Quantum Mechanical Modeling of Sugar Thermochemistry

Joshua Engelkemier  
Iowa State University

Theresa Lynn Windus  
Iowa State University, twindus@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/chem_pubs

Part of the Chemistry Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/chem_pubs/915. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Quantum Mechanical Modeling of Sugar Thermochemistry

Abstract
The recently developed homodesmotic hierarchy for hydrocarbons is extended to include oxygen so that accurate thermochemical quantities for sugars and sugar polymers may be computed with relatively small computational cost. In particular, the method will allow for the determination of heats of formation, which can be used to determine bond strengths important in the decomposition of sugars in, for example, the pyrolysis of biomass. This chapter includes a brief review of the current methodology for calculating thermodynamic properties using electronic structure methods and a description of the proposed extensions. Preliminary results using the lowest members of the hierarchy give a standard heat of formation value of β-D-glucopyranose-gg to be approximately 250 to 260 kcal/mol. These results are promising, and future work will include the calculation of highly accurate building blocks on which this method is based.

Disciplines
Chemistry

Comments
The recently developed homodesmotic hierarchy for hydrocarbons is extended to include oxygen so that accurate thermochemical quantities for sugars and sugar polymers may be computed with relatively small computational cost. In particular, the method will allow for the determination of heats of formation, which can be used to determine bond strengths important in the decomposition of sugars in, for example, the pyrolysis of biomass. This chapter includes a brief review of the current methodology for calculating thermodynamic properties using electronic structure methods and a description of the proposed extensions. Preliminary results using the lowest members of the hierarchy give a standard heat of formation value of β-D-glucopyranose-gg to be approximately 250 to 260 kcal/mol. These results are promising, and future work will include the calculation of highly accurate building blocks on which this method is based.

Many reports describe the multiple challenges associated with the composition of biomass to useful fuels (1). Of critical importance to the conversion is understanding the decomposition of lignocellulose whose main components – cellulose, hemicellulose and lignin – are difficult to break into constituent sugar components because of the polymeric nature of the material that acts to “harden” the material and prevent its decomposition. Even when the sugars are released from the biomass, there is still the challenge of converting the sugar to fuel (primarily ethanol and biodiesel) in an energetically and...
environmentally conservative manner (i.e., using the least amount of energy to accomplish the conversion in a way that will not produce more environmental issues). To understand the decomposition of the source materials requires a detailed understanding of the thermodynamics and kinetics of each of the building blocks (sugars and hydrocarbons) that the source material is composed of. Additionally, the energy involved in bond breaking during decomposition must be understood to predict product formation and to “disrupt” the current process in such a way as to produce more of the desired products. All of these issues point to the need for accurate thermochemistry for these quite large systems. Unfortunately, there is a distinct lack of such information for these systems from both experimental and computational sources. While some specific species have been examined (2), to the best of our knowledge, a systematic examination of the sugars and their decomposition pathways has not been undertaken. Because of the computationally intense nature of many of the current methods, this is not a trivial undertaking.

We propose an extension to the recent hydrocarbon homodesmotic hierarchy (3) to include oxygen. While our motivation is the examination of sugars and products of biomass conversion, the proposed extensions apply to many oxygen-containing species. The next section discusses the two main quantum mechanical methods, composite and balanced reactions, for obtaining accurate thermochemical quantities. It includes a detailed description of the homodesmotic hierarchy that extends to include oxygen important in sugars, which is discussed in the following section. The section after that reviews the application of the hierarchy to obtain the heat of formation of $\beta$-D-glucopyranose-gg. Finally, conclusions are given.

