

MIT Open Access Articles

Design and implementation of a next-generation software interface for on-the-fly quantum and force field calculations in automated reaction mechanism generation

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Magoon, Gregory R., and William H. Green. "Design and Implementation of a Next-Generation Software Interface for on-the-Fly Quantum and Force Field Calculations in Automated Reaction Mechanism Generation." *Computers & Chemical Engineering* 52 (May 2013): 35–45.

As Published: <http://dx.doi.org/10.1016/j.compchemeng.2012.11.009>

Publisher: Elsevier

Persistent URL: <http://hdl.handle.net/1721.1/101193>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-NonCommercial-NoDerivs License



Design and implementation of a next-generation software interface for on-the-fly quantum and force field calculations in automated reaction mechanism generation

Gregory R. Magoon^a and William H. Green*

Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139

a) Present address: Aerodyne Research, 455 Manning Rd., Billerica, MA 01821

*Corresponding author: whgreen@mit.edu

Abstract

A software interface for performing on-the-fly quantum and force field calculations has been developed and integrated into RMG, an open-source reaction mechanism generation software package, to provide needed estimates of thermodynamic parameters. These estimates based on three-dimensional molecular geometries bypasses the traditional group-additivity-based approach, which can suffer from lack of availability of necessary parameters; this issue is particularly evident for polycyclic species with fused rings, which would require *ad hoc* ring corrections in the group-additivity framework. In addition to making extensive use of open-source tools, the interface takes advantage of recent developments from several fields, including three-dimensional geometry embedding, force fields, and chemical structure representation, along with enhanced robustness of quantum chemistry codes. The effectiveness of the new approach is demonstrated for a computer-constructed model of combustion of the synthetic jet fuel JP-10. The interface also establishes a framework for future improvements in the chemical fidelity of computer-generated kinetic models.

Background

Automated reaction mechanism generation software is an important tool in the generation of detailed chemical kinetic models for complex reacting systems (e.g. pyrolytic and combustion systems). RMG (Reaction Mechanism Generator) is one example of such software¹, which uses a rate-based mechanism construction algorithm², and has been applied to a number of systems, including combustion of butanol³ and JP-10⁴. The estimation of thermochemical parameters (enthalpy, entropy, and heat capacity) is a key component of automated reaction mechanism generation programs like RMG. Such thermochemical parameters affect reaction equilibrium constants, kinetic parameter estimates, and thermal effects, influencing both the mechanism generation process and the behavior of the final resulting model. The scale of the problem is significant, as parameters must be estimated for tens or hundreds of thousands of chemical species in the process of generating a typical detailed chemical kinetic model with hundreds of species.

Traditionally, automated reaction mechanism generation software that requires parameter estimates relies on a group additivity approach, such as the approach developed by Benson⁵, for estimating required thermodynamic quantities (enthalpy, entropy, and head capacity). This process is very fast and usually quite accurate. However, the approach relies on the availability of appropriate

parameters. For many classes of compounds, including acyclic species with C, H, and O, the Benson approach can provide accurate results with a relatively small number of parameters. However, when cyclic species are considered, specialized ring corrections are required; these corrections are not as extensible and *ad hoc* ring corrections specific to a particular polycycle are often required; examples include specific “norbornadiene” and “quadricyclane” ring corrections applied to species with these particular polycyclic structures. Therefore, an approach that bypasses the need for specialized parameters, such as *ad hoc* ring corrections, is highly desirable.

One approach to enhance the generality of the thermodynamics parameter estimation process and reduce the need for specialized parametrization is to utilize explicit three-dimensional molecular structure representations (rather than connectivity-based molecular structure representations). Explicit three-dimensional molecular structures can be used with quantum mechanics or force field calculations to obtain estimates of the desired thermodynamics quantities. A proof-of-principle for on-the-fly quantum calculations in automated reaction mechanism generation was developed by Broadbelt et al. in 1994.⁶ The interface described here extends the general approach of Broadbelt et al., incorporating recent developments in several fields, to make on-the-fly quantum and force field calculations more suitable for routine use in automated reaction mechanism generation. We here refer to this as the quantum mechanics thermodynamic property (QMTP) interface.

Design and implementation of the QMTP interface

Design overview

An overview of the QMTP interface is shown in Figure 1. The workflow starts by estimating a three-dimensional molecular structure using RDKit, followed by calls to an outside program to perform quantum mechanics or force field calculations to refine that geometry and compute its enthalpy and vibrational frequencies; the results of the calculation are then read and used to calculate the desired thermodynamic properties using standard statistical mechanical relationships within the framework of the rigid rotor/harmonic oscillator approximation.⁷

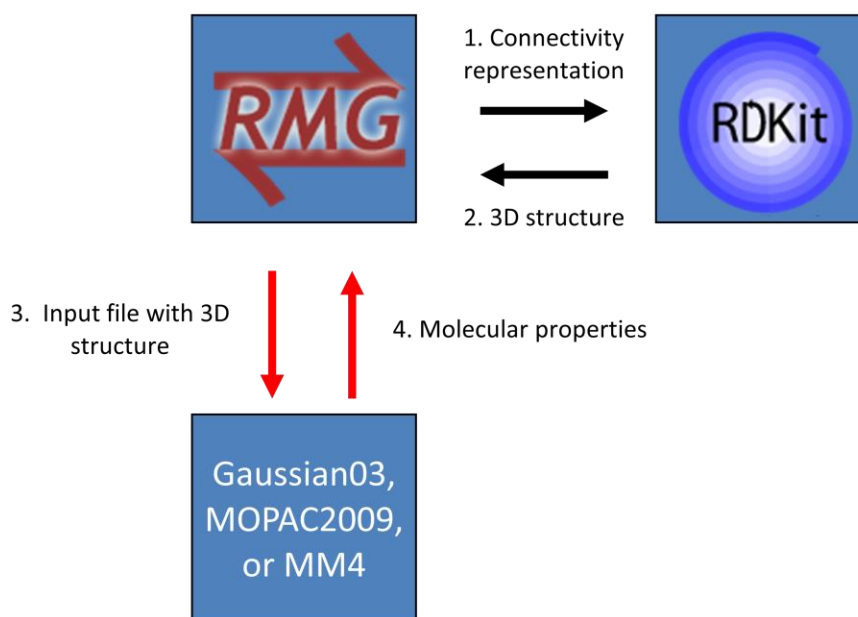


Figure 1. Overview of QMTP interface.

The overview shows that the QMTP interface leverages the large body of existing work in relevant areas of cheminformatics and computational chemistry. In particular, the interface was designed to make extensive use of free and open-source modules, when possible, to mitigate licensing issues and avoid limiting operation to one platform or operating system.

Three-dimensional molecular structure generation

A critical part of the process of performing on-the-fly quantum mechanics or force field calculations is getting some reasonable initial guess for the three-dimensional structure. Several possible approaches were explored. Three-dimensional structure generation procedures tend to fall into two general categories: rule-based approaches and numerical approaches.⁸

Rule-based approaches make use of heuristics and pre-computed ring templates to construct an initial guess for the geometry. In such an approach, chemical structure knowledge of experts is implemented as rules within a computer program; the program may then be used to analyze a compound's connectivity, including factors such as the presence of functional groups and rings, and return a three-dimensional structure. Such methods are relatively fast but are limited in terms of their ability to handle a wide range of species. The program will only handle species for which rules have been programmed.⁹ This can limit, for instance, the types of rings that may be handled by the program. Examples of codes that use these approaches include Frog⁹ and CORINA¹⁰.

An alternative to rule-based approaches is to use distance geometry, a numerical technique. Examples of programs implementing distance geometry methods include Key3D,¹¹ Balloon,¹² RDKit,¹³ and smi23d.¹⁴ Distance geometry refers to the use of specified distance bounds between points to “embed” those points in n -dimensional space.¹⁵ In the context of chemical structure conversion, distance geometry methods apply this technique to atoms in a molecule in order to produce reasonable

atom coordinates in three dimensional space. This embedding is often followed by an energy minimization using a force field to refine the structure. One of the main advantages of these methods is their flexibility; they do not, for instance, need to be preprogrammed with ring structures.

