



# **RMG-Py and CanTherm Documentation**

*Release 1.0.3*

**William H. Green, Richard H. West, and the RMG Team**

February 04, 2016



## CONTENTS

<b>1</b>	<b>RMG User's Guide</b>	<b>3</b>
1.1	Introduction . . . . .	3
1.2	Release Notes . . . . .	4
1.3	Overview of Features . . . . .	5
1.4	Installation . . . . .	5
1.5	Creating Input Files . . . . .	21
1.6	Example Input Files . . . . .	30
1.7	Running a Job . . . . .	37
1.8	Analyzing the Output Files . . . . .	37
1.9	Species Representation . . . . .	38
1.10	Group Representation . . . . .	40
1.11	Databases . . . . .	40
1.12	Thermochemistry Estimation . . . . .	58
1.13	Kinetics Estimation . . . . .	63
1.14	Liquid Phase Systems . . . . .	65
1.15	Guidelines for Building a Model . . . . .	70
1.16	Standalone Modules . . . . .	71
1.17	Frequently Asked Questions . . . . .	79
1.18	Credits . . . . .	80
1.19	How to Cite . . . . .	80
<b>2</b>	<b>CanTherm User's Guide</b>	<b>81</b>
2.1	Introduction . . . . .	81
2.2	Installation . . . . .	82
2.3	Creating Input Files for Thermodynamics and High-Pressure Limit Kinetics Computations . . . . .	83
2.4	Creating Input Files for Pressure Dependent Calculations . . . . .	86
2.5	Running CanTherm . . . . .	91
2.6	Parsing Output Files . . . . .	91
2.7	Frequently Asked Questions . . . . .	92
2.8	Credits . . . . .	93
<b>3</b>	<b>Theory Guide</b>	<b>95</b>
3.1	RMG Theory Guide . . . . .	95
3.2	Pressure-Dependence Theory Guide . . . . .	100
	<b>Bibliography</b>	<b>111</b>



RMG is an automatic chemical reaction mechanism generator that constructs kinetic models composed of elementary chemical reaction steps using a general understanding of how molecules react. This documentation is for the newer Python version of RMG that we call RMG-Py.

CanTherm is developed and distributed as part of RMG-Py, but can be used as a stand-alone application for Thermochemistry, Transition State Theory, and Master Equation chemical kinetics calculations. Its user guide is also included.

The last section of this documentation covers some of the more in depth theory behind RMG and CanTherm.

Please visit <http://reactionmechanismgenerator.github.io/RMG-Py/> for the most up to date documentation and source code. You may refer to the separate *RMG-Py API Reference* document to view the details of RMG-Py's modules and subpackages.



## RMG USER'S GUIDE

For any questions related to RMG and its usage and installation, please post an issue at <https://github.com/ReactionMechanismGenerator/RMG-Py/issues> and the RMG developers will get back to you as soon as we can. You can also search for your problem on the issues page to see if there are already solutions in development. Alternatively, you can email us at [rmg\\_dev@mit.edu](mailto:rmg_dev@mit.edu).

### 1.1 Introduction

**Reaction Mechanism Generator (RMG)** is an automatic chemical reaction mechanism generator that constructs kinetic models composed of elementary chemical reaction steps using a general understanding of how molecules react. This version is written in Python, and called RMG-Py.

#### 1.1.1 License

RMG is an open source program, available to the general public free of charge. The primary RMG code is distributed under the terms of the [MIT/X11 License](#). However, RMG has a number of dependencies of various licenses, some of which may be more restrictive. **It is the user's responsibility to ensure these licenses have been obtained.**

Copyright (c) 2002-2015 Prof. William H. Green ([whgreen@mit.edu](mailto:whgreen@mit.edu)),  
Prof. Richard H. West ([r.west@neu.edu](mailto:r.west@neu.edu)) and the RMG Team ([rmg\\_dev@mit.edu](mailto:rmg_dev@mit.edu))

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the 'Software'), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED 'AS IS', WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

## 1.2 Release Notes

### 1.2.1 RMG-Py Version 1.0.3

Date: February 4, 2016

This mini release contains the following updates:

- Pdep convergence issues in RMG-Py v1.0.2 are now fixed.
- RMG-database version information and anaconda binary version information is now recorded in RMG log file.

### 1.2.2 RMG-Py Version 1.0.2

Date: January 29, 2016

This new release adds several new features and bug fixes.

- Windows users can rejoice: RMG is now available in binary format on the Anaconda platform. Building by source is also much easier now through the Anaconda managed python environment for dependencies. See the updated [Installation Page](#) for more details
- Reaction filtering for speeding up model generation has now been added. It has been shown to speed up model convergence by 7-10x. See more details about how to use it in your RMG job [here](#). Learn more about the theory and algorithm on the [Rate-based Model Enlarging Algorithm](#) page.
- The RMG *native scripts* are now organized under the `rmgpy.tools` submodule for developer ease and better extensibility in external scripts.
- InChI conversion is now more robust for singlets and triplets, and augmented InChIs and InChI keys are now possible with new radical electron, lone pair, and multiplicity flags.
- Output HTML for visualizing models are now cleaned up and also more functional, including features to display thermo comments, display enthalpy, entropy, and free energy of reaction, as well as filter reactions by species. You can use this new visualization format either by running a job in RMG v1.0.2 or revisualizing your CHEMKIN file and species dictionary using the [visualization web tool](#).

### 1.2.3 RMG-database Version 1.0.2

Date: January 29, 2016

In conjunction with the release of RMG-Py v1.0.2, an updated package for the RMG-database has also been released. This release brings some new additions and fixes:

- New group additivity values for oxitene, oxerene, oexpane, and furan ring groups
- **Improvements to sulfur chemistry:**
  - Restructuring of radical trees in the kinetics families `SubstitutionS` and `intra_substitutionCS_cyclization`
  - A reaction library for di-tert-butyl sulfide
- Improvements for the `R_Addition_Multiple_Bond` kinetics family through new rate rules for the addition of allyl radical to double bonds in ethene, propene, and butene-like compounds, based on CBS-QB3 estimates from K. Wang, S.M. Villano, A.M. Dean, “Reactions of allylic radicals that impact molecular weight growth kinetics”, *PCCP*, 6255-6273 (2015).

- Several new thermodynamic and kinetics libraries for molecules associated with the pyrolysis of cyclopentadiene in the presence of ethene, based off of calculations from the paper A.G. Vandeputte, S.S. Merchant, M.R. Djokic, K.M. Van Geem, G.B. Marin, W. H. Green, “Detailed study of cyclopentadiene pyrolysis in the presence of ethene: realistic pathways from C<sub>5</sub>H<sub>5</sub> to naphthalene” (2016)

## 1.3 Overview of Features

**Thermodynamics estimation using group additivity.** Group additivity based on Benson’s groups provide fast and reliable thermochemistry estimates. A standalone utility for estimating heat of formation, entropy, and heat capacity is also included.

**Rate-based model enlargement** Reactions are added to the model based on their rate, fastest first.

**Rate-based termination.** The model enlargement stops when all excluded reactions are slower than a given threshold. This provides a controllable error bound on the kinetic model that is generated.

**Extensible libraries** Ability to include reaction models on top of the provided reaction families.

**Pressure-dependent reaction networks.** Dissociation, combination, and isomerization reactions have the potential to have rate coefficients that are dependent on both temperature and pressure, and RMG is able to estimate both for networks of arbitrary complexity with a bounded error.

**Simultaneous mechanism generation for several conditions.** Concurrent generation of a reaction mechanism over multiple temperature and pressure conditions. Mechanisms generated this way are valid over a range of reaction conditions.

**Dynamic simulation to a target conversion or time.** Often the desired simulation time is not known *a priori*, so a target conversion is preferred.

**Transport properties estimation using group additivity** The Lennard-Jones sigma and epsilon parameters are estimated using empirical correlations (based on a species’ critical properties and acentric factor). The critical properties are estimated using a group-additivity approach; the acentric factor is also estimated using empirical correlations. A standalone application for estimating these parameters is provided, and the output is stored in CHEMKIN-readable format.

## 1.4 Installation

---

**Note:** It is recommended that RMG be installed with Python 2.7, although it has been previously tested that Python 2.5 and 2.6 may also work. Dependency issues render it incompatible with Python 3.x releases.

---

For any questions related to RMG and its usage and installation, please post an issue at <https://github.com/ReactionMechanismGenerator/RMG-Py/issues> and the RMG developers will get back to you as soon as we can. You can also search for your problem on the issues page to see if there are already solutions in development. Alternatively, you can email us at [rmg\\_dev@mit.edu](mailto:rmg_dev@mit.edu)

### 1.4.1 For Basic Users: Binary Installation Using Anaconda

It is highly recommended to use the Python platform Anaconda to perform the installation of RMG-Py. A binary installation is recommended for users who want to use RMG out of the box, and are not interested in changing or recompiling the RMG code or making many additions to RMG’s thermodynamic and kinetics databases.

### Binary Installation Using Anaconda for Unix-Based Systems: Linux and Mac OSX

- Download and install the [Anaconda Python Platform](#) for Python 2.7 (make sure not to install Python 3.0+, which is incompatible with RMG). When prompted to append Anaconda to your PATH, select or type Yes.
- Install both RMG and the RMG-database binaries through the Terminal. Dependencies will be installed automatically. It is safest to make a new Anaconda environment for RMG and its dependencies. Type the following command into the Terminal to create the new environment named 'rmg\_env' containing the latest stable version of the RMG program and its database.

```
conda create -c rmg --name rmg_env rmg rmgdatabase
```

Whenever you wish to use it you must first activate the environment:

```
source activate rmg_env
```

- Optional: If you wish to use the *QMTP interface* with [MOPAC](#) to run quantum mechanical calculations for improved thermochemistry estimates of cyclic species, please obtain a legal license through the [MOPAC License Request Form](#). Once you have it, type the following into your Terminal

```
mopac password_string_here
```

- You may now run an RMG test job. Save the [Minimal Example Input File](#) to a local directory. Use the Terminal to run your RMG job inside that folder using the following command

```
rmg.py input.py
```

You may now use RMG-Py, CanTherm, as well as any of the *Standalone Modules* included in the RMG-Py package.

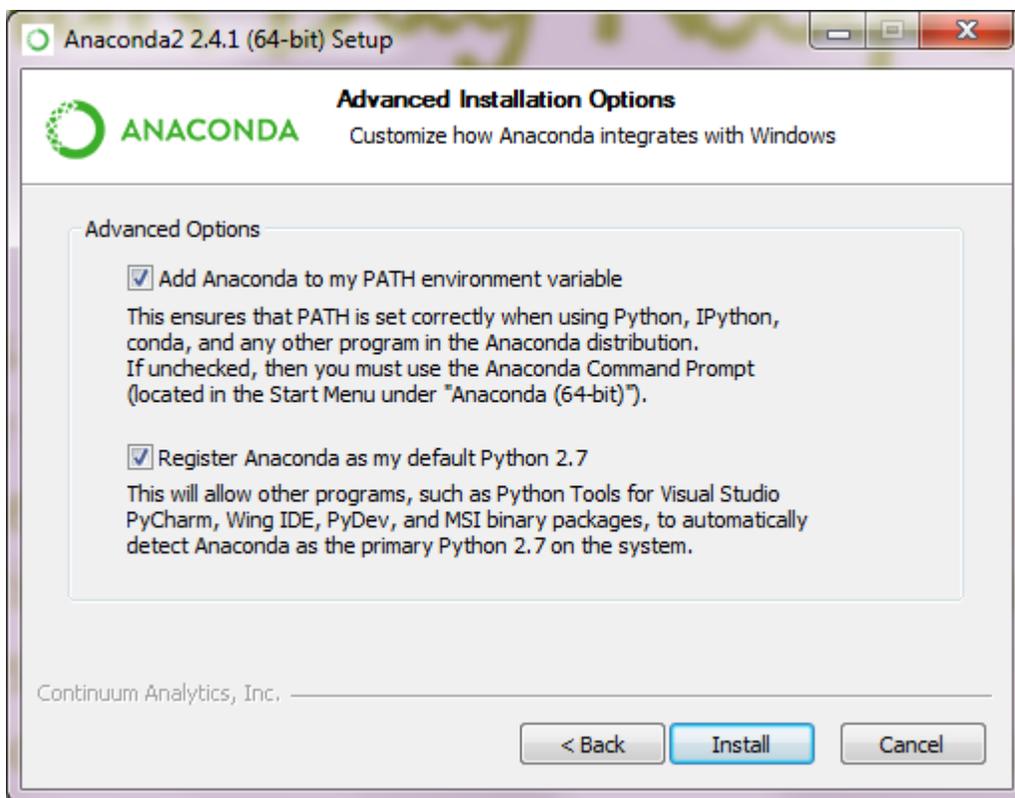
### Updating your binary installation of RMG in Linux or Mac OSX

If you had previously installed a binary version of the RMG package, you may check and update your installation to the latest stable version available on Anaconda by typing the following command on the Terminal

```
source activate rmg_env  
conda update rmg rmgdatabase -c rmg
```

### Binary Installation Using Anaconda for Windows

- Download and install the [Anaconda Python Platform](#) for Python 2.7 (make sure not to install Python 3.0+, which is incompatible with RMG). We recommend changing the default install path to C:\Anaconda\ in order to avoid spaces in the install path and be easily accessible. It is recommended to append Anaconda to your PATH as well as setting it as your default Python executable. All other settings can remain as their defaults.



- Now we want to install both RMG and the RMG-database binaries via the command prompt. Dependencies will be installed automatically. It is safest to make a new Anaconda environment for RMG and all its dependencies. Open a command prompt (either by finding it in your Program Files or by searching for cmd.exe) and type the following to create the new environment named 'rmg\_env' containing the latest stable version of the RMG program and its database.

```
conda create -c rmg --name rmg_env rmg rmgdatabase
```

- Whenever you wish to use it you must first activate the environment in the command prompt by typing:

```
activate rmg_env
```

- Optional: If you wish to use the *QMTP interface* with *MOPAC* to run quantum mechanical calculations for improved thermochemistry estimates of cyclic species, please obtain a legal license through the [MOPAC License Request Form](#). Once you have it, type the following into your command prompt (while the environment is activated)

```
mopac password_string_here
```

- Now you must *set the RMG environment variable in Windows* to allow your system to find the RMG python files more easily.
- If you set any new environment variables, you must now close and reopen the command prompt so that those environment variables can be refreshed and used.
- You may now run an RMG test job. Save the [Minimal Example Input File](#) to a local directory. Use the command prompt to run your RMG job inside that folder by using the following command

```
activate rmg_env
python %RMGPy%\rmg.py input.py
```

You may now use RMG-Py, CanTherm, as well as any of the *Standalone Modules* included in the RMG-Py package.

### Updating your binary installation of RMG for Windows

If you had previously installed a binary version of the RMG package, you may check and update your installation to the latest stable version available on Anaconda by typing the following command in a Command Prompt

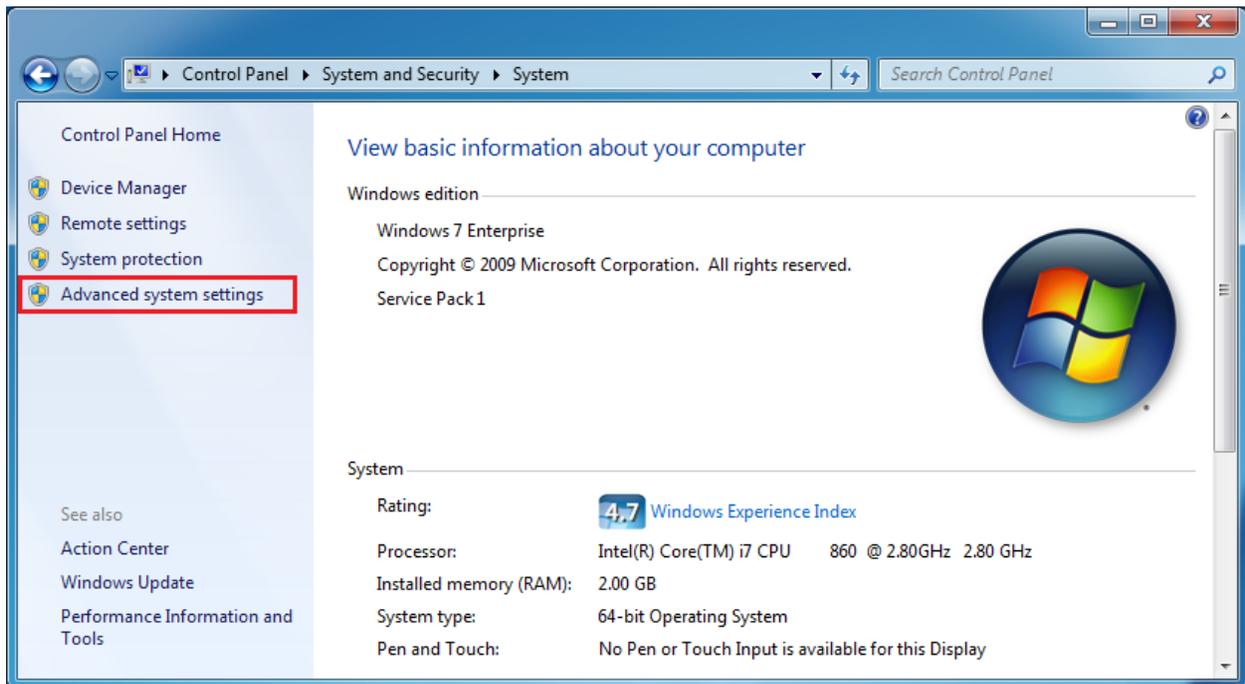
```
source activate rmg_env
conda update rmg rmgdatabase -c rmg
```

### Setting up Windows Environment Variables for RMG

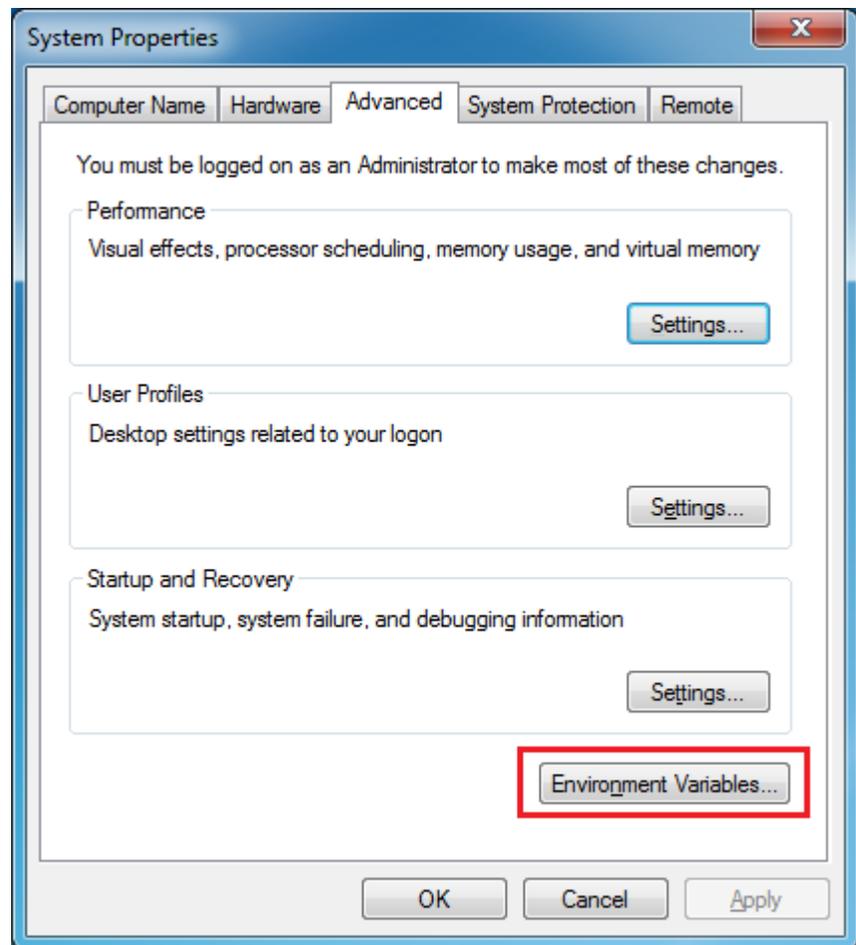
Setting environment variables in Windows allows for easier shortcutting and usage of RMG scripts and packages.

#### Setting the RMGPy variable

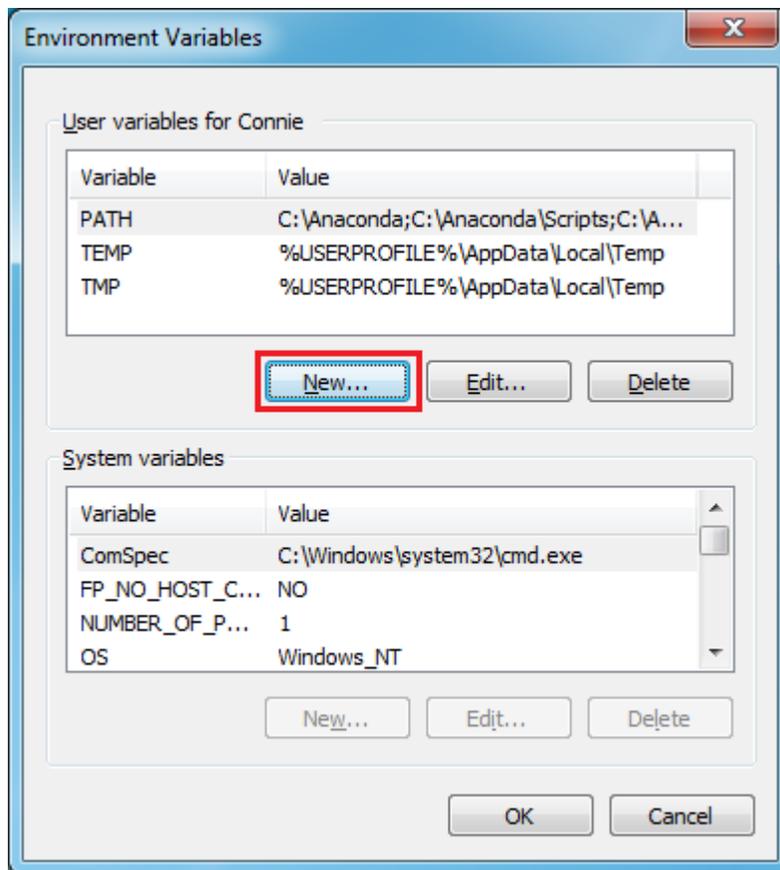
- If you have a search bar available in your start menu, search for “environment variables” and select “Edit environment variables for your account”. Alternatively, navigate to this settings window by first going to “Control Panel > System”, then clicking “Advanced system settings”.



- Once the “System Properties” window opens, click on “Environment Variables...” in the “Advanced” tab.



- Once the “Environment Variables” window opens, click on “New” under the “User variables”.

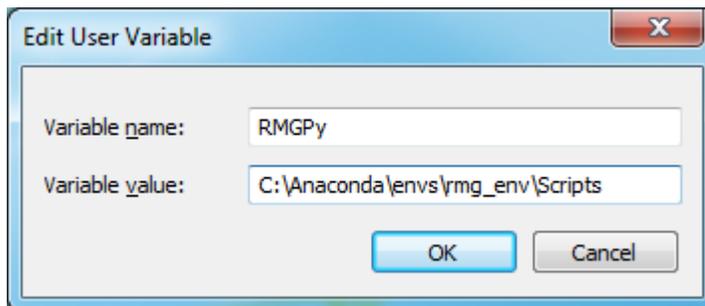


- Set a new variable with the name RMGPy with the appropriate value directed at your RMG path.  
If you installed the binary version of RMG, the environment value should be set to:

```
C:\Anaconda\envs\rmg_env\Scripts\
```

where C:\Anaconda can be replaced by wherever your Anaconda was installed.

Your screen might look like this:



If you are installing RMG by source, you can similarly set your RMGPy variable to the source directory, such as

```
C:\Code\RMG-Py
```

- Click “Ok” on all screens to confirm the changes.

**Note:** If you set any new environment variables, you must close and reopen any command prompts previously open before the changes can take effect.

### Optional: Setting a Permanent Anaconda Environment for RMG

If you use Anaconda solely for RMG, it may be more convenient to set your PATH variable to be permanently directed to the RMG environment. This will allow you to run RMG easily without having to type `activate rmg_env` in the command prompt every time.

Similarly to setting the environment variable for RMGPy, go to “Edit environment variables for your account” and click edit on the PATH variable. Replace the paths containing the Anaconda main directory with the RMG environment in Anaconda.

For example a path such as

```
C:\Anaconda\Scripts\
```

should be changed to:

```
C:\Anaconda\envs\rmg_env\Scripts\
```

Note that `C:\Anaconda` should be wherever your Anaconda was installed.

## 1.4.2 For Developers: Installation by Source Using Anaconda Environment

RMG-Py can now be built by source using the Anaconda Python Platform to assist in installing all necessary dependencies. This is recommended for a developer who may be altering the RMG source code or someone who expects to manipulate the databases extensively. You will also be able to access the latest source code updates and patches through Github.

### Installation by Source Using Anaconda Environment for Unix-based Systems: Linux and Mac OSX

- Download and install the [Anaconda Python Platform](#) for Python 2.7 (make sure not to install Python 3.0+, which is incompatible with RMG). When prompted to append Anaconda to your PATH, select or type Yes.
- Install [Git](#), the open source version control package through the Terminal. **For Mac OS X:** Git is already packages with OS X 10.9 or later, but requires installation of Xcode’s Command Line Tools. Skip the git installation and run it through the terminal, where you will be prompted to install the Command Line Tools if they are not already installed.

```
sudo apt-get install git
```

- Install the latest versions of RMG and RMG-database through cloning the source code via Git. Make sure to start in an appropriate local directory where you want both RMG-Py and RMG-database folders to exist.

```
git clone https://github.com/ReactionMechanismGenerator/RMG-Py.git
git clone https://github.com/ReactionMechanismGenerator/RMG-database.git
```

- Compile RMG-Py

```
cd RMG-Py
conda env create
source activate rmg_env
make
```

- Modify environment variables. Add RMG-Py to the PYTHONPATH to ensure that you can access RMG modules from any python prompt. Modify your `~/.bashrc` file by adding the following line

```
export PYTHONPATH=$PYTHONPATH:YourFolder/RMG-Py/
```

- If you wish to always be able to run RMG-Py, you can modify the anaconda path to point to the RMG environment. Modify the following line in your `~/ .bashrc` file

```
export PATH=~/anaconda/bin:$PATH
```

by changing it to the following line

```
export PATH=~/anaconda/envs/rmg/bin:$PATH
```

be sure to either close and reopen your terminal to refresh your environment variables, or type the following command

```
source ~/.bashrc
```

- Optional: If you wish to use the *QMTP interface* with *MOPAC* to run quantum mechanical calculations for improved thermochemistry estimates of cyclic species, please obtain a legal license through the [MOPAC License Request Form](#). Once you have it, type the following into your Terminal

```
mopac password_string_here
```

You may now use RMG-Py, CanTherm, as well as any of the *Standalone Modules* included in the RMG-Py package.

### Test Suite

There are a number of basic tests you can run on the newly installed RMG. It is recommended to run them regularly to ensure the code and databases are behaving normally.

- **Unit test suite:** this will run all the unit tests in the `rmgpy` package

```
cd RMG-Py
make test
```

- **Database test suite:** this will run the database unit tests to ensure that groups, rate rules, and libraries are well formed

```
cd RMG-Py
make test-database
```

### Running Examples

A number of basic examples can be run immediately. Additional example input files can be found in the `RMG-Py/examples` folder. Please read more on *Example Input Files* in the documentation.

- **Minimal Example:** this will run an Ethane pyrolysis model. It should take less than a minute to complete. The results will be in the `RMG-Py/testing/minimal` folder:

```
cd RMG-Py
make eg1
```

- **Hexadiene Example:** this will run a Hexadiene model with pressure dependence and QMTP. Note that you must have MOPAC installed for this to run. The results will be in the `RMG-Py/testing/hexadiene` folder:

```
cd RMG-Py
make eg2
```

- **Liquid Phase Example:** this will run a liquid phase RMG model. The results will be in the RMG-Py/testing/liquid\_phase folder

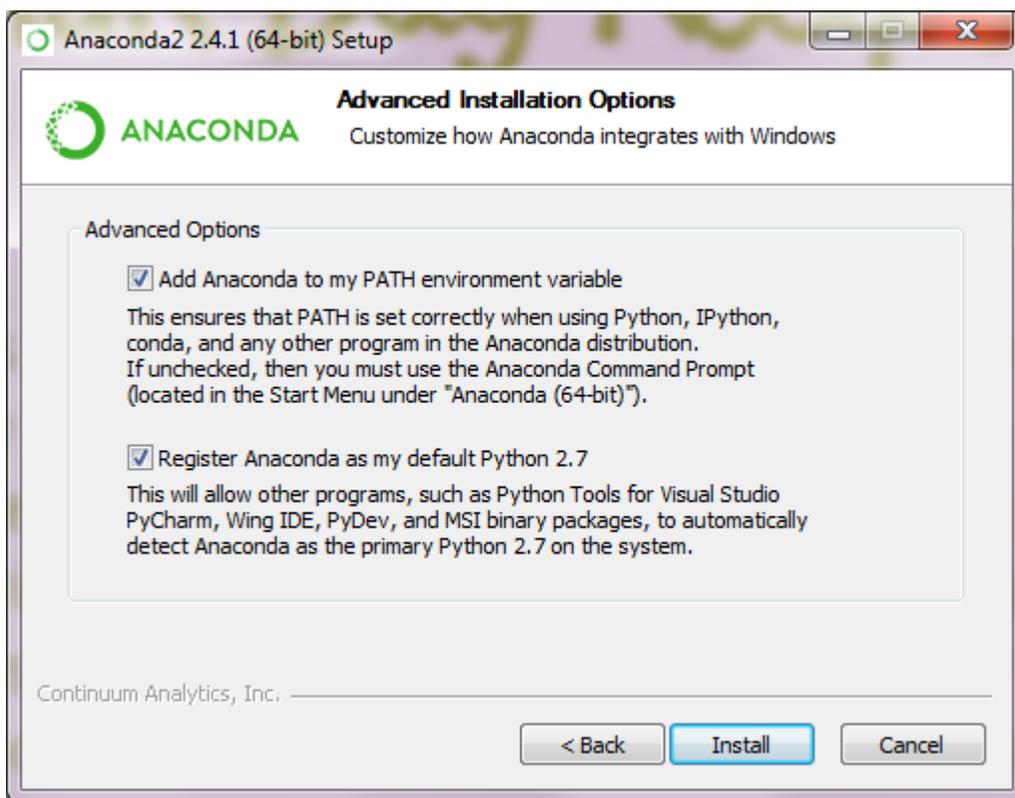
```
cd RMG-Py
make eg3
```

- **ThermoEstimator Example:** this will run the *Thermo Estimation Module* on a few molecules. Note that you must have MOPAC installed for this to run completely. The results will be in the RMG-Py/testing/thermoEstimator folder

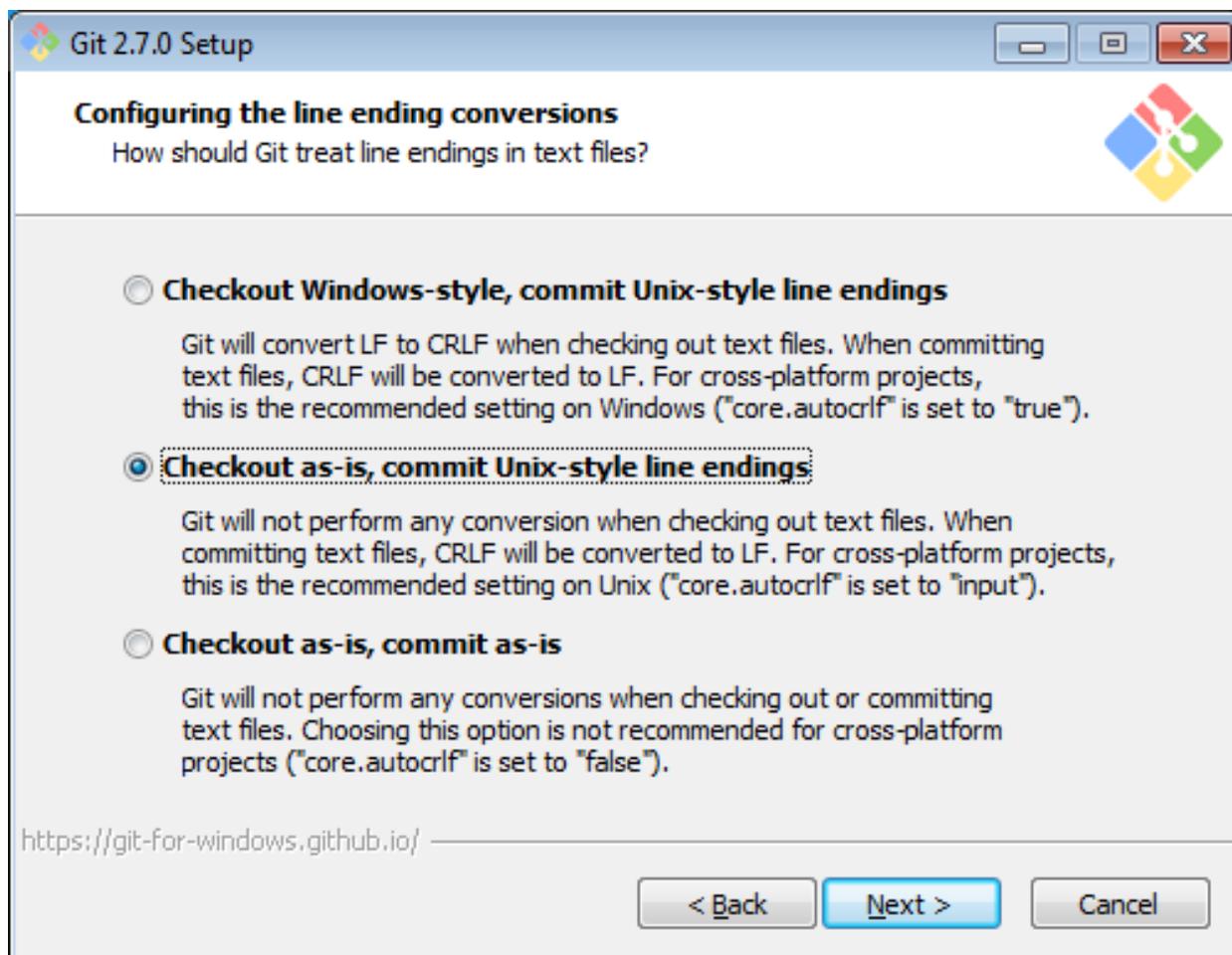
```
cd RMG-Py
make eg4
```

### Installation by Source Using Anaconda Environment for Windows

- Download and install the [Anaconda Python Platform](#) for Python 2.7 (make sure not to install Python 3.0+, which is incompatible with RMG). We recommend changing the default install path to C:\Anaconda\ in order to avoid spaces in the install path and be easily accessible. It is recommended to append Anaconda to your PATH as well as setting it as your default Python executable. All other settings can remain as their defaults.



