Published in final edited form as:

Energy (Oxf). 2012 July ; 43(1): 161–171. doi:10.1016/j.energy.2012.01.072.

JTHERGAS: Thermodynamic Estimation from 2D Graphical Representations of Molecules

Edward Blurock, V. Warth, X. Grandmougin, R. Bounaceur, P.A. Glaude, and F. Battin-Leclerc

CNRS-LRGP, ENSIC, Nancy, France

Abstract

JTHERGAS is a versatile calculator (implemented in JAVA) to estimate thermodynamic information from two dimensional graphical representations of molecules and radicals involving covalent bonds based on the Benson additivity method. The versatility of JTHERGAS stems from its inherent philosophy that all the fundamental data used in the calculation should be visible, to see exactly where the final values came from, and modifiable, to account for new data that can appear in the literature. The main use of this method is within automatic combustion mechanism generation systems where fast estimation of a large number and variety of chemical species is needed. The implementation strategy is based on meta-atom definitions and substructure analysis allowing a highly extensible database without modification of the core algorithms. Several interfaces for the database and the calculations are provided from terminal line commands, to graphical interfaces to web-services. The first order estimation of thermodynamics is based summing up the contributions of each heavy atom bonding description. Second order corrections due to steric hindrance and ring strain are made. Automatic estimate of contributions due to internal, external and optical symmetries are also made. The thermodynamical data for radicals is calculated by taking the difference due to the lost of a hydrogen radical taking into account changes in symmetry, spin, rotations, vibrations and steric hindrances. The software is public domain and is based on standard libraries such as CDK and CML.

Keywords

Thermodynamic estimation; enthalpy; entropy; heat capacity; Benson rules; automatic generation

1 Introduction

The needed developments of new types of engines, e.g. HCCI Engines[1] and fuels, e.g. biofuels[2] [3], have prompted a recent growth of the development of detailed chemical kinetic models for the combustion of a wide range of hydrocarbons. A crucial set of parameters for modeling combustion processes in detailed mechanisms are the set of temperature dependent thermodynamic constants associated with each species. For a typical mechanism (for a typical example see Westbrook[4]), the standard enthalpy, the standard entropy and a set of temperature dependent heat capacities are needed. Some of these values can be found in compilations on the web, for example from the National Institute of Standards (see <http://www.nist.org>) which includes the information from the JANAF tables, Burcat's compilation[5], compilations of IUPAC[6]], or for example the books of Pedley[7] or Stull[8] However, for larger mechanisms, especially those generated automatically (see

for example, Moréac and Blurock[9], Pierucci and Ranzi[10], or Buda et al.[11]) new species are created which are not in the *standard* databases. For this reason methods are needed to compute the thermodynamics of a class of species “on the fly”. Furthermore, especially for species generated automatically, only a minimal amount of structural information, typically just the bonding information held in a two-dimensional graphical (see Balaban[9] and discussion below), is available. The use of ab initio calculations to calculate chemical thermodynamics (see for example Simmie et al.[13] or Bozzelli et al.[14]) is time consuming (due to the high level of basis sets needed) and needs detailed three dimensional structures as input (though progress is being made here also[15]).

One class of methods which has been used successfully for obtaining the thermodynamic values needed for combustion mechanisms is that of group additivity. The central idea of group additivity methods is that a first order approximation to the thermodynamics of the molecule is just the sum of its individual components. The central assumption is that thermodynamics “effects” are local. They don’t extend beyond one or a few bonds. Two of the earliest group additivity models come from Benson[16],[17],[18] in which the groups is atom based, and Laidler[19],[20],[21], in which the groups are based on bonds. The basis for these methods is to take the “average” effect of the particular group within the same class of molecules. Zavistas et al.[22] proposed a simple atom-based technique (also for radicals) in which a simple linear form is used and cited reasonable results for 360 species and radicals. This method determines the individual coefficients from model species and the radical values are obtain with the bond dissociation energy (with a differentiation between primary, secondary and tertiary hydrogen atoms). This method shows that even with very simple group additivity rules, reasonable results are obtained. Bojesen[23] did a systematic study using ab initio studies to calculate molecule frequency distributions (MFD) and group frequency distributions and zero-point energies. From a study using 126 organic molecules from 11 homologous series (same Benson group) he postulated that “group additivity is observed when the MFD is a sum of group frequency distributions”.

But there are other structural features that are not entirely local. One of these are the non-nearest neighbor interactions (NNI). These are steric effects of groups that are at least two atoms away. The Benson method[18] recognizes several of these interactions: alkane, 1,4-gauche, alkane 1,5, alkene 1,4-gauche, alkene single and, double cis, ene-yne cis and ortho interactions.

There are other methods based on more complex structures. For example, Constantanui and Gani[24], propose an extension of the Benson atom group method by having first order groups (similar to the Benson groups) that are supplemented with second order structures extending through more bonds. Another non-local group energetic correction that needs to be made is ring strain corrections. These are also accounted for in the Benson method by having corrections for every type of ring structure, both in terms of number of atoms in the ring and in terms of the atoms and bond types within the ring. Both of these corrections are accounted for before the “averaging” of group effects. A more general extension of group additivity was proposed by Rabitz and Alis[25] with his high-dimensional model representation (HDMR) and was used by Hayes et al.[26]for the prediction of enthalpies. In this model group additivity is the first order of a hierarchical representation of structures. With each level in the hierarchy, the influences of structures further away are introduced. In this representation, for example, the “special case” corrections of non-nearest neighbor interactions and ring strain used in the Benson method are covered by higher order corrections.

A crucial point in the accuracy of the Benson group additivity method is the accuracy of the fundamental database. From Benson’s original values in 1958[16], the group additivity

values have steadily improved to those proposed in 1969[17], those listed in Benson's book[15] and in Cohen's later review (limited to enthalpies of formation) in 1992[27]. Although ab initio calculations are not used directly for the calculation of all classes thermodynamic species, the method is ideal to supplement missing experimental values. Several groups are working to supplement and improve Benson's original list of group additivity values with high level ab initio studies.:

1. Nitrogen containing molecules from Ashcraft and Green(2008)[28]
2. Lower Alkyl Hydroperoxides and related Hydroperoxy and Alkoxy Radicals from Simmie et al.(2008)[13]
3. Hydrocarbons and Hydrocarbon Radicals from Sabbe et al. (2008)[29]
4. Furans and ortho-arynes from Yu et al. (2006)[30]
5. n-Aldehydes (Acetaldehyde, Propanal, Butanal, Pentanal, Hexanal, and Heptanal) and their Radicals from da Silva and Bozzelli (2006)[14]
6. Unsaturated oxygenated hydrocarbons and radicals resulting from reaction of vinyl and phenyl radical systems with O₂ from Sebbar et al. (2005)[31]
7. Large polyunsaturated hydrocarbons from Sumathi et al.(2002)[32]
8. Hydrocarbons and Hydrocarbon Radicals from Sabbe et al. (2005)[33]
9. Vinyl, Phenyl, Ethynyl, and Allyl Peroxides from Sebbar et al. (2005)[31]
10. polycyclic aromatic hydrocarbons Yu et al. (2004)[34]
11. n-Alkanes (using internal rotations) from Vansteenkiste et al.(2003)[35]
12. Ketenes from Sumathi et al. (2002)[36]
13. Multichloro Alkanes and Alkenes from Chen et al. (1998)[37]
14. Alkyl peroxides and trioxides from Lay and Bozzelli (1997)[38]

This list shows not only that the Benson additivity method continues to be a viable method for the calculation of thermodynamic quantities, but also that the group additivity values are not static and are continually being updated with new values for new groups. This updating includes not only the group additivity values themselves, but also the second order corrections such as ring strain energies and non-nearest neighbor interactions. This means that the algorithms that implement these values must allow ease of modification.