Quantum Mechanical Methodology

The chemical community has long been concerned with calculating accurate thermochemical quantities such as heats of formation, ionization energies, proton affinities, and dissociation energies because these are the basic building blocks for understanding the stability of reactants and products and their reactions. In addition, these quantities are of interest to experimentalists and can be directly compared to experimental values when they are available. Therefore, much effort has been put into evaluating these quantities to obtain “chemical accuracy.” For example, heats of formation are one of the most fundamental quantities because many other properties (such as bond dissociation energies) can be derived from them. Chemical accuracy for this property usually refers to being within 1 kcal/mol of the actual experimental value. Unfortunately, very accurate experimental values are not always available, making benchmarking methods a challenge. However, using available experimental values allows the theoretical community to develop several different methodologies that can roughly be separated into two different categories: additive or composite methods and balanced equation methods. Both of these methods and the common theories in each category are described below.
Composite Methods

Calculating molecular energies is the first step in determining heats of formation to chemical accuracy and requires extremely accurate calculations, usually at great computational expense. These molecular energies can then be used with atomic information to determine the overall heats of formation. However, for anything other than very small systems, calculating the molecular energy accurately is challenging. To overcome this, multiple methods have been determined to either add together information of many lower-level computations or to extrapolate to a molecular energy that includes both complete basis sets and high levels of electron correlation. Because the idea is to approximate the results of a very high theory level with a lower level, the pieces in the composite method should be affordable on the current generation of computer architectures. However, computer architectures are continuing to advance, so we can either use a higher level of theory for a smaller molecule or apply the lower levels of theory with a larger molecule. In general, researchers prefer the latter approach because the larger molecular systems usually do not have very accurate experimental thermochemical data available. However, as will be described below, there is still much activity around using higher levels of theory to obtain even sub-chemical accuracy (~0.1 kcal/mol error from the correct value). Because several reviews are available (4), only a brief description of the most-used methods are described here.

Perhaps the most well-known composite method is the “Gaussian” models of Curtiss and coworkers (5). The fundamental versions of these are denoted as G\(n\), where \(n=1-4\). Because several of the other composition methods use similar concepts, and because the method is often used, an example using G2 theory (5c) is described here. In G2, the first step is to compute a geometry at the MP2(FU)/6-31G(d) level (6, 7), where FU means using all of the electrons in the system (i.e., not using a frozen core). This geometry is used for all of the subsequent steps in the calculation. The second step is to calculate an energy at the baseline, higher level calculation, which in this case is MP4/6-311G(d,p) – denoted E[MP4/6-311G(d,p)]. Next, a correction is made for diffuse functions, which are especially important for anions and molecules with extensive wavefunctions, \(\Delta E(+)^{n}\), and is given by:

\[
\Delta E(+) = E[MP4/6 - 311 + G(d, p)] - E[MP4/6 - 311G(d, p)]
\] (1)

The fourth step is the addition of extra polarization functions on heavy atoms (all but hydrogen), \(\Delta E(2df)\):

\[
\Delta E(2df) = E[MP4/6 - 311G(2df, p)] - E[MP4/6 - 311G(d, p)]
\] (2)

The fifth step is a correction for an additional d function on heavy elements and a p function on hydrogen, \(\Delta\):

\[
\Delta = E[MP2/6 - 311 + G(3df, 2p)] - E[MP2/6 - 311G(2df, p)]
\]

\[
- E[MP2/6 - 311 + G(d, p)] + E[MP2/6 - 311G(d, p)]
\] (3)
Up to this point, all of the corrections have improved the basis set limits. The next one, the sixth step, improves the electron correlation from MP4 to QCI (8), \( \Delta E(QCI) \):

\[
\Delta E(QCI) = E[QCI / 6 - 311G(d, p)] - E[MP4 / 6 - 311G(d, p)]
\]

The seventh step is the addition of a higher level correction (HLC) that takes into account other basis set errors:

\[
HLC = -0.00481 n_\beta - 0.00019 n_\alpha
\]

where \( n_\beta \) and \( n_\alpha \) are the number of beta and alpha electrons in the valence on the molecule, respectively. The first coefficient (-0.00481) was optimized to give a zero mean deviation of the calculated atomization energies of 55 molecules from well-known experimental values. The second coefficient (-0.00019) is a correction in the energy associated with the hydrogen atom.

For the eighth step, the zero-point energy must be included to get the total energy, \( E_0 \). Because there is a specific scaling relationship with Hartree-Fock calculations (9), the frequencies from a HF/6-31G(d) calculation are scaled by 0.893, \( E(ZPE) \).