- Install [Git](#), the open source version control package. When asked, append Git tools to your Command Prompt. It is also recommended to commit Unix-style line endings:



- Open Git CMD or a command prompt (either by finding it in your Program Files or by searching for `cmd.exe`). Install the latest versions of RMG and RMG-database through cloning the source code via Git. Make sure to start in an appropriate local directory where you want both RMG-Py and RMG-database folders to exist. We recommend creating a folder such as `C:\Code\`

```
git clone https://github.com/ReactionMechanismGenerator/RMG-Py.git
git clone https://github.com/ReactionMechanismGenerator/RMG-database.git
```

- Create the RMG Anaconda environment

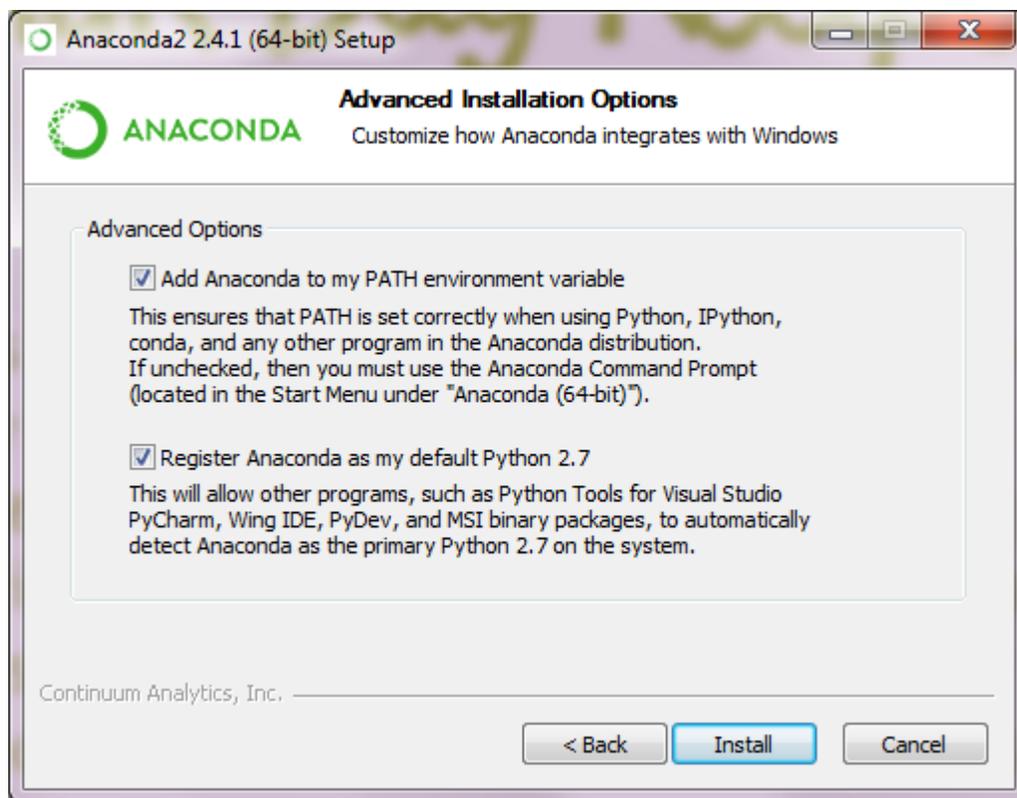
```
cd RMG-Py
conda env create -f environment_windows.yml
```

- Now you can compile RMG-Py

```
cd RMG-Py
mingw32-make
```

- Now it is recommended to modify your system's environment variables. Please see *Setting the RMG environment variable in Windows* for more information.

Additionally, set the `PYTHONPATH` environment variable to the path of your RMG-Py source folder to ensure that you can access RMG modules from any python prompt. The prompt might look like this:



- If you set any new environment variables, you must now close and reopen the command prompt so that those environment variables can be refreshed and used.
- Optional: If you wish to use the *QMTP interface* with MOPAC to run quantum mechanical calculations for improved thermochemistry estimates of cyclic species, please obtain a legal license through the [MOPAC License Request Form](#). Once you have it, type the following into your command prompt

```
mopac password_string_here
```

You may now use RMG-Py, CanTherm, as well as any of the *Standalone Modules* included in the RMG-Py package.

### Test Suite

There are a number of basic tests you can run on the newly installed RMG. It is recommended to run them regularly to ensure the code and databases are behaving normally.

- **Unit test suite:** this will run all the unit tests in the rmgpy package

```
cd RMG-Py
mingw32-make test
```

- **Database test suite:** this will run the database unit tests to ensure that groups, rate rules, and libraries are well formed

```
cd RMG-Py
mingw32-make test-database
```

### Running Examples

A number of basic examples can be run immediately. Additional example input files can be found in the RMG-Py\examples folder. Please read more on *Example Input Files* in the documentation.

- **Minimal Example:** this will run an Ethane pyrolysis model. It should take less than a minute to complete. The results will be in the RMG-Py\testing\minimal folder:

```
cd RMG-Py
mingw32-make eg1
```

- **Hexadiene Example:** this will run a Hexadiene model with pressure dependence and QMTP. Note that you must have MOPAC installed for this to run. The results will be in the RMG-Py\testing\hexadiene folder:

```
cd RMG-Py
mingw32-make eg2
```

- **Liquid Phase Example:** this will run a liquid phase RMG model. The results will be in the RMG-Py\testing\liquid\_phase folder

```
cd RMG-Py
mingw32-make eg3
```

- **ThermoEstimator Example:** this will run the *Thermo Estimation Module* on a few molecules. Note that you must have MOPAC installed for this to run completely. The results will be in the RMG-Py\testing\thermoEstimator folder

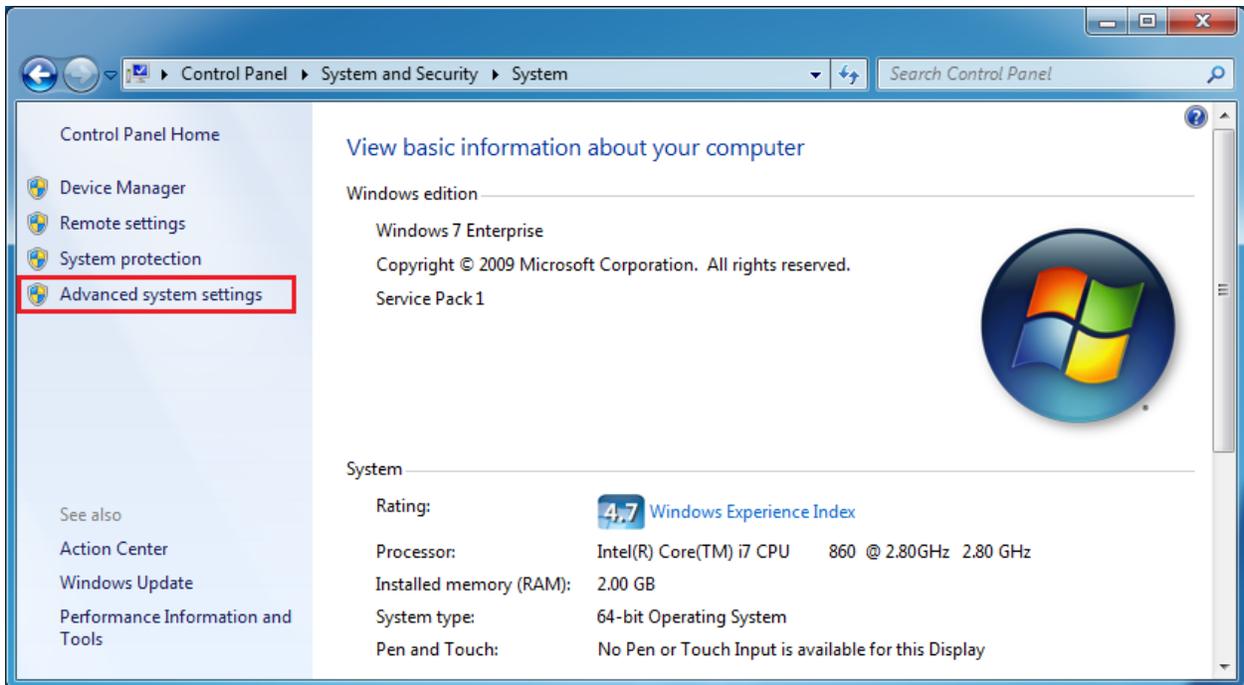
```
cd RMG-Py
mingw32-make eg4
```

### Setting up Windows Environment Variables for RMG

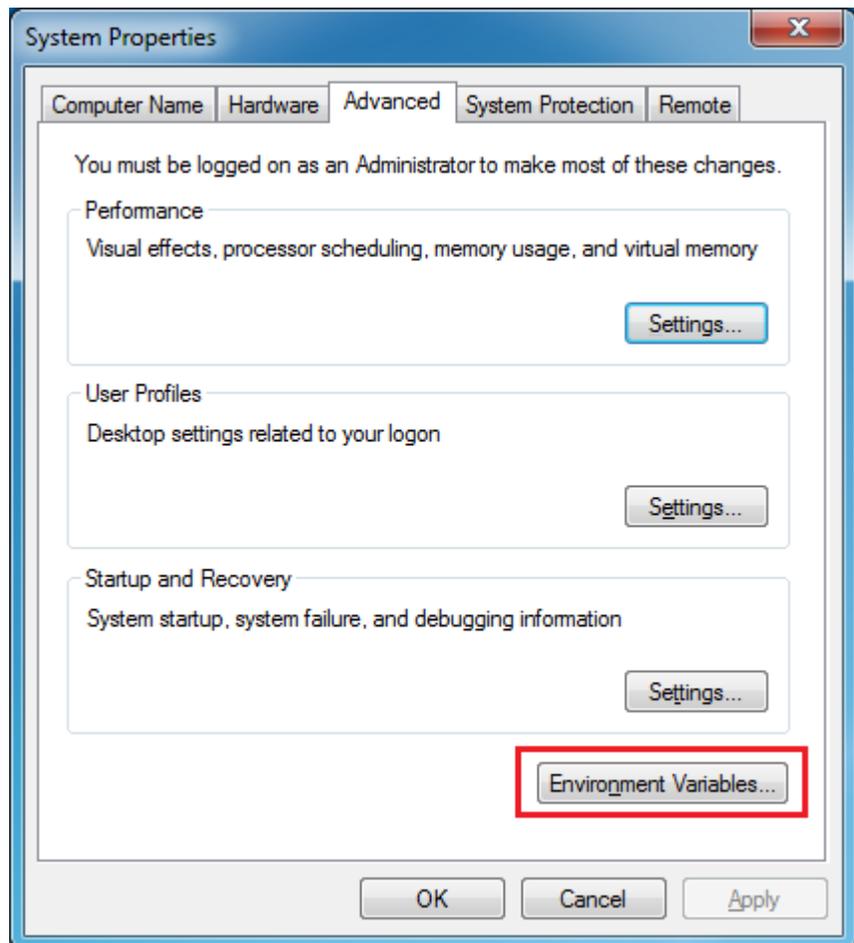
Setting environment variables in Windows allows for easier shortcutting and usage of RMG scripts and packages.

#### Setting the RMGPy variable

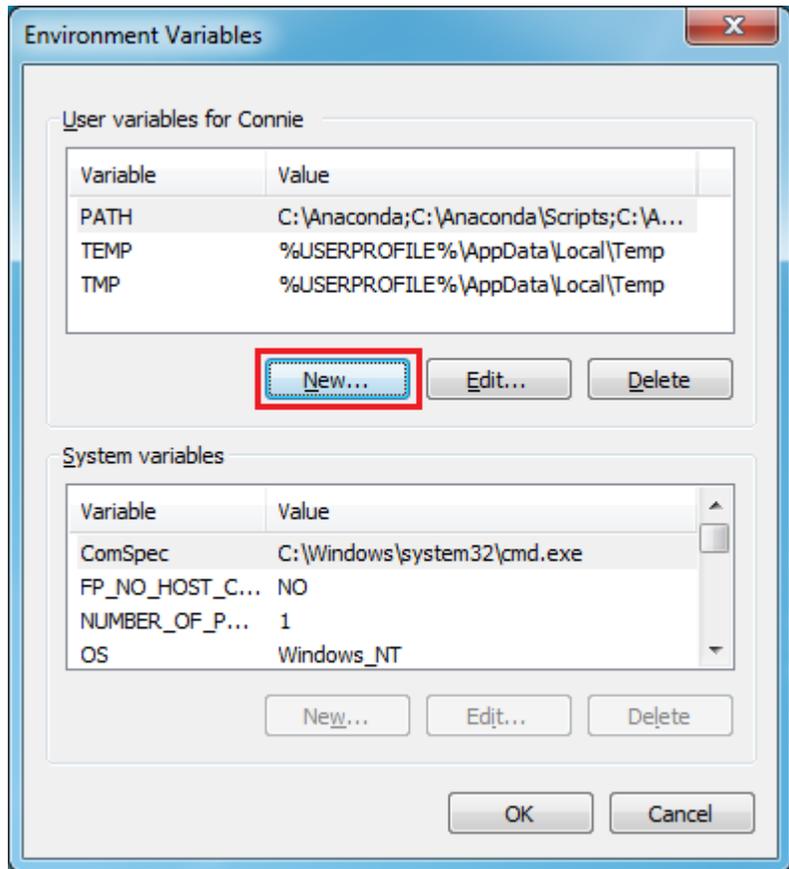
- If you have a search bar available in your start menu, search for “environment variables” and select “Edit environment variables for your account”. Alternatively, navigate to this settings window by first going to “Control Panel > System”, then clicking “Advanced system settings”.



- Once the “System Properties” window opens, click on “Environment Variables...” in the “Advanced” tab.



- Once the “Environment Variables” window opens, click on “New” under the “User variables”.



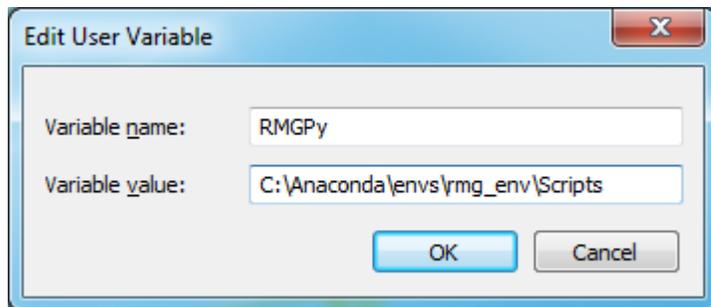
- Set a new variable with the name RMGPy with the appropriate value directed at your RMG path.

If you installed the binary version of RMG, the environment value should be set to:

C:\Anaconda\envs\rmg\_env\Scripts\

where C:\Anaconda can be replaced by wherever your Anaconda was installed.

Your screen might look like this:



If you are installing RMG by source, you can similarly set your RMGPy variable to the source directory, such as

C:\Code\RMG-Py

- Click “Ok” on all screens to confirm the changes.

**Note:** If you set any new environment variables, you must close and reopen any command prompts previously open

before the changes can take effect.

---

### Optional: Setting a Permanent Anaconda Environment for RMG

If you use Anaconda solely for RMG, it may be more convenient to set your PATH variable to be permanently directed to the RMG environment. This will allow you to run RMG easily without having to type `activate rmg_env` in the command prompt every time.

Similarly to setting the environment variable for RMGPy, go to “Edit environment variables for your account” and click edit on the PATH variable. Replace the paths containing the Anaconda main directory with the RMG environment in Anaconda.

For example a path such as

```
C:\Anaconda\Scripts\
```

should be changed to:

```
C:\Anaconda\envs\rmg_env\Scripts\
```

Note that C:\Anaconda should be wherever your Anaconda was installed.

### Updating the RMG-Py Source Code

It is recommended to keep yourself up to date with the latest patches and bug fixes by RMG developers, which is maintained on the official repository at <https://github.com/ReactionMechanismGenerator/RMG-Py/>. You can view the latest changes by viewing the commits tab on the repository. To update your source code, you can “pull” the latest changes from the official repo by typing the following command in the Command Prompt

```
cd RMG-Py
git pull https://github.com/ReactionMechanismGenerator/RMG-Py.git master
```

We also recommend updating the RMG-database regularly. The repo itself can be found at <https://github.com/ReactionMechanismGenerator/RMG-database/>

```
cd RMG-database
git pull https://github.com/ReactionMechanismGenerator/RMG-database.git master
```

For more information about how to use the Git workflow to make changes to the source code, please refer to the handy [Git Tutorial](#)

## 1.4.3 Dependencies

Please visit the page below for detailed information on all of RMG’s dependencies and their license restrictions

### Dependencies

#### List of Dependencies

Briefly, RMG depends on the following packages, almost all of which can be found in the [RMG anaconda channel](#) as binary packages.

- **boost**: portable C++ source libraries

- **cairo:** a 2D vector graphics library with support for multiple backends including image buffers, PNG, PostScript, PDF, and SVG file output. Used for molecular diagram generation
- **cairocffi:** a set of Python bindings and object-oriented API for cairo
- **coverage:** code coverage measurement for Python
- **cython:** compiling Python modules to C for speed up
- **gaussian:** (optional) commercial software program for quantum mechanical calculations. Must be installed separately.
- **gcc:** GNU compiler collection for C, C++, and Fortran. (MinGW is used in windows)
- **gprof2dot:** converts Python profiling output to a dot graph
- **graphviz:** generating flux diagrams
- **guppy:** memory profiling tools
- **jinja2:** Python templating language for html rendering
- **markupsafe:** implements XML/HTML/XHTML markup safe strings for Python
- **matplotlib:** library for making plots
- **mencoder:** (optional) used to encode videos, necessary for generating video flux diagrams. Must be installed separately.
- **mopac:** semi-empirical software package for QM calculations
- **nose:** advanced unit test controls
- **numpy:** fast matrix operations
- **openbabel:** chemical toolbox for speaking the many languages of chemical data
- **psutil:** system utilization diagnostic tool
- **pydas:** differential algebraic system solver
- **pydot:** interface to Dot graph language
- **pydqed:** constrained nonlinear optimization
- **pparsing:** a general parsing module for python
- **pyzmq:** Python bindings for zeroMQ
- **quantities:** unit conversion
- **rdkit:** open-source cheminformatics toolkit
- **scipy:** fast mathematical toolkit
- **scoop:** parallelization of Python code
- **setuptools:** for packaging Python projects
- **sphinx:** documentation generation
- **symmetry:** calculating symmetry numbers of chemical point groups
- **xlwt:** generating Excel output files

## License Restrictions on Dependencies

All of RMG's dependencies except the ones listed below are freely available and compatible with RMG's open source MIT license (though the specific nature of their licenses vary).

- **pydas**: The DAE solvers used in the simulations come from [Linda Petzold's research group](#) at UCSB. For running sensitivity analysis in RMG, the DASPK 3.1 solver is required, which "is subject to copyright restrictions" for non-academic use. Please visit their website for more details. To run RMG without this restriction, one may switch to compiling with the DASSL solver instead in RMG, which is "available in the public domain."

If you wish to do on-the-fly quantum chemistry calculations of thermochemistry (advisable for fused cyclic species in particular, where the ring corrections to group additive estimates are lacking), then you will need the third-party software for the QM calculations:

- **gaussian**: Gaussian03 and Gaussian09 are currently supported and commercially available. See <http://www.gaussian.com> for more details.
- **mopac** MOPAC can be found at <http://openmopac.net/>. Though it is not free for industrial use, it is free for non-profit and academic research use.

## 1.5 Creating Input Files

The syntax and parameters within an RMG input file are explained below. We recommend trying to build your first input file while referencing one of the *Example Input Files*. Alternatively, you can use our web form found at <http://rmg.mit.edu/input> to assist in creating an input file.

### 1.5.1 Syntax

The format of RMG-Py input .py is based on Python syntax.

Each section is made up of one or more function calls, where parameters are specified as text strings, numbers, or objects. Text strings must be wrapped in either single or double quotes.

### 1.5.2 Datasources

This section explains how to specify various reaction and thermo data sources in the input file.

#### Thermo Libraries

By default, RMG will calculate the thermodynamic properties of the species from Benson additivity formulas. In general, the group-additivity results are suitably accurate. However, if you would like to override the default settings, you may specify the thermodynamic properties of species in the ThermoLibrary. When a species is specified in the ThermoLibrary, RMG will automatically use those thermodynamic properties instead of generating them from Benson's formulas. Multiple libraries may be created, if so desired. The order in which the thermo libraries are specified is important: If a species appears in multiple thermo libraries, the first instance will be used.

Now in RMG, you have two types of thermo libraries: gas and liquid thermo libraries. As species thermo in liquid phase depends on the solvent, those libraries can only be used in liquid phase simulation with the corresponding solvent. Gas phase thermo library can be used either in gas phase simulation or in liquid phase simulation. (see more details on the two *thermo library types* and *how to use thermo libraries in liquid phase simulation*)

Please see Section *editing thermo database* for more details. In general, it is best to leave the ThermoLibrary set to its default value. In particular, the thermodynamic properties for H and H<sub>2</sub> must be specified in one of the primary thermo libraries as they cannot be estimated by Benson's method.

For example, if you wish to use the GRI-Mech 3.0 mechanism [*GRIMech3.0*] as a ThermoLibrary in your model, the syntax will be:

```
thermoLibraries = ['primaryThermoLibrary', 'GRI-Mech3.0']
```

This library is located in the \$RMG/RMG-database/input/thermo/libraries directory. All "Locations" for the ThermoLibrary field must be with respect to the \$RMG/RMG-database/input/thermo/libraries directory.

---

**Note:** Checks during the initialization are made to avoid users to use "liquid thermo libraries" in gas phase simulations or to use "liquid phase libraries" obtained in another solvent than the one defined in the input file in liquid phase simulations.

---

### Reaction Libraries

The next section of the `input.py` file specifies which, if any, Reaction Libraries should be used. When a reaction library is specified, RMG will first use the reaction library to generate all the relevant reactions for the species in the core before going through the reaction templates. Unlike the Seed Mechanism, reactions present in a Reaction Library will not be included in the core automatically from the start.

You can specify your own reaction library in the location section. In the following example, the user has created a reaction library with a few additional reactions specific to n-butane, and these reactions are to be used in addition to the Glarborg C3 library:

```
reactionLibraries = [('Glarborg/C3', False)],
```

The keyword `False/True` permits user to append all unused reactions (= kept in the edge) from this library to the chemkin file. `True` means those reactions will be appended.

The reaction libraries are stored in \$RMG-database/input/kinetics/libraries/ and the *Location*: should be specified relative to this path.

Because the units for the Arrhenius parameters are given in each mechanism, the different mechanisms can have different units.

---

**Note:** While using a Reaction Library the user must be careful enough to provide all instances of a particular reaction in the library file, as RMG will ignore all reactions generated by its templates. For example, suppose you supply the Reaction Library with `butyl_1 -> butyl_2`. Although RMG would find two unique instances of this reaction (via a three- and four-member cyclic Transition State), RMG would only use the rate coefficient supplied by you in generating the mechanism.

RMG will not handle irreversible reactions correctly, if supplied in a Reaction Library.

---

### Seed Mechanisms

The next section of the `input.py` file specifies which, if any, Seed Mechanisms should be used. If a seed mechanism is passed to RMG, every species and reaction present in the seed mechanism will be placed into the core, in addition to the species that are listed in the *Species Representation* section.

For details of the kinetics libraries included with RMG that can be used as a seed mechanism, see *Reaction Libraries*.

You can specify your own seed mechanism in the location section. Please note that the oxidation library should not be used for pyrolysis models. The syntax for the seed mechanisms is similar to that of the primary reaction libraries.

```
seedMechanisms = ['GRI-Mech3.0']
```

The seed mechanisms are stored in `RMG-database/input/kinetics/libraries/`

As the units for the Arrhenius parameters are given in each mechanism, different mechanisms can have different units. Additionally, if the same reaction occurs more than once in the combined mechanism, the instance of it from the first mechanism in which it appears is the one that gets used.

## Kinetics Depositories

```
kineticsDepositories = ['training']
```

## Kinetics Families

In this section users can specify the particular reaction families that they wish to use to generate their model. for example you can use only `Intra_RH_Add_Endocyclic` family to build the model by:

```
kineticsFamilies = ['Intra_RH_Add_Endocyclic']
```

Otherwise, by typing 'default' (and excluding the brackets that are shown in the example above), RMG will use recommended reaction families to generate the mechanism. The recommended reaction families can be found in `RMG-database/input/families/recommended.py`.

## Kinetics Estimator

The last section is specifying that RMG is estimating kinetics of reactions from rate rules. For more details on how kinetic estimations is working check [Kinetics Estimation](#):

```
kineticsEstimator = 'rate rules'
```

The following is an example of a database block, based on above chosen libraries and options:

```
database(
    thermoLibraries = ['primaryThermoLibrary', 'GRI-Mech3.0'],
    reactionLibraries = [('Glarborg/C3', False)],
    seedMechanisms = ['GRI-Mech3.0'],
    kineticsDepositories = ['training'],
    kineticsFamilies = 'default',
    kineticsEstimator = 'rate rules',
)
```

### 1.5.3 List of species

Species to be included in the core at the start of your RMG job are defined in the species block. The label, reactive or inert, and structure of each reactant must be specified. The label field will be used throughout your mechanism to identify the species. Inert species in the model can be defined by setting `reactive` to be `False`, for all other species the `reactive` status must be set as `True`. The structure of the species can be defined using either by using SMILES or `adjacencyList`.

The following is an example of a typical species item, based on methane using SMILE or adjacency list to define the structure:

```

species(
    label='CH4',
    reactive=True,
    structure=SMILES("C"),
)

species(
    label='CH4',
    reactive=True,
    structure=adjacencyList(
        """
            1 C 0
        """
    )
)

```

### 1.5.4 Reaction System

Every reaction system we want the model to be generated at must be defined individually. Currently, RMG can only model constant temperature and pressure systems. Future versions will allow for variable temperature and pressure. To define a reaction system we need to define the temperature, pressure and initial mole fractions of the reactant species. The initial mole fractions are defined using the label for the species in the species block. Every reaction system can have its termination criterion based on species conversion or termination time or both. When both termination criterion are specified the model generation will stop when either of the termination criterion is satisfied.

The following is an example of a simple reactor system:

```

simpleReactor(
    temperature=(1350, 'K'),
    pressure=(1.0, 'bar'),
    initialMoleFractions={
        "CH4": 0.104,
        "H2": 0.0156,
        "N2": 0.8797,
    },
    terminationConversion={
        'CH4': 0.9,
    },
    terminationTime=(1e0, 's'),
    sensitivity=['CH4', 'H2'],
    sensitivityThreshold=0.001,
)

```

Troubleshooting tip: if you are using a goal conversion rather than time, the reaction systems may reach equilibrium below the goal conversion, leading to a job that cannot converge physically. Therefore it is may be necessary to reduce the goal conversion or set a goal reaction time.

For sensitivity analysis, RMG-Py must be compiled with the DASPK solver, which is done by default but has some dependency restrictions. (See [License Restrictions on Dependencies](#) for more details.) The sensitivity and sensitivityThreshold are optional arguments for when the user would like to conduct sensitivity analysis with respect to the reaction rate coefficients for the list of species given for sensitivity.

Sensitivity analysis is conducted for the list of species given for sensitivity argument in the input file. The normalized concentration sensitivities with respect to the reaction rate coefficients  $\ln(C_i)/\ln(k_j)$  are saved to a csv file with the file name sensitivity\_1\_SPC\_1.csv with the first index value indicating the reactor system and the second naming the index of the species the sensitivity analysis is conducted for. Sensitivities to thermo of individual species is also saved as semi normalized sensitivities  $\ln(C_i)/d(G_j)$  where the units are given in 1/(kcal mol<sup>-1</sup>).

