Computers and Chemical Engineering 52 (2013) 35–45

Contents lists available at SciVerse ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

# 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[1], William H. Green*

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

ARTICLE INFO

*Article history:*
Received 12 April 2012
Received in revised form
10 November 2012
Accepted 29 November 2012
Available online 22 December 2012

*Keywords:*
Automated reaction mechanism generation
Parameter estimation
Thermodynamic properties
Quantum chemistry
Molecular structure
Cheminformatics

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.

## 1. 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 software (Allen et al., 2009), which uses a rate-based mechanism construction algorithm (Susnow, Dean, Green, Peczak, & Broadbelt, 1997), and has been applied to a number of systems, including combustion of butanol (Harper, Van Geem, Pyl, Marin, & Green, 2011) and JP-10 (Magoon et al., 2010). 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 Benson (Benson & Buss, 1958; Benson, 1976; Cohen & Benson, 1992, 1993) 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

---

* Corresponding author. Tel.: +1 617 253 4580; fax: +1 617 258 8992.
*E-mail address:* whgreen@mit.edu (W.H. Green).

[1] Present address: Aerodyne Research, 45 Manning Rd., Billerica, MA 01821, United States.

0098-1354/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2012.11.009

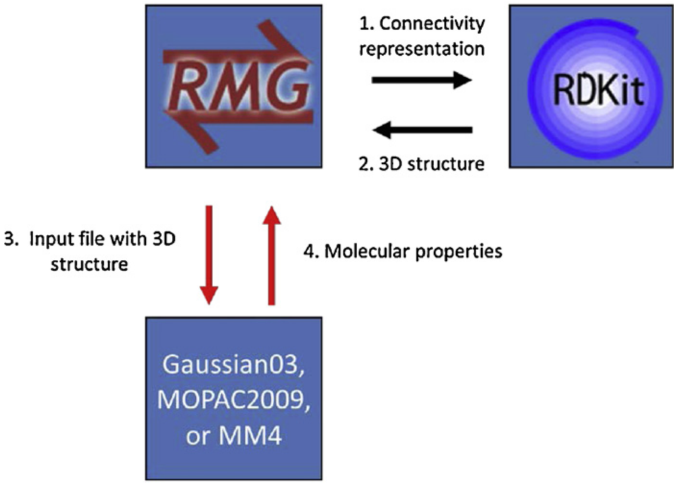

**Fig. 1.** Overview of QMTP interface.

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, Stark, and Klein (1994). 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 *q*uantum *m*echanics *t*hermodynamic *p*roperty (QMTP) interface.

## 2. Design and implementation of the QMTP interface

### 2.1. Design overview

An overview of the QMTP interface is shown in Fig. 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 (Irikura & Frurip, 1998).

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.

### 2.2. 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 (Sadowski & Gasteiger, 1993).

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 (Leite et al., 2007). This can limit, for instance, the types of rings that may be handled by the program. Examples of codes that use these approaches include Frog (Leite et al., 2007) and CORINA (Molecular Networks, 2012).

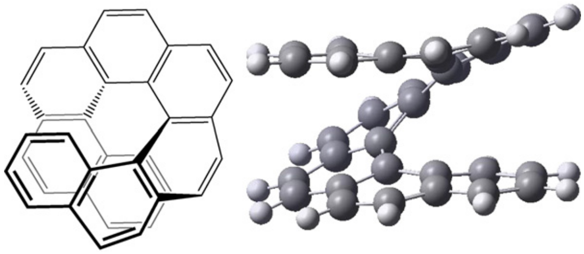

**Fig. 2.** Three dimensional structure for a helical polycyclic aromatic hydrocarbon (octahelicene) obtained with *RDKit*-based approach (including UFF refinement).

An alternative to rule-based approaches is to use distance geometry, a numerical technique. Examples of programs implementing distance geometry methods include Key3D (Mizutani, Nakamura, Ichinose, & Itai, 2006), Balloon (Vainio & Johnson, 2007), RDKit (Landrum, 2012), and smi23d (Gilbert & Guha, 2011). Distance geometry refers to the use of specified distance bounds between points to "embed" those points in *n*-dimensional space (Crippen, 1991; Svrcek-Seiler, 2004). 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 Fig. 2, whereas a typical rule-based approach would produce a nonsensical planar geometry for this case. Recently, Ebejer, Morris, and Deane (2012) 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.

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 (Spellmeyer, Wong, Bower, & Blaney, 1997). 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