absolute error.

T1 and experimental conformational energy differences for a small selection of cyclic molecules are compared in Table 5. As with the acyclic molecules, the T1 calculations properly assign the lowest-energy conformer in all cases, and the mean absolute is only 1.7 kJ mol. G3(MP2) results are very similar, and only for the fluorocyclohexane is the conformer assignment incorrect. Note, however, that the difference in conformer energies is <1 kJ/mol.

Overall, T1 provides the best performance of any *practical* procedure that we have examined for these two test suites of simple flexible molecules. Given the uncertainties in the experimental data that do exist, it is unlikely that errors can be significantly reduced. Because the experimental data are very limited, these comparisons do not provide an adequate account of the efficacy of T1 for conformational energy comparisons. However, the high level of agreement lends credence (or at least encouragement) to the ability of the recipe to properly assign the lowest-energy conformer (or low-energy conformer) for more complex systems where experimental data are not available.

Illustrative Applications. Several applications make use of heats of formation from T1 calculations. These are intended to illustrate the types of questions that can be posed and answered were accurate heats of formation readily available. All the molecules considered are included in the Spartan Molecular Database.[9]

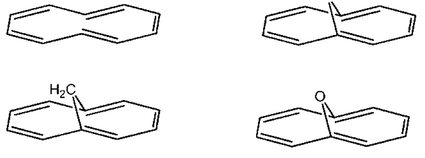
"Quantifying" Aromaticity. While the concept of aromaticity is deeply rooted in chemistry, the degree to which a particular molecule is *aromatic* is, at best, imprecise. The absence of a metric makes it difficult to compare the aromaticity of different molecules or to design a molecule with high aromaticity. One reasonable measure of the aromatic stabilization of hydrocarbons is the difference in the energy of adding one equivalent of hydrogen to an aromatic and the energy of adding a second equivalent to the initial hydrogenation product. While both hydrogenation reactions lead to two new CH bonds at the expense of an HH bond and a CC π bond, only the first reaction destroys aromaticity. For example, experimental heats of formation show that addition of hydrogen to benzene to give 1,3-cyclohexadiene is endothermic by 22 kJ/mol while the addition of hydrogen to 1,3-cyclohexadiene to give cyclohexene is exothermic by 109 kJ/mol. (The final hydrogenation step, taking cyclohexene to cyclohexane, is exothermic by 120 kJ/mol.)



The 131 kJ/mol difference in energy between the two hydrogenation reactions can be interpreted as due to the aromaticity of benzene.

Experimental data to allow the aromaticity of naphthalene, anthracene, and phenanthrene to be quantified are lacking. The T1 model shows a difference between the first and second hydrogenation energies in benzene of 140 kJ/mol (vs 131 kJ/mol from experimental heats). They show lesser effects due to the loss of one of the fused aromatic rings in naphthalene (81 kJ/mol), anthracene (58 kJ/mol), and phenanthrene (99 kJ/mol).

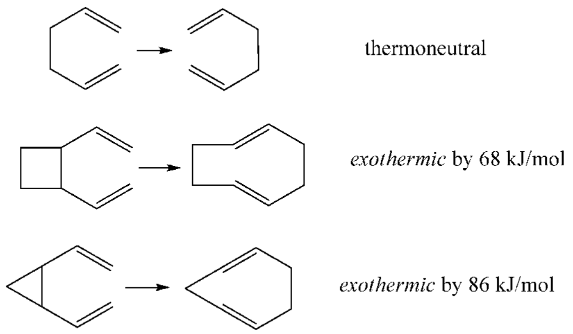
The calculated geometry of (1*Z*,3*E*,5*Z*,7*E*,9*Z*)-cyclodeca-1,3,5,7,9-pentaene exhibits localized single and double bonds. It is not surprising, therefore, that the difference in first and second hydrogenation energies is only 17 kJ/mol. (The first hydrogenation reaction is actually more exothermic, suggesting relief of strain.) Clearly, the molecule is not aromatic. However, CH_2, NH, and O bridged analogues all show delocalized bonding and significant differences between first and second hydrogenation energies: 66 kJ/mol for 1,6-methano[10]annulene, 80 kJ/mol for 1,6-imido[10]annulene, and 55 kJ/mol for 1,6-oxido[10]annulene.