The `sensitivityThreshold` is set to some value so that only sensitivities for  $d\ln(C_i)/d\ln(k_j) > \text{sensitivityThreshold}$  or  $d\ln C_i/d(G_j) > \text{sensitivityThreshold}$  are saved to this file.

Note that in the RMG job, after the model has been generated to completion, sensitivity analysis will be conducted in one final simulation (sensitivity is not performed in intermediate iterations of the job).

### 1.5.5 Simulator Tolerances

The next two lines specify the absolute and relative tolerance for the ODE solver, respectively. Common values for the absolute tolerance are  $1e-15$  to  $1e-25$ . Relative tolerance is usually  $1e-4$  to  $1e-8$ :

```
simulator(
  atol=1e-16,
  rtol=1e-8,
  sens_atol=1e-6,
  sens_rtol=1e-4,
)
```

The `sens_atol` and `sens_rtol` are optional arguments for the sensitivity absolute tolerance and sensitivity relative tolerances, respectively. They are set to a default value of  $1e-6$  and  $1e-4$  respectively unless the user specifies otherwise. They do not apply when sensitivity analysis is not conducted.

### 1.5.6 Model Tolerances

Model tolerances dictate how species get included in the model. For more information, see the theory behind how RMG builds models using the *Flux-based Algorithm*. For running an initial job, it is recommended to only change the `toleranceMoveToCore` and `toleranceInterruptSimulation` values to an equivalent desired value. We find that typically a value between  $0.01$  and  $0.05$  is best. If your model cannot converge within a few hours, more advanced settings such as *reaction filtering* or *pruning* can be turned on to speed up your simulation at a slight risk of omitting chemistry.

```
model(
  toleranceMoveToCore=0.1,
  toleranceInterruptSimulation=0.1,
)
```

- `toleranceMoveToCore` indicates how high the edge flux ratio for a species must get to enter the core model. This tolerance is designed for controlling the accuracy of final model.
- `toleranceInterruptSimulation` indicates how high the edge flux ratio must get to interrupt the simulation (before reaching the `terminationConversion` or `terminationTime`). This value should be set to be equal to `toleranceMoveToCore` unless the advanced *pruning* feature is desired.

#### Advanced Setting: Speed Up by Filtering Reactions

For generating models for larger molecules, RMG-Py may have trouble converging because it must find reactions on the order of  $(n_{\text{reaction sites}})^{n_{\text{species}}}$ . Thus it can be further sped up by pre-filtering reactions that are added to the model. This modification to the algorithm does not react core species together until their concentrations are deemed high enough. It is recommended to turn on this flag when the model does not converge with normal parameter settings. See *Filtering Reactions within the Flux-based Algorithm*. for more details.

```
model(
  toleranceMoveToCore=0.1,
  toleranceInterruptSimulation=0.1,
```

```
    filterReactions=True,  
)
```

**Additional parameters:**

- `filterReactions`: set to `True` if reaction filtering is turned on. By default it is set to `False`.

**Advanced Setting: Speed Up by Pruning**

For further speed-up, it is also possible to perform mechanism generation with pruning of “unimportant” edge species to reduce memory usage.

A typical set of parameters for pruning is:

```
model(  
    toleranceMoveToCore=0.5,  
    toleranceInterruptSimulation=1e8,  
    toleranceKeepInEdge=0.05,  
    maximumEdgeSpecies=200000  
    minCoreSizeForPrune=50,  
    minSpeciesExistIterationsForPrune=2,  
)
```

**Additional parameters:**

- `toleranceKeepInEdge` indicates how low the edge flux ratio for a species must be to keep on the edge. This should be set to zero, which is its default.
- `maximumEdgeSpecies` indicates the upper limit for the size of the edge. The default value is set to 1000000 species.
- `minCoreSizeForPrune` ensures that a minimum number of species are in the core before pruning occurs, in order to avoid pruning the model when it is far away from completeness. The default value is set to 50 species.
- `minSpeciesExistIterationsForPrune` is set so that the edge species stays in the job for at least that many iterations before it can be pruned. The default value is 2 iterations.

**Recommendations:**

We recommend setting `toleranceKeepInEdge` to not be larger than 10% of `toleranceMoveToCore`, based on a pruning case study. In order to always enable pruning, `toleranceInterruptSimulation` should be set as a high value, e.g. `1e8`. `maximumEdgeSpecies` can be adjusted based on user’s RAM size. Usually 200000 edge species would cause memory shortage of 8GB computer, setting `maximumEdgeSpecies = 200000` (or lower values) could effectively prevent memory crash.

**Additional Notes:**

Note that when using pruning, RMG will not prune unless all reaction systems reach the goal reaction time or conversion without exceeding the `toleranceInterruptSimulation`. Therefore, you may find that RMG is not pruning even though the model edge size exceeds `maximumEdgeSpecies`, or an edge species has flux below the `toleranceKeepInEdge`. This is a safety check within RMG to ensure that species are not pruned too early, resulting in inaccurate chemistry. In order to increase the likelihood of pruning you can try increasing `toleranceInterruptSimulation` to an arbitrarily high value.

As a contrast, a typical set of parameters for non-pruning is:

```
model(  
    toleranceKeepInEdge=0,  
    toleranceMoveToCore=0.5,
```

```
toleranceInterruptSimulation=0.5,
)
```

where `toleranceKeepInEdge` is always 0, meaning all the edge species will be kept in edge since all the edge species have positive flux. `toleranceInterruptSimulation` equals to `toleranceMoveToCore` so that ODE simulation get interrupted once discovering a new core species. Because the ODE simulation is always interrupted, no pruning is performed.

Please find more details about the theory behind pruning at *Pruning Theory*.

### 1.5.7 On the fly Quantum Calculations

This block is used when quantum mechanical calculations are desired to determine thermodynamic parameters. These calculations are only run if the molecule is not included in a specified thermo library. The `onlyCyclics` option, if `True`, only runs these calculations for cyclic species. In this case, group additive estimates are used for all other species.

Molecular geometries are estimated via RDKit [*RDKit*]. Either MOPAC (2009 and 2012) or GAUSSIAN (2003 and 2009) can be used with the semi-empirical pm3, pm6, and pm7 (pm7 only available in MOPAC2012), specified in the software and method blocks. A folder can be specified to store the files used in these calculations, however if not specified this defaults to a *QMfiles* folder in the output folder.

The calculations are also only run on species with a maximum radical number set by the user. If a molecule has a higher radical number, the molecule is saturated with hydrogen atoms, then quantum mechanical calculations with subsequent hydrogen bond incrementation is used to determine the thermodynamic parameters.

The following is an example of the quantum mechanics options

```
quantumMechanics(
    software='mopac',
    method='pm3',
    fileStore='QMfiles',
    scratchDirectory = None,
    onlyCyclics = True,
    maxRadicalNumber = 0,
)
```

### 1.5.8 Pressure Dependence

This block is used when the model should account for pressure dependent rate coefficients. RMG can estimate pressure dependence kinetics based on Modified Strong Collision and Reservoir State methods. The former utilizes the modified strong collision approach of Chang, Bozzelli, and Dean [*Chang2000*], and works reasonably well while running more rapidly. The latter utilizes the steady-state/reservoir-state approach of Green and Bhatti [*Green2007*], and is more theoretically sound but more expensive.

The pressure dependence block should specify the following:

#### Method used for estimating pressure dependent kinetics

To specify the modified strong collision approach, this item should read

```
method='Modified Strong Collision'
```

To specify the reservoir state approach, this item should read

```
method='Reservoir State'
```

For more information on the two methods, consult the following resources :

### Grain size and minimum number of grains

Since the  $k(E)$  requires discretization in the energy space, we need to specify the number of energy grains to use when solving the Master Equation. The default value for the minimum number of grains is 250; this was selected to balance the speed and accuracy of the Master Equation solver method. However, for some pressure-dependent networks, this number of energy grains will result in the pressure-dependent  $k(T, P)$  being greater than the high-P limit

```
maximumGrainSize=(0.5, 'kcal/mol')
minimumNumberOfGrains=250
```

### Temperature and pressure for the interpolation scheme

To generate the  $k(T, P)$  interpolation model, a set of temperatures and pressures must be used. RMG can do this automatically, but it must be told a few parameters. We need to specify the limits of the temperature and pressure for the fitting of the interpolation scheme and the number of points to be considered in between this limit. For typical combustion model temperatures of the experiments range from 300 - 2000 K and pressure 1E-2 to 100 bar

```
temperatures=(300, 2000, 'K', 8)
pressures=(0.01, 100, 'bar', 5)
```

### Interpolation scheme

To use logarithmic interpolation of pressure and Arrhenius interpolation for temperature, use the line

```
interpolation=('PDepArrhenius',)
```

The auxillary information printed to the Chemkin chem.inp file will have the “PLOG” format. Refer to Section 3.5.3 of the CHEMKIN\_Input.pdf document and/or Section 3.6.3 of the CHEMKIN\_Theory.pdf document. These files are part of the CHEMKIN manual.

To fit a set of Chebyshev polynomials on inverse temperature and logarithmic pressure axes mapped to [-1,1], use the line

```
interpolation=('Chebyshev', 6, 4)
```

You should also specify the number of temperature and pressure basis functions by adding the appropriate integers. For example, the following specifies that six basis functions in temperature and four in pressure should be used

```
interpolation=('Chebyshev', 6, 4)
```

The auxillary information printed to the Chemkin chem.inp file will have the “CHEB” format. Refer to Section 3.5.3 of the CHEMKIN\_Input.pdf document and/or Section 3.6.4 of the CHEMKIN\_Theory.pdf document.

### Maximum size of adduct for which pressure dependence kinetics be generated

By default pressure dependence is run for every system that might show pressure dependence, i.e. every isomerization, dissociation, and association reaction. In reality, larger molecules are less likely to exhibit pressure-dependent behavior than smaller molecules due to the presence of more modes for randomization of the internal energy. In certain cases involving very large molecules, it makes sense to only consider pressure dependence for molecules smaller than some user-defined number of atoms. This is specified e.g. using the line

```
maximumAtoms=16
```

to turn off pressure dependence for all molecules larger than the given number of atoms (16 in the above example).

The following is an example of pressure dependence options

```
pressureDependence(
    method='modified strong collision',
    maximumGrainSize=(0.5,'kcal/mol'),
    minimumNumberOfGrains=250,
    temperatures=(300,2000,'K',8),
    pressures=(0.01,100,'bar',5),
    interpolation=('Chebyshev', 6, 4),
    maximumAtoms=16,
)
```

Regarding the number of polynomial coefficients for Chebyshev interpolated rates, please refer to the `rmgpy.kinetics.Chebyshev` documentation. The number of pressures and temperature coefficients should always be smaller than the respective number of user-specified temperatures and pressures.

### 1.5.9 Miscellaneous Options

Miscellaneous options:

```
options(
    units='si',
    saveRestartPeriod=(1,'hour'),
    generateOutputHTML=True,
    generatePlots=False,
    saveSimulationProfiles=True,
    verboseComments=False,
    saveEdgeSpecies=True,
)
```

The `units` field is set to `si`. Currently there are no other unit options.

The `saveRestartPeriod` indicates how frequently you wish to save restart files. For very large/long RMG jobs, this process can take a significant amount of time. In such cases, the user may wish to increase the time period for saving these restart files.

Setting `generateOutputHTML` to `True` will let RMG know that you want to save 2-D images (png files in the local species folder) of all species in the generated core model. It will save a visualized HTML file for your model containing all the species and reactions. Turning this feature off by setting it to `False` may save memory if running large jobs.

Setting `generatePlots` to `True` will generate a number of plots describing the statistics of the RMG job, including the reaction model core and edge size and memory use versus execution time. These will be placed in the output directory in the `plot/` folder.

Setting `saveSimulationProfiles` to `True` will make RMG save csv files of the simulation in `.csv` files in the `solver/` folder. The filename will be `simulation_1_26.csv` where the first number corresponds to the reaction system, and the second number corresponds to the total number of species at the point of the simulation. Therefore, the highest second number will indicate the latest simulation that RMG has completed while enlarging the core model. The information inside the csv file will provide the time, reactor volume in  $m^3$ , as well as mole fractions of the individual species.

Setting `verboseComments` to `True` will make RMG generate chemkin files with complete verbose commentary for the kinetic and thermo parameters. This will be helpful in debugging what values are being averaged for the kinetics. Note that this may produce very large files.

Setting `saveEdgeSpecies` to `True` will make RMG generate chemkin files of the edge reactions in addition to the core model in files such as `chem_edge.inp` and `chem_edge_annotated.inp` files located inside the chemkin folder. These files will be helpful in viewing RMG's estimate for edge reactions and seeing if certain reactions one expects are actually in the edge or not.

### 1.5.10 Species Constraints

RMG can generate mechanisms with a number of optional species constraints, such as total number of carbon atoms or electrons per species. These are applied to all of RMG's reaction families.

```
generatedSpeciesConstraints(  
    allowed=['input species','seed mechanisms','reaction libraries'],  
    maximumCarbonAtoms=10,  
    maximumHydrogenAtoms=10,  
    maximumOxygenAtoms=10,  
    maximumNitrogenAtoms=10,  
    maximumSiliconAtoms=10,  
    maximumSulfurAtoms=10,  
    maximumHeavyAtoms=10,  
    maximumRadicalElectrons=10,  
    allowSingletO2 = False,  
)
```

An additional flag `allowed` can be set to allow species from either the input file, seed mechanisms, or reaction libraries to bypass these constraints. Note that this should be done with caution, since the constraints will still apply to subsequent products that form.

Note that under all circumstances all forbidden species will still be banned unless they are manually removed from the database. See *Kinetics Database* for more information on forbidden groups.

By default, the `allowSingletO2` flag is set to `False`. See *Representing Oxygen* for more information.

## 1.6 Example Input Files

Perhaps the best way to learn the input file syntax is by example. To that end, a number of example input files and their corresponding output have been given in the `examples` directory. Two of the RMG jobs are shown below.

### 1.6.1 Ethane pyrolysis (Minimal)

This is the minimal example file characterizing a very basic system for ethane pyrolysis and should run quickly if RMG is set up properly. It does not include any calculation of pressure-dependent reaction rates.

```
# Data sources  
database(  
    thermoLibraries = ['primaryThermoLibrary'],  
    reactionLibraries = [],  
    seedMechanisms = [],  
    kineticsDepositories = ['training'],  
    kineticsFamilies = 'default',  
    kineticsEstimator = 'rate rules',  
)  
  
# List of species  
species(  

```

```

    label='ethane',
    reactive=True,
    structure=SMILES("CC"),
)

# Reaction systems
simpleReactor(
    temperature=(1350,'K'),
    pressure=(1.0,'bar'),
    initialMoleFractions={
        "ethane": 1.0,
    },
    terminationConversion={
        'ethane': 0.9,
    },
    terminationTime=(1e6,'s'),
)

simulator(
    atol=1e-16,
    rtol=1e-8,
)

model(
    toleranceKeepInEdge=0.0,
    toleranceMoveToCore=0.1,
    toleranceInterruptSimulation=0.1,
    maximumEdgeSpecies=100000,
)

options(
    units='si',
    saveRestartPeriod=None,
    generateOutputHTML=True,
    generatePlots=False,
    saveEdgeSpecies=True,
    saveSimulationProfiles=True,
)

```

## 1.6.2 1,3-hexadiene pyrolysis

This example models the pyrolysis of 1,3-hexadiene and demonstrates the effect of turning on the pressure-dependence module within RMG.

```

# Data sources
database(
    thermoLibraries = ['primaryThermoLibrary', 'GRI-Mech3.0'],
    reactionLibraries = [],
    seedMechanisms = [],
    kineticsDepositories = ['training'],
    kineticsFamilies = 'default',
    kineticsEstimator = 'rate rules',
)

# Constraints on generated species
generatedSpeciesConstraints(
    maximumRadicalElectrons = 2,
)

```

```
)  
  
# List of species  
species(  
    label='HXD13',  
    reactive=True,  
    structure=SMILES("C=CC=CCC"),  
)  
species(  
    label='CH4',  
    reactive=True,  
    structure=SMILES("C"),  
)  
species(  
    label='H2',  
    reactive=True,  
    structure=adjacencyList(  
        ""  
        1 H u0 p0 {2,S}  
        2 H u0 p0 {1,S}  
        ""),  
)  
species(  
    label='N2',  
    reactive=False,  
    structure=InChI("InChI=1/N2/c1-2"),  
)  
  
# Reaction systems  
simpleReactor(  
    temperature=(1350,'K'),  
    pressure=(1.0,'bar'),  
    initialMoleFractions={  
        "HXD13": 6.829e-4,  
        "CH4": 0.104,  
        "H2": 0.0156,  
        "N2": 0.8797,  
    },  
    terminationConversion={  
        'HXD13': 0.9,  
    },  
    terminationTime=(1e0,'s'),  
)  
  
simulator(  
    atol=1e-16,  
    rtol=1e-8,  
)  
  
model(  
    toleranceKeepInEdge=0.0,  
    toleranceMoveToCore=0.5,  
    toleranceInterruptSimulation=0.5,  
    maximumEdgeSpecies=100000  
)  
  
quantumMechanics(  
    software='mopac',
```

```

method='pm3',
# fileStore='QMfiles', # relative to where it is run. Defaults within the output folder.
scratchDirectory = None, # not currently used
onlyCyclics = True,
maxRadicalNumber = 0,
)

pressureDependence(
method='modified strong collision',
maximumGrainSize=(0.5,'kcal/mol'),
minimumNumberOfGrains=250,
temperatures=(300,2000,'K',8),
pressures=(0.01,100,'bar',5),
interpolation=('Chebyshev', 6, 4),
)

options(
units='si',
saveRestartPeriod=(1,'hour'),
generateOutputHTML=False,
generatePlots=False,
)

```

### 1.6.3 Commented input file

This is a fully commented input file with all optional blocks for new users to better understand the options of rmg input files

```

# Data sources
database(
#overrides RMG thermo calculation of RMG with these values.
#libraries found at http://rmg.mit.edu/database/thermo/libraries/
#if species exist in multiple libraries, the earlier libraries overwrite the
#previous values
thermoLibraries = ['KlippensteinH2O2','primaryThermoLibrary','DFT_QCI_thermo','CBS-QB3_1dHR'],
#overrides RMG kinetics estimation if needed in the core of RMG.
#list of libraries found at http://rmg.mit.edu/database/kinetics/libraries/
#input each library as a ('library_name',True/False) where a True means that all
#unused reactions will be automatically added to the chemkin file
reactionLibraries = [],
#seed mechanisms are reactionLibraries that are forced into the initial mechanism
#in addition to species listed in this input file.
#This is helpful for reducing run time for species you know will appear in
#the mechanism.
seedMechanisms = ['KlippensteinH2O2','ERC-FoundationFuelv0.9'],
#this is normally not changed in general RMG runs. Usually used for testing with
#outside kinetics databases
kineticsDepositories = 'default',
#lists specific families used to generate the model. 'default' uses a list of
#families from RMG-Database/input/families/recommended.py
#a visual list of families is available in PDF form at RMG-database/families
kineticsFamilies = 'default',
#specifies how RMG calculates rates. currently, the only option is 'rate rules'
kineticsEstimator = 'rate rules',
)

# List of species

```

```

#list initial and expected species below to automatically put them into the core mechanism.
# 'structure' can utilize method of SMILES("put_SMILES_here"),
# adjacencyList("put_adj_list_here"), or InChI("put_InChI_here")
# for molecular oxygen, use the smiles string [O][O] so the triplet form is used
species(
    label='butane',
    reactive=True,           #this parameter is optional if true
    structure=SMILES("CCCC"),
)
species(
    label='O2',
    structure=SMILES("[O][O]"),
)
species(
    label='N2',
    reactive=False,
    structure=adjacencyList("""
1 N u0 p1 c0 {2,T}
2 N u0 p1 c0 {1,T}
"""),
)
# You can list species not initially in reactor to make sure RMG includes them in the mechanism
species(
    label='QOOH',
    reactive=True,
    structure=SMILES("OCC[CH]C")
)
species(
    label='CO2',
    reactive=True,
    structure=SMILES("O=C=O")
)

#Reaction systems
#currently RMG models only constant temperature and pressure as homogeneous batch reactors.
#two options are: simpleReactor for gas phase or liquidReactor for liquid phase
#use can use multiple reactors in an input file for each condition you want to test.
simpleReactor(
    #specifies reaction temperature with units
    temperature=(700,'K'),
    #specifies reaction pressure with units
    pressure=(10.0,'bar'),
    #list initial mole fractions of compounds using the label from the 'species' label.
    #RMG will normalize if sum/=1
    initialMoleFractions={
        "N2": 4,
        "O2": 1,
        "butane": 1./6.5,
    },
    #the following two values specify when to determine the final output model
    #only one must be specified
    #the first condition to be satisfied will terminate the process
    terminationConversion={
        'butane': .99,
    },
    terminationTime=(40,'s'),
    #the next two optional values specify how RMG computes sensitivities of
    #rate coefficients with respect to species concentrations.

```

```

    #sensitivity contains a list of species' labels to conduct sensitivity analysis on.
    #sensitivityThreshold is the required sensitivity to be recorded in the csv output file
#   sensitivity=['CH4'],
#   sensitivityThreshold=0.0001,
)

# liquidReactor(
#     temperature=(500,'K'),
#     initialConcentrations={
#         "N2": 4,
#         "O2": 1,
#         "CO": 1,
#     },
#     terminationConversion=None,
#     terminationTime=(3600,'s'),
#     sensitivity=None,
#     sensitivityThreshold=1e-3
# )

    #liquid reactors also have solvents, you can specify one solvent
    #list of solvents available at : http://rmg.mit.edu/database/solvation/libraries/solvent/
# solvation('water')

#determines absolute and relative tolerances for ODE solver and sensitivities.
#normally this doesn't cause many issues and is modified after other issues are
#ruled out
simulator(
    atol=1e-16,
    rtol=1e-8,
#   sens_atol=1e-6,
#   sens_rtol=1e-4,
)

#used to add species to the model and to reduce memory usage by removing unimportant additional species.
#all relative values are normalized by a characteristic flux at that time point
model(
    #determines the relative flux to put a species into the core.
    #A higher value will result in a larger, more complex model
    #when running a new model, it is recommended to start with higher values and then decrease to converge
    toleranceMoveToCore=0.1,
    #comment out the next three terms to disable pruning
    #determines the relative flux needed to not remove species from the model.
    #Lower values will keep more species and utilize more memory
    toleranceKeepInEdge=0.01,
    #determines when to stop a ODE run to add a species.
    #Lower values will improve speed.
    #if it is too low, may never get to the end simulation to prune species.
    toleranceInterruptSimulation=1,
    #number of edge species needed to accumulate before pruning occurs
    #larger values require more memory and will prune less often
    maximumEdgeSpecies=100000,
    #minimum number of core species needed before pruning occurs.
    #this prevents pruning when kinetic model is far away from completeness
    minCoreSizeForPrune=50,
    #make sure that the pruned edge species have existed for a set number of RMG iterations.
    #the user can specify to increase it from the default value of 2
    minSpeciesExistIterationsForPrune=2,
    #filter the reactions during the enlarge step to omit species from reacting if their
    #concentration are deemed to be too low

```

```

    filterReactions=False,
)
options(
    #only option is 'si'
    units='si',
    #how often you want to save restart files.
    #takes significant amount of time. comment out if you don't want to save
    saveRestartPeriod=None,
    #Draws images of species and reactions and saves the model output to HTML.
    #May consume extra memory when running large models.
    generateOutputHTML=True,
    #generates plots of the RMG's performance statistics. Not helpful if you just want a model.
    generatePlots=False,
    #saves mole fraction of species in 'solver/' to help you create plots
    saveSimulationProfiles=False,
    #gets RMG to output comments on where kinetics were obtained in the chemkin file.
    #useful for debugging kinetics but increases memory usage of the chemkin output file
    verboseComments=False,
    #gets RMG to generate edge species chemkin files. Uses lots of memory in output.
    #Helpful for seeing why some reaction are not appearing in core model.
    saveEdgeSpecies=False,
)

# optional module allows for correction to unimolecular reaction rates at low pressures and/or temperatures.
pressureDependence(
    #two methods available: 'modified strong collision' is faster and less accurate than 'reservoir state'
    method='modified strong collision',
    #these two categories determine how fine energy is discretized.
    #more grains increases accuracy but takes longer
    maximumGrainSize=(0.5,'kcal/mol'),
    minimumNumberOfGrains=250,
    #the conditions for the rate to be output over
    #parameter order is: low_value, high_value, units, internal points
    temperatures=(300,2200,'K',2),
    pressures=(0.01,100,'bar',3),
    #The two options for interpolation are 'PDepArrhenius' (no extra arguments) and
    #'Chebyshev' which is followed by the number of basis sets in
    #Temperature and Pressure. These values must be less than the number of
    #internal points specified above
    interpolation=('Chebyshev', 6, 4),
    #turns off pressure dependence for molecules with number of atoms greater than the number specified be
    #this is due to faster internal rate of energy transfer for larger molecules
    maximumAtoms=15,
)

#optional block adds constraints on what RMG can output.
#This is helpful for improving the efficiency of RMG, but wrong inputs can lead to many errors.
generatedSpeciesConstraints(
    #allows exceptions to the following restrictions
    allowed=['input species','seed mechanisms','reaction libraries'],
    #maximum number of each atom in a molecule
    maximumCarbonAtoms=4,
    maximumHydrogenAtoms=10,
    maximumOxygenAtoms=7,
    maximumNitrogenAtoms=0,
    maximumSiliconAtoms=0,
    maximumSulfurAtoms=0,
)

```

```

        #max number of non-hydrogen atoms
        #maximumHeavyAtoms=20,
        #maximum radicals on a molecule
        maximumRadicalElectrons=1,
        #If this is false or missing, RMG will throw an error if the more less-stable form of O2 is entered
        #which doesn't react in the RMG system. normally input O2 as triplet with SMILES [O][O]
        #allowSingletO2=False,
    )

#optional block allows thermo to be estimated through quantum calculations
# quantumMechanics(
#     #the software package for calculations...can use 'mopac' or 'gaussian' if installed
#     software='mopac',
#     #methods available for calculations. 'pm2' 'pm3' or 'pm7' (last for mopac only)
#     method='pm3',
#     #where to store calculations
#     fileStore='QMfiles',
#     #where to store temporary run files
#     scratchDirectory = None,
#     #onlyCyclics allows linear molecules to be calculated using bensen group additivity...need to verify
#     onlyCyclics = True,
#     #how many radicals should be utilized in the calculation.
#     #If the amount of radicals is more than this, RMG will use hydrogen bond incrementation method
#     maxRadicalNumber = 0,
# )

```

## 1.7 Running a Job

Submitting a job is easy

```
python rmg.py input.py
```

We recommend you make a job-specific directory for each RMG simulation. Some jobs can take quite a while to complete, so we also recommend using a job scheduler (if working in an linux environment).

## 1.8 Analyzing the Output Files

You will see that a successfully executed RMG job will create multiple output files and folders: `output.html` (if `generateOutputHTML=True` is specified) /chemkin /pdep /plot /solver /species restart.pkl RMG.log

### 1.8.1 The Chemkin Folder

The /chemkin folder will likely have a large number of chemkin formatted files. In general, these can be disregarded, as you will be mainly interested in `chem.inp`, the chemkin formatted input file with a species list, thermochemical database, and a list of elementary reactions. All of `inp` files appended with numbers are those that have been generated by RMG as it runs and the mechanism is still in progress of enlarging. `chem_annotated.inp` is provided as a means to help make sense of species syntax and information sources (i.e., how RMG estimated individual kinetic and thermodynamic parameters). In addition, a species dictionary, `species_dictionary.txt`, is generated containing all the species in your mechanism in the format of an adjacency list. Either chemkin file, in addition to the dictionary, may be used as inputs in the `tools` section of this website to better visualize the species and reactions. Alternatively, if the input option `generateOutputHTML=` is set to `True`, you will be able to visualize 2D images of all species and reactions in your mechanism as image files and in an html file, `output.html`. Once you are able to visualize the

mechanism, several useful tools exist. For example, in the *Reaction Details* section, you'll see the following with check-box fields beside them:

- Kinetics
- Comments
- Chemkin strings

If you check the last box, chemkin strings, you can then search for strings corresponding to seemingly nonsensical named species (e.g. S(1234)) that may show up in any analyses/simulations you perform (e.g., with Cantera or Chemkin). Further, under *Reaction Families*, you can selectively view the reactions that been generated based on a particular RMG reaction family or library.

### 1.8.2 The Species Folder

If `generateOutputHTML=True` is specified as an RMG input *option*, the species folder will be populated with png files with 2D pictures of each species in your final mechanism. Otherwise, it will contain no files, or files generated from pressure dependent jobs.

### 1.8.3 The Pdep Folder

The `/pdep` folder will contain files associated with the pressure-dependent reactions that RMG has generated, if you requested such a job. These files are formatted as input files for *CanTherm*, which can be run independently. This can be useful if one wants to visualize the potential energy surface corresponding to any particular network.

### 1.8.4 The Solver Folder

RMG currently includes a solver for isothermal batch reactors. This is in fact a critical part of the model enlargement algorithm. If you have included simulations in your input file, the solutions will be located in `/solver`. You will probably only be interested in the files with the largest number tags.

## 1.9 Species Representation

Species objects in RMG contain a variety of attributes, including user given names, thermochemistry, as well as structural isomers. See the `rmgpy.species.Species` class documentation for more information.

RMG considers each species to be unique, and comprised of a set of molecular structural isomers, including resonance isomers. RMG uses the list of resonance isomers to compare whether two species are the same. Each molecular structure is stored in RMG using graph representations, using vertices and edges, where the vertices are the atoms and the edges are the bonds. This form of representation is known as an adjacency list. For more information on adjacency lists, see the `rmgpy.molecule.adjlist` page.

Species objects in the input file can also be constructed using other common representations such as SMILES, SMARTS, and InChIs. The following can all be used to represent the methane species:

```
species(  
    label='CH4',  
    reactive=True,  
    structure=SMILES("C"),  
)
```

Replacing the structure with any of the following representations will also produce the same species:

```

structure=adjacencyList("
1 C u0 p0 c0 {2,S} {3,S} {4,S} {5,S}
2 H u0 p0 c0 {1,S}
3 H u0 p0 c0 {1,S}
4 H u0 p0 c0 {1,S}
5 H u0 p0 c0 {1,S}
"),

structure=SMARTS("[CH4]"),

structure=SMILES("C"),

structure=InChI("InChI=1S/CH4/h1H4"),

```

To quickly generate any adjacency list, or to generate an adjacency list from other types of molecular representations such as SMILES, InChI, or even common species names, use the Molecule Search tool found here: [http://rmg.mit.edu/molecule\\_search](http://rmg.mit.edu/molecule_search)

### 1.9.1 Representing Oxygen

Special care should be taken when constructing a mechanism that involves molecular oxygen. The ground electronic state of molecular oxygen,  $^3\Sigma_g^-$ , does *not* contain a double bond, but instead a single bond and two lone electrons. In RMG's adjacency list notation the ground state of oxygen is represented as

```

1 0 u1 p2 {2,S}
2 0 u1 p2 {1,S}

```

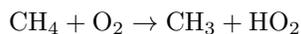
You should use the above adjacency list to represent molecular oxygen in your condition files, seed mechanisms, etc. The triplet form is 22 kcal/mol more stable than the first singlet excited state,  $^1\Delta_g$ , which does contain a double bond. The adjacency list for singlet oxygen is

```

1 0 u0 p2 {2,D}
2 0 u0 p2 {1,D}

```

Selecting the correct structure for oxygen is important, as the reactions generated from a double bond are significantly different than those generated from a radical or diradical. For example, the reaction



would occur for both triplet and singlet oxygen, but in entirely different families. For triplet oxygen the above represents a hydrogen abstraction, while for singlet oxygen it represents the reverse of a disproportionation reaction.