Radicals can be calculated in one of two ways. The first way is the definition of group additivity values for radical centers (see Table 2.14 in Benson's book[18]). The second method is to consider the loss of hydrogen from a parent molecule (the radical center with a hydrogen added). Both methods can consider specific types of differences between the parent molecule and the radical molecules. Cohen[27], in defining the updated radical group additivity values, listed that vibrational modes, internal rotations, electronic degeneracy, and symmetry have to be considered between the radical and the parent molecule. These effects can be taken into account in both methods (they just have to be done consistently).

One implementation of the Benson additivity method which is still in widespread use in the combustion community is the THERM implementation of Ritter and Bozzelli[39][40] (for example, all the thermodynamic constants in the detailed mechanism for a gasoline surrogate mixture published in 2011 by Mehl et al.[41]). In this implementation, the external and internal symmetry numbers must be determined by hand. However, the group additivity values are in an ASCII file. Thermodynamic values for radicals are either deduced from the

hydrogen bond dissociation energy at the radical center or with group additivity values at the radical center. Another implementation of the Benson additivity method is by NIST[42] in the NIST webbook (<http://webbook.nist.gov/chemistry/grp-add>). Here one can draw the species or give the species in MDL mol file format[40]. The group additivity values are basically from Benson's book[18]. Radicals are calculated using the group additivity values with radical centers (Table 2.14 in Benson's book[15]). The symmetry of the molecule is determined automatically.

The automatic detailed mechanism generator, EXGAS[44], uses the program THERGAS[45], to calculate the thermodynamic constants for the species generated. The program is fully automated and no user input (for example, symmetries) other than the 2D graphical representation of the molecule (given in Nancy Linear Form, see below) is needed. The symmetries of the species are estimated automatically. Radicals are calculated using the difference with the parent molecule (radical center with hydrogen added). In addition to the bond dissociation energy, corrections due to loss of internal and external symmetries and loss of vibrational and rotational modes are taken into account.

The implementation described in this paper, JTHERGAS, is based on the THERGAS implementation. The inherent method (and at this stage, the database) is essentially the same. However, the one main design difference is that all data, both the raw data of the fundamental constants fundamental structures involved are visible and modifiable. A significant part of this design is that every aspect of the calculation is based on molecular substructures stored in a database. Very little of the fundamental structures are "hard-coded". This makes the introduction of new functional groups and their associated data straightforward enough that a "chemist" can do it without dealing with the actual program code. The modifiable structures start at the level of defining specialized atoms (for example the atom types used in the Benson rules), to fundamental symmetry structures, energetic correction structures (such as for ring strain or steric hindrances) and vibrational structures. In addition, the interfacing of the program, including the database management, is accessible in many forms, from "terminal" mode (typing commands at the terminal) to graphical interfacing (web pages with GUI's) to web services (functionality over the web). The main purpose of this paper is to illustrate the fundamental flexible design that has been implemented.

2 Concepts and Representations

The estimation for thermodynamic quantities for molecules using the Benson method is based on first order estimations based on adding up the contributions from each heavy atom (non-hydrogen) description, which includes the bonding structure, and second order corrections based on internal, external and optical symmetries, steric hindrances due to rotations and ring strain energies.

2.1 Two Dimensional Representation of molecules

The two dimensional graphical representation[12] is based on a 'colored graph' consisting of a list of nodes with properties and connections between the nodes (also with properties). In the molecular representation, the nodes are the atoms of the molecule and the connections are the bonds of the molecules. In this representation, the three dimensional information is implied by the combination of atoms and bonds.

The basic (essential) information needed for the atom graphical node is first and foremost the atomic number. Additional information, such as charge or whether the atom is a radical, is also given. To facilitate calculations and manipulations additional information may be

supplied to the atom (graphical node). In this document, the graphical atom node will just be referred to as atoms.

In the graphical description of molecular bonds, the two atoms which are bonded and the type of bond, whether single, double or triple is specified (the class of molecules being described is usually restricted to covalent bonds in organic compounds). In the basic 2D graphical description there is no information as to length or direction of the connection. Resonance and aromaticity are described with supplementary information to the atom and/or bond description.

The internal representation of the molecule within JTHERGAS is the *Molecule* class from the Chemistry Development Kit[46] (CDK). This allows the use of many supplemental algorithms, including the more complicated algorithms such as graph isomorphism or ring recognition. The ASCII representation (for input and output and for storage in the SQL database) of the molecule is in two forms. The first is the Chemical Markup Language[47] (CML) which is an extension of the more general XML used throughout the network. The second is a ASCII linear format, what will be called the Nancy Linear Form (NLF), which is an extension of the SMILES representation of a molecule. This format is used, for example, for ASCII input of a molecule, radical or substructure.

2.2 Benson Rules

The key to a Benson rule is the simple description of an atom and the bonding structure around it. The general form is:

$$X - (Y_1)(Y_2) \dots (Y_n) \quad (1)$$

Where X is the *center atom* and the Y's are a description of the atoms (including multiplicity) bonded to the center atom. Associated with each Benson rule is a set of thermodynamic quantities, namely standard enthalpy, standard entropy and heat capacities at several temperatures. The Benson rule represents how the center atom contributes to the thermodynamics of the molecule as a whole. The contributions of all the center atoms (all the atoms in the molecule except for hydrogens) are added to give the total thermodynamics of the molecule. It is the heart of the additivity concept.

The atom descriptions of the center atoms and the connected atoms can more than just the simple atomic description of the molecule. The descriptions can also describe the valency of the atom. For example, the carbon atom descriptions can be C_s, C_d, C_t or C_a for carbons which are singly, doubly, triply or aromatic bonded, respectively. In addition, a meta-atom can be defined representing a functional group such as, for example, carboxyl represented as a single atom called *co*. For example, an aldehyde carbon bonded to a singly bonded carbon (and the hydrogen of the aldehyde) is represented as:

The description of the Y's in equation 1 have also this extended description supplemented by how many times (a slash followed by the multiplicity) that atom is bonded to the center atom. For example the Benson structure for a primary carbon is represented as:

2.3 Nancy Linear Form (NLF)

One of the ASCII representations of the molecule is the Nancy Linear Form (NLF). This is similar to the SMILES representation and is used in the original THERGAS[45] and EXGAS[44] programs. Linear, branched and (multiple) ringed structures, aromatics and radicals can be represented by this format. Some examples are given in Table 1.

2.4 Chemical Markup Language

A standard format/language of the internet is the eXtensible Markup Language (XML) developed by the World Wide Web Consortium (W3C, <http://www.w3.org>). XML is meant to be a means for describing and validating complex documents. XML itself is very general, but there exists specializations within various domains. One of these domains is the Chemical Markup Language (CML) which has many structures associated with (organic) chemistry. Being a specialization of XML means that CML can use the extensive set of software for transmitting information over the internet that has been developed for XML. In addition, there are also software packages that have developed specifically for CML. CML has become a *de facto* standard for transmitting chemical information over the web.

JTHERGAS uses CML to store the structural information of molecules, radicals and substructure within the MySQL (a widely used public domain software for the manipulation of databases, <http://www.mysql.com>) which requires ASCII representations within the database tables. Figure 1 shows an example of the CML format.

2.5 Meta-atoms

An important concept within JTHERGAS is that of a *meta-atom*. This is where the basic atom node information (in the graphical representation) is extended beyond the 'normal' molecular atom representation mentioned previously. Meta-atoms are used at many levels within JTHERGAS. All meta-atoms within JTHERGAS are defined through the database and is one of the keys to JTHERGAS's flexibility.

Two examples of meta atoms are the extended atom definitions of the atoms in the definition of a Benson rule and the Nancy linear forms. More complex meta-atoms exist within JTHERGAS such as one more a methyl (CH₃) or a phenyl (aromatic ring) group which can be used to simplify molecular matching.