Finally, the total energy is obtained by adding equations (1)–(5) and \( E(ZPE) \) to the base MP4/6-31G(d,p) energy:

\[
E_0 = E[MP4 / 6 - 311G(d, p)] + \Delta E(+) + \Delta E(2df) + \Delta + \Delta E(QCI) + HLC + E(ZPE)
\]

Once the calculated molecular energy is available, it can be combined with additional information to calculate important thermodynamic information. Using a test set (denoted the G2 test set) of 125 molecular systems with accurate experimental data for dissociation energies, ionization energies, electron affinities, and proton affinities developed by the G2 authors, the G2 method was able to obtain a mean average deviation of 1.21 kcal/mol compared to experiment – very close to the goal of chemical accuracy. In fact, one of the important contributions from the Gn work is the development of a series of test sets where accurate experimental information is available and can be used for multiple thermodynamic calculational methodologies (10). Subsequent to G2, various improvements were made to the composite method. The most recent version, G4 (5m), uses the latest G3/05 test set (11) containing 454 experimental energies. It delivers an average absolute deviation of 0.83 kcal/mol.

In similar research by DeYonker, Cundari, Wilson and co-workers, the correlation-consistent composite approach (ccCA) (12) uses the G3B method (13) as a foundation for additional changes to improve the overall accuracy. The G3B method is similar to G2, except the B3LYP (14) density functional method is used to optimize geometries and determine the zero-point energies. Compared to the G3B method, one of the main differences is that the Pople-style basis sets, such as 6-311G(d,p), are replaced with the correlation-consistent basis set of Dunning and co-workers (15). Because these basis sets have many well-known extrapolations to the complete basis set limit (16), the authors used several of them to obtain
accurate one-electron energies at the MP2 theory level. Based on their results, the extrapolations developed by Peterson and co-workers (16b) and by Wilson and Dunning (16e) gave the best results for the overall thermochemical data. With this extrapolation in place, and after research by others revealed that the triples excitations in the MP4 formalism can cause large electron correlation errors (17), the MP2 theory level using the extrapolated basis results was chosen as the starting single-point level instead of the MP4 in G3B. In addition, the ccCA includes a correction for core-valence correlation in the basis and does not include the HLC. As with the Gn methods, higher levels of electron correlation are included either through QCISD(T) or the coupled-cluster method with singles, doubles, and approximate triples, CCSD(T), with the latter generally being preferred. In later versions of the method, relativistic effects with the Douglas-Kroll-Hess Hamiltonian (18) and first-order spin-orbit coupling were added. For elements in the second row and below, a correction for tight core functions was also added. Using the G3/99 test set, the ccCA method obtains a 0.96 kcal/mol mean absolute deviation, essentially the accuracy of the G3X model, while avoiding the MP4 calculation as well as involving no empirical parameters.

While having the same goals of decreasing the overall errors of the complete energy, Petersson and co-workers take a different approach toward this extrapolation in their complete basis set CBS-4, CBS-Q, and CBS-QCI/APNO methods (19). These methods are similar to the G2 method in both approach and cost. The major difference is that the models use nonlinear pair natural orbital extrapolations to the complete basis set limit. However, nonlinear extrapolations will not produce calculations that are size consistent unless the corrections are applied to localized quantities. Size consistency means that one calculation including non-interacting molecules (i.e., molecules at infinity from one another) is the same as the sum of separate energy calculations on each molecule. This, of course, is an important property for thermochemical accuracy. In the CBS method, the Pipek and Mezey localization method is used to localize populations to correct for this issue (20). In addition, these methods also have an empirical correction that is specific for each model, but it is based on overlaps to obtain a size-consistent generalization of the correction used in Gn theories. The root mean square errors for the 125 chemical energy differences of the G2 test set are 2.5, 1.3, and 0.7 kcal/mol for CBS-4, CBS-Q, and CBS-QCI/APNO, respectively.