The RMG databases have been modified to make all of the oxygen-related chemistry that was present in RMG databases consistent with the single-bonded biradical representation.

Conversion between triplet and singlet forms is possible through the primary reaction library `OxygenSingTrip`; the reactions involved are very slow, however, and are likely to be absent from any mechanisms generated. At this point, no other reactions of singlet oxygen have been included in RMG.

In order to allow the singlet form of O2 to be used in RMG, please allow it explicitly by setting `allowSingletO2` to `True` in the `generateSpeciesConstraints` section of the RMG input file.

```

generatedSpeciesConstraints(
    allowSingletO2 = True,
)

```

## 1.10 Group Representation

Group representations are used to represent molecular substructures within RMG. These are commonly used for identifying functional groups for use in both the thermo and kinetic databases.

For syntax of how to define groups, see `rmgpy.molecule.adjlist`.

## 1.11 Databases

RMG has databases storing thermochemistry and kinetics data. These databases can be visualized on the RMG website here: <http://rmg.mit.edu/database/>

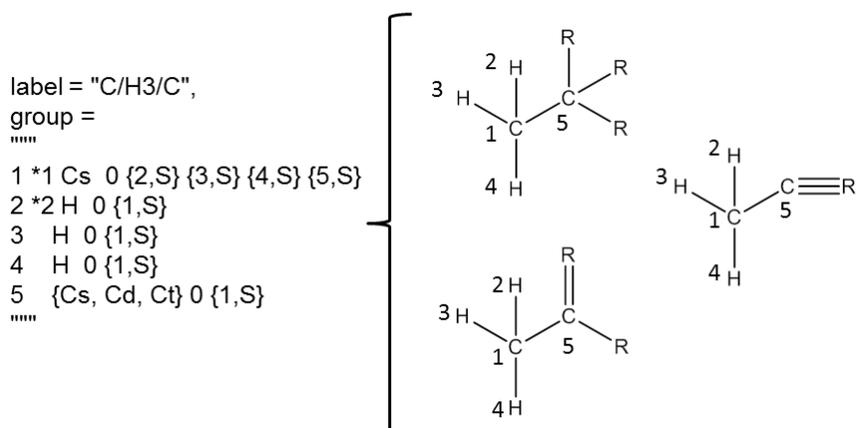
### 1.11.1 Introduction

This section describes some of the general characteristics of RMG's databases.

#### Group Definitions

The main section in many of RMG's databases are the 'group' definitions. Groups are adjacency lists that describe structures around the reacting atoms. Between the adjacency list's index number and atom type, a starred number is inserted if the atom is a reacting atom.

Because groups typically do not describe entire molecules, atoms may appear to be lacking full valency. When this occurs, the omitted bonds are allowed to be anything. An example of a primary carbon group from H-Abstraction is shown below. The adjacency list defined on the left matches any of the three drawn structures on the right (the numbers correspond to the index from the adjacency list).



New **atom types** are also introduced to describe atoms in group definitions. The table below shows all atoms types in RMG.

Atom Type	Chemical Element	Bonding
R	Any	No requirements
R!H	Any except hydrogen	No requirements
H	Hydrogen	No requirements
C	Carbon	No requirements
Cs	Carbon	No double or triple bonds
Cd	Carbon	Exactly one double bond to any non-oxygen atom
Cdd	Carbon	Two double bonds to any atoms
Ct	Carbon	One triple bond
CO	Carbon	Exactly one double bond to an oxygen atom
Cb	Carbon	Exactly two benzene bonds
Cbf	Carbon	Three benzene bonds (Fused aromatics)
O	Oxygen	No requirements
Os	Oxygen	No double bonds
Od	Oxygen	One double bond
Oa	Oxygen	No bonds (Oxygen atom)
S	Sulfur	No requirements
Ss	Sulfur	No double bond
Sd	Sulfur	One double bond
Sa	Sulfur	No bonds (Sulfur atom)

Additionally, groups can also be defined as unions of other groups. For example,:

```
label="X_H_or_Xrad_H",
group=OR{X_H, Xrad_H},
```

## Forbidden Groups

Forbidden groups can be defined to ban structures globally in RMG or to ban pathways in a specific kinetic family.

Globally forbidden structures will ban all reactions containing either reactants or products that are forbidden. These groups are stored in the file located at `RMG-database/input/forbiddenStructures.py`.

To ban certain specific pathways in the kinetics families, a *forbidden* group must be created, like the following group in the `intra_H_migration` family

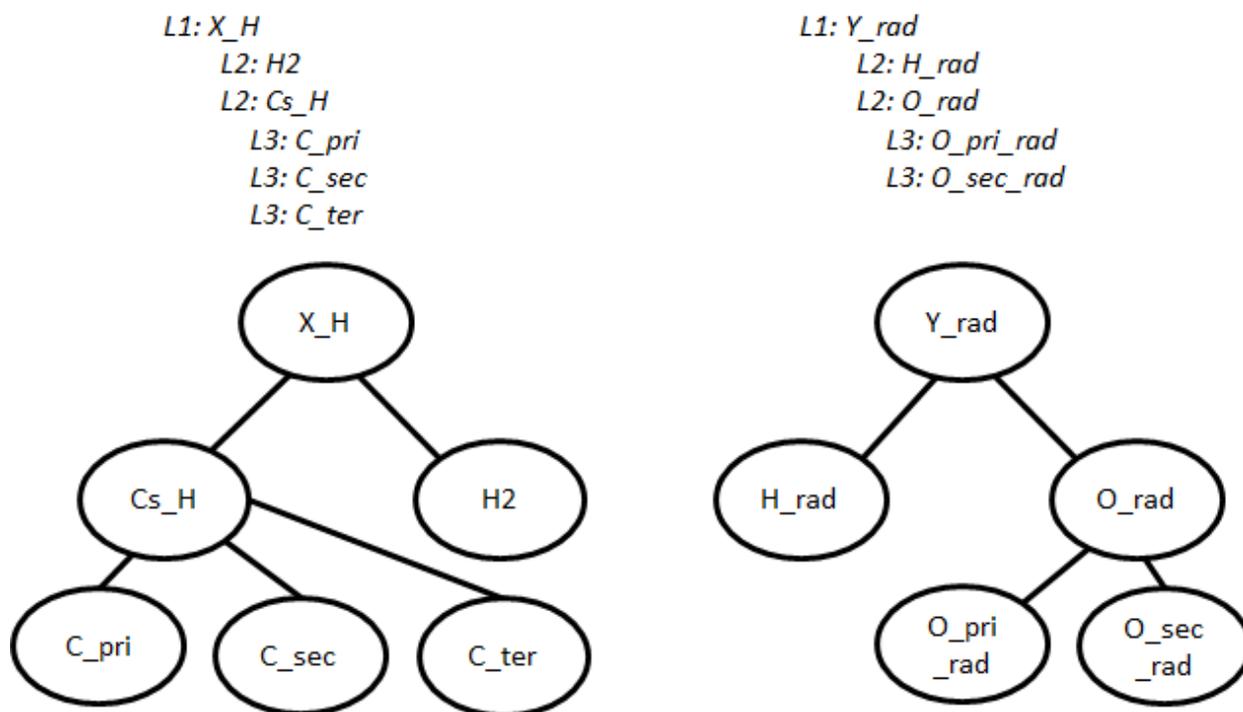
```
forbidden(
    label = "bridged56_1254",
    group =
    """
1 *1 C 1 {2,S} {6,S}
2 *4 C 0 {1,S} {3,S} {7,S}
3   C 0 {2,S} {4,S}
4 *2 C 0 {3,S} {5,S} {8,S}
5 *5 C 0 {4,S} {6,S} {7,S}
6   C 0 {1,S} {5,S}
7   C 0 {2,S} {5,S}
8 *3 H 0 {4,S}
""",
    shortDesc = u"",
    longDesc =
u""
""",
)
```

Forbidden groups should be placed inside the groups.py file located inside the specific kinetics family's folder RMG-database/input/kinetics/family\_name/ alongside normal group entries. The starred atoms in the forbidden group ban the specified reaction recipe from occurring in matched products and reactants.

## Hierarchical Trees

Groups are ordered into the nodes of a hierarchical trees which is written at the end of groups.py. The root node of each tree is the most general group with the reacting atoms required for the family. Descending from the root node are more specific groups. Each child node is a subset of the parent node above it.

A simplified example of the trees for H-abstraction is shown below. The indented text shows the syntax in groups.py and a schematic is given underneath.



Individual groups only describe part of the reaction. To describe an entire reaction we need one group from each tree, which we call **node templates** or simply templates. (C\_pri, O\_pri\_rad), (H2, O\_sec\_rad), and (X\_H, Y\_rad) are all valid examples of templates. Templates can be filled in with kinetic parameters from the training set or rules.

### 1.11.2 Thermo Database

This section describes the general usage of RMG's thermochemistry databases. Thermochemical data in RMG is reported using three different quantities:

1. Standard heat capacity data  $C_p^o(T)$  as a function of temperature  $T$
2. Standard enthalpy of formation at 298K  $\Delta_f H^o(298K)$
3. Standard entropy at 298K  $S^o(298K)$

A heat capacity model based on the Wilhoit equation is used for inter- and extrapolation of the heat capacity data as a function of temperature.

## Libraries

### Library types

Two types of thermo libraries are available in RMG: “gas phase” and “liquid thermo” libraries respectively identified thanks to the absence or presence of the keyword *solvent* = “*solvent\_name*” in the header of a thermo library. Here is an example of a liquid thermo library header:

```
name = "example_liquid_thermo_library"
solvent = "octane"
shortDesc = u"test"
longDesc = u""
```

In this example the library name is “example\_liquid\_thermo\_library” and thermo data provided was obtained in *octane* solvent. The only difference between gas phase and liquid phase thermo libraries is made through this keyword, the rest of the library is similar to gas phase.

**Note:** You can only provide one solvent per library and users should pay attention to not mix thermo of species obtained in different solvent in a same library. RMG will raise an error if users try to load a liquid thermo library obtained in another solvent than the one provided in input file. (in the example provided here, this liquid thermo library can only be used in liquid phase simulation with octane as solvent. RMG will also raise an error if user try to use liquid phase thermo library in gas phase simulations.

### Species thermochemistry libraries

The folder RMG-database/input/thermo/libraries/ in RMG-database is the location to store species thermochemistry libraries. Each particularly library is stored in a file with the extension .py, e.g. ‘DFT\_QCI\_thermo.py’.

An example of a species thermochemistry entry is shown here below:

```
entry(
    index = 1,
    label = "H2",
    molecule =
    """
    1 H 0 0 {2,S}
    2 H 0 0 {1,S}
    """
    ,
    thermo = ThermoData(
        Tdata = ([300,400,500,600,800,1000,1500], 'K'),
        Cpdata = ([6.948,6.948,6.949,6.954,6.995,7.095,7.493], 'cal/(mol*K)'),
        H298 = (0, 'kcal/mol'),
        S298 = (31.095, 'cal/(mol*K)'),
    ),
    shortDesc = u" """,
    longDesc =
    u""
    """
    ,
)
)
```

The text above describes the first entry in the library (index = 1), labeled ‘H2’, through the adjacency list representation. Heat capacity data (‘Cpdata’) is described at 7 different temperatures, along with the standard enthalpy of formation at 298K (‘H298’), and the standard entropy at 298K (‘S298’).

According to the *thermo classes* available in RMG, you can provide different thermo data: NASA, thermodata (as shown above), wilhoit or NASAPolynomial.

## Groups

The folder `RMG-database/input/thermo/groups/` in `RMG-database` is the location to store group contribution databases. Each particularly type of group contribution is stored in a file with the extension `.py`, e.g. `'groups.py'`:

file	Type of group contribution
<code>gauche.py</code>	1,4-gauche non-nearest neighbor interactions (NNIs)
<code>group.py</code>	group additive values (GAVs)
<code>int15.py</code>	1,5-repulsion non-nearest neighbor interactions (NNIs)
<code>other.py</code>	other non-nearest neighbor interactions (NNIs)
<code>polycyclic.py</code>	polycyclic ring corrections (RSCs)
<code>radical.py</code>	hydrogen bond increments (HBIs)
<code>ring.py</code>	monocyclic ring corrections (RSCs)

Like many other entities in RMG, the database of each type of group contribution is organized in a hierarchical tree, and is defined at the bottom of the database file. E.g.:

```
tree(
"""
L1: R
    L2: C
        L3: Cbf
            L4: Cbf-CbCbCbfb
            L4: Cbf-CbCbCbCbfb
            L4: Cbf-CbfCbCbCbfb
        L3: Cb
            L4: Cb-H
            L4: Cb-0s
            L4: Cb-Ss
            L4: Cb-C
                L5: Cb-Cs
                L5: Cb-Cds
                    L6: Cb-(Cds-0d)
            ...
    ...

```

More information on hierarchical tree structures in RMG can be found here: [Introduction](#).

### Group additive values (GAV)

An example of a GAV entry in `group.py` is shown here below:

```
entry(
    index = 3,
    label = "Cbfb-CbCbCbfb",
    group =
    """
1 * Cbf 0 {2,B} {3,B} {4,B}
2 Cb 0 {1,B}
3 Cb 0 {1,B}
4 Cbf 0 {1,B}
    """,
    thermo = ThermoData(
        Tdata = ([300,400,500,600,800,1000,1500], 'K'),

```

```

    Cpdata = ([3.01,3.68,4.2,4.61,5.2,5.7,6.2], 'cal/(mol*K)',
              '+|-', [0.1,0.1,0.1,0.1,0.1,0.1,0.1]),
    H298 = (4.8, 'kcal/mol', '+|-', 0.17),
    S298 = (-5, 'cal/(mol*K)', '+|-', 0.1),
),
shortDesc = u""Cbf-CbCbCb STEIN and FAHR; J. PHYS. CHEM. 1985, 89, 17, 3714""",
longDesc =
u""
Taken from STEIN and FAHR; J. PHYS. CHEM. 1985, 89, 17, 3714
""",
)

```

The text above describes a GAV “Cbf-CbCbCb”, with the central atom denoted by the asterisk in the adjacency list representation. Uncertainty margins are added in the data, after the unit specification. A short description ‘shortDesc’ specifies the origin of the data.

### Ring Strain Corrections (RSC)

RMG distinguishes between monocyclic and polycyclic ring correction databases.

Monocyclic RSCs are used for molecules that contain one single ring. An example of a monocyclic RSC entry in ring.py is shown here below:

```

entry(
    index = 1,
    label = "Cyclopropane",
    group =
    """
1 * Cs 0 {2,S} {3,S}
2  Cs 0 {1,S} {3,S}
3  Cs 0 {1,S} {2,S}
    """,
    thermo = ThermoData(
        Tdata = ([300,400,500,600,800,1000,1500], 'K'),
        Cpdata = ([-3.227,-2.849,-2.536,-2.35,-2.191,-2.111,-1.76], 'cal/(mol*K)'),
        H298 = (27.53, 'kcal/mol'),
        S298 = (32.0088, 'cal/(mol*K)'),
    ),
    shortDesc = u""Cyclopropane ring BENSON""",
    longDesc =
u""
""",
)

```

A molecule may have two or more fused rings that mutually interact. In that case, a polycyclic ring strain correction may be more adequate. RMG identifies molecules with fused ring systems and subsequently searches through polycyclic.py to identify an adequate RSC.

An example of a polycyclic RSC entry in polycyclic.py is shown here below:

```

entry(
    index = 2,
    label = "norbornane",
    group =
    """
1 * Cs 0 {3,S} {4,S} {7,S}
2  Cs 0 {3,S} {5,S} {6,S}
    """
)

```

```

3 Cs 0 {1,S} {2,S}
4 Cs 0 {1,S} {5,S}
5 Cs 0 {2,S} {4,S}
6 Cs 0 {2,S} {7,S}
7 Cs 0 {1,S} {6,S}
"""
    thermo = ThermoData(
        Tdata = ([300,400,500,600,800,1000,1500], 'K'),
        Cpdata = ([-4.5,-3.942,-3.291,-2.759,-2.08,-1.628,-0.898], 'cal/(mol*K)'),
        H298 = (16.14,'kcal/mol'),
        S298 = (53.47,'cal/(mol*K)'),
    ),
    shortDesc = u"""",
    longDesc =
u"""
"""
)

```

### Hydrogen Bond Increments (HBI)

An example of a HBI entry in radical.py is shown here below:

```

entry(
    index = 4,
    label = "CH3",
    group =
    """
1 * C 1 {2,S} {3,S} {4,S}
2 H 0 {1,S}
3 H 0 {1,S}
4 H 0 {1,S}
"""
    thermo = ThermoData(
        Tdata = ([300,400,500,600,800,1000,1500], 'K'),
        Cpdata = ([0.71,0.34,-0.33,-1.07,-2.43,-3.54,-5.43], 'cal/(mol*K)'),
        H298 = (104.81,'kcal/mol', '+|-',0.1),
        S298 = (0.52,'cal/(mol*K)'),
    ),
    shortDesc = u""""Calculated in relation to methane from NIST values""",
    longDesc =
u"""
"""
)

```

### Non-nearest neighbor interactions

The majority of the NNIs groups pertain to small enthalpy of formation corrections. Only a very limited number include entropy or heat capacity corrections. The database other.py contains cis-, ortho- and ketene-corrections.

An example of a NNI entry in gauche.py is shown here below:

```

entry(
    index = 11,
    label = "Cs(Cs(CsCsR)Cs(CsCsR)RR)",

```

```

group =
"""
1 * Cs                0 {2,S} {3,S} {4,S} {5,S}
2  Cs                0 {1,S} {6,S} {7,S} {8,S}
3  Cs                0 {1,S} {9,S} {10,S} {11,S}
4  {Cd,Cdd,Ct,Cb,Cbf,Os,CO,H} 0 {1,S}
5  {Cd,Cdd,Ct,Cb,Cbf,Os,CO,H} 0 {1,S}
6  Cs                0 {2,S}
7  Cs                0 {2,S}
8  {Cd,Cdd,Ct,Cb,Cbf,Os,CO,H} 0 {2,S}
9  Cs                0 {3,S}
10 Cs                0 {3,S}
11 {Cd,Cdd,Ct,Cb,Cbf,Os,CO,H} 0 {3,S}
"""
thermo = ThermoData(
    Tdata = ([300,400,500,600,800,1000,1500], 'K'),
    Cpdata = ([0,0,0,0,0,0,0], 'cal/(mol*K)'),
    H298 = (0.8, 'kcal/mol'),
    S298 = (0, 'cal/(mol*K)'),
),
shortDesc = u" ",
longDesc =
u" "
"""
)

```

### 1.11.3 Kinetics Database

This section describes the general usage of RMG's kinetic database. See *Modifying the Kinetics Database* for instructions on modifying the database.

Pressure independent reaction rates in RMG are calculated using a modified Arrhenius equation, designating the reaction coefficient as  $k(T)$  at temperature  $T$ .

$$k(T) = A \left( \frac{T}{T_0} \right)^n e^{-(E_a + \alpha \Delta H_{rxn}) / (RT)}$$

$R$  is the universal gas constant. The **kinetic parameters** determining the rate coefficient are:

- $A$ : the pre-exponential A-factor
- $T_0$ : the reference temperature
- $n$ : the temperature exponent
- $E_a$ : the activation energy
- $\alpha$ : the Evans-Polanyi coefficient
- $\Delta H_{rxn}$ : the enthalpy of reaction

When Evans-Polanyi corrections are used,  $\Delta H_{rxn}$  is calculated using RMG's thermo database, instead of being specified in the kinetic database.

### Libraries

Kinetic libraries delineate kinetic parameters for specific reactions. RMG always chooses to use kinetics from libraries over families. If multiple libraries contain the same reaction, then precedence is given to whichever library is listed

first in the input.py file.

For combustion mechanisms, you should always use *one* small-molecule combustion library, such as the pre-packaged ERC-Foundation Fuel. The reactions contained in these libraries are poorly estimated by kinetic families and are universally important to combustion systems.

Kinetic libraries should also be used in the cases where:

- A set of reaction rates were optimized together
- You know the reaction rate is not generalizable to similar species (perhaps due to catalysis or aromatic structures)
- No family exists for the class of reaction
- You are not confident about the accuracy of kinetic parameters

## Families

Allowable reactions in RMG are divided up into classes called **reaction families**. All reactions not listed in a kinetic library have their kinetic parameters estimated from the reaction families.

Each reaction family contains the files:

- groups.py containing the recipe, group definitions, and hierarchical trees
- training.py containing a training set for the family
- rules.py containing kinetic parameters for rules

There are currently 45 reaction families in RMG:

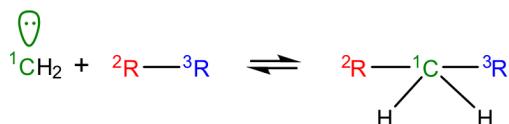
### 1+2\_Cycloaddition



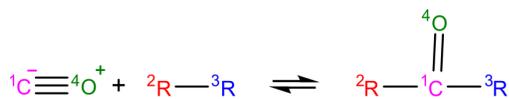
### 1,2-Birad\_to\_alkene



### 1,2\_Insertion\_carbene



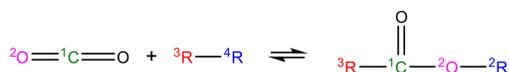
### 1,2\_Insertion\_CO



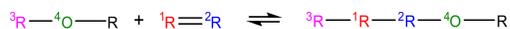
### 1,2\_shiftS



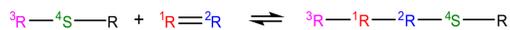
### 1,3\_Insertion\_CO2



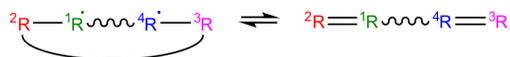
### 1,3\_Insertion\_ROR



### 1,3\_Insertion\_RSR



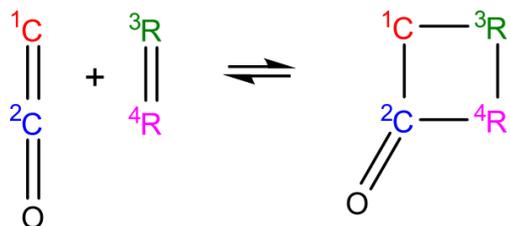
### 1,4\_Cyclic\_birad\_scission



### 1,4\_Linear\_birad\_scission



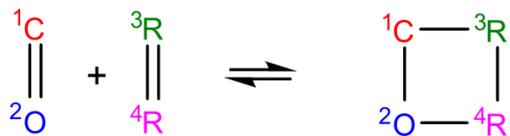
### 2+2\_cycloaddition\_CCO



### 2+2\_cycloaddition\_Cd



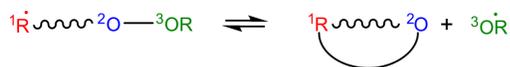
### 2+2\_cycloaddition\_CO



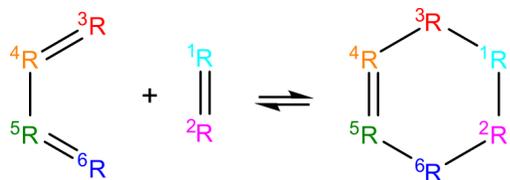
### Birad\_recombination



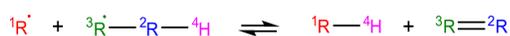
### Cyclic\_Ether\_Formation



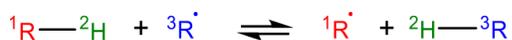
### Diels\_alder\_addition



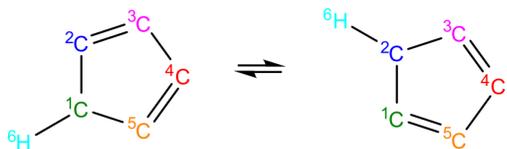
### Disproportionation



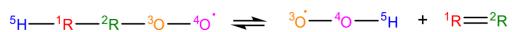
### H\_Abstraction



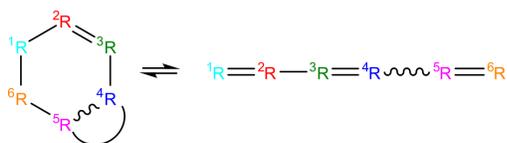
H\_shift\_cyclopentadiene



HO2\_Elimination\_from\_PeroxyRadical



Intra\_Diels\_alder



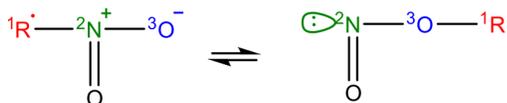
Intra\_Disproportionation



intra\_H\_migration



intra\_NO2\_ONO\_conversion



intra\_OH\_migration



Intra\_R\_Add\_Endocyclic



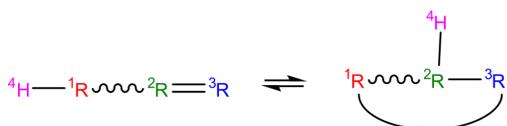
Intra\_R\_Add\_Exocyclic



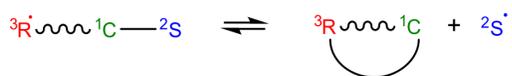
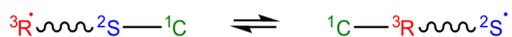
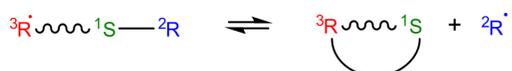
Intra\_R\_Add\_ExoTetCyclic

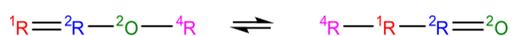
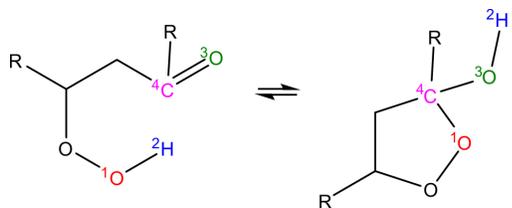
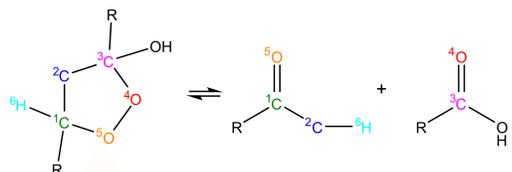
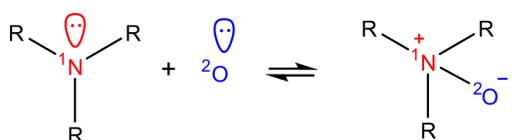


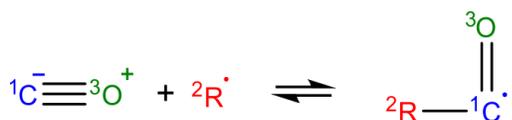
Intra\_RH\_Add\_Endocyclic

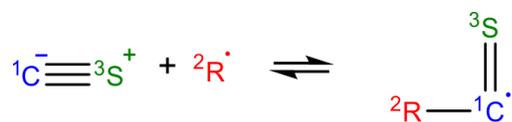
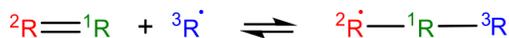


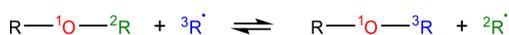
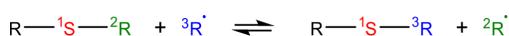
Intra\_RH\_Add\_Exocyclic


**intra\_substitutionCS\_cyclization**

**intra\_substitutionCS\_isomerization**

**intra\_substitutionS\_cyclization**

**intra\_substitutionS\_isomerization**

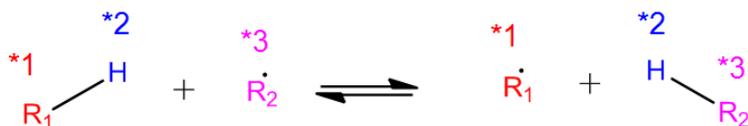
**ketoenol**

**Korcek\_step1**

**Korcek\_step2**

**lone\_electron\_pair\_bond**

**Oa\_R\_Recombination**

**R\_Addition\_COM**

**R\_Addition\_CSm**


**R\_Addition\_MultipleBond**

**R\_Recombination**

**Substitution\_O**

**SubstitutionS**

**Recipe**

The recipe can be found near the top of groups.py and describes the changes in bond order and radicals that occur during the reaction. Reacting atoms are labelled with a starred number. Shown below is the recipe for the H-abstraction family.


**Recipe**

- (1) BREAK\_BOND  $\{^*1, \text{S}, ^*2\}$
- (2) FORM\_BOND  $\{^*2, \text{S}, ^*3\}$
- (3) GAIN\_RADICAL  $\{^*1, 1\}$
- (4) LOSE\_RADICAL  $\{^*3, 1\}$

The table below shows the possible actions for recipes. The arguments are given in the curly braces as shown above. For the order of bond change in the Change\_Bond action, a -1 could represent a triple bond changing to a double bond while a +1 could represent a single bond changing to a double bond.

Action	Argument1	Argument2	Argument3
Break_Bond	First bonded atom	Type of bond	Second bonded atom
Form_Bond	First bonded atom	Type of bond	Second bonded atom
Change_Bond	First bonded atom	Order of bond change	Second bonded atom
Gain_Radical	Specified atom	Number of radicals	
Lose_Radical	Specified atom	Number of radicals	

Change\_Bond order cannot be directly used on benzene bonds. During generation, aromatic species are kekulized to alternating double and single bonds such that reaction families can be applied. However, RMG cannot properly handle benzene bonds written in the kinetic group definitions.

## Training Set vs Rules

The training set and rules both contain trusted kinetics that are used to fill in templates in a family. The **training set** contains kinetics for specific reactions, which are then matched to a template. The kinetic **rules** contain kinetic parameters that do not necessarily correspond to a specific reaction, but have been generalized for a template.

When determining the kinetics for a reaction, a match for the template is searched for in the kinetic database. The three cases in order of decreasing reliability are:

1. Reaction match from training set
2. Node template exact match using either training set or rules
3. Node template estimate averaged from children nodes

The reaction match from training set is accurate within the documented uncertainty for that reaction. A template exact match is usually accurate within about one order of magnitude. When there is no kinetics available for the template in either the training set or rules, the kinetics are averaged from the children nodes as an estimate. In these cases, the kinetic parameters are much less reliable. For more information on the estimation algorithm see *Kinetics Estimation*.

The training set can be modified in `training.py` and the rules can be modified in `rules.py`. For more information on modification see *Adding Training Reactions* and *Adding Kinetic Rules*.

### 1.11.4 Database Modification

Note that the RMG-Py database is written in Python code where line indentions determine the scope. When modifying the database, be sure to preserve all line indentions shown in the examples.

#### Modifying the Thermo Database

##### Creating Thermo Libraries

##### Adding Thermo Groups

##### Adding Thermo to the Depository

#### Modifying the Kinetics Database

For the casual user, it is recommended to use either a kinetic library or add to the training set instead of modifying the kinetic groups.

