

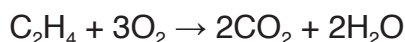
THERMOCHEMICAL RECIPES AND CALCULATING ACCURATE HEATS OF FORMATION

The “energy” of a molecule is most commonly reported as a heat of formation. As noted in the previous topic, this is defined as the enthalpy at 298.15K 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. For example, the heat of formation of ethylene corresponds to the enthalpy of a reaction to yield graphite and molecular hydrogen.



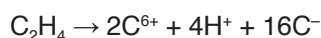
Differences in heats of formation between the products and reactants (reaction enthalpies) indicate the extent to which the reaction will be favorable (*exothermic*) or unfavorable (*endothermic*), and allow thermodynamic product distributions to be established.*

In almost all cases, the heat of formation is obtained from a heat of combustion. For example, the heat of formation of ethylene would likely have been established from its reaction with oxygen to produce carbon dioxide and water.



Experimental heats of formation are available for ~2000 compounds. While much of the data is accurate to within 4-8 kJ/mol, a significant portion is subject to greater uncertainty. The most egregious source of error is that the reported heat actually does not correspond to the reported structure. More common sources of error include impure samples, incomplete combustion and most importantly, poorly characterized combustion products. Hydrocarbons and oxycarbons present fewest problems as combustion leads only to carbon dioxide and water. However, combustion of molecules with other elements may give rise to a complex mixture of products and greater uncertainty, with nitrogen compounds being particularly problematic.

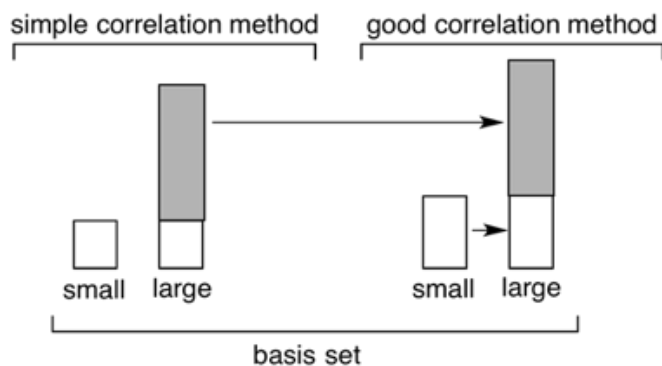
* Of course, other references are equally suitable for examining mass balanced reactions. For example, total energies from quantum chemical calculations are referenced to separated nuclei and electrons, in the case of ethylene.



Despite their fundamental importance, heats of formation are not routinely determined for new compounds. While the combustion experiment is straightforward and does not require particularly expensive instrumentation, accurate measurements may require (and will likely destroy) significant quantities of compound. Very few chemists are willing to part with hundreds of mg of a compound that they have just spent days, weeks or months preparing.

Because of the lack of experimental data and serious concerns over the accuracy of part of the data that do exist, considerable attention has been directed at the use of quantum chemical calculations to obtain heats of formation. One avenue that has been pursued relies on the knowledge that CCSD (T) calculations (coupled cluster singles and doubles with triples corrections introduced perturbatively) reliably reproduce the energetics of a wide variety of organic reactions. The problem is that the singles and doubles contribution scale as $O(\eta^6)$, where η is the number of basis functions, and the triples correction scales as $O(\eta^7)$. This limits practical applications of the method to very small molecules (<15 non-hydrogen atoms).

A simplified account of the “solution” first proposed by John Pople is depicted in the next page. The essential idea is that the energy obtained using a “good” electron correlation method and a large basis set may be closely approximated by combining the energy calculated using the good correlation method with a small basis set with difference in calculated energies for a “simple” correlation method with small and large basis sets. In principle, only three calculations are required: small and large basis set calculations with the simple correlation method and a small basis set calculation with the good correlation method. In practice, some of the schemes that have been formulated do exactly this while others employ a combination of simple correlation methods.