While the methods described above are readily applied to moderately sized molecular systems, some recent computationally intensive composite methods aim for less than 0.25 kcal/mol (or ~ 1 kJ/mol) accuracy and are currently only available for the smallest of molecules. These “calibration accuracy” methods generally fall into two different categories: 1) the Weizmann-n (Wn) (21) methods of Martin and co-workers, and 2) the “high accuracy extrapolated ab initio thermochemistry,” HEAT (22), method developed by a multi-national effort. Both of these methods take advantage of the Active Thermochemical Tables (ATcT) (23), which is a novel source of highly accurate thermochemical data from the best experimental and theoretical data available. Additionally, it includes error bar analysis. In the ATcT approach, thermodynamic networks are used to solve for the thermodynamic quantities of interest (e.g., heats of formation). This approach also determines species that are not well known. In addition, the error analysis is propagated.
through the system with smaller weights in the network to less well characterized data (i.e., trying to minimize the error in each of the individual thermodynamic quantities). This provides a highly accurate database of information for which to compare theoretical results.

In the Wn methods, all elements of the Hamiltonian that can contribute at the kJ/mol level are included in the overall scheme, basis set convergence is determined at each level of theory used (i.e. SCF, CCSD(T)), the smallest basis set possible (still quite large) is used to obtain the target accuracy, and no parameters from fits to experimental data are used. Concerning the last point, several of the Wn methods use a parameter to improve the basis set extrapolations. In addition to the extensive basis set extrapolation and extrapolations for higher-order operators in the coupled cluster series, each level in the Wn hierarchy includes scalar relativistic effects through one-electron Darwin and mass-velocity terms.

The W4 version includes an estimation of the \( \hat{T}^4 \) and \( \hat{T}^5 \) terms in the coupled cluster operator, inner core correlation, and atomic spin-orbit coupling, as well as the diagonal Born-Oppenheimer correction. For a series of 30 molecules, an average accuracy of 0.1 kcal/mol in atomization energies are obtained, compared to those of the ATcT.

The HEAT method has many similarities to the Wn method (some developed before the Wn methods and some developed after). The first significant difference is that the geometries and harmonic zero-point energies in the HEAT method are determined at a very high CCSD(T)/cc-pVQZ level of theory correlating all electrons in the molecule (including the core). The second major difference is that the corrections between the triple and quadruple contributions for the coupled cluster expansion are calculated exactly (i.e., not by approximation) using the double zeta correlation consistent basis set extrapolations. Other terms are similar to those in the Wn methods. The focus of this work is to obtain highly accurate total energies and not atomization energies. While including this data, the research group specifically chose to examine reactions with respect to the elements in their standard states. However, some modifications to the scheme are required because elements such as graphite are prohibitive at this computational level. In this case, the authors chose to use CO\(_2\) as a “substitute” and used appropriate reactions. For the 31 sample molecules chosen from the ATcT, all of the reported heats of formation, with the exception of one, fall within 0.5 kJ/mol of the value given by ATcT. The exception, H\(_2\)O\(_2\), fell within 1.0 kJ/mol of the expected value.

In other research, the “focal point” approach (24) of Wheeler and co-workers provides a general strategy for obtaining very accurate thermochemical information using correlation-consistent basis sets to obtain systematic dual one- and n-particle expansions and includes electron correlation through second-order perturbation and coupled-cluster methods. Combining this method with additional corrections for the anharmonic zero-point energy, the diagonal Born-Oppenheimer contribution, and scalar relativistic effects, these authors have obtained accurate thermochemical information for key small intermediates in soot formation.

In all of the methods described above, there is an assumption that a single reference is sufficient for describing the molecular system. In many molecular systems, this assumption is appropriate. However, it is clear that there are systems...
where multiple references are required (25), such as those that include electronic excited states or complex bond formation or dissociation. Sølling and co-workers have developed a multi-reference equivalent to the G2 and G3 methods to deal with these types of systems (26). In addition, very accurate potentials and vibrational levels (a true challenge!) have been obtained for several diatomic and triatomic systems by Bytautas and Ruedenberg through a novel configuration interaction extrapolation method (27).