Put kinetic parameters into a kinetic library when:

- A set of reaction rates were optimized together
- You know the reaction rate is not generalizable to similar species (perhaps due to catalysis or aromatic structures)
- No family exists for the class of reaction
- You are not confident about the accuracy of kinetic parameters

Put kinetic parameters into the training set when:

- You are confident on the accuracy of the kinetic parameter
- You wish for the reaction to be generalized to similar reactions in your mechanism

## Creating Kinetics Libraries

### Adding New Kinetic Groups and Rate Rules

#### Decide on a Template

First you need to know the template for your reaction to decide whether or not to create new groups:

1. Type your reaction into the kinetics search at <http://rmg.mit.edu/database/kinetics/search/>
2. Select the correct reaction
3. In the results search for “(RMG-Py rate rules)” and select that link. The kinetic family listed is the family of interest.
4. Scroll to the bottom and look at the end of the long description. There may be very long description of the averaging scheme, but the template for the reaction is the very last one listed:

**Reference:**

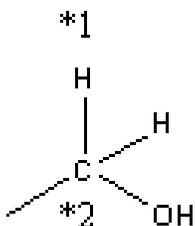
**Reference link:** [http://dev.rmg.mit.edu/database/kinetics/families/H\\_Abstraction/rules](http://dev.rmg.mit.edu/database/kinetics/families/H_Abstraction/rules)

**Reference type:**

**Short description:** Estimated by RMG-Py rate rules

**Long description:** Average of (O\_pri,C\_methyl + Average of (O/H/NonDeC,C\_methyl + O/H/OneDe,C\_methyl). + Average of (O\_pri,C\_rad/H2/Cs). + Average of (O/H/NonDeC,C\_rad/H2/Cs + Average of (H2O2,lnChI=1/C4H9O/c1-2-3-4-5/h5H,1-4H2 + H2O2,lnChI=1/C4H9O/c1-3-4(2)5/h4-5H,2-3H2,1H3 + H2O2,lnChI=1/C4H9O/c1-3-4(2)5/h4-5H,1,3H2,2H3 + H2O2,lnChI=1/C4H9O/c1-4(2,3)5/h5H,1H2,2-3H3). + Average of (H2O2,lnChI=1/C3H5/c1-3-2/h3H,1-2H2). + Average of (ROOH\_pri,C\_rad/H2/CO + ROOH\_sec,C\_rad/H2/CO). ). + Average of (O/H/NonDeC,C\_rad/H/NonDeC + Average of (H2O2,lnChI=1/C4H9O/c1-2-3-4-5/h2,5H,3-4H2,1H3 + H2O2,lnChI=1/C4H9O/c1-2-3-4-5/h3,5H,2,4H2,1H3 + H2O2,lnChI=1/C4H9O/c1-3-4(2)5/h3-5H,1-2H3). + Average of (Average of (H2O2,lnChI=1/C4H9O/c1-2-3-4-5/h4-5H,2-3H2,1H3). ). + Average of (Average of (ROOH\_pri,C\_rad/H/CO/Cs). + Average of (ROOH\_sec,C\_rad/H/CO/Cs). ). ). + Average of (Average of (O/H/NonDeC,C\_rad/Cs3). + Average of (Average of (Average of (H2O2,lnChI=1/C4H9O/c1-3-4(2)5/h5H,3H2,1-2H3). ). + Average of (Average of (ROOH\_pri,C\_rad/OOH/Cs/Cs). ). + Average of (Average of (ROOH\_sec,C\_rad/OOH/Cs/Cs). ). ). ). Estimated using template (O\_H,Cs\_rad) for rate rule (O\_pri,C\_rad/Cs3)

Now you must determine whether the chosen template is appropriate. A good rule of thumb is to see if the all neighbours of the reacting atoms are as specified as possible. For example, assume your species is ethanol



and RMG suggests the group:

```
label = "C_sec",
group =
"""
1 *1 Cs  0 {2,S} {3,S} {4,S}
2 *2 H   0 {1,S}
3   R!H 0 {1,S}
4   R!H 0 {1,S}
""",
```

If you use the suggested groups you will not capture the effect of the alcohol group. Therefore it is better to make a new group.

```
label = "C/H2/Cs0",
group =
"""
1 *1 Cs  0 {2,S} {3,S} {4,S} {5,S}
2 *2 H   0 {1,S}
3   H   0 {1,S}
4   O   0 {1,S}
5   Cs  0 {1,S}
""",
```

If you have determined the suggested groups is appropriate, skip to [Adding Training Reactions](#) or [Adding Kinetic Rules](#). Otherwise proceed to the next section for instructions on creating the new group.

### Creating a New Group

In the family's groups.py, you will need to add an entry of the format:

```
entry(
    index = 61,
    label = "C_sec",
    group =
    """
1 *1 Cs  0 {2,S} {3,S} {4,S} {5,S}
2 *2 H   0 {1,S}
3   C   0 {1,S}
4   H   0 {1,S}
5   R!H 0 {1,S}
""",
    kinetics = None,
    reference = None,
    referenceType = "",
    shortDesc = u"",
    longDesc = u"",
)
```

- The index can be any number not already present in the set
- The label is the name of the group.
- The group is the group adjacency list with the starred reacting atoms.
- The other attributes do not need to be filled for a group

Next, you must enter your new group into the tree. At the bottom of `groups.py` you will find the trees. Place your group in the appropriate position. In the example given in the previous section, the new group would be added under the `C_sec`.

```
L1: X_H
    L2: H2
    L2: Cs_H
        L3: C_pri
        L3: C_sec
            L4: C/H2/Cs0
                L3: C_ter
```

### Adding Kinetic Rules

Rules give generalized kinetic parameters for a specific node template. In most cases, your kinetic parameters describe a specific reaction in which case you will want to add your reaction to the training set.

The rule must be added into `rules.py` in the form:

```
entry(
    index = 150,
    label = "C/H/Cs3;0_rad/NonDe0",
    group1 =
    """
1 *1 Cs 0 {2,S} {3,S} {4,S} {5,S}
2 *2 H 0 {1,S}
3   Cs 0 {1,S}
4   Cs 0 {1,S}
5   Cs 0 {1,S}
    """,
    group2 =
    """
1 *3 0 1 {2,S}
2   0 0 {1,S}
    """,
    kinetics = ArrheniusEP(
        A = (2800000000000.0, 'cm^3/(mol*s)', '*|/', 5),
        n = 0,
        alpha = 0,
        E0 = (16.013, 'kcal/mol', '+|-', 1),
        Tmin = (300, 'K'),
        Tmax = (1500, 'K'),
    ),
    reference = None,
    referenceType = "",
    rank = 5,
    shortDesc = u""Curran et al. [8] Rate expressions for H atom abstraction from fuels.""",
    longDesc =
    u""
    [8] Curran, H.J.; Gaffuri, P.; Pittz, W.J.; Westbrook, C.K. Combust. Flame 2002, 129, 253.
    Rate expressions for H atom abstraction from fuels.
```

pg 257 A Comprehensive Modelling Study of iso-Octane Oxidation, Table 1. Radical:H02, Site: tertiary (c)

Verified by Karma James

```
"""
)

```

- The index can be any number not already used in rules.py.
- The label is the name of the rule.
- The groups must have the adjacency list of the respective groups. Between them they should have all starred atoms from the recipe.
- **The value and units of kinetic parameters must be given.**
  - Multiplicative uncertainty is given as ' $\ast|/$ ', 5 meaning within a factor of 5
  - Additive uncertainty is given as ' $+|/-$ ', 2 meaning plus or minus 2.
- Rank determines the priority of the rule when compared with other rules.
- The short description will appear in the annotated chemkin file.
- The long description only appears in the database.

### Adding Training Reactions

If you know the kinetics of a specific reaction, rather than a rate rule for a template, you can add the kinetics to the database training set. By default, RMG creates new rate rules from this training set, which in turn benefits the kinetics of similar reactions. The new rate rules are formed by matching the reaction to the most most specific template nodes within the reaction's respective family. If you do not want the training depository reactions to create new rate rules in the database, set the option for `kineticsDepositories` within the `database` field in your input file to

```
kineticsDepositories = ['!training'],
```

Currently, RMG's rate rule estimates overrides all kinetics depository kinetics, including training reactions. Unless the training reaction's rate rule ranks higher than the existing node, it will not be used. If you want the training reaction to override the rate rule estimates, you should put the reaction into a reaction library or seed mechanism.

The easiest way to add training reactions to the database is via the RMG website. First, search for the reaction using <http://rmg.mit.edu/database/kinetics/search/>. This will automatically search the existing RMG database for the reaction, as well as identify the reaction family template that this reaction matches. If the reaction does not match any family, then it cannot be added to the training reactions. Click the 'Create training rate from average' button underneath the kinetics plot for the reaction and edit the kinetics and reference descriptions for the reaction. The atom labels marking the reaction recipe actions (lose bond, add radical, etc.) will already be automatically labeled for you. After editing the reaction data, write a short message for the reaction added under the 'Summary of changes' field, then click 'Save.' You will need an account for the RMG website to make an entry.

**Note:** If you are entering the reaction in the reverse direction of the family, you must still label the reactants and products with the `atomLabels` of the original reaction template. Otherwise, RMG will not be able to locate the nodes in the group tree to match the reaction.

Entries added in the reverse direction of the original template will use the current RMG job's thermo database to estimate the kinetics in the forward direction. Therefore this value can differ depending on the order of thermo libraries used when running a job.

If adding the training reaction manually, first identify the reaction family of the reaction, then go to the family's folder in `RMG-database/input/kinetics/families/`. Create a new kinetics entry in the `training.py` file. Make sure to apply the reaction recipe labels properly for the reactants and products.

### Pitfalls

Be careful with the specificity when naming neighbouring atoms. On upper nodes, you should try to be general so that you do not exclude reactions.

Sibling nodes must be exclusive from one another so that there is no question which group a molecule qualifies as. However, you do not need to be exhaustive and list out every possibility.

Make sure your nodes are actually children of their parents. Currently RMG does no atom-by-atom checking and assumes whatever is put into the tree is correct.

Be sure to give errors whenever adding rules. If you don't know the uncertainty, why do you trust the kinetics?

After you are done always check via populate reactions or the website, that your modifications are behaving the way you expect.

Caveat regarding how rate rules are used by RMG and the rate parameters you input: because tunneling is important for many chemical reactions, the rate of a reaction may not be easily represented by a bi-Arrhenius fit. 3-parameter fits are more common. However, the resulting fit may report an 'activation energy' that is much different (possibly by 10+ kcal) than the true barrier height. When RMG is assembling pressure-dependent networks, it will use barrier heights from rate rules. This can lead to very inaccurate rate calculations. To avoid this issue, try to ensure that your fitted arrhenius activation energy truly does reflect the reaction barrier height.

## 1.12 Thermochemistry Estimation

This section gives in-depth descriptions of the methods used for determining thermochemistry of species.

Thermochemistry of species is obtained via three possible ways:

1. Species thermochemistry libraries
2. Group contribution methods
3. On-the-fly Quantum-chemical calculation of Thermochemical Properties (QMTP)

### 1.12.1 Species thermochemistry libraries

These databases contain thermochemical parameters for species. In these databases each entry contains an unambiguous definition of the species (through the adjacency list representation), along with a values for the thermochemistry in a format that allows the evaluation of each thermodynamic variable as a function of temperature.

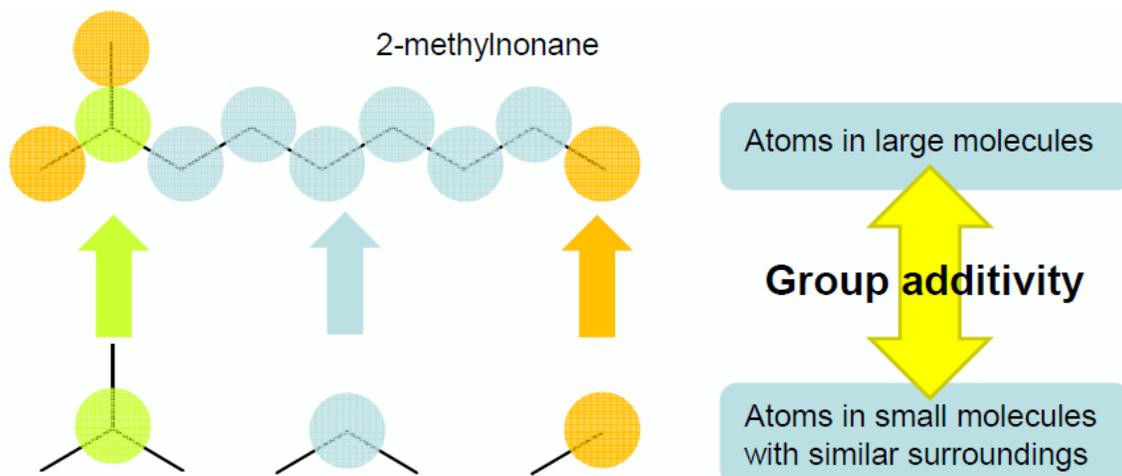
RMG is shipped with a number of species thermochemistry libraries, located in the 'libraries' folder of RMG-database. More information on these species thermochemistry libraries can be found in *Thermo Database*.

### 1.12.2 Group contribution methods

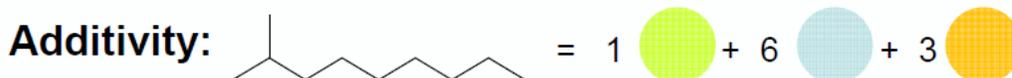
When the thermochemistry of a species is not present in one of the available species thermochemistry libraries, RMG needs to estimate thermochemistry. One way to do so, is by using group contribution methods that estimate the thermochemistry of a molecule based on the sub-molecular fragments present in the molecule. The Benson group additivity framework is such an example of a group contribution method that has proven to provide accurate estimates of the ideal gas thermochemistry for a large range of molecules.

Benson's Group Additivity approach ([*Benson*]), divides a molecule into functional groups, and the contribution of each functional group to the overall thermochemistry is included. For example, the molecule 2-methylnonane consists of three types of groups:

- 1 tertiary carbon atom
- 6 secondary carbon atoms
- 3 primary carbon atoms



**Group definition based on surroundings (ligands)**



Thermochemistry for the molecule X is calculated by summing up the values for each of the contributions  $C_i$ . E.g.:

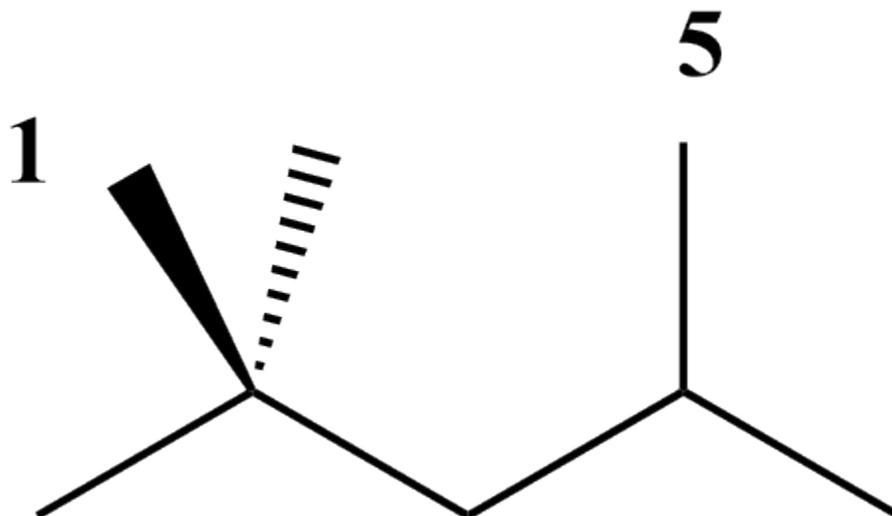
$$\Delta_f H_{298}^{\circ}(X) = \sum_i GAV(C_i)$$

The term 'group additive value' (GAV) denotes a polyvalent (ligancy > 1) monoatomic central atom  $C_i$  surrounded by its nearest-neighbor ligands.

Values for each central atomtype (e.g. "tertiary carbon atom") and its surrounding ligands can be found in the thermo group database, named `group.py`, of RMG. More information can be found here: [Thermo Database](#).

## NNIs

Besides the main group-centered (GAV) contributions, non-next-nearest neighbor interactions (NNI) may also be important to take into account. NNIs are interactions between atoms separated by at least 2 atoms, such as alkane 1,4-gauche, alkane 1,5 (cf. figure), alkene 1,4-gauche, alkene single and double cis, ene-yne cis and ortho interactions.



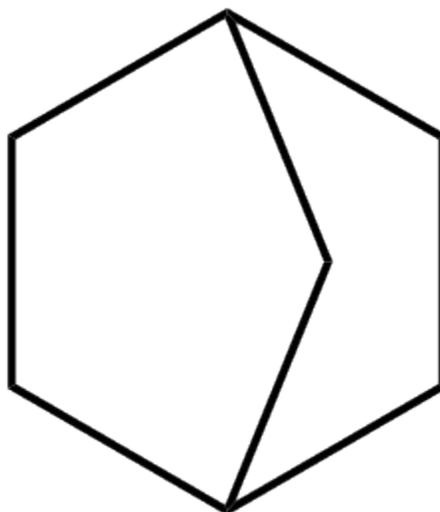
As a result, thermochemistry of the molecule X is determined as :

$$\Delta_f H_{298}^{\circ}(X) = \sum_i GAV(C_i) + \sum_j NNI_j$$

RMG contains a database with NNIs, named `gauche.py` and `int15.py`. More information on the nature on the available NNIs, and corresponding values can be found here: [Thermo Database](#).

## Ring Strain

To account for ring strain, ring strain corrections (RSC) were introduced. Because there is no obvious relation between the RSC and the ring structure, a specific RSC is required for every type of ring. For example, due to the significant ring strain induced in norbornane (cf. figure), a ring correction (RSC) needs to be added to the sum of the GAVs of the individual carbon atoms:



As a result, thermochemistry of the molecule X is determined as :

$$\Delta_f H_{298}^{\circ}(X) = \sum_i GAV(C_i) + RSC$$

RMG contains a database with single-ring corrections, ‘ring.py’ and polycyclic ring corrections, ‘polycyclic.py’. More information on the nature on the available NNIs, and corresponding values can be found here: [Thermo Database](#).

## Hydrogen Bond Increment (HBI) method

Lay et al. [Lay] introduced the hydrogen bond increment (HBI) method to predict thermochemical properties of radicals. In contrast to Benson’s method, the HBI method does not use the group-additivity concept. The HBI enthalpy of formation of a radical ( $R^*$ ) is calculated from the enthalpy of formation of the corresponding parent molecule ( $R-H$ ) by adding a HBI to account for the loss of a hydrogen atom. Hence, for standard enthalpies of formation the HBI is defined as

$$HBI = \Delta_f H_{298}^o(R^*) - \Delta_f H_{298}^o(R-H) = BDE(R-H) - \Delta_f H_{298}^o(H^*)$$

with BDE the bond dissociation enthalpy of the R-H bond at the radical position. Similar expressions are valid for the entropy and heat capacity.

As a result the thermochemistry of the radical is calculated as follows:

$$\begin{aligned}\Delta_f H_{298}^o(R^*) &= HBI(\Delta_f H_{298}^o) + \Delta_f H_{298}^o(R-H) \\ C_p^o(R^*) &= HBI(C_p^o) + C_p^o(R-H) \\ S_{298}^o(R^*) &= HBI(S_{298}^o) + S_{298}^o(R-H)\end{aligned}$$

The HBI method is the default method use to estimate thermochemistry of radicals. Thus, the effect of resonance stabilization on the enthalpy of the radical will be accounted for through the corresponding HBI. For example, the HBI labeled as “C=CC=CCJ” will account for the resonance present in 1,4-pentadien-3-yl radical.

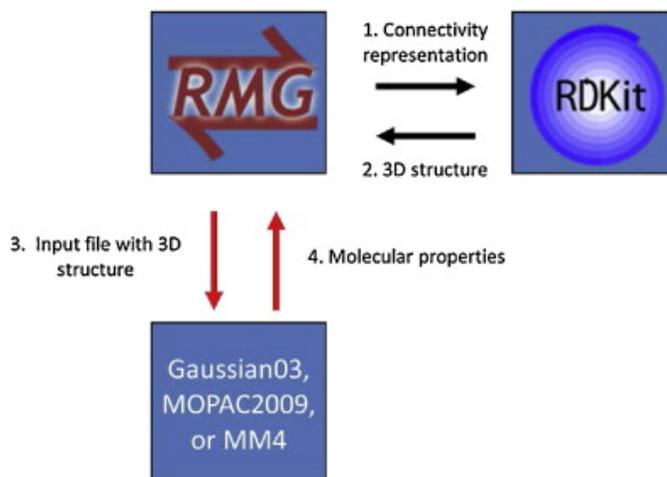
The HBI method can be applied to a variety of saturated compound thermochemistry values. In RMG, library values for saturated compounds are prioritized over group additivity values for saturated compounds. Note that if QMTP is on, the QM saturated value will get priority over group additivity but library value will have priority over QM value. This ensures that there is a systematic HBI correction for values used in the final model: if the saturated molecule thermo uses a library as a source, the radical thermo applies the HBI correction to that same library value.

RMG contains a database for with HBIs, named radical.py. More information on the nature on the available HBIs, and corresponding values can be found here: [Thermo Database](#).

### 1.12.3 On-the-fly Quantum-chemical calculation of Thermochemical Properties (QMTP)

An interface for performing on-the-fly quantum and force field calculations has been developed and integrated into RMG to complement the species thermochemistry databases and group contribution methods [Magoon and Green]\_. This interface is particularly interesting for the estimation of thermochemistry of molecules that are not present in one of the species thermochemistry databases, and which cannot be estimated with sufficient accuracy using the Benson group additivity framework. This pertains specifically to polycyclic fused ring containing species, whose ring strain cannot be modeled using the available ring corrections in RMG’s ring strain correction databases.

The QMTP interface involves a number of steps, summarized in the figure below.



In a first step the connectivity representation is converted into a three-dimensional structure of the molecule through the generation of 3D coordinates for the atoms in the molecule. This is accomplished using a combination of a distance geometry method, followed by an optimization using the UFF force field available in RDKit [RDKit]. Next, an input file is created containing the 3D atomic coordinates along with a number of keywords. This file is sent to a computational chemistry package, either OpenMopac or Gaussian, that calculates the thermochemistry of the given molecule “on-the-fly”. The keywords specify the type of calculation, and the level-of-theory. Finally, the calculated thermochemistry data is sent back to RMG.

The QMTP calculation creates a folder ‘QMfiles’ that contains a number of files that are created during the process. The filename of these files is a combination of the InChI key of the molecule, and a specific filename extension, e.g.

WEEGYLXZBRQIMU-UHFFFAOYSA.out is the output file produced by the QM package for the molecule cineole (SMILES: CC12CCC(CC1)C(C)(C)O2), represented by the InChI key WEEGYLXZBRQIMU-UHFFFAOYSA.

The table belows shows an overview of the used file extensions and their meaning.

File extension	Meaning
.mop	MOPAC input
.out	MOPAC output
.gjf	Gaussian input
.log	Gaussian output
.arc	MOPAC input created by MOPAC
.crude.mol	Mol file using crude, unrefined
.refined.mol	Mol file using UFF refined geometry
.symm	SYMMETRY input
.thermo	thermochemistry output file

For efficiency reasons, RMG minimizes the number of QMTP calculations. As a result, prior to initializing a QMTP routine, RMG checks whether the output files of a specific QMTP calculation are not already present in the QMfiles folder. It does so by comparing the InChI key of the given species to the filenames of the files in the QMfiles folder. If none of the InChI keys of the files correspond to the InChI key of the given species, RMG will initiate a new QMTP calculation.

### Supported QM packages, and levels of theory

The following table shows an overview of the computational chemistry packages and levels of theory that are currently supported in the QMTP interface of RMG.

The MM4 force field software originates from Allinger and Lii. [Allinger].

QM Package	Supported Levels of Theory
OpenMopac	semi-empirical (PM3, PM6, PM7)
Gaussian03	semi-empirical (PM3)
MM4	molecular mechanics (MM4)

## 1.12.4 Symmetry and Chirality

### Symmetry

The notion of symmetry is an essential part of molecules. Molecular symmetry refers to the indistinguishable orientations of a molecule. This is macroscopically quantified as a decrease of the entropy  $S$  by a term  $-R * \ln(\sigma)$  with  $R$  the universal gas constant and  $\sigma$  the global symmetry number, corresponding to the number of indistinguishable orientations of the molecule.

In RMG,  $\sigma$  is calculated as the product of contributions of three symmetry center types : atoms, bonds and axes, cf. below.

$$\sigma = \prod_i \sigma_{atom,i} \cdot \prod_j \sigma_{bond,j} \cdot \prod_k \sigma_{axis,k}$$

More information can be found in the Ph.D Thesis of Joanna Yu [Yu].

For molecules whose thermochemistry is calculated through group contribution techniques, the rotational symmetry number is calculated through graph algorithms of RMG based on the above equation. If the thermochemistry is calculated through the QMTP process, the external, rotational symmetry number is calculated using the open-source software SYMMETRY “Brute Force Symmetry Analyzer” [Patchkovskii]. This program uses the optimized three-dimensional geometry and calculates the corresponding point group.

### Chirality

RMG does not take stereochemistry into account, effectively assuming a racemic mixture of mirror image enantiomers. As a result, a chirality contribution of  $+R * \ln(2)$  is included in the entropy of the molecule.

Chirality for molecules whose thermochemistry is determined using group contribution techniques is detected using graph algorithms similar to those used for determining the symmetry number. If the thermochemistry is calculated through the QMTP process, chirality is detected using the point group information obtained via the software SYMMETRY.

Chiral molecules belong to point groups that lack a superposable mirror image (i.e. point groups lacking  $\sigma_h$ ,  $\sigma_d$ ,  $\sigma_v$ , and  $S_n$  symmetry elements).

## 1.12.5 References

## 1.13 Kinetics Estimation

This section gives in-depth descriptions of algorithms used for determining kinetic parameters. For general usage of the kinetic database see [Kinetics Database](#).

### 1.13.1 Priority of Kinetic Databases

When multiple sources are available for kinetic parameters, the following priority is followed:

1. Seed mechanisms (based on listed order in input.py)

2. Reaction libraries (based on listed order in input.py)
3. Matched training set reactions
4. Exact template matches from rules or matched training groups (based on rank)
5. Estimated averaged rules

In the case where multiple rules or training set reactions fall under the same template node, we use a user-defined rank to determine the priority of kinetic parameters

Rank	Example methods
Rank 1	Experiment
Rank 2	High level calculation: CCSD(T)-F12, MS-Tor, 2d rotors
Rank 3	Mid level calculation: CBS-QB3, 1-d rotors
Rank 4	Low level calculation: no rotor treatment
Rank 5	User's estimates without supporting methods
Rank 10	Averaged value from child nodes' rate rules
Rank 0	Untrusted method and never used in generation

The rank of 0 is assigned to kinetics that are generally default values for top level nodes that we have little faith in. It is never used in generation and its value will in fact be overridden by averages of its child nodes, which generates an averaged rate rule with rank 10.

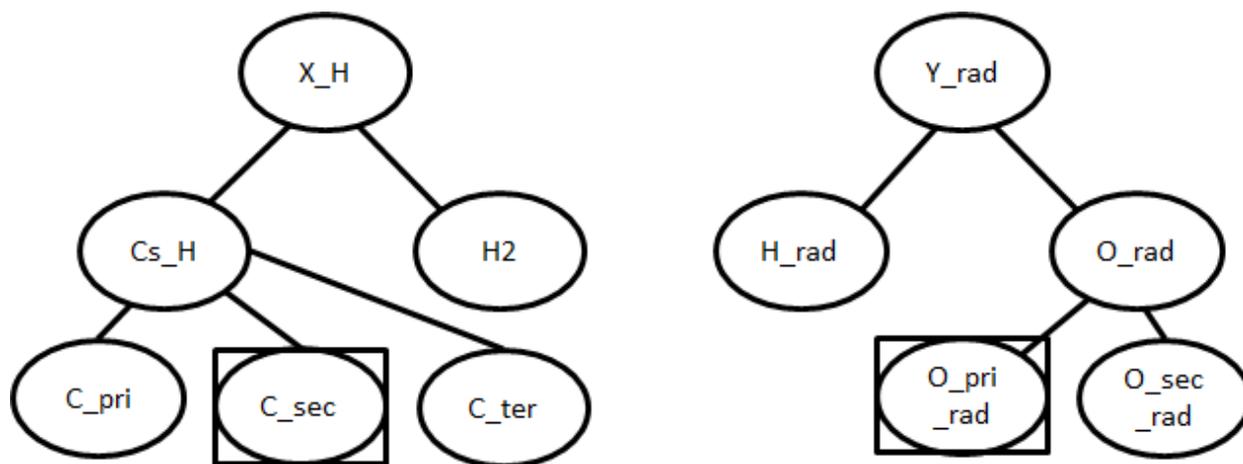
Only non-zero rules are used in generation. A rank of 1 is assigned to the most trustworthy kinetics, while a rank of 10 is considered very poor (ie. averaged kinetics). Thus, a rate rule of rank 3 will be given priority over a rate rule of rank 5.

### 1.13.2 Kinetic Families

To show the algorithm used by kinetic families, the following H-abstraction will be used as an example

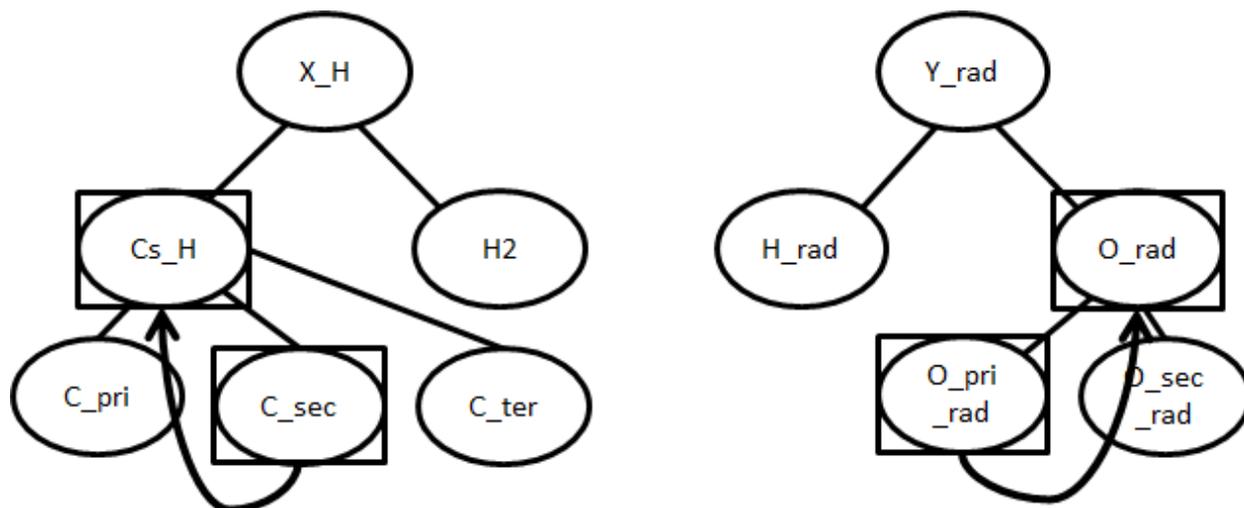


First the reacting atoms will be identified. Then, the family's trees will be descended as far as possible to give the reaction's groups.



Using the sample tree shown above, the desired template is (C\_sec, O\_pri\_rad). The algorithm will then search the database for parameters for the template. If they are present, an exact match will be returned using the kinetics of that template. Note that an exact match refers to the nodes (C\_sec, O\_pri\_rad) and not the molecules (propane, OH).

There may not be an entry for (C\_sec, O\_pri\_rad) in the database. In that case, the rule will attempt to “fall up” to more general nodes:



Now the preferred rule is (Cs\_H, O\_rad). If database contains parameters for this, those will be returned as an estimated match.