RDKit was chosen for its ability to handle radicals, its open-source nature, and its use of flexible distance geometry methods. Performance tests of *RDKit* gave generally quite satisfactory results. For example, the *RDKit*-based approach produces the desired structure for a challenging helical polycyclic aromatic hydrocarbon, as shown in Figure 2, whereas a typical rule-based approach would produce a nonsensical planar geometry for this case. Recently, Ebejer *et al.* have performed a more systematic study of various free three-dimensional structure generation tools and found the accuracy and speed of *RDKit* compared favorably with alternatives.¹⁶

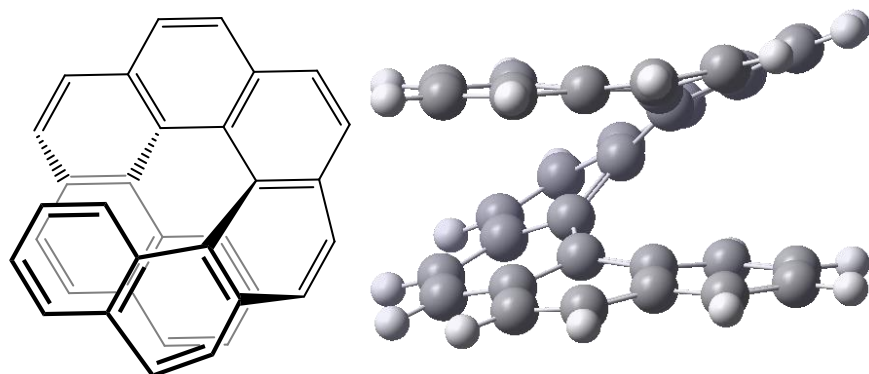


Figure 2. Three dimensional structure for a helical polycyclic aromatic hydrocarbon (octahelicene) obtained with *RDKit*-based approach (including UFF refinement)

Once the tool for three-dimensional structure generation was chosen, several other aspects of the “0D->3D” conversion process needed to be considered. One issue is obtaining an appropriate reference conformation of the molecule for performing the calculations; ideally, we would use the global minimum energy conformation of the molecule; the force field structure refinement will produce a local minimum, which may or may not also be the global minimum. A number of different approaches exist to search for global energy minima, but the distance geometry methods previously described also offer a simple and straightforward means for conformational exploration.¹⁷ With such an approach, different random seeds are used to initialize the distance geometry algorithm; after each of the resulting structures are refined with a force field, the structure with the lowest force field energy following force field refinement represents the most stable conformation identified. As the number of random seeds (iterations) increases, it becomes more likely that the global minimum energy conformation will be identified.¹¹ Unfortunately, with this approach there is no way to tell how many iterations will be required to reach the global minimum, nor will it be obvious that the global minimum has been obtained once it is enumerated (this issue is shared by most conformational search approaches, with the exception of deterministic global optimization using branch-and-bound approaches¹⁸). We have chosen here to use a heuristic for the number of embeddings ($\max[1, 5 \times (N_{\text{atoms}} - 3)]$) that scales linearly with the number of atoms in the molecule; this was chosen to avoid potentially burdensome computational cost associated with exponential scaling, while still accounting for the generally greater conformational

flexibility of larger molecules. It is noted that this will not guarantee that the global minimum has been identified, though we expect that this will produce a reasonable low-lying minimum for most of the cases where this approach would be applied (for purposes of automated reaction mechanism generation for decomposition of small- or medium-sized molecules). Additionally, situations could arise where the use of the lowest lying conformation for a particular molecule is not an appropriate representation of the conformation that arises on the actual reaction path, and in such situations, this approach would tend to produce less reliable results.

Another consideration is the design of the interface between the programs involved in the process. To implement the “0D->3D” structure conversion, a Python script was constructed to provide an interface between RMG and the *RDKit* program. After RMG creates a “two-dimensional” MDL MOL file with the appropriate connectivity, *RDKit* is instructed to read the structure into memory, perform the three-dimensional embedding (a number of times, as discussed above), refine the three-dimensional coordinates using the UFF force field, and write out a MOL file with the three-dimensional coordinates for the UFF-refined structure with the lowest UFF energy, along with a MOL file for the corresponding unrefined structure (the use of this unrefined structure will become apparent in the next section).

Format conversion and output parsing

The OpenBabel command-line utility¹⁹ is used extensively throughout the QMTP interface to create input files and convert between various formats. Python code based on version 1.0 of the cclib libraries²⁰ is used to parse output files.

Calculation method

With the exception of the MM4 method discussed below, the initial incarnation of the interface has been designed around using the semi-empirical PM3 method²¹, as implemented in Gaussian03²² and MOPAC2009²³. The PM3 method was chosen for its relative speed compared to other quantum mechanics-based calculations, for its accuracy relative to similar semi-empirical methods (e.g. AM1), and for its wide availability in popular computer codes. Despite the focus on this method for the initial implementation, and for the discussion in this paper, it is noted that the framework described here can be readily extended to use additional methods and/or quantum mechanics programs. For example, calculations could be performed with the PM6 method or with the newer Gaussian09 with very minor modifications to the source code. The significant number of species encountered during mechanism generation would appear to make calculations on all these species with more advanced methods like density functional theory (DFT) impractical at present; however, with the availability of a large quantity of CPU resources, these more advanced methods could become computationally practical through careful parallelization of the automated reaction mechanism generation and QMTP code. More generally, it is expected that as computer hardware improves and as electronic structure calculation/force field methods are further developed, it will become practical to perform on-the-fly calculations using more accurate and robust methods with the QMTP interface during automated reaction mechanism generation.

In the case of both MM4 and PM3, the implementations directly provide heat of formation at standard temperature (~298 K) and therefore subsequent post-processing to obtain this quantity through, for example, atomization or isodesmic reaction schemes, is not required.

Failure checking and recovery

Although the codes used to perform quantum mechanics or force field calculations are relatively robust, they are not error proof. When calculations of this nature are performed manually, troubleshooting is often required. We have incorporated automated troubleshooting into the design for the QMTP interface, since to automatically generate large kinetic models, QMTP must successfully return a sensible estimate of the thermochemistry for every molecule considered – a failure rate of even 0.01% would be unacceptable.

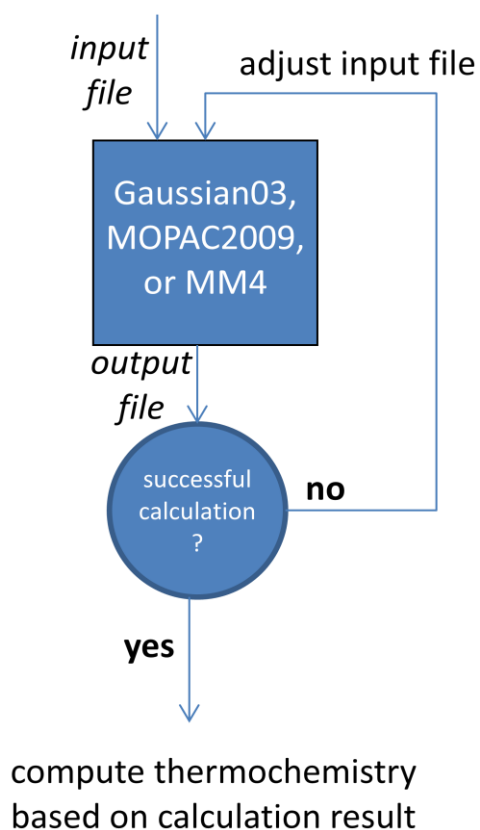


Figure 3. Overview of the automated troubleshooting implemented in the QMTP interface

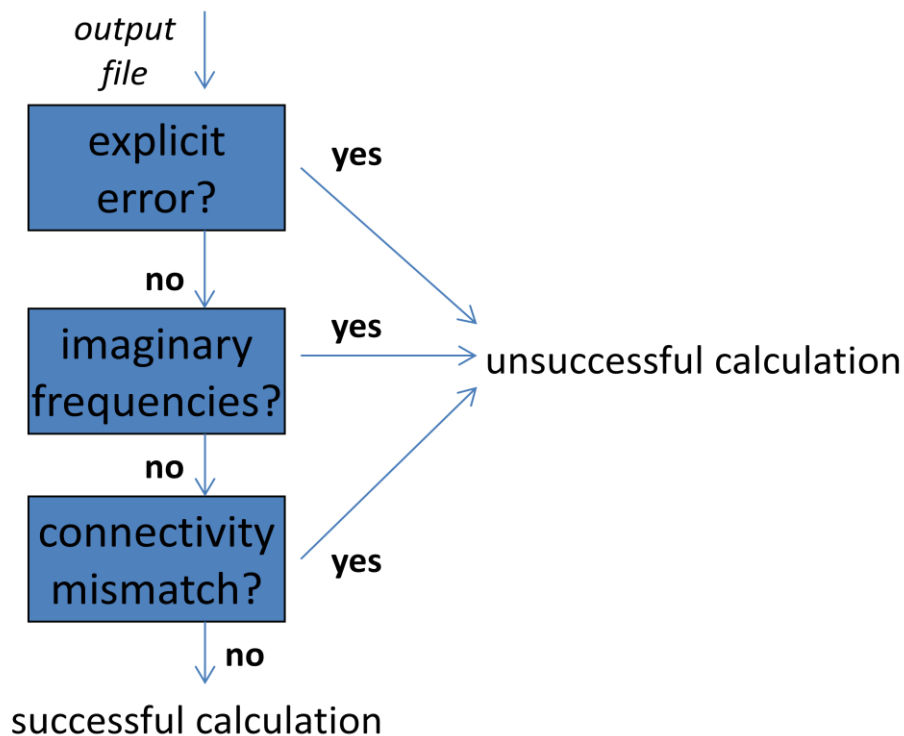


Figure 4. Overview of calculation failure checking in the QMTP interface

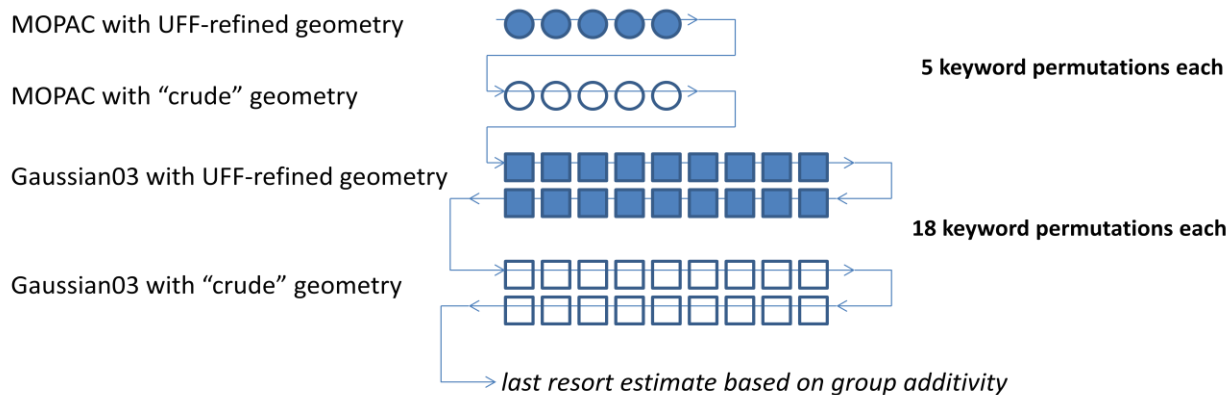


Figure 5. Sequence of input file adjustments currently implemented in QMTP interface for PM3 calculations