One of the flexibilities within the JTHERGAS system is that all the meta-atoms are not hard-coded within the library, but are defined within the extendable database. With each meta-atom there is associated a complete molecular structure. This means that any new meta atom can be defined when needed without modifying the base code. Figure 2 shows the meta atom structures in the database for a doubly bonded carbon and a carboxyl group. The structure on the right is used to describe the doubly bonded carbon, which can have the meta atom name *C_d* and the structure on the left describes a multi-atom carbonyl (ketone) meta atom, *co*.

2.6 Molecule, Structure and Substructure Recognition

An essential operation in the implementation of JTHERGAS is the comparison of molecular structures. Since the molecular representation is a graph, this operation translates to *graph isomorphism*[12]. The comparison of structures occurs at all levels of the implementation. There are two basic types of structural operations. The first operation is a true/false predicate asking, for example, "Are these structures the same or is this a substructure of another?". The second type of operation determines atom-to-atom (or bond-bond) correspondences between the two structures being compared. This operation can be thought of as overlaying

one molecular graph on top of the other and noting which atoms correspond to each other. Given a structure, there can be many ways that it can be overlayed (matched) with another.

2.6.1 Atom-to-atom Correspondences—The result of determining how one structure fits into another is a table of atom-to-atom correspondences. For example Table 2 shows how a primary methyl substructure (to the right in the figure) fits into an ethane molecule. Note that every atom not only has an atomic number label, but also a number. An atom matches if the atomic number matches. Table 2 shows that there are 12 ways to overlay the methyl structure within the ethane.

2.6.2 Redundancy, Equivalency and Symmetry—Table 2 also shows another important point about atom-to-atom correspondences. Without their explicit labels the hydrogen atoms on carbon 1 are equivalent and the hydrogen atoms on carbon 2 are equivalent. Without the explicit labels on the hydrogen atoms, this means that the atom-to-atom correspondences A1-A6 are indistinguishable and the atom-to-atom correspondences B1-B6 are indistinguishable. In other words there are really only two distinguishable sets of atom-to-atom correspondences. The extra 6 matches correspond to the $3!$ ($3*2*1$) combinations of matching two sets of three objects.

An important question that often needs to be answered within JTHERGAS is not where are the atom-to-atom correspondences, but how many matches are there. In the example of the table, the question being asked is how many primary methyl atoms are in ethane? The answer is two, namely 12 divided by the combinations of matching the three equivalent hydrogens in the primary methyl group. This factor $3!$ ($3*2*1$) or six will be referred to as the symmetry factor in matching.

2.6.3 Wild-card atoms—In the previous example, in the methyl structure, a carbon with label 5 was specified. This limits the general question to how many primary carbons are connected to a carbon atom. It would not recognize that the carbon in methanol is also a primary carbon. To make this structure more general, a *wild-card* atom can be defined. In this example, carbon 5 would be replaced by the wild card. Doing this, methanol would be recognized as having one primary carbon. Figure 2 gives an example of wild card atoms in the definitions of a doubly bonded carbon and a carboxyl group.

3 Implementation of Group Additivity Method with Corrections

The calculation of the thermodynamic information for molecules consists of the first order additivity result of applying the basic Benson rules combined with second order corrections adding symmetry and energy corrections.

3.1 Benson Rules

As described previously applying Benson rules consists of identifying all the center atoms (in the form shown in equation 1) and adding up all their contributions. The JTHERGAS system has a library consisting of 366 Benson rules, the majority of which correspond to the tables in Benson's book:

1. Table A.1: Hydrocarbons
2. Table A.2 Oxygen-containing Compounds
3. Table A.3 Nitrogen-containing Compounds
4. Table A.4 Halogen-containing Compounds
5. Table A.5 Sulfur-containing Compounds

6. Table A.6 Organometallic, Organophosphorous, Compounds

In order to implement these rules in a general way, two types of meta-atoms have been defined. The first have to do with the types of carbon bonding, namely, double bonded carbon (C_d), triple bonded carbon (C_t), carbon on a benzene ring (C_B), allenic carbon (C_a) and carbon on a fused ring (C_{bf}). The second type of meta-atom have multiple atoms, namely, CO, NO₂, CN, SO, SO₂, CS, and PO. Associated with each of these atom and meta-atom types is a molecular structure with wild-card atoms. Before the application of the Benson rule, the set of Benson rule meta-atoms are substituted into the structure. The atom type of the atom changes. For example, for double-bonded carbon, the atom type is changed from simple C to C_d .

For multiple atom meta-atoms, the entire group is substituted. For example, the meta-atom CO is bonded to two ligands at the C. Upon substitution, the CO atoms are eliminated and the two ligands are connected with a new atom type labeled 'CO'. In general, within the meta-atom, the distinctly specified atoms are eliminated leaving the bonding to the wild-card atoms.

These meta-atoms are defined within the database. New meta-atom substitutions can be easily implemented by updating the database (for example within the graphical interface). The core JAVA library does not have to be modified by a software specialist.

3.2 Recognition of Symmetry

One of the second order corrections involves the automatic recognition of molecular symmetries: internal and external symmetry and optical isomers. Symmetry is inherently based on the three dimensional structure. However, the information available to JOTHERGAS is a molecular graph. The three dimensional information has to be inferred and assumptions have to be made. This is a design decision that has to be made when creating the database for symmetry structures.

As with the rest of JOTHERGAS, the philosophy of the implementation of symmetry recognition is that all information in the recognition process is within the extendable database. There are basically two pieces of information that are needed, the generalized structure with unspecified ligands and a table specifying the properties of the ligands.

The key to the symmetry recognition process is to match (atom-atom correspondences) the symmetry structure within the target molecule and then identify the substructures corresponding to the unspecified atoms. For example, in Figure 3, the carbon in the general structure can be matched with all the carbons in propane. For the match with C2, the middle carbon in propane, R1 and R2 can be matched the the hydrogens and R3 and R4 can be matched with the methyl groups. Due to the symmetry of the system, there are 4! ($4*3*2*1$) or 24 ways to match the unspecified atoms with the center carbon of propane. For the symmetry determination, the information is condensed to that the carbon has two sets of two matching groups, i.e. the set with hydrogens and the set with the methyl groups. For external symmetry, it is also important that the hydrogens, being single atoms, are considered to be linear structures and that the methyl groups have a symmetry of 3.

The generalized procedure can be said to be:

1. Match the symmetry structure within the target
2. For each match, identify the groups attached to each of the unspecified atoms (labeled R in the example).
3. Identify which ligands are the same and group them together.

4. Identify the symmetry of each group or whether the group is linear or not.
5. Condense this information into a table of groups where each group has the number of equivalent ligands and the symmetry or linearity of ligand.

The table derived from the match is then used compared to symmetry tables in the database to determine the symmetry.

Using this procedure, a new symmetry can be added by just defining a new symmetry structure (with unspecified atoms), the corresponding table and the final derived symmetry value. These are just new entries in the JTHERGAS database. This facilitates the addition of new functional groups which could have their own set of symmetries.

3.2.1 External Symmetry—The external symmetry of a molecule represents the total symmetry of a molecule (taking into account a degree of free rotation). In the example presented, the procedure determines that propane has a symmetry of two because it matches the table where there are two sets of two matching ligands. In this case, the symmetry or the linearity of the ligands does not affect the total external symmetry. Within JTHERGAS, there are four symmetry specifications which yield a symmetry factor greater than one. These are given in the table. For example, using Table 4, 2,2-methylpropane has symmetry 12, $C(NH_2)_4$ has a symmetry of 4, propane has a symmetry 2, $CH(CN)_3$ has a symmetry of 3. For structures which multiple forms which could produce different symmetries assumptions have to be made about the ‘average’ or most prominent structure which may or may not have a symmetry contribution. This is a design decision which has to be made when creating the database.