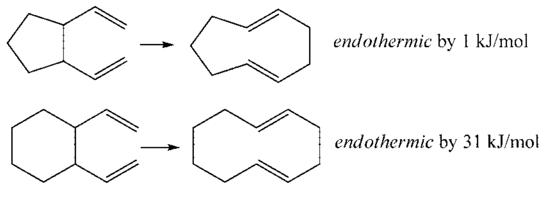
Trans Cycloalkenes. Alkyl groups attached to the two ends of a double bond almost always prefer to be trans rather than cis. For example, *trans*-2-butene is more stable than *cis*-2-butene by 3 kJ/mol, while *trans*-2,2,2,5-tetramethyl-3-hexene is 43 kJ/mol more stable than the corresponding cis isomer. The common explanation is that cis isomers are more crowded and unfavorable steric interactions are of greater detriment. Small- to medium-ring cycloalkenes are exceptions to the rule. The heat of formation of *trans*-cyclooctene is known to be 43 kJ/mol higher than that of the cis isomer, but experimental thermochemical data for other cycloalkenes are not available. It is interesting to ask how unfavorable the trans stereoisomer for cycloheptene is likely to be, and at what ring size would the isomer energies of cycloalkanes mimic those of acyclic alkenes.

T1 calculations show that the cis isomer is favored for cyclododecene and smaller-ring cycloalkenes and provide the following cis−trans energy differences: cycloheptene (114 kJ/mol), cyclooctene (8 kJ/mol smaller than the experimental value), cyclononene (13 kJ/mol), cyclodecene (15 kJ/mol), cycloundecene (2 kJ/mol), and cyclododecene (1 kJ/mol).

Thermodynamic Driving Force. The *Hammond Postulate* states that the more exothermic a reaction, the more its transition state resembles reactants. This is the basis for the notion of *thermodynamic driving force*. The more the geometry of a transition state resembles the geometry of reactants, the closer its energy (and the lower the activation barrier). For example, 1,5-hexadiene undergoes Cope rearrangement (to itself) more slowly than does *cis*-1,2-divinylcyclobutane (to 1,5-cyclooctadiene). The former is thermoneutral, whereas experimental thermochemical data show that the latter is exothermic by 87 kJ/mol. The latter is presumably due to relief of strain. Rearrangement of *cis*-1,2-divinylcyclopropane to 1,4-cycloheptadiene is even more rapid, consistent with the notion that a three-membered ring is more strained than a four-membered ring. Experimental thermochemical data are unavailable. T1 calculations provide support for what is known experimentally and fill in the blanks.

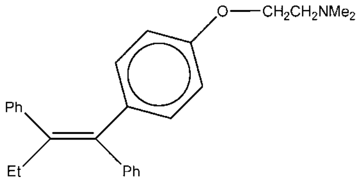


Cope rearrangements of *cis*-1,2-divinylcyclopentane to 1,5-cyclononadiene and *cis*-1,2-divinylcyclohexane to 1,5-cyclodecadiene would not be expected to benefit from loss of ring strain in the reactants. Experimental data are not available, but T1 calculations show that the Cope rearrangement of divinylcyclopentane is essentially thermoneutral while the corresponding rearrangement of divinylcyclohexane is actually endothermic.



Gas- Versus Solid-Phase Conformations. The gas-phase structures and conformations of ~1000 small molecules have been established experimentally, primarily from microwave spectroscopy. For the most part, the data are limited to very simple systems with one or at most two degrees of conformational freedom. Far more is known about the conformations of molecules as crystalline solids. X-ray diffraction is now a routine experiment, and the structures (conformations) of ~250 000 organic molecules have been determined and are available in the Cambridge Structural Database.[24] What is not evident, however, is whether the conformations in the solid-phase structures are the same as (or very similar to) those of isolated molecules, or whether environmental factors (packing forces) are large enough to override any native preferences. Unfortunately, very few molecules have been characterized both in the gas and in the solid phase.