If there is still no kinetics for the template, the entire set of children for Cs\_H and O\_rad will be checked. For this example, this set would include every combination of {C\_pri, C\_sec, C\_ter} with {O\_pri\_rad, O\_sec\_rad}. If any these templates have kinetics, an average of their parameters will be returned as an estimated match. The average for  $A$  is a geometric mean, while the average for  $n$ ,  $E_a$ , and  $\alpha$  are arithmetic means.

If there are still no “sibling” kinetics, then the groups will continue to fall up to more and more general nodes. In the worst case, the root nodes may be used.

A *Full List of the Kinetics Families* in RMG is available.

## 1.14 Liquid Phase Systems

To simulate liquids in RMG requires a module in your input file for liquid-phase:

```
solvation(
    solvent='octane'
)
```

Your reaction system will also be different (liquidReactor rather than simpleReactor):

```
liquidReactor(
    temperature=(500, 'K'),
    initialConcentrations={
        "octane": (6.154e-3, 'mol/cm^3'),
        "oxygen": (4.953e-6, 'mol/cm^3')
    },
    terminationTime=(5, 's'),
    sensitivity=['octane', 'oxygen'],
    sensitivityThreshold=0.001,
)
```

For sensitivity analysis, RMG-Py must be compiled with the DASPK solver, which is done by default but has some dependency restrictions. (See *License Restrictions on Dependencies* for more details.) Like for the simpleReactor, the

sensitivity and sensitivityThreshold are optional arguments for when the user would like to conduct sensitivity analysis with respect to the reaction rate coefficients for the list of species given for sensitivity.

Sensitivity analysis is conducted for the list of species given for sensitivity argument in the input file. The normalized concentration sensitivities with respect to the reaction rate coefficients  $\text{dln}(C_i)/\text{dln}(k_j)$  are saved to a csv file with the file name sensitivity\_1\_SPC\_1.csv with the first index value indicating the reactor system and the second naming the index of the species the sensitivity analysis is conducted for. Sensitivities to thermo of individual species is also saved as semi normalized sensitivities  $\text{dln}(C_i)/\text{d}(G_j)$  where the units are given in  $1/(\text{kcal mol}^{-1})$ . The sensitivityThreshold is set to some value so that only sensitivities for  $\text{dln}(C_i)/\text{dln}(k_j) > \text{sensitivityThreshold}$  or  $\text{dln}C_i/\text{d}(G_j) > \text{sensitivityThreshold}$  are saved to this file.

Note that in the RMG job, after the model has been generated to completion, sensitivity analysis will be conducted in one final simulation (sensitivity is not performed in intermediate iterations of the job).

### 1.14.1 Equation of state

Specifying a liquidReactor will have two effects:

1. disable the ideal gas law renormalization and instead rely on the concentrations you specified in the input file to initialize the system.
2. prevent the volume from changing when there is a net stoichiometry change due to a chemical reaction ( $A = B + C$ ).

### 1.14.2 Solvation thermochemistry

The next correction for liquids is solvation effects on the thermochemistry. By specifying a solvent in the input file, we load the solvent parameters to use.

The free energy change associated with the process of transferring a molecule from the gas phase to the solvent phase is defined as the free energy of solvation ( $\Delta G$ ). Many different methods have been developed for computing solvation energies among which continuum dielectric and force field based methods are popular. Not all of these methods are easy to automate, and many are not robust i.e. they either fail or give unreasonable results for certain solute-solvent pairs. CPU time and memory (RAM) requirements are also important considerations. A fairly accurate and fast method for computing  $\Delta G$ , which is used in RMG, is the LSER approach described below.

#### Use of thermo libraries in liquid phase system

As it is for gas phase simulation, thermo libraries listed in the input files are checked first to find thermo for a given species and return the first match. As it exists two types of thermo libraries, (more details on *thermo libraries*), thermo of species matching a library in a liquid phase simulation is obtained following those two cases:

If library is a “liquid thermo library”, thermo data are directly used without applying solvation on it.

If library is a “gas thermo library”, thermo data are extracted and then corrections are applied on it using the *LSER method* for this specific species-solvent system.

---

**Note:** Gas phase libraries can be declared first, liquid thermo libraries will still be tested first but the order will be respected if several liquid libraries are provided.

---

#### Use of Abraham LSER to estimate thermochemistry

The Abraham LSER provides an estimate of the the partition coefficient (more specifically, the log (base 10) of the partition coefficient) of a solute between the vapor phase and a particular solvent ( $K_{vs}$ ) (also known as gas-solvent

partition coefficient) at 298 K:

$$\log K_{vs} = c + eE + sS + aA + bB + lL \quad (1.1)$$

The Abraham model is used in RMG to estimate  $\Delta G$  which is related to the  $K_{vs}$  of a solute according to the following expression:

$$\begin{aligned} \Delta G &= -RT \ln K_{vs} \\ &= -2.303RT \log K_{vs} \end{aligned} \quad (1.2)$$

The variables in the Abraham model represent solute ( $E, S, A, B, V, L$ ) and solvent descriptors ( $c, e, s, a, b, v, l$ ) for different interactions. The  $sS$  term is attributed to electrostatic interactions between the solute and the solvent (dipole-dipole interactions related to solvent dipolarity and the dipole-induced dipole interactions related to the polarizability of the solvent) [Vitha2006], [Abraham1999], [Jalan2010]. The  $lL$  term accounts for the contribution from cavity formation and dispersion (dispersion interactions are known to scale with solute volume [Vitha2006], [Abraham1999]). The  $eE$  term, like the  $sS$  term, accounts for residual contributions from dipolarity/polarizability related interactions for solutes whose blend of dipolarity/polarizability differs from that implicitly built into the  $S$  parameter [Vitha2006], [Abraham1999], [Jalan2010]. The  $aA$  and  $bB$  terms account for the contribution of hydrogen bonding between the solute and the surrounding solvent molecules. H-bonding interactions require two terms as the solute (or solvent) can act as acceptor (donor) and vice versa. The descriptor  $A$  is a measure of the solute's ability to donate a hydrogen bond (acidity) and the solvent descriptor  $a$  is a measure of the solvent's ability to accept a hydrogen bond. A similar explanation applies to the  $bB$  term [Vitha2006], [Abraham1999], [Poole2009].

The solvent descriptors ( $c, e, s, a, b, l$ ) are largely treated as regressed empirical coefficients. Parameters are provided in RMG's database for the following solvents:

1. acetonitrile
2. benzene
3. butanol
4. carbontet
5. chloroform
6. cyclohexane
7. decane
8. dibutylether
9. dichloroethane
10. dimethylformamide
11. dimethylsulfoxide
12. dodecane
13. ethanol
14. ethylacetate
15. heptane
16. hexadecane
17. hexane
18. isooctane
19. nonane
20. octane

21. octanol
22. pentane
23. toluene
24. undecane
25. water

### Group additivity method for solute descriptor estimation

Group additivity is a convenient way of estimating the thermochemistry for thousands of species sampled in a typical mechanism generation job. Use of the Abraham Model in RMG requires a similar approach to estimate the solute descriptors ( $A$ ,  $B$ ,  $E$ ,  $L$ , and  $S$ ). Platts et al. ([Platts1999]) proposed such a scheme employing a set of 81 molecular fragments for estimating  $B$ ,  $E$ ,  $L$ ,  $V$  and  $S$  and another set of 51 fragments for the estimation of  $A$ . Only those fragments containing C, H and O are implemented in order to match RMG's existing capabilities. The value of a given descriptor for a molecule is obtained by summing the contributions from each fragment found in the molecule and the intercept associated with that descriptor.

### Mintz model for enthalpy of solvation

For estimating  $\Delta G$  at temperatures other than 298 K, the enthalpy change associated with solvation,  $\Delta H$  must be calculated separately and, along with  $\Delta S$ , assumed to be independent of temperature. Recently, Mintz et al. ([Mintz2007], [Mintz2007a], [Mintz2007b], [Mintz2007c], [Mintz2007d], [Mintz2008], [Mintz2008a], [Mintz2009]) have developed linear correlations similar to the Abraham model for estimating  $\Delta H$ :

$$\Delta H(298K) = c' + a'A + b'B + e'E + s'S + l'L \quad (1.3)$$

where  $A$ ,  $B$ ,  $E$ ,  $S$  and  $L$  are the same solute descriptors used in the Abraham model for the estimation of  $\Delta G$ . The lowercase coefficients  $c'$ ,  $a'$ ,  $b'$ ,  $e'$ ,  $s'$  and  $l'$  depend only on the solvent and were obtained by fitting to experimental data. In RMG, this equation is implemented and together with  $\Delta G(298\text{ K})$  can be used to find  $\Delta S(298\text{ K})$ . From this data,  $\Delta G$  at other temperatures is found by extrapolation.

### 1.14.3 Diffusion-limited kinetics

The next correction for liquid-phase reactions is to ensure that bimolecular reactions do not exceed their diffusion limits. The theory behind diffusive limits in solution phase reactions is well established ([Rice1985]) and the effective rate constant of a bimolecular reaction is given as:

$$k_{\text{eff}} = \frac{4\pi RDk_{\text{int}}}{4\pi RD + k_{\text{int}}} \quad (1.4)$$

where  $k_{\text{int}}$  is the intrinsic reaction rate,  $R$  is the sum of radii of the reactants and  $D$  is the sum of the diffusivities of the reacting species. RMG uses the McGowan method for estimating radii, and diffusivities are estimated with the Stokes-Einstein equation using experimental solvent viscosities ( $\eta$  (T)). In a unimolecular to bimolecular reaction, for example, the forward rate constant ( $k_f$ ) can be slowed down if the reverse rate ( $k_{r,\text{eff}}$ ) is diffusion limited since the equilibrium constant ( $K_{\text{eq}}$ ) is not affected by diffusion limitations. In cases where both the forward and the reverse reaction rates are bimolecular, both diffusive limits are estimated and RMG uses the direction with the larger magnitude.

The viscosity of the solvent is calculated Pa.s using the solvent specified in the command line and a correlation for the viscosity using parameters  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$ :

$$\ln \eta = A + \frac{B}{T} + C \log T + DT^E \quad (1.5)$$

To build accurate models of liquid phase chemical reactions you will also want to modify your kinetics libraries or correct gas-phase rates for intrinsic barrier solvation corrections (coming soon).

#### 1.14.4 Example liquid-phase input file

This is an example of an input file for a liquid-phase system:

```
# Data sources
database(
    thermoLibraries = ['primaryThermoLibrary'],
    reactionLibraries = [],
    seedMechanisms = [],
    kineticsDepositories = ['training'],
    kineticsFamilies = 'default',
    kineticsEstimator = 'rate rules',
)

# List of species
species(
    label='octane',
    reactive=True,
    structure=SMILES("C(CCCCC)CC"),
)

species(
    label='oxygen',
    reactive=True,
    structure=SMILES("[O][O]"),
)

# Reaction systems
liquidReactor(
    temperature=(500,'K'),
    initialConcentrations={
        "octane": (6.154e-3,'mol/cm^3'),
        "oxygen": (4.953e-6,'mol/cm^3')
    },
    terminationTime=(5,'s'),
)

solvation(
    solvent='octane'
)

simulator(
    atol=1e-16,
    rtol=1e-8,
)

model(
    toleranceKeepInEdge=1E-9,
    toleranceMoveToCore=0.001,
    toleranceInterruptSimulation=0.1,
    maximumEdgeSpecies=100000
)

options(
```

```
units='si',
saveRestartPeriod=None,
generateOutputHTML=False,
generatePlots=False,
saveSimulationProfiles=True,
)
```

## 1.15 Guidelines for Building a Model

RMG has been designed to build kinetic models for gas phase pyrolysis and combustion of organic molecules made of C, H, O and S. By kinetic model, we mean a set of reactions and associated kinetics that represent the chemical transformations occurring in the system of interest. These systems could be the combustion of fuels, pyrolysis of hydrocarbon feedstocks, etc. The total number of reactions and species typically required to describe some of these processes can run into the thousands making these models difficult and error-prone to build manually. This is the main motivation behind using software like RMG that build such models automatically in a systematic reproducible manner.

In RMG, the user is expected to provide an input file specifying the conditions (temperature, pressure, etc.) under which one desires to develop kinetic models. The following are some tips for setting up your input/condition file.

### 1.15.1 Start with a good seed mechanism

RMG is a useful tool in elucidating important pathways in a given process but may not capture certain special reaction types which may be specific to the system you are interested in. However, if you already have a good idea of these reactions that are important and are not available in the standard RMG library, you can create a ‘seed mechanism’ and include it in the input file to RMG. This will directly include these in the model core and add other reactions from the RMG library on top of it using our rate based algorithm. (Similarly, you can specify your own thermodynamic parameters for species using thermochemistry libraries which are similar in concept to seed mechanisms. In order to build these libraries, you will need to specify all species in the RMG adjacency list format.) In a combustion system, RMG tends to do a decent job filling in the termination and propagation steps of a mechanism if it is guided with the initiation and chain branching steps using a seed mechanism. Ideally, RMG should be able to find all the right chemistry through our kinetics database but holes in current kinetic databases can make this task difficult. A good seed mechanism can address this issue for the system of interest and also reduce the size, cost and time taken to arrive at a converged model.

### 1.15.2 Setting up the right termination criterion

Start with a relatively large tolerance (such as 0.1) when building your first model to make sure that RMG can converge the model to completion without any hiccups, then begin tightening the tolerance if you are able to converge the initial model. For large molecules such as tetradecane (C<sub>14</sub>), even a tolerance of 0.1 may be too tight for RMG to work with and lead to convergence problems. Note that a good seed mechanism allows for faster convergence.

### 1.15.3 Restricting the number of carbon atoms, oxygen atoms, and radical sites per species

Options to tune the maximum number of carbon or oxygen atoms, or number of radical sites per species can be specified at the beginning of the condition file. In most systems, we do not expect large contributions from species with more than 1 radical center (i.e. biradicals, etc.) to affect the overall chemistry, thus it may be useful to limit the maximum number of radicals to 2 (to allow for O<sub>2</sub>). The same applies for the maximum number of oxygens you want to allow per species. Restricting the number of carbon atoms in each species may also be worthwhile to prevent very large molecules from being generated if many such species appear in your model. Using any of these options requires

some prior knowledge of the chemistry in your system. It is recommended that an initial model be generated without turning these options on. If many unlikely species show up in your model (or if your model has trouble converging and is generating many unlikely species on the edge), you can begin tuning these options to produce a better model.

#### 1.15.4 Adding key species into the initial condition file

Sometimes, chain branching reactions like dissociation of ROOH species do not make it to the core directly because if their fluxes are very small and the tolerance is not tight enough. In these cases, seeding the condition file with these species (with zero concentration) is helpful. By adding these species to the initial set of species in the condition file, the reactions involving those species will be automatically added to the core. (Putting these reactions in the seed mechanism has the same effect.) Thus, if a species is known to be a part of your system and RMG is having trouble incorporating it within your model, it should be added to the condition file with 0.0 set as the concentration.

#### 1.15.5 Starting with a single molecule when generating a model for a mixture

For modeling the combustion of fuel mixtures, you may want to start with determining their composition and starting with a kinetic study of the dominant compound. It is possible to model the combustion of fuel mixtures but they are more challenging as well as harder to converge in RMG because RMG will automatically generate all cross reactions between the reacting species and intermediates. Starting with single species is always a good idea and is also useful when thinking about fuel mixtures. In order to build a better background in chemical kinetic model development and validation, please look at a recent paper from our group on butanol combustion available [here](#). This should give you some idea about how RMG can be put to use for the species of interest to you.

### 1.16 Standalone Modules

There are several standalone modules that can be run separately from RMG. These scripts can be found in the `RMG-Py/scripts` folder. The database scripts are located in the `RMG-database/scripts` folder.

#### 1.16.1 Model Comparison

The script `diffModels` compares two RMG generated models to determine their differences. To use this method you will need the chemkin and species dictionary outputs from RMG. These can be found in the `chemkin` folder from the directory of the `input.py` file used for the RMG run. The syntax is as follows:

```
python diffModels.py CHEMKIN1 SPECIESDICT1 --thermo1 THERMO1 CHEMKIN2 SPECIESDICT2 --thermo2 THERMO2 --web
```

where CHEMKIN represents the chemkin input file (`chem00XX.inp`), SPECIESDICT is the species dictionary from RMG (`species_dictionary.txt`) and the optional `--thermo` flag can be used to add separate thermo CHEMKIN files THERMO. The numbers (1 and 2) represent which model to each file is from. The optional `--web` flag is used for running this script through the RMG-website.

Running the script without any optional flags looks like:

```
python diffModels.py CHEMKIN1 SPECIESDICT1 CHEMKIN2 SPECIESDICT2
```

Output of each comparison is printed, and the method then produces a html file (`diff.html`) for easy viewing of the comparison.

This method is also available to use with a web browser from the RMG website: [Model Comparison Tool](#).

## 1.16.2 Merging Models

This script combines up to 5 RMG models together. The thermo and kinetics from common species and reactions is taken from the first model with the commonality. To better understand the difference in two models, use `diffModels.py`. To use this method type:

```
python mergeModels.py --model1 chemkin1 speciesdict1 --model2 chemkin2 speciesdict2
```

where `chemkin` specifies the chemkin input file from the RMG run and `speciesdict` represents the species dictionary from the RMG run. These can be found in the `chemkin` folder from the directory of the `input.py` file used for the RMG run. The numbers are for different models that you want to merge. To merge more than two files, you can add `--model3 chemkin3 speciesdict3`. Up to 5 models can be merged together this way

Running this method will create a new species dictionary (`species_dictionary.txt`) and chemkin input file (`chem.inp`) in the parent directory of the terminal.

This method is also available to use with a web browser from the RMG website: [Model Merge Tool](#).

## 1.16.3 Generate Reactions

The script `generateReactions.py` generates reactions between all species mentioned in an input file. To call this method type:

```
python generateReactions.py Input_File
```

where `Input_File` is a file similar to a general RMG input file which contains all the species for RMG to generate reactions between. An example file is placed in `$RMGPy/examples/generateReactions/input.py`

```
# Data sources for kinetics
database(
    thermoLibraries = ['KlippensteinH2O2', 'primaryThermoLibrary', 'DFT_QCI_thermo', 'CBS_QB3_1dHR'],
    reactionLibraries = [],
    seedMechanisms = [],
    kineticsDepositories = 'default',
    #this section lists possible reaction families to find reactions with
    kineticsFamilies = ['!Intra_Disproportionation', '!Substitution_0'],
    kineticsEstimator = 'rate rules',
)

# List all species you want reactions between
species(
    label='ethane',
    reactive=True,
    structure=SMILES("CC"),
)

species(
    label='H',
    reactive=True,
    structure=SMILES("[H]"),
)

species(
    label='butane',
    reactive=True,
    structure=SMILES("CCCC"),
)
```

```

# you must list reactor conditions (though this may not effect the output)
simpleReactor(
    temperature=(650,'K'),
    pressure=(10.0,'bar'),
    initialMoleFractions={
        "ethane": 1,
    },
    terminationConversion={
        'butane': .99,
    },
    terminationTime=(40,'s'),
)

#optional module if you want to get pressure dependent kinetics.

#pressureDependence(
#    method='modified strong collision',
#    maximumGrainSize=(0.5,'kcal/mol'),
#    minimumNumberOfGrains=250,
#    temperatures=(300,2200,'K',2),
#    pressures=(0.01,100,'bar',3),
#    interpolation=('Chebyshev', 6, 4),
#    maximumAtoms=15,
#)

#optional module if you want to limit species produced in reactions.

#generatedSpeciesConstraints(
#    allowed=['input species','seed mechanisms','reaction libraries'],
#    maximumCarbonAtoms=4,
#    maximumHydrogenAtoms=10,
#    maximumOxygenAtoms=7,
#    maximumNitrogenAtoms=0,
#    maximumSiliconAtoms=0,
#    maximumSulfurAtoms=0,
#    maximumHeavyAtoms=20,
#    maximumRadicalElectrons=1,
#)

```

This method will produce an `output.html` file in the directory of `input.py` which contains the all the reactions produced between the species.

This method is also available to use with a web browser from the RMG website: [Populate Reactions](#).

### 1.16.4 Sensitivity Analysis

For sensitivity analysis, RMG-Py must be compiled with the DASP solver, which is done by default but has some dependency restrictions. (See [License Restrictions on Dependencies](#) for more details.) Sensitivity analysis can be conducted in a standalone system for an existing kinetics model in Chemkin format.

To use the sensitivity analysis standalone module:

```
python sensitivity.py input.py chem.inp species_dictionary.txt
```

where `chem.inp` is the CHEMKIN file and the `species_dictionary.txt` contains the dictionary of species associated with the CHEMKIN file. `input.py` is an input file similar to one used for an RMG job but does not generate a

RMG job. See the following `input.py` example file found under the `$RMGPy/examples/sensitivity/input.py` folder

```
# Data sources
database(
    thermoLibraries = ['primaryThermoLibrary'],
    reactionLibraries = [],
    seedMechanisms = [],
    kineticsDepositories = ['training'],
    kineticsFamilies = ['!Intra_Disproportionation', '!Substitution_0'],
    kineticsEstimator = 'rate rules',
)

# Constraints on generated species
generatedSpeciesConstraints(
    maximumRadicalElectrons = 2,
)

# List of species
species(
    label='ethane',
    reactive=True,
    structure=SMILES("CC"),
)

# Reaction systems
simpleReactor(
    temperature=(1350, 'K'),
    pressure=(1.0, 'bar'),
    initialMoleFractions={
        "ethane": 1.0,
    },
    terminationConversion={
        'ethane': 0.9,
    },
    terminationTime=(1e6, 's'),
    sensitivity=['ethane'],
    sensitivityThreshold=0.01,
)

simulator(
    atol=1e-16,
    rtol=1e-8,
    sens_atol=1e-6,
    sens_rtol=1e-4,
)

model(
    toleranceKeepInEdge=0.0,
    toleranceMoveToCore=0.1,
    toleranceInterruptSimulation=0.1,
    maximumEdgeSpecies=100000
)

options(
    units='si',
    saveRestartPeriod=None,
    saveSimulationProfiles=True,
    generateOutputHTML=False,
```

```
generatePlots=False,
)
```

The names of species named in the input file must coincide with the name specified in the CHEMKIN file.

Sensitivity analysis is conducted for the list of species given for `sensitivity` argument in the input file. The normalized concentration sensitivities with respect to the reaction rate coefficients  $\text{dln}(C_i)/\text{dln}(k_j)$  are saved to a csv file with the file name `sensitivity_1_SPC_1.csv` with the first index value indicating the reactor system and the second naming the index of the species the sensitivity analysis is conducted for. Sensitivities to thermo of individual species is also saved as semi normalized sensitivities  $\text{dln}(C_i)/\text{d}(G_j)$  where the units are given in  $1/(\text{kcal mol}^{-1})$ . The `sensitivityThreshold` is set to some value so that only sensitivities for  $\text{dln}(C_i)/\text{dln}(k_j) > \text{sensitivityThreshold}$  or  $\text{dln}C_i/\text{d}(G_j) > \text{sensitivityThreshold}$  are saved to this file.

### 1.16.5 Generating Flux Diagrams

The script, `generateFluxDiagrams.py`, will create a movie out of a completed RMG model that shows interconnected arrows between species that represent fluxes.

To use this method, you just need a completed RMG run. The syntax is as follows:

```
python generateFluxDiagram.py input.py
```

where `input.py` is the input file for the completed RMG run. The program will use the automatically generated file structure to find the other necessary files to create the movie.

This method is also available to use with a web browser from the RMG website: [Generate Flux Diagram](#).

### 1.16.6 Thermo Estimation Module

The thermo estimation module can be run stand-alone. An example input file for this module is shown below:

```
database(
    thermoLibraries = ['primaryThermoLibrary', 'GRI-Mech3.0']
)

species(
    label='Cineole',
    structure=SMILES('CC1CCC(CC1)C(C)(C)O2'),
)

quantumMechanics(
    software='mopac', #mopac or gaussian
    method='pm3', #pm3, pm6, pm7
    fileStore='QMfiles', # defaults to inside the output folder.
    onlyCyclics = True, #True, False
    maxRadicalNumber = 0, # 0, 1
)
```

The `database` block is used to specify species thermochemistry libraries. Multiple libraries may be created, if so desired. The order in which the thermo libraries are specified is important: If a species appears in multiple thermo libraries, the first instance will be used.

Please see Section *Thermo Database* for details on editing the thermo library. In general, it is best to leave the `ThermoLibrary` set to its default value. In particular, the thermodynamic properties for H and H2 must be specified in one of the primary thermo libraries as they cannot be estimated by Benson's method.

For example, if you wish to use the GRI-Mech 3.0 mechanism [*GRIMech3.0*] as a ThermoLibrary in your model, the syntax will be:

```
thermoLibraries = ['primaryThermoLibrary', 'GRI-Mech3.0']
```

This library is located in the `RMG-database/input/thermo/libraries` directory. All “Locations” for the ThermoLibrary field must be with respect to the `RMG-database/input/thermo/libraries` directory.

The optional `quantumMechanics` block is used when quantum mechanical calculations are desired to determine thermodynamic parameters. These calculations are only run if the molecule is not included in a specified thermo library. The `software` option accepts either the `mopac` or `gaussian` string. The `method` option refers to the level-of-theory, which can either be `pm3`, “`pm6`”, or `pm7`. A folder can be specified to store the files used in these calculations, however if not specified this defaults to a *QMfiles* folder in the output folder. The `onlyCyclics` option, if `True`, only runs these calculations for cyclic species. In this case, group contribution estimates are used for all other species. The calculations are also only run on species with a maximum radical number set by the user. If a molecule has a higher radical number, the molecule is saturated with hydrogen atoms, then quantum mechanical calculations with subsequent hydrogen bond incrementation is used to determine the thermodynamic parameters.

Submitting a job is easy:

```
python thermoEstimator.py input.py
```

We recommend you make a job-specific directory for each thermoEstimator simulation.

Note that the RMG website also provides thermo estimation through the [Molecule Search](#).

### 1.16.7 Convert FAME to CanTherm Input File

This module is utilized to convert FAME file types (used in RMG-Java) to CanTherm objects (used in RMG-Py) for pressure dependent calculations.

FAME is an early version of the `pdep` code in CanTherm written in Fortran and used by RMG-Java. This script enables importing FAME input files into CanTherm. Note that it is mostly designed to load the FAME input files generated automatically by RMG-Java, and may not load hand-crafted FAME input files. If you specify a *moleculeDict*, then this script will use it to associate the species with their structures.

```
python convertFAME.py fame_object
```

where `fame_object` is the FAME file used to be converted into the CanTherm object.

Some additional options involve adding an RMG dictionary to process with the file. The syntax for this is

```
python convertFAME.py -d RMG_dictionary.txt fame_object
```

where `RMG_dictionary.txt` is the dictionary to process with the file.

A max energy cutoff is also possible when converting the file formats.

```
python convertFAME.py -d RMG_dictionary.txt -x value units value units fame_object
```

where `value` represents the max energy amount and `units` represents its units

### 1.16.8 Database Scripts

This section details usage for scripts available in `RMG-database/scripts` folder.

### evansPolanyi.py

This script will generate an Evans-Polanyi plot for a single kinetics depository.

Usage:

```
python evansPolanyi.py [-h] DEPOSITORY
```

Positional arguments:

```
DEPOSITORY    the depository to use
```

Optional arguments:

```
-h, --help    show help message and exit
```

### exportKineticsLibraryToChemkin.py

This script exports an individual RMG-Py kinetics library to a chemkin and dictionary file. Thermo is taken from RMG's estimates and libraries. In order to use more specific thermo, you must tweak the thermoLibraries and estimators in use when loading the database. The script will save the chem.inp and species\_dictionary.txt files in the local directory.

Usage:

```
python exportKineticsLibrarytoChemkin.py [-h] LIBRARYNAME
```

Positional arguments:

```
LIBRARYNAME    the libraryname of the RMG-Py format kinetics library
```

Optional arguments:

```
-h, --help    show help message and exit
```

### exportOldDatabase.py

This script exports the database to the old RMG-Java format. The script requires two command-line arguments: the path to the database to import, and the path to save the old RMG-Java database to.

Usage:

```
python exportOldDatabase.py OUTPUT
```

Positional arguments:

```
OUTPUT    path to the directory where the RMG-Java database should be saved
```

### importChemkinLibrary.py

This script imports a chemkin file (along with RMG dictionary) from a local directory and saves a set of RMG-Py kinetics library and thermo library files. These py files are automatically added to the input/kinetics/libraries and input/thermo/libraries folder under the user-specified *name* for the chemkin library.

Usage:

```
python importChemkinLibrary.py [-h] CHEMKIN DICTIONARY NAME
```

Positional arguments:

CHEMKIN	The path of the chemkin file
DICTIONARY	The path of the RMG dictionary file
NAME	Name of the chemkin library to be saved

Optional arguments:

```
-h, --help show help message and exit
```

### importJavaKineticsLibrary.py

This script imports an individual RMG-Java kinetics library from a local directory and saves the output kinetics library py file into a path of the user's choosing. This library will be automatically added to the 'libraryname' folder in the input/kinetics/libraries directory and can be used directly as an RMG-Py kinetics library.

Usage:

```
python importJavaKineticsLibrary.py [-h] INPUT LIBRARYNAME
```

Positional arguments:

INPUT	the input path of the RMG-Java kinetics library directory
LIBRARYNAME	the libraryname for the RMG-Py format kinetics library

Optional arguments:

```
-h, --help show help message and exit
```

### importJavaThermoLibrary.py

This script imports an individual RMG-Java thermo library from a local directory and saves the output thermo library py file into a path of the user's choosing. This library will be automatically saved to libraryname.py in the input/thermo/libraries directory and can be used directly as an RMG-Py thermo library.

Usage:

```
python importJavaThermoLibrary.py [-h] INPUT LIBRARYNAME
```

Positional arguments:

INPUT	the input path of the RMG-Java thermo library directory
LIBRARYNAME	the libraryname for the RMG-Py format thermo library

Optional arguments:

```
-h, --help show help message and exit
```

### importOldDatabase.py

This script imports an RMG-Java database from the output directory and saves it in the input directory. Only recommended for use in extreme circumstances.

Usage:

```
python importOldDatabase.py [-h] INPUT OUTPUT
```

Positional arguments:

INPUT	the input path of the RMG-Java database directory
LIBRARYNAME	output path for the desired RMG-Py database directory

Optional arguments:

-h, --help	show help message and exit
------------	----------------------------

### 1.16.9 Standardize Model Species Names

This script enables the automatic renaming of species names of two or more Chemkin files (and associated species dictionaries) so that they use consistent, matching names. Simply pass the paths of the Chemkin files and species dictionaries on the command-line, e.g.:

```
python standardizeModelSpeciesNames.py --model1 /path/to/chem1.inp /path/to/species_dictionary1.txt --model2 /p
```

The resulting files are saved as chem1.inp and species\_dictionary1.txt, chem2.inp, species\_dictionary2.txt and so forth (depending on how many models you want to standardize) and will be saved in the execution directory.

## 1.17 Frequently Asked Questions

For any other questions related to RMG and its usage and installation, please post an issue at <https://github.com/ReactionMechanismGenerator/RMG-Py/issues> and the RMG developers will get back to you as soon as we can. You can also search for your problem on the issues page to see if there are already solutions in development. Alternatively, you can email us at [rmg\\_dev@mit.edu](mailto:rmg_dev@mit.edu).