Other similar symmetry tables are and can be defined for linear, doubly and triply bonded and saturated and unsaturated ringed structures. It can be that several tables (with structure) can match within a given target. The external symmetry is chosen to be the largest symmetry number.

For the determination of external symmetries, there are two sets of tables. The primary table is what has just been described. But a secondary table exists which describes the symmetry of the isolated ligands corresponding to the unspecified atoms.

3.2.2 Internal Symmetry—The internal symmetry involves counting the number of times a symmetry structure exists within the target molecule. Currently the internal symmetry is calculated by the number of symmetry groups matching a carbon atom where three groups are the same and the fourth different. Each contributes a multiplicative factor of 3. This structure is represented by the general carbon atom bonding shown in Figure 3. The associated table (similar to Table 4) has two groups, one group denoting three equivalent connecting substructures and one group with the remaining substructure. In the notation of the previous section this is $C(B_1)(B_2)^3$ where B_1 and B_2 are arbitrary connecting groups. For example, terminal methyl groups and the second carbon in 2,2-dimethylpentane match this symmetry description.

3.2.3 Optical Isomers—There currently are two optical isomer specifications within JTHERGAS. The first is that around a carbon atom. This means that the general carbon atom shown in Figure 3 (right hand side) is used. Each carbon atom in which every ligand is different (corresponding to four groups of one element in each group) contributes one optical isomer. A second optical isomer is defined as a hydroperoxy group (OOH) which induces an optical isomer due to the high rotation energy barrier. The symmetry structure is ROOH, where R is left unspecified. A given molecule can have several optical isomer contributions.

3.3 Energetic Corrections

Another set of energetic corrections has to do with ring strain and rotational hindrances. In both cases, the correction is recognized through representative substructures in the database. Associated with each structure is a set of thermodynamic corrections similar to the Benson rules.

3.3.1 Ring Strain Corrections—Ring strain energies are recognized through representative ring substructures. These substructures are usually represented by the heavy atom (meaning no hydrogen) skeletal structure of the ring. For multiple ring structures, the largest matching ring is chosen. Currently there are 65 single ringed cyclic corrections corresponding to the cyclic structures in Benson's tables[18]: Hydrocarbons (Table A.1), Oxygen-containing rings (Table A.2), nitrogen-containing rings (Table A.3) and organosulfur ring compounds (Table A.5). Data have been up dated and supplemented from Domalski and Hearing[48] and Cohen[27]. Once again, the exact energetic contribution of a particular ring is a design decision as to which is the prominent or average configuration.

3.3.2 Steric Hindrances—Steric hindrances are caused by rotation around single bonds. The corrections for cis-alkene are not included because this is not represented in the current molecular form. The 2D-graphical representation has no spatial component cis-trans information would have to be given as supplementary flag to the double bond. Steric hindrances are determined also by substructure matching. In these cases, the spatial information is implied by the structure. When a substructure is found in the molecule, then the thermodynamic quantities associated with that structure are applied.

4 Method for Calculating Radicals

The method of calculating radical thermodynamics in JATHERGAS is based on the difference between the radical and the parent molecule with a hydrogen added. The full Benson method is applied to the parent molecule (Benson rules with corrections). For the radical only the corrective terms are calculated:

1. Internal Symmetry
2. External Symmetry
3. Optical Isomers
4. Steric Corrections

These are calculated for the radical in the same way as for the parent molecule. The difference of these corrections are added into the final result. In addition, for the radical species the following are estimated:

1. Dissociation Energy
2. Loss of Vibrational modes
3. Loss of Rotational modes

The vibrational and rotational modes are translated to corrections to the temperature dependent heat capacities and to entropies. In this implementation, the loss of rotational modes is neglected. This analysis will appear in later implementations.

4.1 Dissociation Energy

The key to the estimation of the disassociation energy is a database of radicals with associated dissociation energies. The dissociation energy of the target radical is estimated by

finding a the largest (defined by the number of atoms in the structure) radical structure in this database that is a substructure. The database is based on the article of Luo[49] and consists of 109 radical species: 10 alkyl radicals up to 4 carbons, 23 alkene and alkyne radicals, 35 alkyl radicals with oxygen, 2 containing sulfur, 12 alkyl radicals containing nitrogen, 23 alkyl radicals containing halogens and 4 alkyl radicals containing silicon.

4.2 Loss of Vibrational States

For the calculation of the thermodynamics of radicals within the JTHERGAS system, the loss of vibrational states due to the loss of a hydrogen atom to form the radical is computed. The vibrational states lost are translated to corrections to the temperature dependent heat capacities. The estimation of vibrational states is also based on a table of structures. A vibrational mode is identified by a given structure. Associated with the structure is a vibrational frequency (For example, see Table A.13 in Benson's book[18]). This vibrational frequency is then translated to contributions to the entropies and the heat capacities as a function of temperature (see Table A.15 and A.17) in Benson's book.

Also associated with each vibrational structure is a numerical factor. This factor serves two purposes. The first is to take care of symmetries in the vibrational structure. For example, the bend structure, H-C-H, due to its symmetry, matches twice as many times. To ensure the counting, the number of matches has to be divided by two. The other purpose of the numerical factor is to determine whether the structure should be added (positive factor) or subtracted (negative factor). In counting the vibrational structures both matches in the parent molecule and the radical are made. The number of times a certain vibrational mode appears in the parent mode and the number of times a vibrational mode appears in the radical are subtracted. A positive numerical factor keeps this relationship and a negative numerical factor reverses this relationship. This was incorporated to give more flexibility in the definition of which modes should be counted.

Within JTHERGAS, the implementation goes beyond what is outlined in Table A.13 in Benson's book[18] for carbon radicals. The loss of tetrahedral modes is calculated for loss of a hydrogen atom in primary, secondary and tertiary carbon configurations. This replaces the simple CCH and HCH bond bends listed in Benson's tables. In addition, corrections due to resonance are also included. For example, for the calculation of a carbon next to a double bond, i.e. a resonant structure, the structure $\cdot\text{CH}_2=\text{C}$ is matched. The associated frequencies are: 1300 (resonant stretch), two 1150 (twist and rock primary bends) and 1450 (primary bend). These frequencies are then translated to corrections in entropies and temperature dependent heat capacities.

5 Foundation of Standard Libraries

In order to make JTHERGAS more extendible and usable by other systems, an effort has been made to use standard packages within the industry. The main advantages are that other independent projects, both within chemistry and the information technologies, particularly web technologies, provide additional libraries and interfaces based on these standard packages. For example, the use of a *standardised* molecular representation, such as that in CDK, facilitates interfacing to other chemical applications.

5.1 JTHERGAS Core Algorithm JAVA Library

The heart of the JTHERGAS system is the core JAVA library. This is where all the primary algorithms lie for the calculation of the thermodynamic quantities. Along with SQL database, the JTHERGAS core system is all that is needed to calculate thermodynamic values. Other libraries, built in separate packages provide other interfaces such as to GUI's, web pages and web-services.

5.2 mySQL Database Tables

The mySQL (www.mysql.com) database management system was chosen for a variety of reasons. The first is that it “runs on more than 20 platforms including Linux, Windows, Mac OS, Solaris, HP-UX, IBM AIX” and “has become the world’s most popular open source database because of its high performance, high reliability and ease of use”. In addition, many interfaces exist to access a mySQL database. Not only are there numerous GUI interfaces to directly accessing the information in the database, but there are programming interfaces for the many programming languages, such as JAVA used for JTHERGAS and web languages such as javascript and php, which used for the web interfaces in JTHERGAS.