The solid-phase structure of the breast cancer drug tamoxifen has been established.[25] Although this molecule is fairly large (molecular weight 372 amu), it has only 216 single-bond conformers. The T1 procedure, which limits the number of conformers examined to 100, should be acceptable.



T1 heats of formation for the "best" (gas-phase) conformer and that found in the crystal differ by only 5 kJ/mol. The two arrangements differ only by torsion about a single bond.

Conclusion

Heats of formation make it possible to determine whether a particular chemical reaction will be favorable (exothermic) or unfavorable (endothermic) and quantify the effects that changes in structure have on thermochemistry. However, experimental data are limited, primarily because the measurements (calorimetry) typically require and destroy significant amounts of material. The T1 procedure allows routine calculation of the heats of formation of uncharged, closed-shell molecules comprising H, C, N, O, F, S, Cl, and Br with molecular weights up to 400−500 amu. It is based on the G3(MP2) recipe but requires 2−3 orders of magnitude less computation. This has been achieved by substituting a HF/6-31G* geometry for the MP2/6-31G* geometry, eliminating the HF/6-31G* frequency and the QCISD(T)/6-31G* calculations and approximating the large basis set MP2 calculation by a dual basis set RI-MP2 calculation. Atom counts and Mulliken bond orders, together with the HF and RI-MP2 energies, are treated as variables in a linear regression fit of T1 to G3(MP2) heats for 1128 molecules resulting in mean absolute and rms errors of 1.8 and 2.5 kJ/mol, respectively. A comparison of T1 heats of formation with experimental values for 1817 drawn from the full NIST thermochemical database results in mean absolute and rms errors of 8.7 and 11.9 kJ/mol, respectively.

The T1 recipe has been applied to conformationally flexible molecules as well as to rigid molecules. These have been treated by randomly sampling up to 100 diverse conformers, performing a T1 calculation on each and selecting the conformer with the lowest heat of formation.

Times for T1 Calculations. As previously commented, computation times for T1 calculations are 2−3 orders of magnitude less than those for G3(MP2) calculations. While all four simplifications discussed earlier contribute to the savings, the primary reason for molecules with molecular weights >150 amu is the elimination of the QCISD(T)/6-31G* calculation in the G3(MP2) recipe. This scales as $O(N^7)$, where N is a measure of size, whereas the highest-order step in T1 (the RI-MP2 calculation) scales as $O(N^5)$.

T1 is routinely applicable for organic molecules of moderate size on "consumer grade" personal computers running Windows, Macintosh, or LINUX. A T1 energy calculation for cyclododecene, $C_{12}H_{22}$, with 628 basis functions with the G3MP2large basis set, requires on the order of 30 min using one processor of a 2.2 GHz Intel Core Two Duo chip, whereas the comparable calculation on tamoxifen, $C_{26}H_{29}NO$, with 1242 basis functions, requires on the order of 4 h.[26]

Implementation of T1 in Spartan. A preliminary version of the T1 recipe was implemented in the *Spartan'06* electronic structure program.[27] This paper describes an improved procedure that is provided with *Spartan'08*[28,29] and which leads to slightly different heats.

T1 Database of Heats of Formation. A database of heats of formation obtained from the T1 recipe has been developed. At the present time, it comprises ~40 000 closed-shell uncharged molecules involving H, C, N, O, F, S, Cl, and Br only, including ~2000 molecules for which experimental data are available. This collection is part of the Spartan Molecular Database (SMD). Conformationally flexible molecules are handled as detailed in this paper (considering up to 100 randomly selected conformers), but only the lowest-energy conformer is included in SMD. Note that the T1 collection differs from the other (much larger) collections of organic molecules that make up SMD, in that it assigns the lowest-energy conformer based on the T1 recipe and not on MMFF molecular mechanics. We suggest that this provides much more realistic representation of conformational preferences than previously available.