Balanced Equations/Homodesmotic Reactions

As described above, the composite methods have the goal of obtaining very accurate overall energies that can then be used to compute thermodynamic quantities. The next type of methods rely heavily on cancellation of error in reaction calculations using several different types of reaction definitions. The advantage of these methods is that relatively low levels of theory can be used to obtain accurate thermochemical data, even on large chemical systems. Again, researchers have introduced several different methodologies. This chapter describes only the commonly used methods.

Pople et al. developed the first electronic-structure-based method, and it defines an isogyric reaction as one that leaves the number of unpaired electron spins unchanged (28). Dissociation energies for simple hydrides were developed using a hydrogen molecule (e.g., BH2 → BH + H or BH3 + H → BH2 + H2). Combining these results with the exact dissociation energy of H2 gives the desired result. Researchers used these types of reactions to determine singlet-triplet separations, ionization energies, and enthalpies of formation for a series of small molecules where experimental data was available. Using MP4 methods with small basis sets to calculate the total energies, the authors obtained ionization energies within 0.1 eV ≈ 2.3 kcal/mol for most cases.

Since those initial isogyric schemes, there have been many additional definitions of balanced equations, including homodesmotic (equal bonds) (29), isodesmotic (28a), and hyperhomodesmotic (30), to name only a few. All of them depend on balancing different parts of the chemical equation – bond types, hybridization of the atoms, etc. All of these methods rely on having very accurate heats of formation available for a set of relatively small molecules or “standard” molecules for which the reactions can be built. Therefore, the composition methods are of critical importance in this effort as well. While several of these balanced equation methods are fairly well defined, extensive confusion in the literature exists between the different methods.

Wheeler and colleagues clearly define a standard homodesmotic hierarchy for unstrained hydrocarbons (3) to clarify the balanced reaction types. In this work, the authors also describe the relationship of their work to other methods in the literature. Their scheme contains five progressive classes of homodesmotic reactions, in which the higher orders are systematically more rigorously balanced, yet also more expensive, than the classes below them. Indeed, what was a product in the previous class, becomes a reactant in the next highest one. (The reactions are formed in such a manner that the species of interest, usually larger than any of the standard species, is combined with small reactants and broken down
into substituents that are larger than the reactants, but smaller than the original molecule.) The following list paraphrases their definitions:

**RC1 (Isogyric):** the number of total electron pairs is balanced.

**RC2 (Isodesmic):** (a) the number of total electron pairs and (b) the number of carbon-carbon single, double, and triple bonds are balanced.

**RC3 (Hypohomodesmotic):** (a) the number of carbon atoms in sp³, sp², and sp hybridization and (b) the number of carbon atoms with zero, one, two, or three attached hydrogens are balanced.

**RC4 (Homodesmotic):** (a) the number of each combination of two, bonded, separately hybridization-specific carbon atoms and (b) the number of sp³, sp², and sp hybridized carbon atoms with zero, one, two, or three attached hydrogens are balanced.

**RC5 (Hyperhomodesmotic):** (a) the number of each combination of two, bonded carbons each with zero, one, two, or three attached hydrogens and connected by a single, double, or triple bond and (b) the number of sp³, sp², and sp carbon atoms with zero, one, two, or three attached hydrogens are balanced.

Using these definitions, they were also able to define all possible elemental reactants and products necessary to satisfy the requirements of each reaction class for any hydrocarbon. The visual interpretation of this system is especially helpful in understanding the inherent logic. The hypohomodesmotic definition can be thought of as breaking the molecule into smaller products based on each non-terminal carbon center. For example, a carbon attached to three carbons by two single bonds and a double bond is represented on the product side by 2-methylpropene. To counter-balance the extra terminal sp³ and sp² carbons (−CH₃, −CH₂) now in the products, ethane and ethylene must be added to the reactants. The homodesmotic definition is identical except in cases when the study molecule has a conjugated pi system. RC4 preserves conjugating pairs of pi bonds in the products, which Wheeler and colleagues found to produce significant improvement over RC3 calculations. The hyperhomodesmotic definition is analogous to the hypohomodesmotic one, but it involves two non-terminal carbon atoms, the bond between them, and then the remaining, truncated bonds of each to any other carbons as in RC3. Thus, RC4 behaves like RC5, but only in situations with conjugation. As Wheeler and colleagues mentioned in their publication, the hierarchy could theoretically be expanded even further so that each product molecule contained the bonds of each three atoms and so forth, until the reactants and products were exactly identical. However, in this extreme case, the technique ceases to be useful.