The QMTP interface employs several approaches for checking and dealing with calculation failures of various types. The QMTP interface has been programmed with a list of alternative inputs to provide to the quantum mechanics/force field program for cases in which the initial default inputs do not produce a successful result; this is depicted schematically in Figure 3.

For each calculation, the QMTP interface will check whether the calculation has completed without an obvious (explicit) error from the called program. Assuming this passes, the QMTP interface will also confirm that there are no imaginary frequencies, which would indicate that a saddle point (rather than a minimum) had been obtained. If requested by the user, the QMTP interface can also

check that the apparent connectivity of the optimized structure matches the desired connectivity (this is discussed further below). These checks are summarized schematically in Figure 4. If all these checks pass, the key results from the output file are parsed and used to compute the desired thermochemical properties. On the other hand, if any of these checks fail, the next input combination in the list will be attempted. The process is repeated until a successful result is obtained or all of the programmed input combinations have been attempted. (In the latter case, the QMTP interface will print a warning to the user and fall back to the conventional group-additivity based estimates for the molecule.)

The different input combinations include variations on keywords to adjust the geometry optimization algorithm, change the how the Hessian is guessed or updated, or change the self-consistent field (SCF) algorithm; other variations in input include attempts to use initial guess geometry, or even try an alternative program implementing the same method. This is schematically depicted in Figure 5. The diagram shows how the QMTP interface will start with MOPAC, trying five different keyword combinations using the UFF refined geometry. If all these five attempts fail, the QMTP interface will retry the same keyword combinations using the “crude”, unrefined geometry from *RDKit*. If these five new attempts also fail, the QMTP interface will move on to trying 18 different keyword combinations with Gaussian03, first with the UFF-refined geometry, and then with the unrefined geometry, should the initial 18 attempts fail. The result is up to 46 different input combinations that are attempted for each molecule. For additional flexibility, the interface is designed to allow the user to request only MOPAC or only Gaussian be used, allowing the interface to be used when only one of the two software packages is available.

The different input combinations used have been designed with the goal of increasing the robustness of the interface and avoiding the need to fall back to alternative, less reliable estimation procedures. In particular, the default list of keyword combinations that have been programmed into the QMTP interface has been developed by manually investigating cases that fail all earlier attempts and trying to find a combination of keywords that would produce a successful result. In the process of investigating such cases, it was also found that the UFF-refined geometry was sometimes unreasonable in some fashion (e.g. by the introduction of a spurious symmetry to the molecule); however, the corresponding “crude”, unrefined coordinates from *RDKit*'s distance geometry routine were more reasonable and readily processed using the default list of keywords. Thus, in addition to storing the UFF-refined coordinates, the QMTP interface also stores the corresponding crude, unrefined coordinates. These coordinates may then be used to provide the initial guess geometry, should the initial attempts with the UFF-refined geometry prove to be unsuccessful.

One might imagine the possibility for more sophisticated troubleshooting approaches that attempt to diagnose the actual problem with failed jobs and adapt the input keywords accordingly, or do extensive cross-checking between thermochemistry computed using different methods. Although this might reduce the time needed to obtain a successful result in many cases, the extra layers of complication in developing the appropriate checks and logic were outside the scope of the present study.

Storage of results and use of modified InChI and InChIKey

Advances in disk storage capabilities have made it practical to store a library of calculation input and output files for hundreds of thousands of molecular species. As the time to read to the calculation results is much smaller than the time to actually perform the calculation, it is efficient to store the calculation results between runs so that they may be reused without time-consuming recalculation. To do this, a unique and consistent shorthand notation for each molecule is desired so that results for a particular molecule will be recognized from run to run, regardless of atom numbering differences or other representation discrepancies between the runs. Recent developments in chemical structure representation, have produced such a unique chemical identifier, known as the InChI (International Chemical Identifier).²⁴ We have found that the InChI (with some adjustments described below) is well-suited for labeling of stored quantum chemistry results (at least to the extent that atomic connectivity is unambiguous), as it does not encode electron position or bond type, so, for example, different resonance forms are “correctly” represented by the same InChI. However, as the InChI can be quite long and includes characters such as “/”, it is not well-suited for file names for the stored results. Even so, a hash of the InChI, known as the InChIKey, has more desirable properties for filenames, being only 25 characters in length and consisting only of upper-case letters and a single hyphen (“-”).²⁵ Consequently, input and output files for calculations performed by the QMTP interface are named using a version of the InChIKey, so they may be readily retrieved during later runs, obviating the need to re-perform the calculation. The stored results are kept in a folder that may be preserved between runs so that a library of hundreds of thousands of results can be accumulated.

As there is a non-zero (though very small) chance for the InChI strings for two molecules in the library mapping to the same InChIKey (an “InChIKey collision”), the QMTP interface places the InChI as the molecule name in the input file for the calculation so that it can be checked.

As alluded to above, a couple of special considerations are applied to the InChI/InChIKey strings. Firstly, it is noted that the recently-introduced “standard” InChI employs an option that does not localize tautomeric hydrogen atoms in certain species, and is therefore not well suited to representing species in the gas phase. Consequently, we use a “non-standard” InChI with the “FixedH” layer that localizes such tautomeric hydrogen atoms. Additionally, no versions of InChI that we are aware of currently allow for representing the electron spin multiplicity of the species. In order to represent this, we here employ a modified version of the InChI/InChIKey wherein an additional “multn” layer/string is appended to the InChI/InChIKey for species with multiplicity of three or higher. This is best illustrated by example; as shown below, CH₂ singlet and CH₂ triplet would be represented by InChI and InChIKey as shown in Table 1.

Table 1. Illustrative examples of modified InChI/InChIKey system to distinguish molecules with different electron spin multiplicities.

	Modified InChI	Modified InChIKey (v1.02beta)
CH ₂ (singlet)	InChI=1/CH2/h1H2	HZVOZRGWRWCICA-UHFFFAOYAZ
CH ₂ (triplet)	InChI=1/CH2/h1H2/mult3	HZVOZRGWRWCICA-UHFFFAOYAZmult3

Connectivity checking

Ideally, the initial guess geometry and optimized geometry will correspond to the intended target molecule. However, it is recognized that there is no inherent guarantee for this in the process described above. It can be imagined that in certain cases, the optimizer will converge to a potential energy surface minimum corresponding to an entirely different molecule than the intended species. The consequence would be non-representative (and thus unreliable and likely inaccurate) thermochemical properties being computed for the molecule of interest. In the QMTP interface, we have implemented connectivity checking safeguard features to guard against this. The options associated with this feature and process is described in further detail below.

With the connectivity checking feature, the QMTP interface will attempt to perceive connectivity (using methods described in greater detail below) for calculation results that are otherwise successful; when there appears to be a connectivity mismatch, the calculation will be treated as a failure, similar to jobs with imaginary frequencies or other sorts of errors, and input is adjusted until either a successful result, including an apparent connectivity match with the target molecule, is obtained or all of the various attempt options have been exhausted (in which case, the QMTP interface will fall back to the group additivity estimation approach). This is shown schematically at the bottom of Figure 4.

Connectivity checking has been implemented with two alternative approaches to connectivity perception. The primary approach is to use the OpenBabel's¹⁹ built-in connectivity perception algorithm on the final optimized geometry. This produces a .MOL file with connectivity information for the optimized geometry. For added robustness and greater flexibility (as well as for the ability to process MM4 results), a backup connectivity perception approach (which is used in cases where the primary check fails due to OpenBabel crash or due to connectivity mismatch) has been implemented via our MoleCoor utility.²⁶ The MoleCoor utility perceives connectivity using an algorithm described elsewhere.²⁷ The idea of the algorithm is to consider two atoms to be bonded when the distance between them falls below the sum of their covalent radii plus a pre-set tolerance, while also being above a minimum threshold (0.40 Å). (The lower bound prevents atoms that are abnormally close to each other from being considered to be bonded.) In our approach, the tolerance has been set somewhat loosely at 0.50 Å to accommodate apparent O-O bonding in a strained peroxide test case (apparently not recognized by OpenBabel's algorithm). This backup connectivity perception process also produces a .MOL file with connectivity information.