Appendix A: Derivation of T1 Parameters

A linear regression has been performed in order to fit 1128 G3(MP2) heats of formation. Parameters are the HF/6-31G* and MP2/G3MP2large energies, atom counts, and Mulliken bond orders (from HF/6-31G* calculations). The regression involves a total of 67 terms, although only the leading constant, HF/6-31G* and MP2/G3MP2large energies and 47 of the bond-order terms are relevant to the fit. Eight terms are atom counts and together with the leading constant merely

TABLE A1: Atomic Parameters

atom	parameter
hydrogen	1 441.645 8
carbon	94 905.481 6
nitrogen	136 359.910 1
oxygen	187 186.165 4
sulfur	992 554.649 9
fluorine	248 572.715 7
chlorine	1 147 206.4
bromine	6 420 784.9

TABLE A2: Mulliken Bond-Order Parameters

bond	bond order	(bond order)^2	(bond order)^3	(bond order)^4	(bond order)^5
CH	−	8.334 058 7	−	−	−
NH	−	95.930 156 6	−70.978 139 3	−3.611 542 9	−
OH	−	−579.752 762 4	1 805.010 9	−1 282.553 0	
SH	−	8.728 423 9			
CC	−73.674 702 2	171.809 443 4	−147.132 791 9	50.171 505 2	−5.916 149 3
CN	11.700 617 4	16.417 486 4	−18.011 766 3	4.197 480 8	−
CO	21.334 624 0	7.557 530 6	−4.662 641 8	−	−
CF	−2.222 871 5	−	−	−	−
CS	0.121 763 691 0	−	−	−	−
CCl	−16.685 134 9	−	−	−	−
CBr	26.033 730 7	−	−	−	−
NN	−60.853 359 4	202.438 240 3	−137.641 015 1	28.516 181 6	−
NO	−7.623 534 4	88.002 765 0	−36.224 265 3	−	−
NF	13.227 751 8	−	−	−	−
NS	19.577 963 9	−	−	−	−
NCl	−2.772 988 0	−	−	−	−
NBr	46.735 315 4	−	−	−	−
OO	49.551 766 9	−	−	−	−
OF	20.911 386 0	−	−	−	−
OS	30.142 011 8	1.472 619 5	−	−	−
OCl	4.904 372 0	−	−	−	−
OBr	−81.626 915 8	−	−	−	−
FS	−6.242 945 4	−	−	−	−
SS	5.381 767 4	−	−	−	−
SCl	−17.657 519 2	−	−	−	−
SBr	33.387 663 9	−	−	−	−

TABLE A3: Mulliken Bond-Order Parameters for Unique Molecules

bond	bond order
H_2	8.619 982 2
HF	1.353 762 0
HCl	−16.646 097 4
HBr	31.010 632 1
N_2	2.45
F_2	−6.026 766 6
FCl	−19.858 617 7
FBr	−111.515 395 5
Cl_2	−43.821 173 68
ClBr	11.397 204 4
Br_2	150.576 194 7

serve to move the zero in total energy (infinitely separated nuclei and electrons) to the zero in heat of formation (standard molecules at 298 K). The remaining 11 terms refer to unique molecules (H_2, HF, HCl, HBr, N_2, F_2, FCl, FBr, Cl_2, ClBr, and Br_2) and, with the possible exception of the term for H_2, could be eliminated.

The leading constant and the terms multiplying the HF/6-31G* and RI-MP2/G3MP2large energy are as follows:

$$-2.895\,264\,9 + 314.924\,095 * E(\text{HF/6-31G*}) + 2\,180.815\,87 * E(\text{RI-MP2/G3MP2large})$$

The eight parameters associated with atom counts are given in Table A1. Bond-order parameters (except those for unique molecules) are given in Table A2. These are the coefficients of a polynomial involving the Mulliken bond order as obtained form the HF/6-31G* wave function. Coefficients multiplying the Mulliken bond order for 11 unique molecules (H_2, HF, HCl, HBr, N_2, F_2, FCl, FBr, Cl_2, ClBr, and Br_2) are provided in Table A3.

References and Notes

(1) Burgess, D. R.; Linstrom, P. J.; Mallard, W. G. *NIST Chemistry Webbook*; NIST Standard Reference Data: Gaithersburg, MD, 2005; p *69.*

(2) (a) Stewart, J. J. P. J. *Phys. Chem. Ref. Data* **2004**, *33*, 713. (b) Stewart, J. J. P. J. *Mol. Modell.* **2004**, *10*, 6.