Of particular interest in this work are the results of calculations of 38 hydrocarbons containing five or six carbon atoms. The RC4 and RC5 reactions give bond separation enthalpies with errors consistently less than 0.4 kcal/mol using several levels of theory including HF, DFT, MP2, and CCSD(T). Even RC2 and RC3 results are consistently below 8 kcal/mol and 2 kcal/mol, respectively. These results are promising in that highly expensive calculations can be avoided except in computing the constituent standard molecules, which are small in
comparison to the systems of interest. Enthalpies of formation for large polyynes were also computed (C_{10}H_{2}-C_{26}H_{2}) using DFT, showing the use of these methods to obtain accurate values for large molecular systems.

This chapter adapts these standards to include oxygen-based functional groups, namely alcohols, ethers, ketones, and the combinations of them. Because oxygen is the most common heteroatom in organic molecules and sugars, the usefulness of homodesmotic reactions is greatly enhanced by the capability to include oxygen. The current research was tested with β-D-glucopyranose-gg (Figure 1).

![Figure 1. β-D-glucopyranose-gg](image)

**Extension for Sugars**

Just as the hydrocarbon homodesmotic hierarchy is self-consistent in that the higher classes’ parameters are expansions, not revisions, of the lower ones, so too must be a hierarchy that includes oxygen. Because carbon and oxygen are large atoms compared to hydrogen, it is reasonable to give oxygen equal priority to carbon, not merely as a substituent of it. We show that any sp³ or sp² carbon atom from Wheeler and coworker’s elemental molecules may be substituted by a sp³ or sp² oxygen atom, respectively.

Oxygen-oxygen bonding was not considered at this time because oxygen significantly differs from carbon in its ability to form long chains of bonds. However, oxygen can still be part of a primarily carbon chain in the form of ethers. The lack of side branching in ethers structurally differentiates them from interstitial carbons. Likewise, an alcohol, (primary, secondary, or tertiary) is analogous to a terminal carbon, whether it is bonded to a primary, secondary, or tertiary carbon. They are both heavy atoms connected to the appropriate number of hydrogens and just one carbon.

Concerning the sp² hybridized functionalities, a ketone is structurally like an alkene except with an oxygen as one of the pi-bonding pairs, rather than a carbon. A notable difference in this situation is that the ketone is necessarily terminal, whereas the alkene may be in the middle of a carbon chain. The ketone may, however, be involved in conjugation. Oxygens are rarely, if ever, species? hybridized at equilibrium conditions, so there is no suitable comparison with alkynes.

Substitution of a carbon atom with a similarly hybridized oxygen atom does not significantly change the molecule’s basic form, except when it would cause fragmentation as noted above. Therefore, the patterns that Wheeler and colleagues have already established can be modified by this principle to include oxygen. The adapted definitions are:
RC1: the number of total electron pairs is balanced.
RC2: (a) the number of total electron pairs and (b1) the number of carbon-carbon single, double, and triple bonds and (b2) the number of oxygen-carbon single and double bonds are balanced.
RC3: (a1) the number of carbon atoms in sp³, sp², and sp hybridization and (a2) the number of oxygen atoms in sp³ and sp² hybridization and (b1) the number of carbon atoms with zero, one, two, or three attached hydrogens and (b2) the number of oxygen atoms with zero or one attached hydrogen are balanced.
RC4: (a) the number of each combination of two, bonded, separately hybridization-specific carbon atom and carbon-or-oxygen atom pairs and (b1) the number of sp³, sp², and sp hybridized carbon atoms with zero, one, two, or three attached hydrogens and (b2) the number of sp³ and sp² hybridized oxygen atoms with zero or one attached hydrogen are balanced.
RC5: (a) the number of each combination of two, bonded carbon atom and carbon-or-oxygen atom pair each with zero, one, two, or three attached hydrogens and connected by a single, double, or triple bond and (b1) the number of sp³, sp², and sp hybridized carbon atoms with zero, one, two, or three attached hydrogens and (b2) the number of sp³ and sp² hybridized oxygen atoms with zero or one attached hydrogen are balanced.