In either case, processing this .MOL file through the InChI utility produces an InChI string, which, following removal of stereochemical layers, may be directly compared to the InChI of the target molecule based on RMG's internal connectivity-based representation of the species. An InChI mismatch corresponds to an apparent connectivity mismatch, while an InChI match corresponds to an apparent connectivity match. In cases where both the primary and backup checks fail, the calculation result is assessed to be a failure, and fallback procedures are followed as described previously.

An example illustrating how this connectivity checking process can help to ensure reliable thermochemical parameter estimation, while maintaining robustness, is shown in Figure 6. The figure shows how the initial attempt using the UFF-refined initial guess geometry can converge to a structure that does not correspond to the desired cyclopentyne target molecule. However, in this case, the “crude” initial guess geometry is much more reasonable and will readily converge to a structure corresponding to cyclopentyne. Without connectivity checking, the results from the incorrect structure would be used, while with connectivity checking, the results that are used will be based on the structure with the desired connectivity.

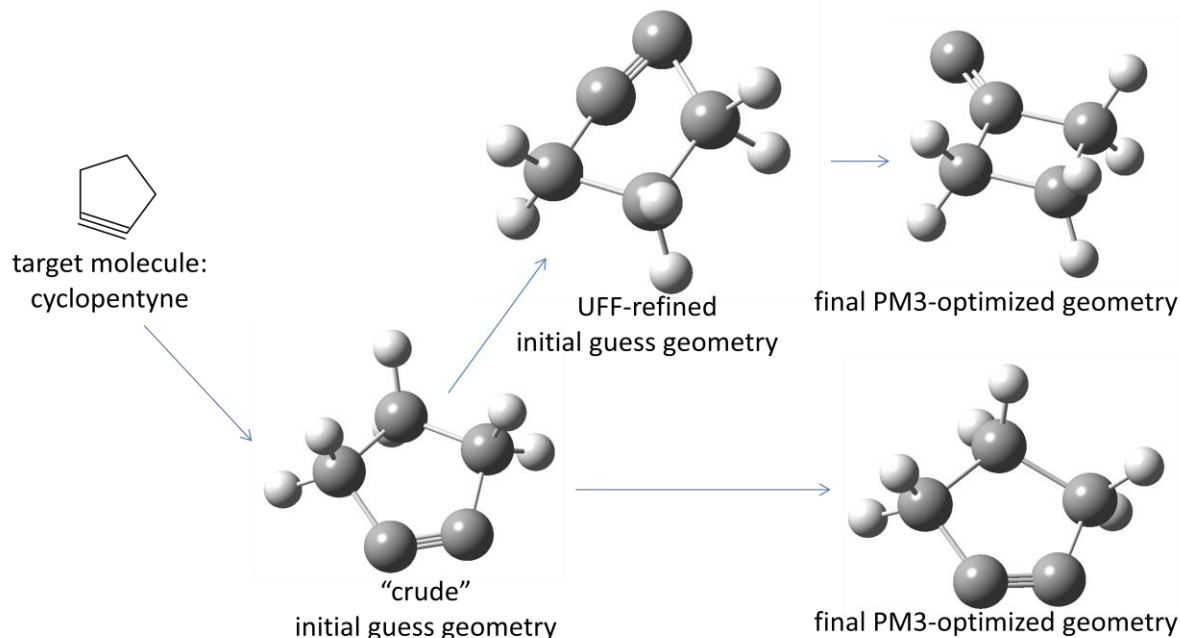


Figure 6. PM3 optimization results for different initial guess geometries for cyclopentyne.

It should be noted, however, that connectivity is not always clear-cut, and ambiguous cases can arise. It is quite possible that connectivity checking approach used here could indicate false positive or false negative matches. Even so, we have made a reasonable attempt to make the connectivity checking process reliable and robust, and our experience thus far suggests that the use of this feature is superior to alternatives.

External symmetry number and chirality calculation

The symmetry number for molecular rotation (also known as the external symmetry number) and chirality are key factors in the entropy of a molecular species, and these contributions can affect the equilibrium constants for reactions by factors often two or greater. Information about the point group of the molecular structure is sufficient to determine both the symmetry number and chirality corrections. Consequently, we make use of the open-source SYMMETRY program²⁸ by passing it the optimized three-dimensional geometry as well as a tolerance to allow small deviations from exact symmetry; SYMMETRY then calculates the point group and returns the result to RMG. The symmetry

number is determined from the point group based on published relationships.⁷ A chirality contribution of $+R \ln 2$ is included for point groups that lack a superposable mirror image (i.e. point groups lacking σ_h , σ_d , σ_v , and S_n symmetry elements), effectively assuming a racemic mixture of mirror image enantiomers.

Force field and rotor scan capabilities

An interface with the MM4 force field software²⁹ has also been implemented in RMG's QMTP interface. MM4 force field calculations are much less computationally demanding than semi-empirical electronic structure calculations, and these methods have been found to produce highly reliable thermodynamic property estimates for classes of molecules for which they have been parametrized (e.g. alkanes³⁰, alkenes³¹, and conjugated hydrocarbons³²).

As alluded to above, the default behavior is to compute thermochemical properties within the framework of the rigid rotor/harmonic oscillator approximation. However, the speed of the MM4 calculations allows us to gather additional information about the potential energy surfaces of molecules with rotors to better account for conformational flexibility without an inordinate amount of added computational cost. In particular, we have implemented rotor scan capabilities when using the MM4 force field. In this approach, a relaxed scan of each rotor is performed in 5° increments; each rotor is considered independently, so that each scan gives a one-dimensional profile of energy and moment of inertia for the mode of interest. Rotor symmetry number is estimated based on connectivity. This information is provided to a Python-language code based on CanTherm v1.0³³ to compute the desired thermochemical quantities using a separable hindered rotor treatment (and also accounting for variation in the moment of inertia with the rotor angle).

Scope of QMTP calculations

Although on-the-fly PM3 calculations based on explicit three-dimensional geometries are expected to be useful in many cases, it is recognized that there are many cases where alternative approaches, such as the original method based on group additivity, are still desirable. For example, we expect that the traditional group additivity approach to be more reliable (and much faster) than the QMTP approach for acyclic hydrocarbons. Therefore, we have implemented a switch that will further restrict calls to the QMTP estimation routines to cyclic species, for which the traditional group additivity based estimates are expected to be inaccurate. Also, the user may wish to specify "known" thermochemistry parameters for a set of molecular species. Therefore, the QMTP methods will only be called for molecules without user-specified thermodynamic parameters.

Also, it is recognized that many methods, including the previously-discussed PM3 and MM4, may be less accurate and/or less robust for treating radical systems with unpaired electrons. Consequently, we have implemented an option where the user can specify the maximum number of radical sites that a molecule can have for it to be directly processed by the QMTP interface. So, if a user sets this option to zero, only closed shell molecules are directly processed by the QMTP interface. In other cases (monoradicals, biradicals, etc.) hydrogen bond increments (HBI)³⁴ are applied to QMTP results for the saturated, closed-shell parent molecule. (It is noted that the HBI adjustments applied are the same that are used for conventional group additivity-based estimates for radicals.)

For routine use for systems with only carbon, hydrogen, and oxygen, we currently suggest settings that use QMTP only for cyclic species and closed-shell species (applying HBI corrections for radicals).

It should be noted however, that the approach of using generic HBI corrections does not always achieve a desirable level of accuracy. In particular, it has been found that the use of the generic “C=CJC=C” radical correction in 1,3-cyclopentadien-2-yl underestimates the HBI by a significant amount, resulting in a heat of formation value (79.4 kcal/mol) that differs from a CBS-QB3-based result (95.9 kcal/mol) by over 16 kcal/mol. In this case, the error most likely due to the fact that the allenic resonance form of this radical is severely strained by the presence of a ring, in contrast to unstrained acyclic molecules such as 1,3-butadien-2-yl, which would have been used to determine this particular radical correction. The alternative for these cases would be either to apply more specific ring HBI corrections, such as those published in the literature for particular cyclic radicals (e.g. Ref. ³⁵), or use the QMTP interface to directly perform calculations on monoradicals. In the case of 1,3-cyclopentadien-2-yl, the latter approach using PM3 produces an enthalpy of formation at 298.15 K of 86.4 kcal/mol, which is in closer agreement with the CBS-QB3 value, though still deviates by about 9.5 kcal/mol.

Testing of the QMTP interface

In addition to extensive testing during the development of the QMTP interface (including development of a library of keywords for failure recovery) several additional tests were performed on the current version of the interface, to more formally characterize its accuracy, robustness, and speed, as well as its influence on kinetic models.

Accuracy of estimates

In addition to informal comparisons to ensure accuracy, the QMTP interface was tested on a challenging test set of polycyclic hydrocarbons from Osmont et al., who had compiled experimental values for standard enthalpy of formation at 298.15 K for 47 molecules.³⁶ The QMTP interface (with PM3 and with the MM4 approach with separable hindered rotor treatment) was applied to 43 of these molecules (four of the molecules had two geometric isomers; in these cases, the QMTP result was applied to the isomer with the lower enthalpy of formation.)