The original Pople implementation and subsequent developments primarily by Larry Curtiss are now collectively known as Gx methods. G1 relied on MP4 as the “good” correlation method with a gradual shift to the more costly and more accurate CCSD(T) method. To some extent, the overall objective has also broadened, from a means to approximate correlation energy to a stand-alone source of thermochemical data (298° heats of formation). Our focus is on the former, as possible low-cost replacements of the CCSD(T) correlation method.

The two most widely-used methods are G3(MP2) and G3, and the two most recent additions to the series are G4(MP2) and G4. A summary of the procedures involved in calculating electronic energy using these four methods is provided on the next page. The full recipes lead to heat of formation in addition to electronic energy, and require specifying an equilibrium geometry and a procedure for obtaining vibrational frequencies in addition to “empirical” corrections to account for the standard states of each of the elements. While each the four procedures involves several steps, these can be carried out in proper sequence without user intervention.

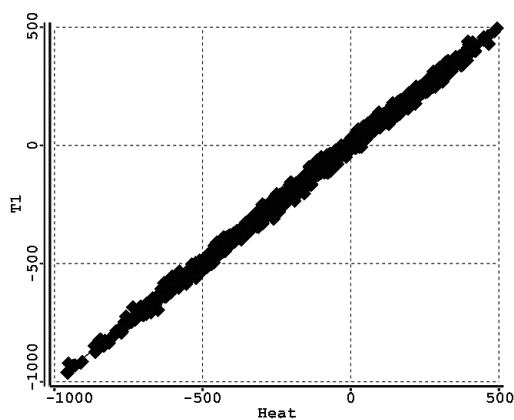
	simple correlation method	good correlation method	limiting steps
G3(MP2)	MP2	QCISD(T)/6-31G*	T in QCISD(T)
G3	MP2 & MP4	QCISD(T)/6-31G*	MP4 and T in QCISD(T)
G4(MP2)	MP2	CCSD(T)/6-31G*	T in CCSD(T)
G4	MP2 & MP4	CCSD(T)/6-31G*	MP4 and T in CCSD(T)

The T1 Thermochemical Recipe

The simplest of the GX recipes, termed G3(MP2), involves several costly steps, most significantly an MP2/6-31G* geometry calculation, a HF/6-31G* frequency calculation and a QCISD(T)/6-31G* energy calculation. In practice, G3(MP2) scales as the 7th power of size and is applicable only to molecules with molecular weights less than 150 amu. Clearly, even simpler procedures are required for routine application to larger molecules.

The goal behind the T1 recipe was to maintain the accuracy of G3(MP2) but at significantly reduced computation cost. It is limited to uncharged, closed-shell molecules comprising H, C, N, O, F, Si, P, S, Cl and Br. T1 substitutes the MP2/6-31G* geometry used in G3(MP2) by a HF/6-31G* geometry, eliminates both the HF/6-31G* frequency and the QCISD(T)/6-31G* energy calculations and approximates the MP2 energy calculation with the G3MP2 large basis set by an analogous calculation using a dual basis set RI-MP2 model. Taken together, these changes reduce computation time by 2-3 orders of magnitude, and T1 calculations on molecules in the molecular weight range of 400-500 amu are practical.

The T1 recipe, unlike G3(MP2), involves parameters, specifically atom counts, Mulliken bond orders and HF/6-31G* and RI-MP2 energies. These have been determined using linear regression as a best fit to G3(MP2) (not experimental) heats of formation for >1100 small molecules. It reproduces these values with mean absolute and RMS errors of 1.8 and 2.5 kJ/mol, respectively. More important, the T1 recipe reproduces experimental heats of formation for a set of >1800 diverse organic molecules from the NIST thermochemical database with mean absolute and RMS errors of 8.5 and 11.5 kJ/mol, respectively. The plot provided below covers the data from -1000 to +500 kJ/mol.



Heats of formation from the T1 recipe are included as a property in the Spartan Spectra and Properties Database (SSPD), a database that presently comprises more than a 275,000 entries.