(3) Purely empirical procedures such as the Benson additivity rules (Benson, S. W.; Cruickshank, F. R.; Golden, D. M.; Haugen, G. R.; O'Neal, H. E.; Rodgers, A. S.; Shaw, R.; Walsh, R. *Chem. Rev.* **1969**, *69*, 279.) have long been available and are moderately successful for small molecules. However, nonbonded interactions in larger systems are not likely to be additive, and simple rules that do not take them explicitly into account are not likely to provide an adequate account. .

(4) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1990**, *110*, 4703.

(5) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1991**, *91*, 7221.

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

(7) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *8*, 84108.

(8) The decision to reproduce heats of formation from G3(MP2) calculations as opposed to experiment is deliberate. The G3(MP2) calculations provide a much more uniform account that is unaffected by uncertainties in the experimental measurements.

(9) The Spartan Molecular Database (SMD) is a collection of ~150 000 (primarily) organic molecules, the structures, energies, properties, and spectra for which have been obtained from one or more quantum chemical models. Models include Hartree−Fock models with 3-21G, 6-31G*, and 6-311+G** basis sets, the EDF1/6-31G* density functional model, B3LYP density functional models, and MP2 models with 6-31G* and 6-311+G** basis sets and the G3(MP2) model. For information on access and availability, see http//:www.wavefun.com/products/databases.html.

(10) Hehre, W. J. A. *Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Inc.: Irvine, CA, 2003; Chapter 5.

(11) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703–4709.

(12) Liang, W. Z.; Head-Gordon, M. J. *Phys. Chem. A* **2004**, *108*, 3206.

(13) (a) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, 143. (b) Head-Gordon, M. *Mol. Phys.* **1999**, *96*, 673.

(14) (a) DeStasio, R. A.; Jung, Y.; Head-Gordon, M. J. *Chem. Theor. Comput.* **2005**, *1*, 862. (b) Steele, R. P.; DeStasio, R. A.; Shao, Y.; Kan, J.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 74108.

(15) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(16) Total energy is defined as the energy of a hypothetical reaction in which the molecule is split up into its constituent nuclei and electrons at infinite separation.

(17) Most of these molecules are rigid (or have a favored conformer that is easily identified).

(18) Note that the average value of a property other than the energy will not likely be properly described by the value of the property for the lowest-energy conformation. Consider the dipole moment of 1,3-butadiene as a case in point. The measured value is small but nonzero, even though the dipole moment of the lowest-energy (trans) conformer is zero due to symmetry. Contributions from the higher-energy (gauche) conformer are responsible.

(19) Metropolis, N.; Rosenbluth, A. E.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1057.

(20) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

(21) The reason for this is that molecular mechanics in general and the MMFF method in particular (Halgren, T. A. *J. Comput. Chem.* **1996**, *17,* 490 and following papers in this issue) overestimates the strengths of (intramolecular) hydrogen bonds. In some cases, this leads to conformers that lack hydrogen bonds (or with fewer than the maximum possible number of hydrogen bonds) to not be recognized as energy minima. .

(22) This assumes that all of the starting conformers obtained from PM3 are energy minima at HF/6-31G*.

(23) The MOGADOC database (www.uni-ulm.de/strudo/mogadoc/index) contains a large collection of small molecules.

(24) Allen, F. A. *Acta Crystallogr.* **2003**, *B58*, 380.

(25) Precigoux, G.; Courseille, C.; Geoffre, S.; Hospital, M. *Acta Crystallogr.* **1979**, *B35*, 3070.

(26) The total computation time includes structure determination using the HF/6-31G* model. For small molecules (<150 amu), this may dominate depending on the quality of the initial guess.

(27) *Spartan '06 Tutorial and User's Guide*; Wavefunction, Inc: Irvine, CA, 2006.

(28) *Spartan '08 Tutorial and User's Guide*; Wavefunction, Inc.: Irvine, CA, 2008.

(29) Since the submission of this paper, T1 has been extended to include Si and P. This extension and ∼5000 new database entries for silicon and phosphorus-containing compounds will be made available in the coming months.

JP810144Q