Table 2. Accuracy of RMG estimation approaches, using experimental enthalpy of formation values compiled by Osmont et al. as reference. Two molecules with three-membered rings failed the attempts with the MM4 approach and were excluded from averaging. N=43, with the exception of the MM4 case.

	Mean abs. error (kcal/mol)	Root-mean-squared error (kcal/mol)
Original RMG group-additivity approach	40.0	55.8
QMTP interface with PM3 approach	7.0	10.6
QMTP interface with MM4 approach	29.6	53.1

Results of the comparison are shown in Table 2. The results show that the error in the PM3 approach is, on average, more than a factor of five lower than the original RMG approach based on

group additivity. However, the MM4 approach, on average, does noticeably worse than the PM3 approach. A closer analysis of the results from the MM4 approach shows that the largest errors are concentrated in species with three-membered rings, for which the MM4 force field currently lacks parameters.³⁷

Table 3. Accuracy of RMG estimation approaches, using experimental enthalpy of formation values compiled by Osmont et al. as reference for a subset of molecules without three-membered rings. N=23

	Mean abs. error (kcal/mol)	Root-mean-squared error (kcal/mol)
Original RMG group-additivity approach	22.8	40.9
QMTP interface with PM3 approach	8.4	12.9
QMTP interface with MM4 approach	1.3	3.0

Comparison on the subset of molecules without three-membered rings produces results that are much more favorable with the MM4 approach, as shown in Table 3. In fact, the MM4 approach significantly outperforms the PM3 approach on this subset.

It is noted that although this test set encompasses a wide variety of ring structures, some important classes of rings, including those with double bonds and oxygenates, are not assessed here.

Effect on kinetic models

To assess the influence of the use of QMTP estimates on an actual detailed chemical kinetic model, we took an existing combustion model, JP-10 combustion mechanism version 0.19³⁸ and re-generated thermodynamic properties for the species using two approaches. The first approach (“no QMTP”) is the traditional approach used by RMG with use of group additivity in most cases and with results taken from built-in default and GRI-Mech 3.0 thermodynamic libraries for selected compounds where estimation is not required. The second approach (“QMTP”) relies on the QMTP interface described here for cyclic and polycyclic compounds; cyclic radical thermochemistry is computed using hydrogen bond increment corrections to the saturated molecule as computed using the QMTP interface; as with the first approach, results for selected molecules can be taken from the built-in default and GRI-Mech 3.0 thermodynamic libraries. The mechanism includes a number of polycyclic compounds, including JP-10 itself, for which the estimates obtained by the two approaches are likely to disagree significantly. In total, thermodynamic parameters for 256 of the 320 species in this model (80%) are based on PM3. In both cases, the model enforces thermodynamic consistency; kinetic parameters are specified for only one reaction direction and are the same in the two approaches.

Error! Reference source not found., below, shows the time to maximum modeled CH concentration (used here as a proxy for ignition delay) versus shock tube results from Davidson et al.³⁹ at 45 different conditions.

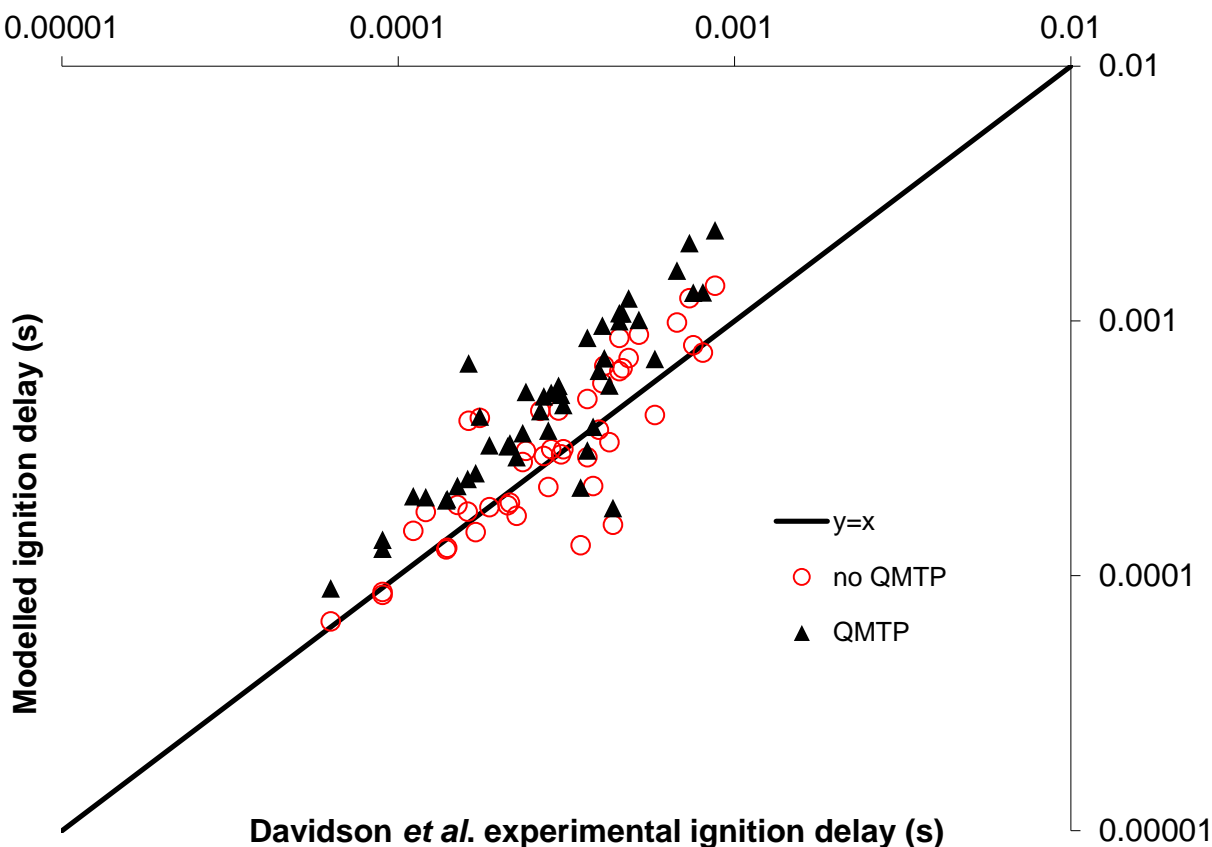


Figure 7. Parity plot showing the time to maximum modeled CH mole fraction (used here as a proxy for the modeled ignition delay) versus experimental ignition delay from shock tube data from Davidson *et al.*³⁹

The plot shows that the use of the QMTP thermochemistry (which should, in general, be more reliable) has a significant effect, sometimes changing the time to maximum CH concentration by more than 70%. More specifically, the use of QMTP tends to shift to longer modeled ignition delays. Although the points with QMTP tend to overestimate the ignition delay, the agreement with experiment is still acceptable considering other uncertainties in the model (e.g. kinetic parameters) and the uncertainty in experimental results which we estimate at roughly a factor of two; more importantly, the model is now more justifiable as its thermodynamic parameters should be more reliable, as noted above.

It should be emphasized that the underlying mechanism and kinetic parameters are the same in both approaches; the effect of using the QMTP approach should be even more pronounced when considering its use in the context of the entire RMG mechanism generation approach. Because detailed chemical kinetic models generated by RMG using the QMTP approach will involve different thermochemistry estimates for core species and potential intermediates, the actual species and reactions that are important enough to be included in the mechanism will differ. Also, even if the same reaction appears in both models, the kinetic parameter estimates themselves can be affected through equilibrium and Evans-Polanyi-type considerations (and also potential energy surface effects when obtaining pressure-dependent kinetic estimates via Master Equation calculations).

Robustness and speed

To test the performance of the QMTP interface in a mechanism generation setting, RMG was run for 120 hours (5 days) with the QMTP interface turned on to use PM3 results for cyclic molecules¹. The interface that was considered involved high-temperature oxidation of the polycyclic molecule, JP-10.

Over this 120 hour period, RMG used the QMTP interface to successfully obtain PM3 results for 22,277 cyclic molecules (22,244 with the primary MOPAC approach and 53 with Gaussian03), in addition to performing traditional mechanism generation functions (e.g. simulating the time evolution of the model, generating reactions, and estimating kinetic parameters). An additional 20 cyclic species were attempted using QMTP but failed all attempts; for all 20 of these cases, the failure of QMTP was due to an apparent connectivity mismatch from the intended species, and without the connectivity checking option turned on, these would have been considered successes. This corresponds to a success rate of greater than 99.9% and demonstrates the robustness of this approach. Graphs for the 20 failures are shown in Figure 8; many of these structures appear to highly strained and there may not exist minima on the PM3 potential energy surface that have the desired connectivity. Among the successful results were 7 molecular species for which results from initial attempts failed the connectivity check but later attempts were able to produce a result with the desired connectivity. An additional 6 cases failed the primary connectivity check, but the backup connectivity check suggested that the molecule had the desired connectivity, and the attempt was considered a success.