Figure 2 includes a chart of all possible elemental reactants and products of this redefined homodesmotic hierarchy that are in addition to the purely hydrocarbon sets proposed by Wheeler and colleagues. Obviously, there are many more elemental pieces in the oxygen-inclusive homodesmotic hierarchy. There are about 150 in all, more than 100 of which are RC5-only. Carboxylic acids first appear as RC3 products, but esters only manifest in a few RC5 products. Many of these theoretical combinations seem highly unstable, which helps explain why experimental data is often scarce. Some molecules that look absurd, however, may seem more reasonable when reassembled into the study molecule. Gem polyols, for example, could represent one or more ether linkages after being broken down according to the requirements of the reaction class. Even so, the sheer number of possibilities, already ignoring stereochemistry, makes the task of finding suitably accurate enthalpies of formation for all the elements quite daunting. It’s useful to note here that this scheme does not include radical species that are certainly of interest in the decomposition of sugars (especially in pyrolysis). This deficiency will be addressed in future work.
RC1 Products/RC2 Reactants
H₂O

RC2 Products/RC3 & RC4 Reactants

RC3 & RC4 Products/RC4 & RC5 Reactants*

RC4 & RC5 Products

* dotted underline indicates that this molecule is a product in RC4; mixed underline indicates that this molecule is both a product and reactant in RC4
RC5-only Products
Wheeler and co-workers showed that even density functional methods with a modest basis set level can be used to give excellent results, especially for levels RC3 and above. β-D-glucopyranose-gg, and all of the elemental reactants and products needed to satisfy the five classes of the hierarchy, were prepared using the graphical user interface Ecce (31) and energetically optimized by NWChem (32). We used restricted density functional theory (B3LYP) and TZVP DFT orbitals without coulomb or exchange fitting. The resulting balanced reactions, (7) through (10), are given below:

\[
\text{RC1 (7)} \quad \text{HO} + 12 \text{H}_2 \rightarrow 6 \text{CH}_4 + 6 \text{H}_2\text{O}
\]
The zero-point corrected total energy values are used to calculate the enthalpy of each reaction according to the general equation:

\[ \Sigma \text{Energy}_{\text{products}} - \Sigma \text{Energy}_{\text{reactants}} = \Delta H^\circ_{\text{rxn}} \]  

(11)

Standard enthalpy of formation values (33) were then used to calculate the enthalpy of formation of the study molecule, \( \beta \)-glucopyranose-gg, for reactions (7) and (8) according to the following equation:

\[ \Delta H^\circ_{\text{study}} = \Sigma (\Delta H^\circ_{\text{prod}}) - \Sigma (\Delta H^\circ_{\text{other react}}) - \Delta H^\circ_{\text{rxn}} \]  

(12)

At this time, accurate values for the standard enthalpies of formation of all the elemental molecules for reactions (9) and (10) are unavailable. We are in the process of determining accurate enthalpies of formation for these species using the composite methods described earlier.
Table 1. Thermodynamic data from computation and literature