5.3 Standard JAVA libraries and molecular representations

The Chemistry Development Kit[46] (CDK, <https://sourceforge.net/projects/cdk/>) is a library of JAVA classes for “chemo- and bioinformatics” and is becoming a standard package for chemical applications in JAVA. Though primarily maintained by Egon Willighagen, CDK has contributions from over 50 developers. JTHERGAS uses CDK primarily for the molecular representations, ring recognition and the isomorphism algorithms.

An extension of CDK, Jumbo, is a standardized package for converting molecular representations, including that of CDK to CML that is used by JTHERGAS for database and web interfaces.

6 Calculation example

The following is an example showing how the flexible substructure approach is used within JTHERGAS to calculate the temperature dependent thermodynamic properties of 3-pentanone-1-yl radical ($\text{CH}_3\text{CH}_2\text{COCH}_2\text{CH}_2\cdot$) from the NLF representation, `ch2(./ch2/co/ch2/ch3`.

The first stage is to convert the ASCII NLF string to the internal CDK JAVA class Molecule. In the automatic interpretation of the NLF string, one sees an ambiguity when meta-atoms, in this case `co`, are used. This could be interpreted as a single carbon oxygen bond. The first step is to convert the original NLF string to a more strict form, `ch2(./ch2/‘co’/ch2/ch3`, where multiple character meta atoms are put in quotes. In the JTHERGAS database, there is a list of all the possible NLF meta atoms, each with their structure and corresponding ASCII string. In this example, the structure of the general ketone (as in the right hand side of Figure 2) is associated with the string ‘co’. The original string is then transformed with the matches of the NLF meta atoms.

The second step is to transform the NFL string (with quoted meta atoms) into the CDK Molecule class. The first representation is with the meta atom as a single atom with the name ‘co’ (CDK Atom Class) within the molecule. The next step is to substitute the meta atom structure into the molecule. The substructures connected to the meta atom are isolated, noting where they connect to the meta atom. The meta atom is then removed from the molecule description and the meta atom structure, the specified atoms, is added to the molecule. In this example, the doubly bonded carbon and oxygen are added. Within the meta atom description is where it should be connected within the molecule (the bonds to the ‘starred’ atoms in Figure 2). Bond are added between the meta atom structure and the isolated substructures. This procedure is repeated for all meta atoms in the molecule.

The above is an example of how an arbitrary meta atom in the NLF form can be defined. All that is needed is a structure definition and a corresponding ASCII string. At a later stage of the calculation, to apply the Benson rules (on the parent molecule), also invokes a similar

process. A set of Benson meta atoms (which are often similar, but not necessarily the same, as the NLF meta atoms) are defined for this process.

The calculation of the thermodynamics of a radical consists doing a full calculation of the parent molecule, i.e. the molecule where a hydrogen is added at the radical center, and finding the differences in symmetry, both external and internal, finding the losses of vibrational and rotational modes due to the radical center and determining the dissociation energy. The parent molecule is formed by making a copy of the original molecule, adding a hydrogen atom to the copy, bonding it to the molecule at the radical center and eliminating the radical flag at that center.

The next stage is to compute the thermodynamics of the parent molecule using Benson additivity rules. A Benson additivity rule consists of a center atom (other than hydrogen), a list of connecting atoms and the corresponding thermodynamic information. The atoms can also be meta atoms. The molecule is processed by making a list of applicable Benson rules based on the set of center atoms found in the molecule. Since the 'atoms' can also be Benson meta atoms, the first step is to isolate (and substitute) the Benson meta atoms into the molecule. This is done by collecting all the Benson meta atom structures in the database and to try to see if they match (starting with the largest structures) into the molecule. In the example, the Benson atom structure that matches is the ketone group (as before, but not associated with the Benson meta atom tables). The (unstarred) matched atoms, the carbon and the oxygen in the example molecule, are eliminated from the molecule, once again noting the bonding with the remaining molecule. A Benson meta atom is created, with the ASCII string name of the Benson meta atom group, and bonded into the molecule.

Once all the Benson meta atoms are substituted in, then all the non-hydrogen atoms in the molecule are transformed into a list of Benson rules. The thermodynamic database is then searched for each of these rules to extract the thermodynamic information. The thermodynamic contributions from the five rules for the parent molecule are shown in Table 5.

The next stage in the calculation is the determination of the symmetries, internal, external and optical, of the parent molecule and the radical. As described previously, this process consists of finding a symmetry structure within the molecule and then applying the table of rules.

To determine the external symmetry of the parent molecule, matches between the list of external symmetry structures and the parent molecules are made. The structures in the list are tried from largest to smallest. The only structure which matches the example is that of ketone symmetry structure (once again the right hand side of Figure 2). The associated table says that if the two connecting groups are the same, then the molecule has a external symmetry of 2. Within the parent molecule, the connecting groups (CH_3CH_2) are then isolated. In general, the algorithm then determines, through graph isomorphism, which groups are the same and groups them together. In the example, the two connecting substructures are the same, so one group with two identical structures is formed. This matches the ketone symmetry table, thus the parent molecule has a symmetry of two. Doing the same analysis on the radical, it is found that the two connecting substructures are not the same (CH_3CH_2 and $\text{CH}_2\text{CH}_2\cdot$) and thus two groups of one structure is formed, which does not match the ketone symmetry structure. The radical does not match any other symmetry structures so the radical has an external symmetry of one. The contribution to the entropy due to the differences in symmetry is:

$$R \ln \frac{\sigma_{RH}}{\sigma_R} = R \ln \frac{2}{1} = R \ln 2 \quad (4)$$

To determine the internal symmetry, the algorithm counts how many times the internal symmetry structure matches within the molecule. Currently there is one structure, a carbon atom where three of the connecting groups are equal. The internal symmetry structure is the general carbon atom (see Figure 3 left hand structure). In the parent molecule, this structure matches four carbons (all carbons except the ketone) producing $4 \times 4! = 96$ ($4! = 4 \times 3 \times 2 \times 1$) atom-to-atom correspondence lists. With each one of these correspondence lists, the connecting substructures are isolated. In this case, the connecting substructures to the center carbon. The substructures are compared and those which are the same are grouped. A group table is formed. For the terminal carbons, each of the 48 group tables has three equivalent hydrogen atoms and the remaining substructure ($\text{CH}_2\text{COCH}_2\text{CH}_3$), but each with a different atom-to-atom correspondence list. Two of these tables are considered equivalence if the set of matching atoms within the molecule are the same set. This equivalency reduces the 48 group tables down to two, one for each end of the molecule. This table also matches the table structure of the internal symmetry table. The tables which are set up for the other 48 secondary carbon atom-to-atom correspondences have the structure of three groups, two equivalence hydrogens and two more individual structures (CH_3 and COCH_2CH_3). This does not match the internal symmetry table, so does not contribute to the internal symmetry. The two internal symmetry structures give a total internal energy $3^2 = 9$. The same analysis for the radical gives a match for only the terminal methyl. The other methyl is now a radical. This gives an internal symmetry of the radical to be $3^1 = 3$. The contribution to the entropy with this analysis is:

$$R \ln \frac{\sigma_{RH}}{\sigma_R} = R \ln \frac{9}{3} = R \ln 3 \quad (5)$$

Optical isomers are also found through basically the same technique. Carbon atoms are only optical isomers when all four groups are different and as was shown in the last analysis, none of the carbon atoms fit this symmetry. In addition, there are no hydrogen peroxy groups. Thus there is no contribution to the entropy.

The vibrational analysis is based on finding the loss of vibrational modes due to the formation of the radical. The analysis proceeds by searching both the parent molecule and the radical molecule for vibrational structure matches and counting the number of matches. The number of each match in the parent molecule is subtracted from the modes found in the radical molecules. For the primary radical in the example, the vibrational modes are found with a primary radical structure, a carbon hydrogen bond and a carbon radical hydrogen bond and the information in Table 6.