The results show that the QMTP approach calculates the desired thermodynamic quantities for each molecule, on average, in a matter of seconds, which, while significantly longer than the millisecond time scale for group-additivity based calculations, is sufficiently fast for routine use in mechanism generation.

¹ In order to avoid an apparent bug specific to the Linux version of Gaussian03 used here, four of the Gaussian03 keyword alternatives involving the use of "opt=calcall" (corresponding to eight of the 36 Gaussian03 attempt options) were removed from the programmed keyword list. It is not expected that the removal of 8 attempts has a significant effect on the results.

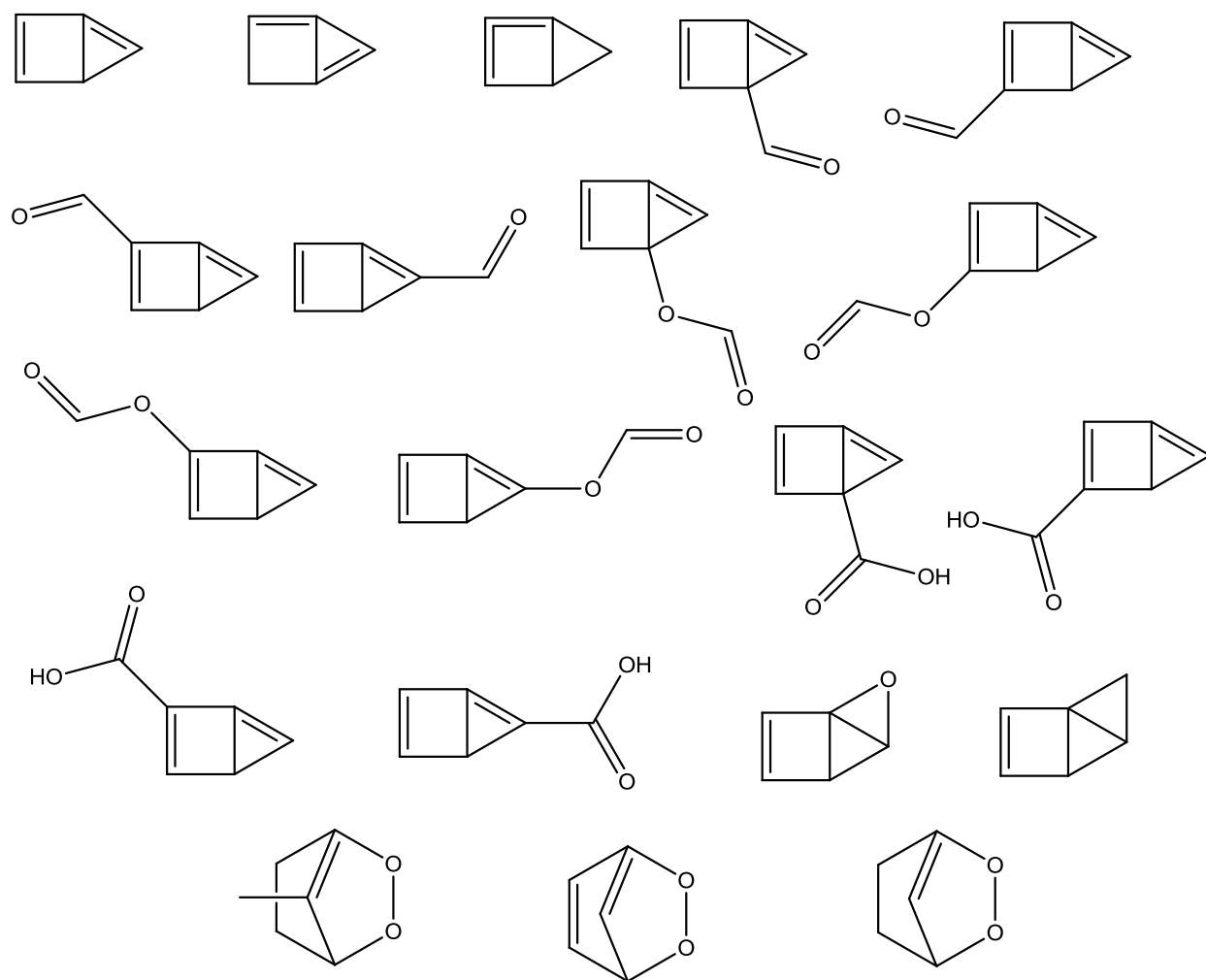


Figure 8. Twenty species that failed all QMTP attempts in JP-10 high-temperature oxidation test run. All failed due to apparent connectivity mismatch between the desired molecule and the PM3-optimized geometry.

Figure 9 shows the number of species (on a logarithmic scale) that first succeeded at each of the attempts. The plot shows that in the vast majority of cases, the first MOPAC attempt was able to produce a successful result. However, the number of cases where the backup attempts proved to be useful still number in the hundreds; the results demonstrate the utility of using the alternative “crude” geometry, as well as the alternative PM3 implementation offered by Gaussian03. Even so, these backup attempts reach a point of diminishing returns, and the graph shows that many of the later attempts did not prove to be useful in this particular test run.

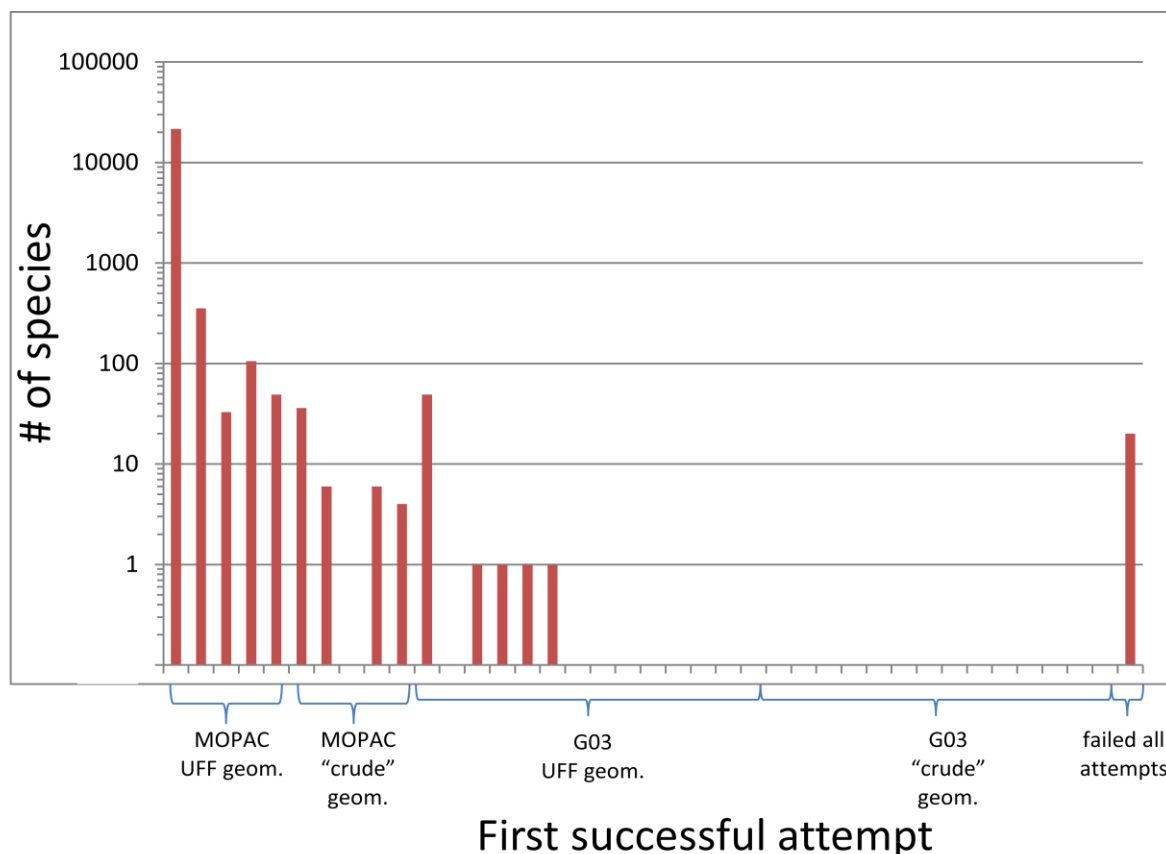


Figure 9. Distribution of outcomes for QMTP interface in JP-10 high-temperature oxidation test run. Each bin on the abscissa corresponds to a particular “attempt” by the QMTP interface to perform PM3 calculations with various choices of input keywords, initial guess geometries, and PM3 implementations. The first five attempts correspond to the use of the MOPAC implementation of PM3 with the UFF-refined initial guess geometry; the next five use the MOPAC implementation with the “crude” (non-UFF-refined) initial guess geometry; the next 14 correspond to the use of the Gaussian03 implementation of PM3 with UFF-refined geometry; the final 14 attempts correspond to the use of the Gaussian03 implementation with the “crude” geometry.

Opportunities for improvement and applying the approach to other areas

Selective use of more accurate methods

As shown in Table 3, MM4 is significantly more accurate than PM3 for certain types of molecules. Similarly, in some situations it would be better to run a DFT or high-level quantum chemistry calculation rather than relying on PM3 or MM4. It would be easy to modify QMTP to use these other types of calculations – the challenge is writing the logic which would decide which type of calculation to perform for each molecule.