<table>
<thead>
<tr>
<th></th>
<th>Total Energy + ZPVE (E&lt;sub&gt;h&lt;/sub&gt;)</th>
<th>ΔH&lt;sub&gt;r&lt;/sub&gt;° (kJ/mol) (33)</th>
<th>ΔH&lt;sub&gt;r&lt;/sub&gt;° (E&lt;sub&gt;h&lt;/sub&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-D-glucopyranose gauche gauche</td>
<td>-687.227422055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrogen</td>
<td>-1.1691918625</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Methane</td>
<td>-40.4913470200</td>
<td>-74.6</td>
<td>-0.02841</td>
</tr>
<tr>
<td>Water</td>
<td>-76.4379808374</td>
<td>-241.8</td>
<td>-0.092097</td>
</tr>
<tr>
<td>Ethane</td>
<td>-79.7856244897</td>
<td>-84.0</td>
<td>-0.03199</td>
</tr>
<tr>
<td>Methanol</td>
<td>-115.717518491</td>
<td>-201.0</td>
<td>-0.076557</td>
</tr>
</tbody>
</table>

Table 2. Derived thermodynamic data from calculation

<table>
<thead>
<tr>
<th></th>
<th>ΔE&lt;sub&gt;rxn&lt;/sub&gt; (E&lt;sub&gt;h&lt;/sub&gt;)</th>
<th>ΔH&lt;sub&gt;r&lt;/sub&gt;° (E&lt;sub&gt;h&lt;/sub&gt;)</th>
<th>ΔH&lt;sub&gt;r&lt;/sub&gt;° (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC1</td>
<td>-0.318242739</td>
<td>-0.4048</td>
<td>-254.0</td>
</tr>
<tr>
<td>RC2</td>
<td>-0.119468227</td>
<td>-0.4107</td>
<td>-257.7</td>
</tr>
</tbody>
</table>

The necessary information for reactions (7) and (8) are given in Table 1. Using this data and equations (11) and (12), the reaction energy and the heats of formation can be calculated as in Table 2.

It is also worth noting that the RC1 and RC2 values from Table 2 for the standard enthalpy of formation of β-D-glucopyranose-gg are reasonably in agreement with each other. The reaction energy approaches zero going up the hierarchy, which makes sense because the difference between reactants and products is also decreasing.

Conclusions

In this work, we presented an extension to the hydrocarbon hierarchy of homodesmotic reactions developed by Wheeler and co-workers to include oxygen. More than 150 new elements need to have accurate heats of formation available before each hierarchical level can be fully exploited. Obviously, much work remains to be done. Neither our research nor Wheelers took stereochemical effects, such as rotamers of butane-2,3-dione (one of the products of the RC5 reaction for β-D-glucopyranose-gg), into consideration. This is problematic because preliminary DFT calculations indicate the difference can be quite significant – in excess of 1 kcal/mol. To verify the soundness of the adapted homodesmotic scheme presented here, more calculations at higher levels of theory need to be performed, as well as comparisons with experimental data. In addition, incorporating this data into an ATcT scheme would improve the overall accuracy of the developed information. If the current scheme is successful, room
exists for the hierarchy to expand, to include radicals and more heteroatoms, such as nitrogen.

It is too early to determine if the theories presented here are successful. As noted above, there are many barriers to implementing this technique in a practical manner; but the possibilities are great, and the early results are promising. Even with the crude RC1 and RC2 methods, general agreement between the two gave the standard heat of formation value of β-D-glucopyranose-gg to be approximately 250 to 260 kcal/mol. This is not precise enough to use in experiment, but it is hopeful that RC3, RC4, and RC5 will provide even better results. It will be especially interesting to see if RC5 is dramatically better than RC4, because it has the capability to preserve more complex oxygen functional groups, such as esters.

In particular, however, a relatively low level of theory is required to examine large molecular systems such as the sugars. Using the obtained heats of formation acquired through new extensions to the homodesmotic hierarchy, accurate bond dissociation energies and other critical thermochemical data may help us understand the processes involved in the decomposition of sugars and biomass.

Acknowledgments

Work at the Ames Laboratory was supported by the Department of Energy Office of Biological and Environmental Research (BER) and Office of Advanced Scientific Computing Research (ASCR) divisions as part of the Scientific Discovery through Advanced Computing (SciDAC) under Contract No. DE-AC02-07CH11358. In addition, the authors thank Ames Laboratory and Iowa State University for the computing resources used in this work.

References