The dissociation energy is calculated with the *best guess*. For each structure in the library of dissociation energy structures, starting with the largest (defined by the number of atoms) a match is attempted within the radical molecule structure. If a match is found, then the associated dissociation energy is used. For the example the best match found is based on the CH_2CH_3 radical, namely 100.9 kcal/mol.

This example has shown the role that substructures play in the thermodynamic analysis of a structure. In this example, the ketone functional group was used in the definition of the NLF, the Benson rule and in the symmetry stages of the calculation. Using the ketone functional group as an example, one can see what structures and tables have to be defined when introducing a new functional group.

Table 7 shows that the JTHERGAS estimates are close to the literature values. The table shows several types of molecules. C_3H_8 and n-heptane are simple straight chained molecules. nC_3H_7 and sC_3H_7 are two simple radicals. Neopentane is a branched molecule with a external symmetry of 12. Isooctane (2,2,4-trimethylpentane) and 2,2,3,3-tetramethylbutane are two examples with gauche corrections for the methyl groups. 1-butanol and 2-butanal are two alcohols. Butanal is an example of an aldehyde and 2-butanone is an example of a ketone. 2-chloropentane is simple molecule with a halogen. Methycyclopentane and 1,4-dioxarane are two with ring corrections.

7 Interface Implementation

JTHERGAS has been designed to have multiple interfaces, not only those provided by the standard libraries on which it is based, but also on implementations designed specially for JTHERGAS. The goal of the interfaces is to facilitate the use of JTHERGAS by chemist and eliminate the need for an information technology specialist to, for example, update the database. The interfaces range in complexity and ease of use from simple line commands for basic use and database management, to web-services for access to functionality over the internet, to web page GUI interfaces with forms and molecular drawing facilities.

The overall schematic for the JTHERGAS interface implementation is shown in Figure 4. The Jthermodynamic Core System is the heart of the JTHERGAS system and has all the functionality needed for computations and database management. It resides on a local system or *Server* (meaning it can provide services to other computers through *Client* interfaces. Local use of the JTHERGAS system can be through line commands (part of the JTHERGAS system) or a local GUI interface (such as a JAVA package developed by other users – as a note, the GUI interfaces developed within JTHERGAS are implemented through web pages). Through *natural language interfaces*, meaning interfaces that allow libraries to be exchanged between two different programming languages, the functionality of JTHERGAS can be compiled into other software running on the same local computer.

Acting as a *Server*, the JTHERGAS system can provide interfacing to other *Client* computers connected on the web. This is done through two types of interfaces: web-services and a web page server (such as Apache or Glassfish). Web-services essentially provide a function that can be called over the web. The web page server manages the web pages provided by JTHERGAS to compute and manage the database. a central server can be maintained having a *standard* database that can be used by multiple users on other systems.

7.1 Operating System Independent

The use of JAVA as the primary programming language was the first design decision to ensure operating system independence. JAVA is available on all standard operating systems including all Windows, LINUX and MAC systems (the JAVA runtime environment is installed by the user). This means that the local installation of the JTHERGAS system can be made on any operating system.

The second design decision to ensure operating system independence was the reliance on standardized external libraries and packages that have implementations on all operating systems. In addition, the use of the Client-Server implementation provides another level of operating system independence, namely, that the operating system of the clients is independent of the operating system of the server system.

7.2 Line Command

The line command interface is based on the core JTHERGAS library and is the minimal system having all the available functionalities. It is designed for easy implementation of a

local copy of the JTHERGAS system were the individual user make thermodynamic queries and management and update the chemical database. Database update is done through human readable and editable ASCII files using primarily the Nancy linear form for molecules and substructures.

7.3 Web Services

In order to provide interfacing of the primary functionality of the JTHERGAS system to applications over the internet, web-services are provided. Web-services are essentially functions which can be called from one machine, evaluated on another machine somewhere in cyberspace and the result returns to the calling machine. The concept of web-services is a W3C (www.w3.org) standard and is used commonly by implementations of web technologies.

The advantage of using web-services in JTHERGAS is that it provides a standard interface to a multitude of languages and platforms. A given application could access the JTHERGAS system from, for example, the *main* server and take advantage of all the latest updates within the database without the user having to maintain the database himself. The same interface, with a parameter adjustment specifying source, could access a local version.

7.4 Graphical Interfaces over the internet

In the long run, graphical interfaces using web pages will be the primary means to make computations and update and manage the JTHERGAS database. The design philosophy of these pages is to provide the ability of non-information technology specialists to be able update and manage the JTHERGAS database, i.e. only chemical knowledge of the information needed by the JTHERGAS system will be needed.

The web page interface has the advantage that a central system running the JTHERGAS system can serve a range of computers. However, it should be noted that the installation of the full system with the web-service and web page servers involves system privileges on the local computer.

8 Conclusion

This paper presents an outline of the JTHERGAS system for the computation of thermodynamic properties given a two dimensional molecular representation. This type of computation provides not only a good *first guess* for temperature dependent thermodynamic quantities, but is also an essential component in automatic combustion mechanism generators which generate a large quantity of a large variety of chemical molecules and radicals.

JTHERGAS is unique on several levels. The first is that the estimation of the thermodynamics goes beyond simple Benson rules and introduces automatic symmetry and steric corrections. The second is that a general method is provided for the calculation of radical species.

JTHERGAS is unique at the implementation level also for several reasons. The implementation philosophy is that of flexibility and ease and transparency of database extension. The recognition of symmetry and the structures contributing to steric corrections and ring strain are all based on database objects and not on hard coded. This means that the data used to make all the calculations are visible and that extensions to new structural features is a matter of updating only the database. Another uniqueness of the implementation strategy of JTHERGAS is the multitude of interfaces available from line commands to web interfaces. The design philosophy at this level was that non-information technology

specialists can update the database with only chemical knowledge of the system. In addition, the interfaces provide a platform independent implementation.

The initial use of the platform independent, easily extendable JTHERGAS is to provide a reasonable first guess of a large range of organic (nitrogen, sulfur and oxygen containing groups) molecules and radicals that appear in combustion processes from a two dimensional graphical representation. It is especially useful for systems for the automatic generation of combustion mechanisms.