### 1.17.1 Why can't my adjacency lists be read any more?

The adjacency list syntax changed in July 2014. The minimal requirement for most translations is to prefix the number of unpaired electrons with the letter *u*.

Example old syntax:

```
HXD13
1 C 0 {2,D}
2 C 0 {1,D} {3,S}
3 C 0 {2,S} {4,D}
4 C 0 {3,D} {5,S}
5 *1 C 0 {4,S} {6,S}
6 *2 C 0 {5,S}
```

Example new syntax:

```
HXD13
1 C u0 {2,D}
2 C u0 {1,D} {3,S}
3 C u0 {2,S} {4,D}
4 C u0 {3,D} {5,S}
5 *1 C u0 {4,S} {6,S}
6 *2 C u0 {5,S}
```

The new syntax, however, allows much greater flexibility, including definition of lone pairs, partial charges, wildcards, and molecule multiplicities, and was necessary to allow us to add Nitrogen chemistry. See `rmgpy.molecule.adjlist` for details of the new syntax.

### 1.18 Credits

RMG is based upon work supported by the Department of Energy, Office of Basic Energy Sciences through grant DE-FG02-98ER14914 and by the National Science Foundation under Grants No. 0312359 and 0535604.

Project Supervisors:

- Prof. William H. Green ([whgreen@mit.edu](mailto:whgreen@mit.edu))
- Prof. Richard H. West ([r.west@neu.edu](mailto:r.west@neu.edu))

Current Developers: ([rmg\\_dev@mit.edu](mailto:rmg_dev@mit.edu))

- Pierre L. Bhoorasingh
- Connie Gao
- Kehang Han
- Fariba Seyedzadeh Khanshan
- Belinda Slakman
- Dr. Nick M. Vandewiele
- Nathan Yee

Previous Developers:

- Dr. Joshua W. Allen
- Jacob Barlow
- Dr. Beat A. Buesser
- Dr. Caleb A. Class
- Victor Lambert
- Dr. Shamel S. Merchant
- Sean Troiano
- Dr. Aaron Vandeputte

### 1.19 How to Cite

Connie W. Gao, Joshua W. Allen, William H. Green, Richard H. West, “Reaction Mechanism Generator: automatic construction of chemical kinetic mechanisms.” *Computer Physics Communications* (2016).

- `genindex`
- `modindex`
- `search`

## CANTHERM USER'S GUIDE

### 2.1 Introduction

**CanTherm** is a tool for computing the thermodynamic properties of chemical species and high-pressure-limit rate coefficients for chemical reactions using the results of a quantum chemistry calculation. Thermodynamic properties are computed using the rigid rotor-harmonic oscillator approximation with optional corrections for hindered internal rotors. Kinetic parameters are computed using canonical transition state theory with optional tunneling correction.

CanTherm can also estimate pressure-dependent phenomenological rate coefficients  $k(T, P)$  for unimolecular reaction networks of arbitrary complexity. The approach is to first generate a detailed model of the reaction network using the one-dimensional master equation, then apply one of several available model reduction methods of varying accuracy, speed, and robustness to simplify the detailed model into a set of phenomenological rate coefficients. The result is a set of  $k(T, P)$  functions suitable for use in chemical reaction mechanisms.

#### 2.1.1 About CanTherm

CanTherm is written in the [Python](#) programming language to facilitate ease of development, installation, and use.

#### 2.1.2 License

CanTherm is provided as free, open source code under the terms of the [MIT/X11 License](#). The full, official license is reproduced below

```
Copyright (c) 2002-2015 Prof. William H. Green (whgreen@mit.edu),  
Prof. Richard H. West (r.west@neu.edu) and the RMG Team (rmg_dev@mit.edu)
```

```
Permission is hereby granted, free of charge, to any person obtaining a  
copy of this software and associated documentation files (the 'Software'),  
to deal in the Software without restriction, including without limitation  
the rights to use, copy, modify, merge, publish, distribute, sublicense,  
and/or sell copies of the Software, and to permit persons to whom the  
Software is furnished to do so, subject to the following conditions:
```

```
The above copyright notice and this permission notice shall be included in  
all copies or substantial portions of the Software.
```

```
THE SOFTWARE IS PROVIDED 'AS IS', WITHOUT WARRANTY OF ANY KIND, EXPRESS OR  
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,  
FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE  
AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER  
LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING
```

```
FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER  
DEALINGS IN THE SOFTWARE.
```

## 2.2 Installation

### 2.2.1 Dependencies

Python versions 2.6+ are recommended for CanTherm. Currently Python 3.x versions are incompatible with CanTherm.

CanTherm relies on a number of Python packages for certain functionality. The following lists the required Python packages that are not part of the Python standard library:

- **NumPy**. Provides efficient array and matrix operations.
- **SciPy**. Provides efficient linear algebra functions and special functions.
- **Cython v0.19+**. C-extensions for Python
- **RDKit**. Cheminformatics libraries and functions
- **Cairo**. Cairo graphics rendering for Python for drawing reaction networks
- **Quantities**. For converting between different scientific units
- **Argparse**. For parsing input arguments when running scripts

You can install these dependencies on Linux in the following fashion.

```
sudo apt-get install python-numpy python-scipy python-cairo  
sudo pip install cython>=0.19 quantities argparse
```

It is recommended that RDKit be installed manually with InChI capabilities on.

### 2.2.2 Installing CanTherm

Once you have obtained the required dependencies, CanTherm can be installed by first obtaining the RMG-Py source code by either downloading into a directory using git:

```
git clone git@github.com:ReactionMechanismGenerator/RMG-Py.git
```

or by downloading the zip file of the current RMG-Py master source code found [here](#) and unzipping into the appropriate directory.

Inside the root package directory for RMG-Py, execute the following make command:

```
make cantherm
```

The appropriate cythonization and compilation steps will now begin. When they are completed, you can run a test example by going into the examples folder to run a sample Cantherm job to verify the installation:

```
cd examples/cantherm/networks/acetyl+O2  
python ../../../../cantherm.py input.py
```

This will allow you to use the cantherm.py as a Python script anytime you point to it.

## 2.3 Creating Input Files for Thermodynamics and High-Pressure Limit Kinetics Computations

### 2.3.1 Syntax

The format of CanTherm input files is based on Python syntax. In fact, CanTherm input files are valid Python source code, and this is used to facilitate reading of the file.

Each section is made up of one or more function calls, where parameters are specified as text strings, numbers, or objects. Text strings must be wrapped in either single or double quotes.

### 2.3.2 Model Chemistry

The first item in the input file should be a `modelChemistry()` function, which accepts a string describing the model chemistry. Currently the allowed model chemistries are: 'CBS-QB3' 'G3' 'M08S0/MG3S\*' \* indicates that the grid size used in the [QChem] electronic structure calculation utilized 75 radial points and 434 angular points 'CCSD(T)-F12/cc-pVnZ-F12' n = D, T, Q 'CCSD(T)-F12/aug-cc-pVnZ-F12' n = D, T, Q 'MP2\_rmp2\_pVnZ' n = D, T, Q 'FCI/cc-pVnZ' n = D, T, Q 'DFT\_G03\_b3lyp' a B3LYP calculation with a moderately large basis set 'BMK/cbsb7' or 'BMK/6-311G(2d,d,p)'

CanTherm uses this information to adjust the computed energies to the usual gas-phase reference states by applying atom, bond and spin-orbit coupling energy corrections. This is particularly important for the `thermo()` calculations (see below). The example below demonstrates how to specify CBS-QB3 as a model chemistry:

```
modelChemistry("CBS-QB3")
```

### 2.3.3 Species

Each species of interest must be specified using a `species()` function, which accepts the following parameters:

Parameter	Description
label	A unique string label used as an identifier
geomLog	The path to the Gaussian log file containing the optimized geometry
statesLog	The path to the Gaussian log file containing the computed frequencies
extSymmetry	The external symmetry number for rotation
freqScaleFactor	A factor by which to scale all frequencies
linear	True if the molecule is linear, False if not
rotors	A list of <code>HinderedRotor()</code> objects describing the hindered rotors
atoms	A dict associating atom symbols with the number of each atom in the molecule
bonds	A dict associating bond types with the number of each bond in the molecule

The geometry and states log files can be identical if you computed them in the same Gaussian output. Allowed atom symbols for the `atoms` parameter are 'C', 'H', 'N', 'O', and 'P'. Allowed bond types for the `bonds` parameter are 'C-H', 'C-C', 'C=C', 'C#C', 'O-H', 'C-O', 'C=O', 'C#O', 'N#N', 'O=O', 'H-H', and 'C#N'. In both cases you can omit atoms and bonds not present in your species, and their counts will be automatically set to zero.

Each `HinderedRotor()` object requires the following parameters:

Parameter	Description
scanLog	The path to the Gaussian/Qchem log file containing the scan
pivots	The indices of the atoms in the hindered rotor torsional bond
top	The indices of all atoms on one side of the torsional bond (including the pivot atom)
symmetry	The symmetry number for the torsional rotation

The following is an example of a typical species item, based on ethane:

```
species(
  label = 'ethane',
  geomLog = 'ethane_cbs.log',
  statesLog = 'ethane_cbs.log',
  extSymmetry = 2,
  freqScaleFactor = 0.99,
  linear = False,
  rotors = [
    HinderedRotor(scanLog='ethane_scan_1.log', pivots=[0,4], top=[0,1,2,3], symmetry=3),
  ]
  atoms = {'C': 2, 'H': 6},
  bonds = {'C-C': 1, 'C-H': 6},
)
```

Note that the atoms identified within the rotor section should correspond to the geometry indicated by geomLog.

### 2.3.4 Transition State

Each transition state of interest must be specified using a `transitionState()` function, which accepts exactly the same parameters as the `species()` function described above. This is only required if you wish to perform a kinetics computation.

The following is an example of a typical transition state item:

```
transitionState(
  label = 'TS',
  geomLog = 'H+C2H4.log',
  statesLog = 'H+C2H4.log',
  extSymmetry = 2,
  freqScaleFactor = 0.99,
  linear = False,
  rotors = [],
  atoms = {'C': 2, 'H': 5},
  bonds = {'C-C': 1, 'C-H': 5},
)
```

### 2.3.5 Reaction

Each reaction of interest must be specified using a `reaction()` function, which accepts the following parameters:

Parameter	Description
label	A unique string label used as an identifier
reactants	A list of strings indicating the labels of the reactant species
products	A list of strings indicating the labels of the product species
transitionState	The string label of the transition state

This is only required if you wish to perform a kinetics computation. The following is an example of a typical reaction item:

```
reaction(
  label = 'H + C2H4 <=> C2H5',
  reactants = ['H', 'C2H4'],
  products = ['C2H5'],
  transitionState = 'TS',
)
```

```
tunneling='Eckart'
)
```

Note that in the above example, ‘Wigner’ is also an acceptable method of estimating the quantum tunneling factor.

## 2.3.6 Thermodynamics Computations

Use a `thermo()` function to compute the thermodynamic parameters for a species. Pass the string label of the species you wish to compute the thermodynamic parameters for and the type of thermodynamics model to generate (either ‘Wilhoit’ or ‘NASA’ for a Wilhoit polynomial model or NASA polynomial model). A table of thermodynamic parameters will also be displayed in the output file.

Below is a typical `thermo()` function:

```
thermo('ethane', model='Wilhoit')
```

## 2.3.7 Kinetics Computations

Use a `kinetics()` function to compute the high-pressure limit kinetic parameters for a reaction. If desired, define a desired temperature range and number of temperatures at which the high-pressure rate coefficient will be tabulated and saved to the output file. 3-parameter modified Arrhenius coefficients will automatically be fit to the computed rate coefficients. The quantum tunneling factor will also be displayed

Below is a typical `kinetics()` function:

```
kinetics(
label = 'H + C2H4 <=> C2H5',
Tmin = (400, 'K'), Tmax = (1200, 'K'), Tcount = 6,
Tlist = ([400, 500, 700, 900, 1100, 1200], 'K'),
)
```

This is also acceptable:

```
kinetics('H + C2H4 <=> C2H5')
```

## 2.3.8 Examples

Perhaps the best way to learn the input file syntax is by example. To that end, a number of example input files and their corresponding output have been given in the `examples` directory.

## 2.3.9 Troubleshooting and FAQs

1) The network that CanTherm generated and the resulting pdf file show abnormally large absolute values. What’s going on?

This can happen if the number of atoms and atom types is not properly defined or consistent in your input file(s).

## 2.3.10 Cantherm User Checklist

Using cantherm, or any rate theory package for that matter, requires careful consideration and management of a large amount of data, files, and input parameters. As a result, it is easy to make a mistake somewhere. This checklist was made to minimize such mistakes for users:

- Do correct paths exist for pointing to the files containing the electronic energies, molecular geometries and vibrational frequencies?

For calculations involving pressure dependence:

- Does the network pdf look reasonable? That is, are the relative energies what you expect based on the input?

For calculations using internal hindered rotors:

- Did you check to make sure the rotor has a reasonable potential (e.g., visually inspect the automatically generated rotor pdf files)?
- Within your input files, do all specified rotors point to the correct files?
- Do all of the atom label indices correspond to those in the file that is read by the logger (GaussianLog, Qchem-Log, etc.)?
- Why do the fourier fits look so much different than the results of the ab initio potential energy scan calculations? This is likely because the initial scan energy is not at a minimum. One solution is to simply shift the potential with respect to angle so that it starts at zero and, instead of having CanTherm read a Qchem or Gaussian output file, have CanTherm point to a 'ScanLog' file. Another problem can arise when the potential at  $2\pi$  is also not [close] to zero.

## 2.4 Creating Input Files for Pressure Dependent Calculations

### 2.4.1 Syntax

There are four parts to a pressure-dependent calculation input file, giving the species, transition states, path reactions, reaction network, and algorithm parameters. The species section must come before the reaction section. Before discussing each of these sections, a brief word on the general input file syntax will be given.

### 2.4.2 Species Parameters

Each species in the network must be specified using a `species()` block. This includes all unimolecular isomers, bimolecular reactants and products, and the bath gas(es). A species that appears in multiple bimolecular channels need only be specified with a single `species()` block.

There are a number of required and optional parameters associated with a species block:

Parameter	Required?	Description
<code>label</code>	all species	A unique string label used as an identifier
<code>structure</code>	all species except bath gas	A chemical structure for the species defined using either SMILES or InChI
<code>E0</code>	all species	The ground-state energy (including zero-point energy)
<code>modes</code>	all species	The molecular degrees of freedom (see below)
<code>spinMultiplicity</code>	all species	The ground-state spin multiplicity (degeneracy), sets to 1 by default if not used
<code>opticalIsomers</code>	all species	The number of optical isomers of the species, sets to 1 by default if not used
<code>molecularWeight</code>	all species	The molecular weight, if not given it is calculated based on the structure
<code>collisionModel</code>	optional	Transport data for the species, if available
<code>energyTransferModel</code>	optional	Assigned with <code>SingleExponentialDown</code> model if available
<code>thermo</code>	optional	Thermo data for the species

The `label` parameter should be set to a string with the desired user name for the species.

```
label = 'nButanol'
```

The `structure` parameter is defined by either SMILES or InChI. For instance, either representation is acceptable for the acetone molecule:

```
structure = SMILES('CC(C)=O')
structure = InChI('InChI=1S/C3H6O/c1-3(2)4/h1-2H3')
```

The `E0` ground state energy should be given in the quantity format (`value`, `'units'`), using units of either kJ/mol, kcal/mol, J/mol, or cal/mol:

```
E0 = (-34.6, 'kcal/mol')
```

The `modes` parameter is required for all unimolecular isomers and all bimolecular reactant channels. When specifying the `modes` parameter, define a list with the following types of degrees of freedom. To understand how to define these degrees of freedom, please click on the links below:

#### Translational degrees of freedom

Class	Description
IdealGasTranslation	A model of three-dimensional translation of an ideal gas

#### Rotational degrees of freedom

Class	Description
LinearRotor	A model of two-dimensional rigid rotation of a linear molecule
NonlinearRotor	A model of three-dimensional rigid rotation of a nonlinear molecule
KRotor	A model of one-dimensional rigid rotation of a K-rotor
SphericalTopRotor	A model of three-dimensional rigid rotation of a spherical top molecule

#### Vibrational degrees of freedom

Class	Description
HarmonicOscillator	A model of a set of one-dimensional harmonic oscillators

#### Torsional degrees of freedom

Class	Description
HinderedRotor	A model of a one-dimensional hindered rotation

The `spinMultiplicity` is defined using an integer, and is set to 1 if not indicated in the species block.

```
spinMultiplicity = 2
```

Similarly, the `opticalIsomers` is also defined using an integer, and is set to 1 if not used in the species block.

```
opticalIsomers = 6
```

The `molecularWeight` parameter should be defined in the quantity format (`value`, `'units'`), for example:

```
molecularWeight = (44.04, 'g/mol')
```

If the `molecularWeight` parameter is not given, it is calculated by CanTherm based off the chemical structure.

The `collisionModel` is defined with the transport data, if available, using a `TransportData` object:

```
collisionModel = TransportData(sigma=(3.70, 'angstrom'), epsilon=(94.9, 'K'))
```

The `energyTransferModel` model available is a `SingleExponentialDown`.

- SingleExponentialDown - Specify  $\alpha_0$ ,  $T_0$  and  $n$  for the average energy transferred in a deactivating collision

$$\langle \Delta E_{\text{down}} \rangle = \alpha_0 \left( \frac{T}{T_0} \right)^n$$

An example of a typical energyTransferModel block is:

```
energyTransferModel = SingleExponentialDown(
    alpha0 = (0.5718, 'kcal/mol'),
    T0 = (300, 'K'),
    n = 0.85,
)
```

The following is an example of a typical species item, based on the acetylperoxy radical  $\text{CH}_3\text{C}(=\text{O})\text{OO}\cdot$ :

```
species(
    label = 'acetylperoxy',
    structure = SMILES('CC(=O)O[O]'),
    E0 = (-34.6, 'kcal/mol'),
    modes = [
        IdealGasTranslation(mass=(75.04, "g/mol")),
        NonlinearRotor(inertia=(54.2977,104.836,156.05),"amu*angstrom^2"), symmetry=1),
        HarmonicOscillator(frequencies=(319.695,500.474,536.674,543.894,727.156,973.365,1037.77,1119.72,1181.5),
        HinderedRotor(inertia=(7.38359, "amu*angstrom^2"), symmetry=1, fourier=[[-1.95191, -11.8215, 0.740041, -0.],
        HinderedRotor(inertia=(2.94723, "amu*angstrom^2"), symmetry=3, fourier=[[0.130647, 0.0401507, -2.54582, -0.],
    ],
    spinMultiplicity = 2,
    opticalIsomers = 1,
    molecularWeight = (75.04, "g/mol"),
    collisionModel = TransportData(sigma=(5.09, 'angstrom'), epsilon=(473, 'K')),
    energyTransferModel = SingleExponentialDown(
        alpha0 = (0.5718, 'kcal/mol'),
        T0 = (300, 'K'),
        n = 0.85,
    ),
)
```

### 2.4.3 Transition States

Transition states for reactions in the pressure dependent network should be defined very similarly to species using a transitionState block, however it has less parameters:

Parameter	Description
label	A unique string label used as an identifier
E0	The ground-state energy (including zero-point energy)
modes	The molecular degrees of freedom (same as for species, see above)
spinMultiplicity	The ground-state spin multiplicity (degeneracy), sets to 1 by default if not used
opticalIsomers	The number of optical isomers of the species, sets to 1 by default if not used
frequency	The negative frequency of the first-order saddle point

An example of a transitionState block is shown below.

```
transitionState(
    label = 'isom1',
    E0 = (-5.8, 'kcal/mol'),
    modes = [
```

```

IdealGasTranslation(mass=(75.04,"g/mol")),
NonlinearRotor(inertia=( [49.3418,103.697,149.682] ,"u*angstrom**2"), symmetry=1, quantum=False),
HarmonicOscillator(frequencies=( [148.551,306.791,484.573,536.709,599.366,675.538,832.594,918.413,1022.2
],
spinMultiplicity = 2,
opticalIsomers = 1,
frequency = (-1679.04,'cm^-1'),
)

```

## 2.4.4 Path Reactions

Each path reaction - a reaction directly connecting two molecular configurations in the network - is specified using a `reaction()` block. The following parameters are available:

Parameter	Re-quired?	Description
label	All reactions	A name for the reaction
reactants	All reactions	A list of reactant species
products	All reactions	A list of product species
transitionState	All reactions	The transition state
kinetics	Optional	The high pressure-limit kinetics for the reaction
tunneling	Optional	The type of tunneling model (either 'Eckhart' or 'Wigner') to use for tunneling through the reaction barrier

A typical reaction block might look like this.

```

reaction(
  label = 'isom1',
  reactants = ['acetylperoxy'],
  products = ['hydroperoxylvinoxy'],
  transitionState = 'isom1',
  kinetics = Arrhenius(A=(2.65e6,'m^3/(mol*s)'), n=0.0, Ea=(0.0,'kcal/mol'), T0=(1,"K")),
  tunneling = 'Eckart',
)

```

Note that the reactants and products must have been previously declared using a `species` block, using the same name labels. Transition states must also be previously declared using a `transitionState` block.

## 2.4.5 Network

A declaration for the overall network must be given using the `network` block.

This includes setting the following parameters:

Parameter	Description
label	A name for the network
isomers	A list of species participating in unimolecular reaction channels
reactants	A list of the species that participate in bimolecular reactant channels
bathGas	A dictionary of bath gases and their respective mole fractions, adding up to 1.0

CanTherm is largely able to determine the molecular configurations that define the potential energy surface for your reaction network simply by inspecting the path reactions. However, you must indicate which unimolecular and bimolec-

ular configurations you wish to include in the master equation formulation; all others will be treated as irreversible sinks.

Note that all species and bath gases used in the `network` block must have been previously declared with the same name labels in a previous `species` block in the input file.

You do not need to specify the product channels (infinite sinks) in this manner, as any configuration not marked as an isomer or reactant channel will be treated as a product channel. An example of the `network` block is shown below.

```
network(
  label = 'acetyl + O2',
  isomers = [
    'acetylperoxy',
    'hydroperoxylvinoxy',
  ],
  reactants = [
    ('acetyl', 'oxygen'),
  ],
  bathGas = {
    'nitrogen': 0.4,
    'argon': 0.6,
  }
)
```

## 2.4.6 Algorithm Parameters

The overall parameters for the pressure-dependence calculation must be defined in a `pressureDependence` block at the end of the input file. The following parameters are necessary:

Parameter	Description
<code>label</code>	Use the name for the <code>network</code> declared previously
<code>method</code>	Method to use for calculating the pdep network. Use either 'modified strong collision', 'reservoir state', or 'chemically-significant eigenvalues'
<code>interpolationModel</code>	Select the output type for the pdep kinetics, either in 'chebyshev' or 'pdeparrhenius' (plog) format
<code>activeKRotor</code>	A flag indicating whether to treat the K-rotor as active or adiabatic
<code>activeJRotor</code>	A flag indicating whether to treat the J-rotor as active or adiabatic

Additionally, temperature/pressure ranges and energy grain sizes must be given.

### Temperature and Pressure Ranges

CanTherm will compute the  $k(T, P)$  values on a grid of temperature and pressure points. `Tmin`, `Tmax`, and `Tcount` values, as well as `Pmin`, `Pmax`, and `Pcount` parameter values must be provided. CanTherm will automatically choose the intermediate temperatures based on the interpolation model you wish to fit. This is the recommended approach.

**Energy Grains** Determine the fineness of the energy grains to be used in the master equation calculations. Dictate the `maximumGrainSize`, and the `minimumGrainCount`.

An example of the algorithm parameters block for the acetyl + O2 network is shown below.

```
pressureDependence(
  label='acetyl + O2',
  Tmin=(300.0,'K'), Tmax=(2000.0,'K'), Tcount=8,
  Pmin=(0.01,'bar'), Pmax=(100.0,'bar'), Pcount=5,
  maximumGrainSize = (1.0,'kcal/mol'),
  minimumGrainCount = 250,
  method = 'modified strong collision',
  #method = 'reservoir state',
```

```
#method = 'chemically-significant eigenvalues',
interpolationModel = ('chebyshev', 6, 4),
#interpolationModel = ('pdeparrrhenius'),
#activeKRotor = True,
activeJRotor = True,
)
```

## 2.4.7 Examples

Perhaps the best way to learn the input file syntax is by example. To that end, a number of example input files and their corresponding output have been given in the `examples/cantherm/networks` directory, which includes both an *acetyl+O2* and *n-butanol* example.

## 2.5 Running CanTherm

To execute a CanTherm job, invoke the command

```
$ python cantherm.py FILE
```

The job will run and the results will be saved to `output.py` in the same directory as the input file. If you wish to save the output elsewhere, use the `-o/-output` option, e.g.

```
$ python cantherm.py INPUTFILE -o OUTPUTFILE
```

### 2.5.1 Drawing Potential Energy Surface

CanTherm contains functionality for automatically generating an image of the potential energy surface for a reaction network. This is done automatically and outputted in pdf format to a file called `network.pdf`.

### 2.5.2 Log Verbosity

You can manipulate the amount of information logged to the console window using the `-q/-quiet` flag (for quiet mode) or the `-v/-verbose` flag (for verbose mode). The former causes the amount of logging information shown to decrease; the latter causes it to increase.

### 2.5.3 Help

To view help information and all available options, use the `-h/--help` flag, e.g.

```
$ python cantherm.py -h
```

## 2.6 Parsing Output Files

### 2.6.1 Thermodynamic and High-pressure Limit Kinetics Calculations

The syntax of CanTherm output files closely mirrors that of the input files. For each `thermo()` function in the input file, there will be a corresponding `thermo()` function in the output file containing the computed thermodynamic

model. Similarly, For each `kinetics()` function in the input file, there will be a corresponding `kinetics()` function in the output file containing the computed kinetics model.

## 2.6.2 Pressure-Dependent Calculations

The output file contains the entire contents of the input file. In addition, the output file contains a block of `pdepreaction()` calls. The parameters of each `pdepreaction()` block match those of the `reaction()` block from the input file, except that no transition state data is given and the `kinetics` are by definition pressure-dependent.

A `pdepreaction()` item is printed for the forward and reverse direction of every reaction involving isomers and reactant channels only. For reactions involving a product channel, there is only a `pdepreaction()` item for the direction in which the product channel is the product of the reaction. To use this output, you must either keep all of the reactions and treat them as irreversible, or discard the duplicate reverse directions and treat the remaining reactions as reversible. This decision is left to the end user.

## 2.6.3 Log File

A log file containing similar information to that displayed on the console during CanTherm execution is also automatically saved. This file has the name `cantherm.log` and is found in the same directory as the output file. The log file accepts logging messages at an equal or greater level of detail than the console; thus, it is often useful (and recommended) to examine both if something unexpected has occurred.

The `examples/cantherm` directory contains both CanTherm input files and the resulting output files.

## 2.7 Frequently Asked Questions

### Are there other software packages for investigating pressure-dependent reaction networks?

Yes. The following is an illustrative list of such packages:

Name	Method(s)	Language	Author(s)
MultiWell	stochastic	Fortran	J. R. Barker <i>et al</i>
UNIMOL	CSE	Fortran	R. G. Gilbert, S. C. Smith
ChemRate	CSE	C++ <sup>1</sup>	V. Mokrushin, W. Tsang
Variflex	CSE	Fortran	S. J. Klippenstein <i>et al</i>
MESMER	CSE (+ RS)	C++	S. H. Robertson <i>et al</i>
CHEMDIS <sup>2</sup>	MSC	Fortran	A. Y. Chang, J. W. Bozzelli, A. M. Dean

(MSC = modified strong collision, RS = reservoir state, CSE = chemically-significant eigenvalues)

Many of the above packages also provide additional functionality beyond the approximate solving of the master equation. For example, Variflex can be used for variational transition state theory calculations, while ChemRate provides a (Windows) graphical user interface for exploring a database of experimental data and physical quantities.

<sup>1</sup>Uses MFC for Windows graphical user interface

<sup>2</sup>No longer distributed

## 2.8 Credits

**Author:** Joshua W. Allen ([joshua.w.allen@gmail.com](mailto:joshua.w.allen@gmail.com))

**P.I.:** Prof. William H. Green ([whgreen@mit.edu](mailto:whgreen@mit.edu))

The author acknowledges the Green group for helping put the software through its paces and providing suggestions for its improvement.

CanTherm is based upon work supported by the [King Abdullah University of Science and Technology](#).

- [genindex](#)
- [modindex](#)
- [search](#)

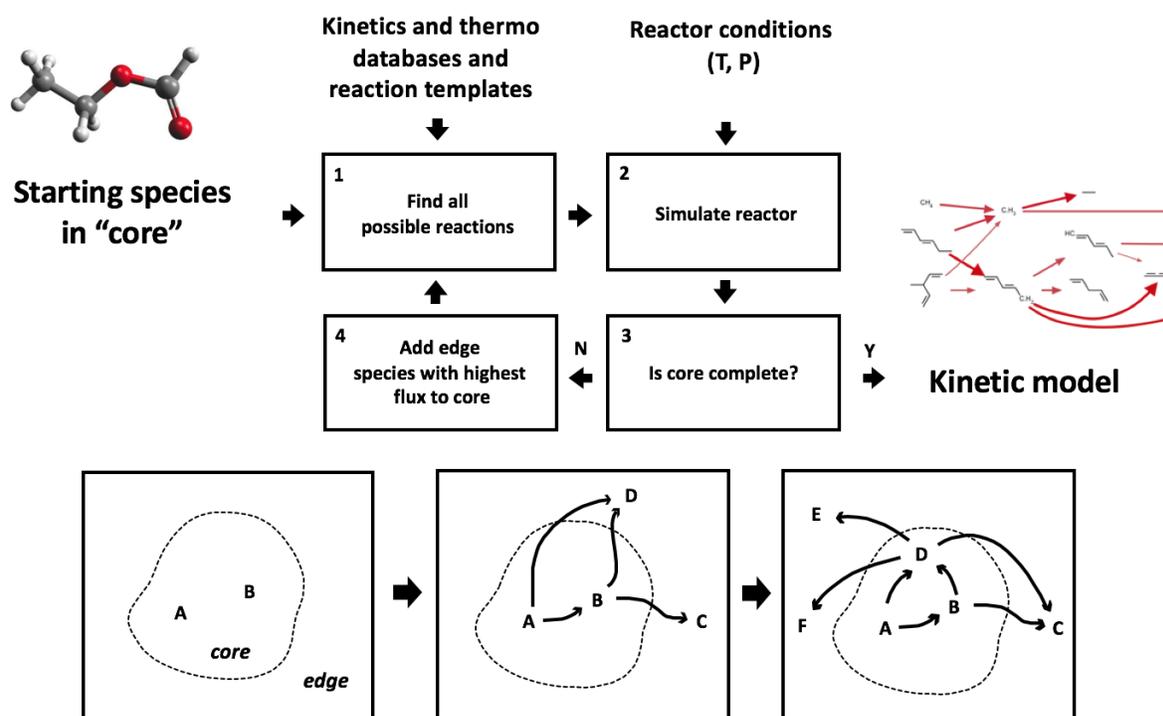


THEORY GUIDE

The theoretical foundations to some of the parts of RMG and Cantherm are described in greater detail in these sections.