Improvements to treating conformational flexibility

As discussed previously, the typical treatment is to assume harmonic behavior for all internal modes; we have implemented rotor scan methods with the MM4 force field that account for conformational flexibility at a higher level (separable hindered rotor treatment), but even this approach

may not achieve a sufficient level of accuracy for certain molecules/applications, and some examples of this appear in the literature.⁴⁰ There is opportunity for improvement of this treatment, particularly when using the relatively computationally inexpensive MM4 approach. One of the more promising approaches is the recently introduced multi-structural method of Truhlar and coworkers⁴¹; this method is reportedly able to account for conformational flexibility via both rotation about bonds as well as other modes such as ring inversion, and obeys desired statistical mechanical limits (e.g. high temperature heat capacity). Implementation of this approach within the QMTP interface for use with MM4 calculations would seem to be the next logical step for improving treatment of conformational flexibility. A key aspect of such an implementation would be accurate and robust enumeration of all the minima on the molecule's potential energy surface, without double counting equivalent conformations. Toward this end, we have developed and implemented a novel conformational equivalence algorithm, designed for this particular application, in the MoleCoor utility.²⁶

Polarizability estimates

In addition to being used in estimation of thermodynamic parameters, on-the-fly quantum mechanics calculations can produce estimates of polarizability, with the appropriate keyword choices. Such polarizability estimates can, for example, be incorporated into approaches to estimate transport properties estimation, and, in fact, an RMG script using the QMTP framework has been created and applied to produce polarizability estimates as part of a post-processing transport property estimation step. With additional software development, it should be possible to integrate on-the-fly polarizability estimates directly into RMG's transport property estimation routines.

Standalone thermodynamic property estimation

Although "black box" *ab initio* methods have enabled reasonably straightforward estimation of thermodynamic properties for a wide range of compounds, the application of these methods requires some initial training and mistakes can easily creep into calculation results. Also, the amount of "human time" associated with input file construction, job monitoring, result analysis, and troubleshooting is non-trivial. As such, a more automated approach for setting up and performing calculations and analyzing the results is highly desirable. The QMTP interface described previously is well-suited to this level of automation, as it performs all necessary steps, from generation of an initial guess three-dimensional structure, to construction of calculation input file, to troubleshooting, to processing the calculation results into the desired thermodynamic property estimates, without human intervention. In fact, the QMTP features have recently been incorporated into a standalone thermodynamic property estimation tool distributed with RMG. Though this tool currently uses the same PM3 and MM4 methods as used by RMG, it should be possible to extend this to use other, more accurate and time-consuming *ab initio* methods, such as the popular CBS-QB3⁴² or Gaussian-*n*⁴³ composite methods.

Kinetic parameter estimation

Currently, the results from the QMTP interface affect kinetic parameters somewhat indirectly through the properties of reactants and products. It could be possible to further improve certain kinetic parameter estimates using information that may be obtained from on-the-fly calculations using explicit

three-dimensional geometries for species minima. For example, a scheme for kinetic parameter estimation for intramolecular disproportionation reactions based on ring strain of stable molecules has been proposed by Herbinet et al.⁴⁴

One can imagine even more direct means of using on-the-fly calculations with explicit three-dimensional geometries to obtain kinetic parameters. In particular, there are opportunities to adapt the approach to locate first-order saddle points; calculations using saddle-points could be used to produce kinetic parameter estimates within the framework of transition-state theory (TST). If realized, these types of capabilities could be particularly useful for reactions with cyclic transition states (e.g. intramolecular hydrogen abstraction) for which the existing connectivity-based estimates can be unreliable.

Summary

An interface for performing on-the-fly quantum mechanics or force field calculations in the context of automated reaction mechanism generation has been described. Testing has demonstrated the impact, accuracy, and robustness of the interface, which make it suitable for routine use during mechanism generation. The interface is particularly useful for obtaining more-reliable thermochemical parameters for cyclic species (for which alternative automated group additivity-based approaches are prone to significant error). Possible avenues for further improvements and alternative applications have also been described.

Acknowledgements

The authors would like to acknowledge helpful discussions with Professor Linda Broadbelt of Northwestern University, Laurent Catoire of CNRS, and Professor Richard H. West, now of Northeastern University. The authors also acknowledge work by Nick Vandewiele of Ghent University for development of the standalone thermodynamic property estimation tool mentioned here. Simon E. Albo, Oluwayemisi O. Oluwole, and Hsi-Wu Wong of Aerodyne Research, Inc. are acknowledged for work with the authors on transport property estimation incorporating polarizability estimates from quantum calculations, mentioned here, as well as JP-10 combustion model generation. Helpful discussions with Dr. Markus Kraft and the Computational Modelling Group at the University of Cambridge are also acknowledged. We would like to acknowledge funding for the development of RMG by the Department of Energy, Office of Basic Energy Sciences through grant DE-FG02-98ER14914. Work on the JP-10 system was funded by the Naval Air Warfare Center under contract number N68335-10-C-0534.

References

1. Allen, J. W.; Ashcraft, R. W.; Beran, G. J.; Goldsmith, C. F.; Harper, M. R.; Jalan, A.; Magoon, G. R.; Matheu, D. M.; Petway, S.; Sumathy, R.; Sharma, S.; Van Geem, K. M.; Song, J.; Wen, J.; West, R. H.; Wong, A.; Wong, H.-W.; Yelvington, P. E.; Yu, J.; Green, W. H. *RMG (Reaction Mechanism Generator)*, 2009.

2. Susnow, R. G.; Dean, A. M.; Green, W. H.; Peczak, P.; Broadbelt, L. J., Rate-Based Construction of Kinetic Models for Complex Systems. *The Journal of Physical Chemistry A* **1997**, *101* (20), 3731-3740.
3. Harper, M. R.; Van Geem, K. M.; Pyl, S. P.; Marin, G. B.; Green, W. H., Comprehensive reaction mechanism for n-butanol pyrolysis and combustion. *Combust Flame* **2011**, *158* (1), 16-41.
4. Magoon, G. R.; Green, W. H.; Oluwole, O. O.; Wong, H.-W.; Albo, S. E.; Lewis, D. K., Updating Our Understanding of JP-10 Decomposition Chemistry: A Detailed JP-10 Combustion Mechanism Constructed Using RMG – an Automatic Reaction Mechanism Generator. In *46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, Nashville, TN, 2010.
5. (a) Benson, S. W.; Buss, J. H., Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties. *J Chem Phys* **1958**, *29*; (b) Benson, S. W., In *Thermochemical Kinetics*, 2nd ed., Wiley: New York, NY, 1976; (c) Cohen, N.; Benson, S. W., The thermochemistry of alkanes and cycloalkanes. In *The Chemistry of Alkanes and Cycloalkanes*, Patai, S.; Rappoport, Z., Eds. John Wiley & Sons Ltd: 1992; (d) Cohen, N.; Benson, S. W., Estimation of Heats of Formation of Organic Compounds by Additivity Methods. *Chemical Reviews* **1993**, *93*, 2419-2438.
6. Broadbelt, L. J.; Stark, S. M.; Klein, M. T., Computer generated reaction networks: on-the-fly calculation of species properties using computational quantum chemistry. *Chemical Engineering Science* **1994**, *49* (24, Part 2), 4991-5010.
7. Irikura, K. K.; Frurip, D. J., Computational thermochemistry: prediction and estimation of molecular thermodynamics. American Chemical Society (via <http://cccbdb.nist.gov/thermo.asp>): 1998.
8. Sadowski, J.; Gasteiger, J., From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. *Chemical Reviews* **1993**, *93* (7), 2567-2581.
9. Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tufféry, P., Frog: a Free Online druG 3D conformation generator. *Nucleic Acids Res* **2007**, *35* (suppl 2), W568-W572.
10. Molecular Networks, G. CORINA, <http://www.molecular-networks.com/software/corina/index.html> Erlangen, Germany.
11. Mizutani, M. Y.; Nakamura, K.; Ichinose, T.; Itai, A., Starting Point to Molecular Design: Efficient Automated 3D Model Builder Key3D. *Chemical & Pharmaceutical Bulletin* **2006**, *54* (12), 1680-1685.
12. Vainio, M. J.; Johnson, M. S., Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *Journal of Chemical Information and Modeling* **2007**, *47* (6), 2462-2474.
13. Landrum, G. RDKit, <http://rdkit.org>.
14. Gilbert, K.; Guha, R. smi23d, <http://www.chembioGRID.org/cheminfo/smi23d/>.
15. (a) Crippen, G. M., Chemical distance geometry: Current realization and future projection. *Journal of Mathematical Chemistry* **1991**, *6* (1), 307-324; (b) Svrcek-Seiler, W. A. 0.02 € on Embedding 2004. <http://www.tbi.univie.ac.at/Bled/Slides04/svrci.pdf>.
16. Ebejer, J.-P.; Morris, G. M.; Deane, C. M., Freely Available Conformer Generation Methods: How Good Are They? *Journal of Chemical Information and Modeling* **2012**, *52* (5), 1146-1158.
17. Spellmeyer, D. C.; Wong, A. K.; Bower, M. J.; Blaney, J. M., Conformational analysis using distance geometry methods. *Journal of Molecular Graphics and Modelling* **1997**, *15* (1), 18-36.
18. (a) Maranas, C. D.; Floudas, C. A., A deterministic global optimization approach for molecular structure determination. *The Journal of Chemical Physics* **1994**, *100* (2), 1247-1261; (b) Lin, Y.; Stadtherr, M. A., Deterministic global optimization of molecular structures using interval analysis. *J Comput Chem* **2005**, *26* (13), 1413-1420.
19. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G., Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011**, *3*, 1-14.
20. O'Boyle, N. M.; Tenderholt, A. L.; Langner, K. M., cclib: A library for package-independent computational chemistry algorithms. *J Comput Chem* **2008**, *29* (5), 839-845.