References

- [1]. Yao M, Zheng Z, Liu H. Progress and recent trends in homogeneous charge compression ignition (hcci) engines. *Progress in Energy and Combustion Science*. 2009; 35:398–437.
- [2]. Gupta KK, Rehman A, Sarviya RM. Bio-fuels for the gas turbine: a review. *Renewable and Sustainable Energy Reviews*. 2010; 14:2946–2955.
- [3]. Kou N, Zhao F. Techno-economical analysis of a thermo-chemical biofuel plant with feedstock and product flexibility under external disturbances. *Energy*. 2011; 36:6745–6752.
- [4]. Westbrook CK, Pitz WJ, Herbinet O, Curran HJ, Silke EJ. A comprehensive detailed chemical kinetic reaction mechanism for combustion of n-alkane hydrocarbons from n-octane to n-hexadecane. *Combustion and Flame*. 2009; 156:181–199.
- [5]. Goos, E.; Burcat, A.; Ruscic, B. Ideal gas thermochemical database with updates from active thermochemical tables. n.d.
- [6]. Ruscic B. Iupac critical evaluation of thermochemical properties of selected radicals. part i. *J. Phys. Chem. Ref. Data*. 2005; 34:573.
- [7]. Pedley, JB. Thermochemical data and structures of organic compounds. 1st ed.. CRC Press; 1994.
- [8]. Stull, DR. The chemical thermodynamics of organic compounds. John Wiley & Sons Inc; 1969.
- [9]. Moréac G, Blurock E, Mauss F. Automatic generation of a detailed mechanism for the oxidation of n-decane. *Combustion Sc. & Tech*. 2006; 178:2025–2038.
- [10]. Pierucci S, Ranzi E. A review of features in current automatic generation software for hydrocarbon oxidation mechanisms. *Computers & Chemical Engineering*. 2008; 32:805–826.
- [11]. Buda F, Bounaceur R, Warth V, Glaude PA, Fournet R, Battin-Leclerc F. Progress toward a unified detailed kinetic model for the autoignition of alkanes from c4 to c10 between 600 and 1200 k. *Combustion and Flame*. 2005; 142:170–186.
- [12]. Balaban AT. Chemical graphs: looking back and glimpsing ahead. *Journal of Chemical Information and Computer Sciences*. 1995; 35:339–350.
- [13]. Simmie JM, Black G, Curran HJ, Hinde JP. Enthalpies of formation and bond dissociation energies of lower alkyl hydroperoxides and related hydroperoxy and alkoxy radicals. *The Journal of Physical Chemistry A*. 2008; 112:5010–5016. [PubMed: 18461912]
- [14]. da Silva G, Bozzelli JW. Enthalpies of formation, bond dissociation energies, and molecular structures of the n-aldehydes (acetaldehyde, propanal, butanal, pentanal, hexanal, and heptanal) and their radicals. *The Journal of Physical Chemistry A*. 2006; 110:13058–13067. [PubMed: 17134166]
- [15]. Andronico A, Randall A, Benz RW, Baldi P. Data-driven high-throughput prediction of the 3-d structure of small molecules: review and progress. *J. Chem. Inf. Model*. 2011; 51:760–776. [PubMed: 21417267]
- [16]. Benson SW, Buss JH. Additivity rules for the estimation of molecular properties. thermodynamic properties. *J. Chem. Phys*. 1958; 29:546.
- [17]. Benson SW, Cruickshank FR, Golden DM, Haugen GR, O'Neal HE, Rodgers AS, Shaw R, Walsh R. Additivity rules for the estimation of thermochemical properties. *Chemical Reviews*. 1969; 69:279–324.
- [18]. Benson, SW. Thermochemical kinetics: methods for the estimation of thermochemical data and rate parameters, 2nd ed. 2nd ed.. John Wiley & Sons; 1976.

- [19]. Laidler KJ. A system of molecular thermochemistry for organic gases and liquids. *Canadian Journal of Chemistry*. 1958; 34:626–648.
- [20]. Lovering EG, Laidler KJ. A system of molecular thermochemistry for organic gases and liquids: part ii. extension to compounds containing sulphur and oxygen. *Can. J. Chem.* 1960; 38:2367–2372.
- [21]. Lovering EG, NOR OBM. A system of molecular thermochemistry for organic gases and liquids .3. extension to aromatic and cyclic compounds. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* n.d. 40:199–204.
- [22]. Zavitsas AA, Rogers DW, Matsunaga N. Heats of formation of organic compounds by a simple calculation. *The Journal of Organic Chemistry*. 2010; 75:6502–6515. [PubMed: 20825149]
- [23]. Bojesen G. A statistical- mechanical analysis of group additivity. calculation of thermochemical values from frequency distributions. *Journal of Physical Organic Chemistry*. 2008; 21:833–843.
- [24]. Constantinou L, Gani R. New group contribution method for estimating properties of pure compounds. *AIChE Journal*. 1994; 40:1697–1710.
- [25]. Alis OF, Rabitz H. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry* n.d. 25:197–233.
- [26]. Hayes MY, Li B, Rabitz H. Estimation of molecular properties by high-dimensional model representation. *J Phys Chem A*. 2006; 110:264–272. [PubMed: 16392864]
- [27]. Cohen N. Thermochemistry of alkyl free radicals. *The Journal of Physical Chemistry*. 1992; 96:9052–9058.
- [28]. Ashcraft RW, Green WH. Thermochemical properties and group values for nitrogen-containing molecules. *J. Phys. Chem. A*. 2008; 112:9144–9152. [PubMed: 18754603]
- [29]. Sabbe MK, De Vleeschouwer F, Reyniers M-F, Waroquier M, Marin GB. First principles based group additive values for the gas phase standard entropy and heat capacity of hydrocarbons and hydrocarbon radicals. *The Journal of Physical Chemistry A*. 2008; 112:12235–12251. [PubMed: 18980365]
- [30]. Yu J, Sumathi R, Green WH. Accurate and efficient method for predicting thermochemistry of furans and ortho-arynes: expansion of the bond-centered group additivity method. *J. Phys. Chem. A*. 2006; 110:6971–6977. [PubMed: 16722711]
- [31]. Sebbar N, Bockhorn H, Bozzelli JW. Thermochemical properties, rotation barriers, and group additivity for unsaturated oxygenated hydrocarbons and radicals resulting from reaction of vinyl and phenyl radical systems with O_2 . *J. Phys. Chem. A*. 2005; 109:2233–2253. [PubMed: 16838995]
- [32]. Sumathi R, Green WH. Missing thermochemical groups for large unsaturated hydrocarbons: contrasting predictions of g2 and cbs-q. *J. Phys. Chem. A*. 2002; 106:11141–11149.
- [33]. Sabbe MK, Saey M, Reyniers M-F, Marin GB, Van Speybroeck V, Waroquier M. Group additive values for the gas phase standard enthalpy of formation of hydrocarbons and hydrocarbon radicals. *The Journal of Physical Chemistry A*. 2005; 109:7466–7480. [PubMed: 16834116]
- [34]. Yu J, Sumathi R, Green WH. Accurate and efficient method for predicting thermochemistry of polycyclic aromatic hydrocarbons – bond-centered group additivity. *J. Am. Chem. Soc.* 2004; 126:12685–12700. [PubMed: 15453803]
- [35]. Vansteenkiste P, Van Speybroeck V, Marin GB, Waroquier M. Ab initio calculation of entropy and heat capacity of gas phase n-alkanes using internal rotations. *J. Phys. Chem. A*. 2003; 107:3139–3145.
- [36]. Sumathi R, Green WH. Thermodynamic properties of ketenes: group additivity values from quantum chemical calculations. *J. Phys. Chem. A*. 2002; 106:7937–7949.
- [37]. Chen C-J, Wong D, Bozzelli JW. Standard chemical thermodynamic properties of multichloro alkanes and alkenes: a modified group additivity scheme. *The Journal of Physical Chemistry A*. 1998; 102:4551–4558.
- [38]. Lay TH, Bozzelli JW. Enthalpies of formation and group additivity of alkyl peroxides and trioxides. *The Journal of Physical Chemistry A*. 1997; 101:9505–9510.