21. (a) Stewart, J. J. P., Optimization of parameters for semiempirical methods I. Method. *J Comput Chem* **1989**, *10* (2), 209-220; (b) Stewart, J. J. P., Optimization of parameters for semiempirical methods II. Applications. *J Comput Chem* **1989**, *10* (2), 221-264.
22. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 03, Revision D.01*, Gaussian, Inc.: Wallingford CT, 2003.
23. Stewart, J. J. P. *MOPAC2009*, <http://OpenMOPAC.net>, Stewart Computational Chemistry: Colorado Springs, CO, USA, 2008.
24. *InChI*, <http://www.iupac.org/inchi/>.
25. Release of InChI Version 1.02 beta; Introducing InChIKey. *Chemistry International* 2007, pp 18-19.
26. Magoon, G. R.; Green, W. H. *MoleCoor*, <https://github.com/gmagoon/MoleCoor>, 2011.
27. (a) Baber, J. C.; Hodgkin, E. E., Automatic assignment of chemical connectivity to organic molecules in the Cambridge Structural Database. *Journal of Chemical Information and Computer Sciences* **1992**, *32* (5), 401-406; (b) Sayle, R., PDB: Cruft to Content. In *MUG 2001*, Santa Fe, NM, 2001.
28. Patchkovskii, S. *SYMMETRY*, <http://www.cobalt.chem.ucalgary.ca/ps/symmetry/>, 2003.
29. Allinger, N. L.; Lii, J.-H. *MM4(2008) and MM4(2003)*.
30. Allinger, N. L.; Chen, K.; Lii, J.-H., An Improved Force Field (MM4) for Saturated Hydrocarbons. *J Comput Chem* **1996**, *17* (5&6), 642-668.
31. Nevins, N.; Chen, K.; Allinger, N. L., Molecular mechanics (MM4) calculations on alkenes. *J Comput Chem* **1996**, *17* (5-6), 669-694.
32. Nevins, N.; Lii, J.-H.; Allinger, N. L., Molecular mechanics (MM4) calculations on conjugated hydrocarbons. *J Comput Chem* **1996**, *17* (5-6), 695-729.
33. Sharma, S.; Harper, M. R.; Green, W. H. *CANTHERM v1.0*, <https://github.com/GreenGroup/CanTherm>, 2010.
34. Lay, T. H.; Bozzelli, J. W.; Dean, A. M.; Ritter, E. R., Hydrogen Atom Bond Increments for Calculation of Thermodynamic Properties of Hydrocarbon Radical Species. *The Journal of Physical Chemistry* **1995**, *99* (39), 14514-14527.
35. (a) Feng, Y.; Liu, L.; Wang, J.-T.; Zhao, S.-W.; Guo, Q.-X., Homolytic C-H and N-H Bond Dissociation Energies of Strained Organic Compounds. *The Journal of Organic Chemistry* **2004**, *69* (9), 3129-3138; (b) Tian, Z.; Fattahi, A.; Lis, L.; Kass, S. R., Cycloalkane and Cycloalkene C-H Bond Dissociation Energies. *J Am Chem Soc* **2006**, *128* (51), 17087-17092; (c) Agapito, F.; Nunes, P. M.; Costa Cabral, B. J.; Borges dos Santos, R. M.; Martinho Simões, J. A., Energetic Differences between the Five- and Six-Membered Ring Hydrocarbons: Strain Energies in the Parent and Radical Molecules. *The Journal of Organic Chemistry* **2008**, *73* (16), 6213-6223.
36. Osmont, A.; Catoire, L.; Gökalp, I., Physicochemical Properties and Thermochemistry of Propellanes. *Energ Fuel* **2008**, *22* (4), 2241-2257.
37. Lii, J.-H., personal communication. 2010.
38. Magoon, G. R.; Aguilera-Iparraguirre, J.; Green, W. H.; Lutz, J. J.; Piecuch, P.; Wong, H.-W.; Oluwole, O. O., Detailed chemical kinetic modeling of JP-10 (exo-tetrahydrodicyclopentadiene) high-

- temperature oxidation: Exploring the role of biradical species in initial decomposition steps. *Int J Chem Kinet* **2012**, *44* (3), 179-193.
39. Davidson, D. F.; Horning, D. C.; Herbon, J. T.; Hanson, R. K., Shock tube measurements of JP-10 ignition. *P Combust Inst* **2000**, *28*, 1687-1692.
40. (a) Sharma, S.; Raman, S.; Green, W. H., Intramolecular Hydrogen Migration in Alkylperoxy and Hydroperoxyalkylperoxy Radicals: Accurate Treatment of Hindered Rotors. *The Journal of Physical Chemistry A* **2010**, *114* (18), 5689-5701; (b) Speybroeck, V. V.; Vansteenkiste, P.; Neck, D. V.; Waroquier, M., Why does the uncoupled hindered rotor model work well for the thermodynamics of n-alkanes? *Chem Phys Lett* **2005**, *402* (4-6), 479-484.
41. Zheng, J.; Yu, T.; Papajak, E.; Alecu, I. M.; Mielke, S. L.; Truhlar, D. G., Practical methods for including torsional anharmonicity in thermochemical calculations on complex molecules: The internal-coordinate multi-structural approximation. *Phys Chem Chem Phys* **2011**, *13* (23), 10885-10907.
42. Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A., A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys* **1999**, *110* (6), 2822-2827.
43. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K., Gaussian-4 theory. *The Journal of Chemical Physics* **2007**, *126* (8), 084108-12.
44. Herbinet, O.; Sirjean, B.; Bounaceur, R.; Fournet, R.; Battin-Leclerc, F.; Scacchi, G.; Marquaire, P. M., Primary mechanism of the thermal decomposition of tricyclodecane. *J Phys Chem A* **2006**, *110* (39), 11298-11314.

Supplementary Information

PM3 keyword sets:

Gaussian03:

1. # pm3 opt=(verytight,gdiis) freq IOP(2/16=3)
2. # pm3 opt=(verytight,gdiis) freq IOP(2/16=3) IOP(4/21=2)
3. # pm3 opt=(verytight,calcfc,maxcyc=200) freq IOP(2/16=3) nosymm
4. # pm3 opt=(verytight,calcfc,maxcyc=200) freq=numerical IOP(2/16=3) nosymm
5. # pm3 opt=(verytight,gdiis,small) freq IOP(2/16=3)
6. # pm3 opt=(verytight,nolinear,calcfc,small) freq IOP(2/16=3)
7. # pm3 opt=(verytight,gdiis,maxcyc=200) freq=numerical IOP(2/16=3)
8. # pm3 opt=tight freq IOP(2/16=3)
9. # pm3 opt=tight freq=numerical IOP(2/16=3)
10. # pm3 opt=(tight,nolinear,calcfc,small,maxcyc=200) freq IOP(2/16=3)
11. # pm3 opt freq IOP(2/16=3)
12. # pm3 opt=(verytight,gdiis) freq=numerical IOP(2/16=3) IOP(4/21=200)
13. # pm3 opt=(calcfc,verytight,newton,notrustupdate,small,maxcyc=100,maxstep=100) freq=(numerical,step=10) IOP(2/16=3) nosymm
14. # pm3 opt=(tight,gdiis,small,maxcyc=200,maxstep=100) freq=numerical IOP(2/16=3) nosymm
15. # pm3 opt=(verytight,gdiis,calcall) IOP(2/16=3)
16. # pm3 opt=(verytight,gdiis,calcall,small,maxcyc=200) IOP(2/16=3) IOP(4/21=2) nosymm
17. # pm3 opt=(verytight,gdiis,calcall,small) IOP(2/16=3) nosymm
18. # pm3 opt=(calcall,small,maxcyc=100) IOP(2/16=3)

MOPAC2009:

(calculations performed as two steps for optimization and frequency/thermo calculation, labeled "a" and "b", below)

1.
 - a. pm3 precise nosym
 - b. pm3 oldgeo thermo nosym precise
2.
 - a. pm3 precise nosym gnorm=0.0 nonr
 - b. pm3 oldgeo thermo nosym precise
3.
 - a. pm3 precise nosym gnorm=0.0
 - b. pm3 oldgeo thermo nosym precise
4.
 - a. pm3 precise nosym gnorm=0.0 bfgs
 - b. pm3 oldgeo thermo nosym precise
5.
 - a. pm3 precise nosym recal=10 dmax=0.10 nonr cycles=2000 t=2000
 - b. pm3 oldgeo thermo nosym precise