- [39]. Ritter ER. Therm: a computer code for estimating thermodynamic properties for species important to combustion and reaction modeling. *J Chem Inf Comput Sci*. 1991; 31:400–408. [PubMed: 1939398]
- [40]. Ritter ER, Bozzelli JW. Therm: thermodynamic property estimation for gas phase radicals and molecules. *International Journal of Chemical Kinetics*. 1991; 23:767–778.
- [41]. Mehl M, Pitz WJ, Westbrook CK, Curran HJ. Kinetic modeling of gasoline surrogate components and mixtures under engine conditions. *Proceedings of the Combustion Institute*. 2011; 33:193–200.
- [42]. Stein, SE.; Brown, RL. Nist chemistry webbook, nist standard reference database. Vol. 69. National Institute of Standards and Technology; Gaithersburg MD: Structures and properties group additivity model; p. 20899n.d
- [43]. Dalby A, Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA, Laufer J. Description of several chemical structure file formats used by computer programs developed at molecular design limited. *Journal of Chemical Information and Computer Sciences*. 1992; 32:244–255.
- [44]. Warth V, Stef N, Glaude PA, Battin-Leclerc F, Scacchi G, Côme GM. Computer-aided derivation of gas-phase oxidation mechanisms: application to the modeling of the oxidation of n-butane. *Combustion and Flame*. 1998; 114:81–102.
- [45]. Muller C, Michel V, Scacchi G, Come GM. Thergas - a computer program for the evaluation of thermochemical data of molecules and free-radicals in the gas phase. *Journal De Chemie Physique Et De Physico-Chemie Biologiques*. 1995; 92:1154–1178.
- [46]. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The chemistry development kit (cdk): an open-source java library for chemo- and bioinformatics. *J. Chem. Inf. Model*. 2003; 43:493–500.
- [47]. Murray-Rust P, Rzepa HS. Chemical markup, xml, and the worldwide web. 1. basic principles. *Journal of Chemical Information and Computer Sciences*. 1999; 39:928–942.
- [48]. Domalski ES, Hearing ED. Estimation of the thermodynamic properties of c-h-n-o-s-halogen compounds at 298.15 k. *J. Phys. Chem. Ref. Data*. 1993; 22:805.
- [49]. Luo, Y-R. Handbook of bond dissociation energies in organic compounds. CRC Press; 2003.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<molecule id="ch(oh)(.)ch3" xmlns="http://www.xml-cml.org/schema">
  <atomArray>
    <atom id="a0" elementType="C" formalCharge="0"/>
    <atom id="a4" elementType="C" formalCharge="0" spinMultiplicity="2"/>
    <atom id="a6" elementType="O" formalCharge="0"/>
  </atomArray>
  <bondArray>
    <bond id="b4" atomRefs2="a4 a0" order="S"/>
    <bond id="b6" atomRefs2="a6 a4" order="S"/>
  </bondArray>
</molecule>
```

Figure 1. Example of CML representation of a molecule

Figure 2. The meta-atom structures associated with a doubly bonded carbon (left) and a carboxyl group (right). Note that a star is used to denote that any atom can be matched. See Section on ‘Wild-Card atoms’).

Figure 3. The structure on the left can be used to define symmetries around a carbon atom. The R atoms are the wild card atoms. Using propane as an example, C2 can be matched and R1 and R2 can be matched to H7 and H6 and R3 and R4 can be matched to the methyl groups with C1 and C5.

Figure 4. Schematic of the interfaces available for JTHERGAS.

Table 1
Example of Nancy Linear FORM (NLF)

Nancy Linear Form	Molecule
ch3/ch2/ch2/ch2/ch2/ch2/ch2/ch3	octane
ch3/c(ch3)2/ch2/ch(ch3)/ch3	2,2,4-trimethylpentane
ch3/ch//ch/ch2/ch2/ch3	2-hexene
ch(#1)&ch&ch&ch&ch&ch&1	Benzene
ch3/'co'/ch2/ch2/ch3	2-pentanone
c(#1)&c(#2)&ch&ch&ch&ch&1,1&ch&ch&ch&ch&2	naphtalene

Table 2
Atom-to-atom correspondences between ethane and primary carbon

	C1	C2	H3	H4	H5	H6	H7	H8
A1	C1	C5	H2	H3	H4			
A2	C1	C5	H2	H4	H3			
A3	C1	C5	H3	H2	H4			
A4	C1	C5	H3	H4	H2			
A5	C1	C5	H4	H2	H3			
A6	C1	C5	H4	H3	H2			
B1	C5	C1				H2	H3	H4
B2	C5	C1				H2	H4	H3
B3	C5	C1				H3	H2	H4
B4	C5	C1				H3	H4	H2
B5	C5	C1				H4	H2	H3
B6	C5	C1				H4	H3	H2

Table 3
The symmetry information table associated with matching C2 of propane with the generalised carbon atom from Figure

Group	Number of Equivalent Ligands	Symmetry or Linearity of Ligand
1 (methyl)	2	3
2 (hydrogen)	2	Linear

Table 4
The external symmetries around a carbon atom

Name of Symmetry	Symmetry	Group	Number of Ligands	Symmetry or Linearity of Group
$C(B)^4-12$	12	B	4	<i>L or 3</i>
$C(B)^4-4$	4	B	4	<i>Not L or 3</i>
$C(B_1)(B_2)^3-3$	3	B₁	1	<i>L or 3</i>
		B₂	3	<i><u>U</u></i>
$C(B_1)^2(B_2)^2$	2	B₁	2	<i>U</i>
		B₂	2	<i>U</i>

Table 5
Benson Rule Thermodynamic values for example. Enthalpy is in kilocalories, the entropy in calories per degree Kelvin and the heat capacities are in calories per degree Kelvin

Rule	Mult	ΔH_{298}	ΔS_{298}	C_{p30}	C_{p400}	C_{p50}	C_{p600}	C_{p800}	C_{p1000}	C_{p1500}
				0		0				
co-(c)/2	1	-31.7	15.367	5.6	6.23	7.09	7.76	8.89	9.61	
c-(h)/3-(c)	2	-10.098	30.423	6.15	7.84	9.4	10.79	13.02	14.77	17.58
c-(co)-(c)-(h)/2	2	-5.219	9.458	5.9	7.7	8.7	9.5	11.1	12.2	

Table 6
Difference in vibrational modes of parent and radical species

Vibrational Mode	Structure	Factor	Frequency	Parent	Radical	Total
CCH(rk) Bend	$\cdot\text{CH}_2\text{C}$	1	1150	0	2	-2
CCH(tw) Bend	$\cdot\text{CH}_2\text{C}$	-1	1150	0	2	2
CHH Bend	$\cdot\text{CH}_2\text{C}$	1	1450	0	2	-2
CH Stretch	CH	1	3100	10	7	3
$\text{C}_{\text{radical}}\text{H}$ Stretch	$\text{C}_{\text{radical}}\text{H}$	1	3100	0	2	2

Table 7
Comparison between JTHERGAS estimates and literature values. The literature values come from two sources, a) Computational Chemistry Comparison and Benchmark Database (<http://cccbdb.nist.gov/>) and b) NIST chemistry webbook (<http://webbook.nist.gov/chemistry/>). ΔH_{298} is in kcal/mole, ΔS_{298} is in cal/mole/K and C_{p300} is in cal/mole/K.

Molecule	JTHERGAS			Literature			Source
	ΔH_{298}	ΔS_{298}	C_{p300}	ΔH_{298}	ΔS_{298}	C_{p300}	
C3H8	-25.13	64.46	17.77	-25.02	64.61	17.59	a
sC3H7	21.37	66.12	13.63	-21.47	67.16	16.55	a
nC3H7	23.67	66.89	16.61	-24.09	67.88	17.36	a
neopentane	-40.16	72.03	28.55	-40.13	73.14	28.88	a
n-heptane	-44.85	101.89	39.65	-45.24		39.48	b
2-methylhexane	-46.44	99.99	39.66	-46.6	100.38	39.52	b
2,2,4-trimethylpentane	-53.26	99.85	44.97	-53.57		45.03	b
2,2,3,3-tetramethylbutane	-53.2	92.26	44.08	-54.06	93.04	44.74	a
1-butanol	-65.89	86.36	26.29	-65.63	86.42	25.82	a
2-butanol	-70.4	85.25	26.91	-69.96			a
2-chloropropane	-33.48	73.53	20.66	-34.61	73.15	20.93	b
butanal	-49.97	82.19	24.45	-49.25	82.16	24.71	a
2-butanone	-57.11	81.31	23.08	-55.83	81.14	24.30	a
1,4-dioxane	-76.61	71.33	28.03	-75.36	71.65	22.01	a
methylcyclopentane	-25.96	80.12	26.33	-25.50	81.24	26.27	a