### 3.1 RMG Theory Guide

#### 3.1.1 Rate-based Model Enlarging Algorithm



To construct a mechanism, the user must specify an initial set of species and the initial conditions (temperature, pressure, species concentrations, etc.). RMG reacts the initial species in all possible ways according to its known reaction families, and it integrates the model in time. RMG tracks the rate (flux) at which each new “edge” species is produced, and species (and the reactions producing them) that are produced with significant fluxes are incorporated into the model (the “core”). These new core species are reacted with all other core species in the model to generate a new set of edge species and reactions. The time-integration restarts, and the expanded list of edge species is monitored for significant species to be included in the core. The process continues until all significant species and reactions have

been included in the model. The definition of a “significant” rate can be specified by the user by taking the following definition for a single species rate:

$$R_i = \frac{dC_i}{dt}$$

and the following definition for the reaction system’s characteristic rate, which is the sum of all **core** species rates:

$$R_{char} = \sqrt{\sum_j R_j^2} \quad \text{species } j \in \text{core}$$

When a species  $i \in \text{edge}$  exceeds a “significant” rate equal to  $\epsilon R_{char}$ , it is added to the core. The parameter  $\epsilon$  is the user-specified `toleranceMoveToCore` that can be adjusted under the *model tolerances* in the *RMG Input File*.

For more information on rate-based model enlargement, please refer to the papers [Gao2016] or [Susnow1997].

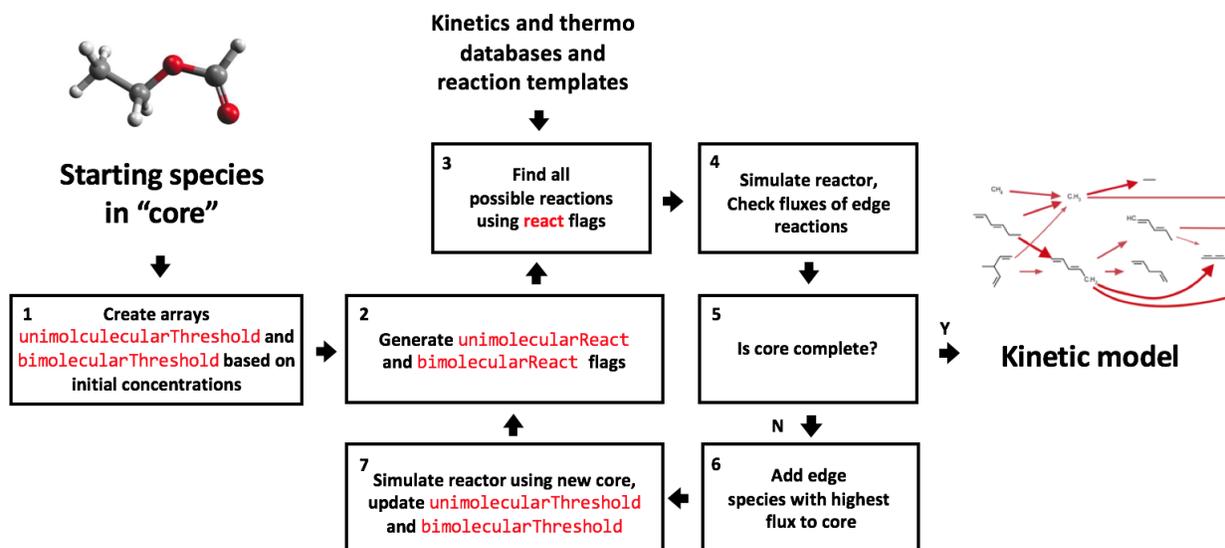
## Filtering Reactions within the Rate-based Algorithm

Filtering reactions in the react step in the flux-based algorithm attempts to speed up model generation by attacking the pain point. RMG has trouble converging when generating models for large molecules because it searches for reactions on the order of  $(n_{reaction\ sites})^{n_{species}}$ .

The original algorithm performs in the following manner:

1. Reacts species together (slow)
2. Determines which reactions are negligible (fast)

By filtering reactions we add a pre-filtering step before step 1 which prevents species from reacting together when the reactions are expected to be negligible throughout the simulation.



Here, `unimolecularThreshold` and `bimolecularThreshold` are binary arrays storing a flag for whether a species or a pair of species are above a reaction threshold. For a unimolecular rate, this threshold is set to True if the unimolecular rate of reaction  $k$  for a species  $A$

$$R_{unimolecular} = k_{threshold} C_A > \epsilon R_{char}$$

at any given time  $t$  in the reaction system, where  $k_{threshold} = \frac{k_B T}{h}$

For a bimolecular reaction occurring between species A and B, this threshold is set to True if the bimolecular rate

$$R_{bimolecular} = k_{threshold}C_A C_B > \epsilon R_{char}$$

at any given time  $t$  in the reaction system, where  $k_{threshold} = 1 \times 10^{13} \frac{cm^3}{mol*s}$  which is considered the diffusion limiting rate.

Two additional binary arrays `unimolecularReact` and `bimolecularReact` store flags for when the `unimolecularThreshold` or `bimolecularThreshold` flag shifts from False to True. RMG reacts species when the flag is set to True.

### 3.1.2 Prune Edge Species

When dealing with complicated reaction systems, RMG calculation would easily hit the computer memory limitation. Memory profiling shows most memory especially during memory limitation stage is occupied by edge species. However, most edge species in fact wouldn't be included in the core (or final model). Thus, it's natural to get rid of some not "so useful" edge species during calculation in order to achieve both low memory consumption and mechanism accuracy. Pruning is such a way.

#### Key Parameters in Pruning

- `toleranceKeepInEdge`  
Any edge species to prune should have peak flux along the whole conversion course lower than `toleranceKeepInEdge * characteristic flux`. Thus, larger values will lead to smaller edge mechanisms.
- `toleranceMoveToCore`  
Any edge species to enter core model should have flux at some point larger than `toleranceMoveToCore * characteristic flux`. Thus, in general, smaller values will lead to larger core mechanisms.
- `toleranceInterruptSimulation`  
Once flux of any edge species exceeds `toleranceInterruptSimulation * characteristic flux`, dynamic simulation will be stopped. Usually this tolerance will be set a very high value so that any flux's exceeding that means mechanism is too incomplete to continue dynamic simulation.
- `maximumEdgeSpecies`  
If dynamic simulation isn't interrupted in half way and total number of the edge species whose peak fluxes are higher than `toleranceKeepInEdge * characteristic flux` exceeds `maximumEdgeSpecies`, such excessive amount of edge species with lowest peak fluxes will be pruned.
- `minCoreSizeForPrune`  
Ensures that a minimum number of species are in the core before pruning occurs, in order to avoid pruning the model when it is far away from completeness. The default value is set to 50 species.
- `minSpeciesExistIterationsForPrune`  
Set the number of iterations an edge species must stay in the job before it can be pruned. The default value is 2 iterations.

#### How Pruning Works

- `genindex`
- `modindex`
- `search`

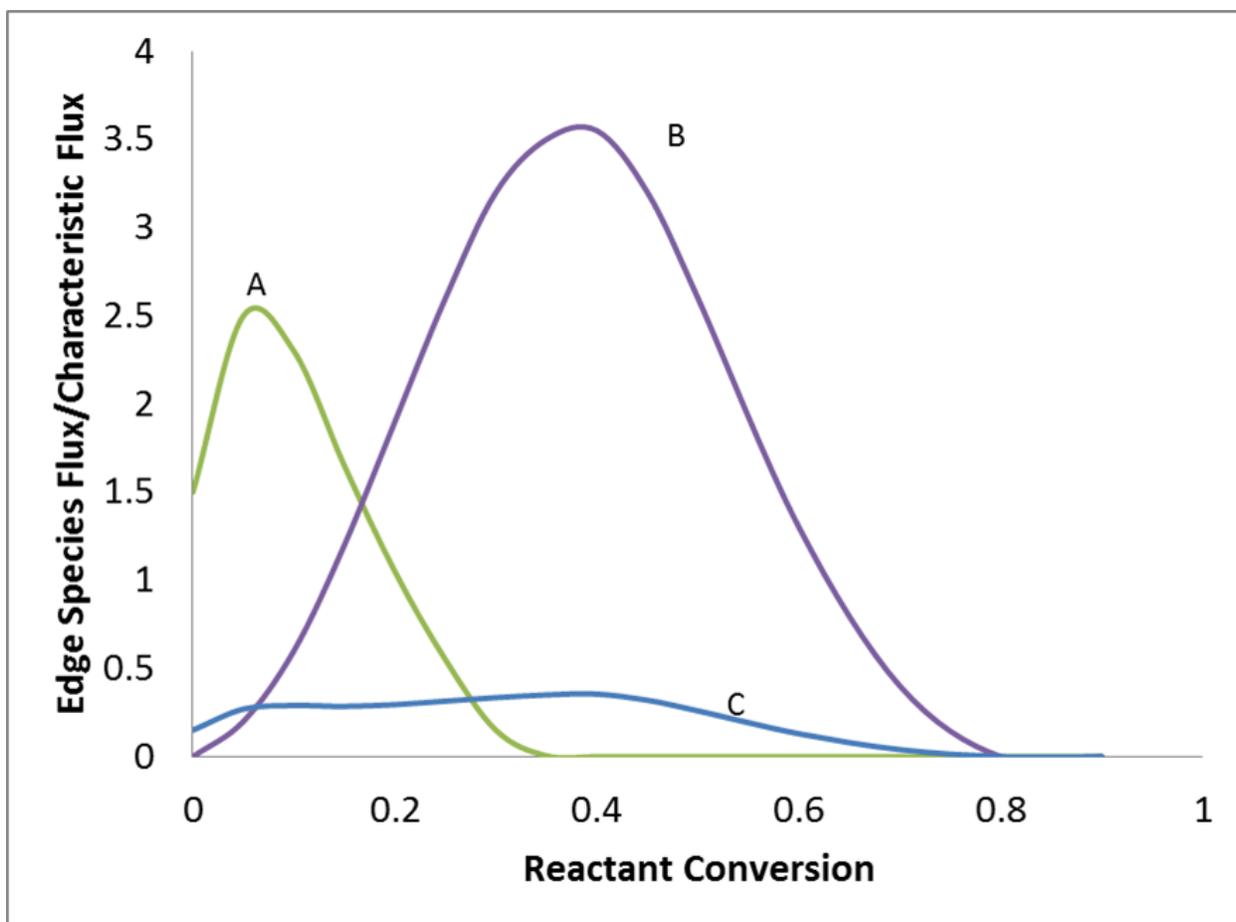


Fig. 3.1: The goal of pruning is to delete those “useless” edge species. So “usefulness” should be defined and it’s natural to have flux as a criterion for “usefulness”. Since flux changes with reactant conversion, peak flux is chosen here to make decision of pruning or not. Every time pruning is triggered, edge species with peak flux lower than  $\text{toleranceKeepInEdge} * \text{characteristic flux}$  will be deleted.

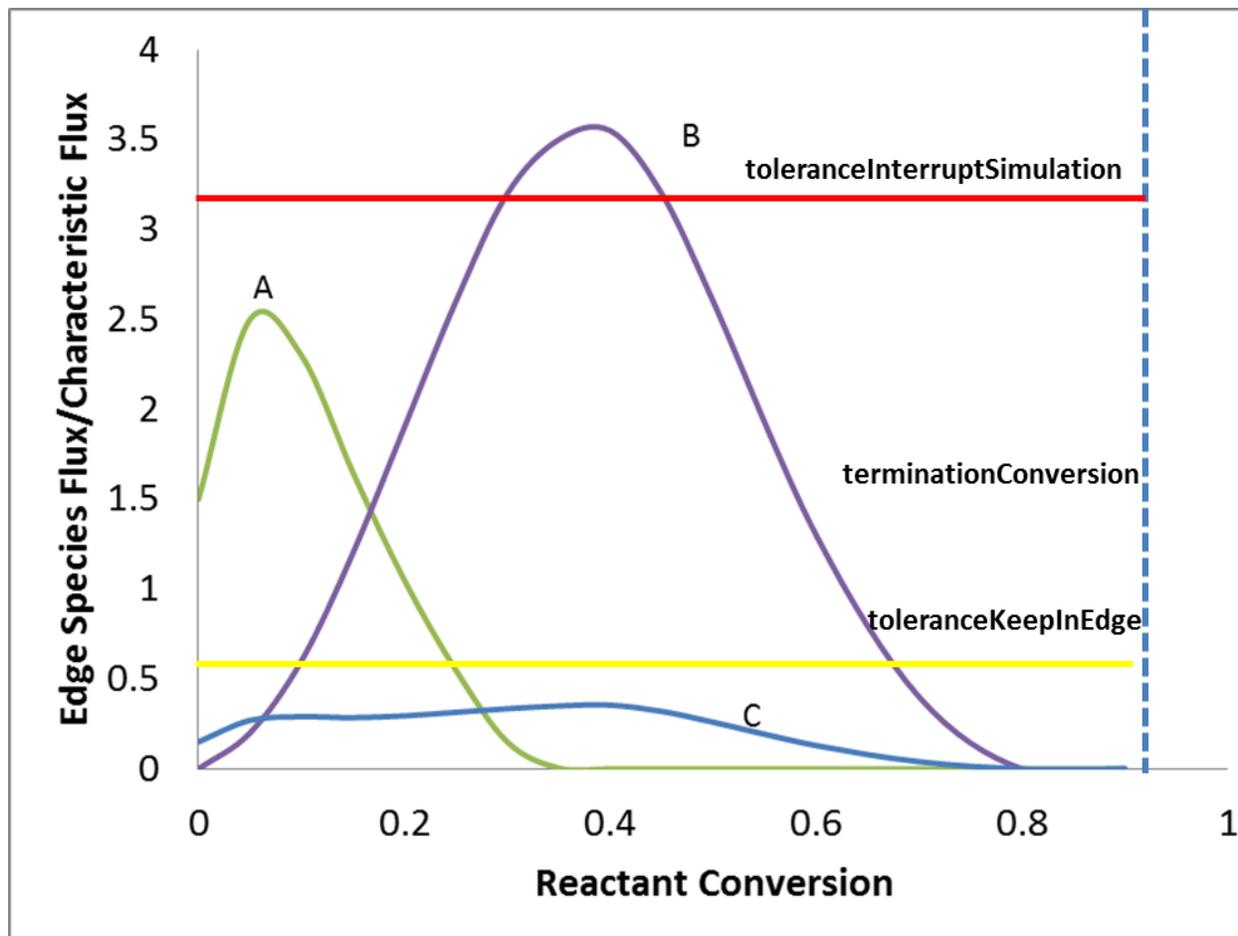


Fig. 3.2: However, pruning is not always triggered because of `toleranceInterruptSimulation`. As mentioned above, in order to prune, RMG needs to figure out the peak flux of each edge species, which requires dynamic simulation to complete. If some run of dynamic simulation is terminated in half way by `toleranceInterruptSimulation`, pruning is rejected although there might be some edge species with peak fluxes lower than `toleranceKeepInEdge` \* characteristic flux. Since pruning requires to complete dynamic simulation, setting `toleranceInterruptSimulation` to be positive infinity, as an extreme case, means always enabling pruning. Another extreme case would be that it has same value as `toleranceMoveToCore` where no pruning occurs.

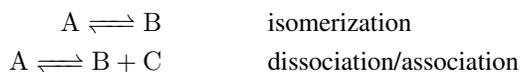
In summary, each run of dynamic simulation will proceed towards `terminationConversion` unless some flux exceeds `toleranceInterruptSimulation` \* characteristic flux. Following complete simulation is the pruning of edge species whose flux is not high enough to be kept in the edge, which is followed by pruning of excessive amount of edge species to make sure total edge species number is no greater than `maximumEdgeSpecies`.

## 3.2 Pressure-Dependence Theory Guide

### 3.2.1 Introduction

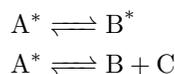
#### Unimolecular Reactions

Unimolecular reactions are those that involve a single reactant or product molecule, the union of isomerization and dissociation/association reactions:



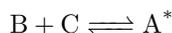
Gas-phase chemical reactions occur as the result of bimolecular collisions between two reactant molecules. This presents a problem when there is only one participating reactant molecule! The conclusion is that the above reactions cannot be elementary as written; another step must be involved.

For a unimolecular reaction to proceed, the reactant molecule  $A$  must first be excited to an energy that exceeds the barrier for reaction. A molecule that is sufficiently excited to react is called an *activated species* and often labeled with an asterisk  $A^*$ . If we replace the stable species with the activated species in the reactions above, the reactions become elementary again:

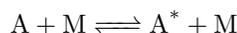


There are a number of ways that an activated species  $A^*$  can be produced:

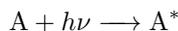
- **Chemical activation.**  $A^*$  is produced as the adduct of an association reaction:



- **Thermal activation.**  $A^*$  is produced via transfer of energy from an otherwise inert species  $M$  via bimolecular collision:



- **Photoactivation.**  $A^*$  is produced as a result of absorption of a photon:



Once an activated molecule has been produced, multiple isomerization and dissociation reactions may become competitive with one another and with collisional stabilization (thermal deactivation); these combine to form a network of unimolecular reactions. The major pathway will depend on the relative rates of collision and reaction, which in turn is a function of both temperature and pressure. At high pressure the collision rate will be fast, and activated molecules will tend to be collisionally stabilized before reactive events can occur; this is called the *high-pressure limit*. At low pressures the collision rate will be slow, and activated molecules will tend to isomerize and dissociate, often traversing multiple reactive events before collisional stabilization can occur.

The onset of the pressure-dependent regime varies with both temperature and molecular size. The figure below shows the approximate pressure at which pressure-dependence becomes important as a function of temperature and molecular size. The parameter  $m \equiv N_{\text{vib}} + \frac{1}{2}N_{\text{rot}}$  represents a count of the internal degrees of freedom (vibrations and hindered rotors, respectively). The ranges of the x-axis and y-axis suggest that pressure dependence is in fact important over a wide regime of conditions of practical interest, particularly in high-temperature processes such as pyrolysis and combustion [Wong2003].

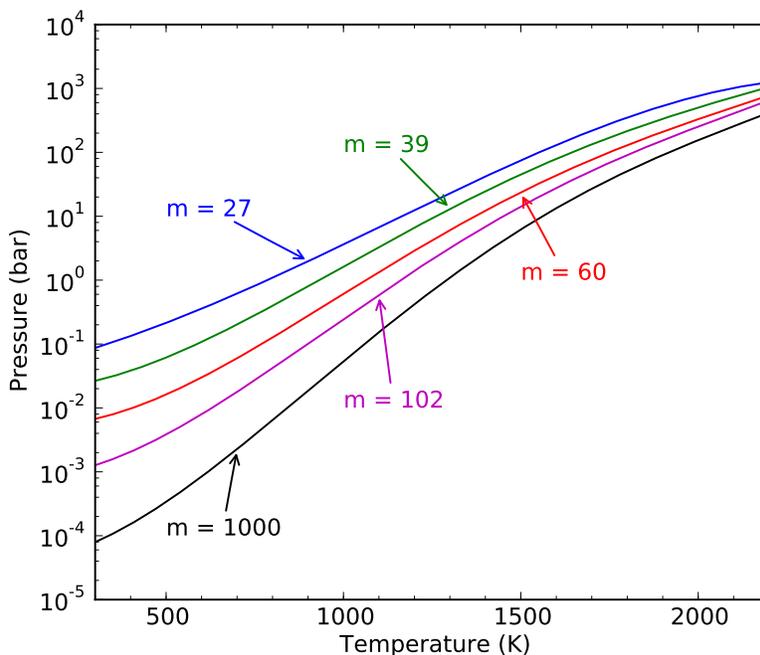


Fig. 3.3: Plot of the switchover pressure – indicating the onset of pressure dependence – as a function of temperature and molecular size. The value  $m \equiv N_{\text{vib}} + \frac{1}{2}N_{\text{rot}}$  represents a count of the internal degrees of freedom. Over a wide variety of conditions of practical interest, even very large molecules exhibit significant pressure dependence. Figure adapted from Wong, Matheu and Green (2003).

### Historical Context

The importance of bimolecular collisions in unimolecular reactions was first proposed by Lindemann in 1922 [Lindemann1922]. It was soon recognized by Hinshelwood and others that a rigorous treatment of these processes required consideration of molecular energy levels [Hinshelwood1926]. The RRKM expression for the microcanonical rate coefficient  $k(E)$  was derived in the early 1950s [Rice1927] [Kassel1928] [Marcus1951]. In the late 1950s master equation models of chemical systems began appearing [Siegert1949] [Bartholomay1958] [Montroll1958] [Krieger1960] [Gans1960], including an early linear integral-differential equation formulation by Widom [Widom1959]. Analytical solutions for a variety of simple models soon followed [Keck1965] [Troel1967] [Troel1973], as did the first numerical approaches [Tardy1966]. Numerical methods – which are required for complex unimolecular reaction networks – became much more attractive in the 1970s with the appearance of new algorithms, including Gear’s method for solving stiff systems of ordinary differential equations [Gear1971] and efficient algorithms for calculating the density of states [Beyer1973] [Stein1973] [Astholz1979]. In the 1990s computing power had increased to the point where it was practical to solve them numerically by discretizing the integrals over energy.

### 3.2.2 The Master Equation

A full treatment of the energy states of each molecule is unfeasible for molecules larger than diatomics, as there are simply too many states. To simplify things we apply the RRKM approximation, which leaves the state of a molecule as a function of two quantities: the total energy  $E$  and total angular momentum quantum number  $J$ . Frequently we will find that even this is too difficult, and will only keep the total energy  $E$  as an independent variable.

## Isomers, Reactants, and Products

Throughout this document we will utilize the following terminology:

- An **isomer** is a unimolecular configuration on the potential energy surface.
- A **reactant channel** is a bimolecular configuration that associates to form an isomer. Dissociation from the isomer back to reactants is allowed.
- A **product channel** is a bimolecular configuration that is formed by dissociation of an isomer. Reassociation of products to the isomer is *not* allowed.

The isomers are the configurations for which we must model the energy states. We designate  $p_i(E, J, t)$  as the population of isomer  $i$  having total energy  $E$  and total angular momentum quantum number  $J$  at time  $t$ . At long times, statistical mechanics requires that the population of each isomer approach a Boltzmann distribution  $b_i(E, J)$ :

$$\lim_{t \rightarrow \infty} p_i(E, J, t) \propto b_i(E, J)$$

We can simplify by eliminating the angular momentum quantum number to get

$$p_i(E, t) = \sum_J p_i(E, J, t)$$

Let us also denote the (time-dependent) total population of isomer  $i$  by  $x_i(t)$ :

$$x_i(t) \equiv \sum_J \int_0^\infty p_i(E, J, t) dE$$

The two molecules of a reactant or product channel are free to move apart from one another and interact independently with other molecules in the system. Accordingly, we treat these channels as fully thermalized, leaving as the only variable the total concentrations  $y_{nA}(t)$  and  $y_{nB}(t)$  of the molecules  $A_n$  and  $B_n$  of reactant channel  $n$ . (Since the product channels act as infinite sinks, their populations do not need to be considered explicitly.)

Finally, we will use  $N_{\text{isom}}$ ,  $N_{\text{reac}}$ , and  $N_{\text{prod}}$  as the numbers of isomers, reactant channels, and product channels, respectively, in the system.

## Collision Models

Bimolecular collisions with an inert species  $M$  are the primary means by which an isomer molecule changes its energy. A reasonable estimate – although generally a bit of an underestimate – of the total rate of collisions  $k_{\text{coll},i}(T)$  for each isomer  $i$  comes from Lennard-Jones collision theory:

$$k_{\text{coll},i}(T) = \sqrt{\frac{8k_B T}{\pi \mu_i}} \pi d_i^2 \Omega_i^{(2,2)*}$$

Above,  $\mu_i$  is the reduced mass,  $d_i$  is the collision diameter, and  $k_B$  is the Boltzmann constant. The collision diameter is generally taken as  $d \approx \frac{1}{2}(\sigma_i + \sigma_M)$ , the arithmetic average of the Lennard-Jones  $\sigma$  parameter for the isomer and the bath gas. The parameter  $\Omega_i^{(2,2)*}$  represents a configurational integral, which is well-approximated by the expression

$$\Omega_i^{(2,2)*} = 1.16145 \tilde{T}^{-0.14874} + 0.52487 e^{-0.7732 \tilde{T}} + 2.16178 e^{-2.43788 \tilde{T}}$$

where  $\tilde{T} \equiv k_B T / \sqrt{\epsilon_i \epsilon_M}$  is a reduced temperature and  $\epsilon_i$  is the Lennard-Jones  $\epsilon$  parameter. Note that we have used a geometric average for the  $\epsilon$  parameters of the isomer and the bath gas in this expression. Assuming the total gas concentration to be constant and that the gas is ideal, we obtain an expression for the collision frequency  $\omega_i(T, P)$ , which makes explicit the pressure dependence:

$$\omega_i(T, P) = k_{\text{coll},i}(T) \frac{P}{k_B T}$$

Now that we have an estimate for the total rate of collisions, we need to develop a model of the effect that these collisions have on the state of the isomer distribution. To this end, we define  $P(E, J, E', J')$  as the probability of a collision resulting in a transfer of a molecule from state  $(E', J')$  to state  $(E, J)$ . There are two mathematical constraints on  $P(E, J, E', J')$ . The first of these is normalization:

$$\sum_{J'} (2J' + 1) \int_0^\infty P(E, J, E', J') dE' = 1$$

The second of these is detailed balance, required in order to obtain the Boltzmann distribution at long times:

$$P(E, J, E', J') b(E', J') = P(E', J', E, J) b(E, J)$$

$$P_i(E', J', E, J) = \frac{\rho_i(E', J')}{\rho_i(E, J)} \exp\left(-\frac{E' - E}{k_B T}\right) P_i(E, J, E', J') \quad E < E'$$

Rather than define models directly for  $P(E, J, E', J')$ , we usually eliminate the angular momentum contribution and instead define  $P(E, E')$ . This can be related to  $P(E, J, E', J')$  via

$$P(E, J, E', J') = P(E, E') \phi(E, J) = P(E, E') (2J + 1) \frac{\rho(E, J)}{\rho(E)}$$

where  $\rho(E) \equiv \sum_J (2J + 1) \rho(E, J)$ .

There are a variety of models used for  $P(E, E')$ . By far the most common is the single exponential down model

$$P(E, E') = C(E') \exp\left(-\frac{E' - E}{\alpha}\right) \quad E < E'$$

where  $C(E')$  is determined from the normalization constraint. Note that this function has been defined for the deactivating direction ( $E < E'$ ) only, as the activating direction ( $E > E'$ ) is then set from detailed balance. The parameter  $\alpha$  corresponds to the average energy transferred in a deactivating collision  $\langle \Delta E_d \rangle$ , which itself is a weak function of temperature.

Other models for  $P(E, J, E', J')$  include the Gaussian down

$$P(E, E') = C(E') \exp\left[-\frac{(E' - E)^2}{\alpha^2}\right] \quad E < E'$$

and the double exponential down

$$P(E, E') = C(E') \left[ (1 - f) \exp\left(-\frac{E' - E}{\alpha_1}\right) + f \exp\left(-\frac{E' - E}{\alpha_2}\right) \right] \quad E < E'$$

The parameters for these simple models generally contain so much uncertainty that more complex functional forms are generally not used.

## Reaction Models

Chemical reaction events cause a change in molecular configuration at constant energy. The rate coefficient for this process must be determined as a function of energy rather than the usual temperature. Such a quantity is called a *microcanonical rate coefficient* and written as  $k(E, J)$ . In the master equation we will differentiate between microcanonical rate coefficients for isomerization, dissociation, and association by using different letters:  $k_{ij}(E, J)$  for isomerization,  $g_{n,j}(E, J)$  for dissociation, and  $f_{im}(E, J)$  for association. (By convention, we use indices  $i$  and  $j$  to refer to unimolecular isomers,  $m$  and  $n$  to refer to bimolecular reactant and product channels, and, later,  $r$  and  $s$  to refer to energy grains.)

As with collision models, the values of the microcanonical rate coefficients are constrained by detailed balance so that the proper equilibrium is obtained. The detailed balance expressions have the form

$$k_{ij}(E, J)\rho_j(E, J) = k_{ji}(E, J)\rho_i(E, J)$$

for isomerization and

$$f_{in}(E, J)\rho_n(E, J) = g_{ni}(E, J)\rho_i(E, J)$$

for association/dissociation, where  $\rho_i(E, J)$  is the density of states of the appropriate unimolecular or bimolecular configuration.

An alternative formulation incorporates the macroscopic equilibrium coefficient  $K_{\text{eq}}(T)$  and equilibrium distributions  $b_i(E, J, T)$  at each temperature:

$$k_{ij}(E, J)b_j(E, J, T) = K_{\text{eq}}(T)k_{ji}(E, J)b_i(E, J, T)$$

for isomerization and

$$f_{in}(E, J)b_n(E, J, T) = K_{\text{eq}}(T)g_{ni}(E, J)b_i(E, J, T)$$

for association/dissociation. Note that these two formulations are equivalent if the molecular degrees of freedom are consistent with the macroscopic thermodynamic parameters. There are multiple reasons to use the latter formulation:

- Only the density of states of the unimolecular isomers need be computed. This is a result of the assumption of thermalized bimolecular channels, which means that we only need to compute the product  $f_{in}b_n$ , and not the individual values of  $f_{in}$  and  $b_n$ .
- Only the reactive rovibrational modes need be included in the density of states. Missing modes will not affect the observed equilibrium because we are imposing the macroscopic equilibrium via  $K_{\text{eq}}(T)$ .
- Constants of proportionality in the density of states become unimportant, as they cancel when taking the ratio  $\rho(E, J)/Q(\beta)$ . For example, if the external rotational constants are unknown then we will include an active K-rotor in the density of states; this property means that the rotational constant of this active K-rotor cancels and is therefore arbitrary.

There are two common ways of determining values for  $k(E, J)$ : the inverse Laplace transform method and RRKM theory. The latter requires detailed information about the transition state, while the former only requires the high-pressure limit rate coefficient  $k_{\infty}(T)$ .

### Inverse Laplace Transform

The microcanonical rate coefficient  $k(E)$  is related to the canonical high-pressure limit rate coefficient  $k_{\infty}(T)$  via a Boltzmann averaging

$$k_{\infty}(T) = \frac{\sum_J \int_0^{\infty} k(E)\rho(E, J)e^{-\beta E} dE}{\sum_J \int_0^{\infty} \rho(E, J)e^{-\beta E} dE}$$

where  $\rho(E, J)$  is the rovibrational density of states for the reactants and  $\beta \equiv (k_{\text{B}}T)^{-1}$ . Neglecting the angular momentum dependence, the above can be written in terms of Laplace transforms as

$$k_{\infty}(T) = \frac{\mathcal{L}[k(E)\rho(E)]}{\mathcal{L}[\rho(E)]} = \frac{\mathcal{L}[k(E)\rho(E)]}{Q(\beta)}$$

where  $Q(\beta)$  is the rovibrational partition function for the reactants. The above implies that  $E$  and  $\beta$  are the transform variables. We can take an inverse Laplace transform in order to solve for  $k(E)$ :

$$k(E) = \frac{\mathcal{L}^{-1}[k_{\infty}(\beta)Q(\beta)]}{\rho(E)}$$

Hidden in the above manipulation is the assumption that  $k_\infty(\beta)$  is valid over a temperature range from zero to positive infinity.

The most common form of  $k_\infty(T)$  is the modified Arrhenius expression

$$k(T) = AT^n \exp\left(-\frac{E_a}{k_B T}\right)$$

where  $A$ ,  $n$ , and  $E_a$  are the Arrhenius preexponential, temperature exponent, and activation energy, respectively. For  $n = 0$  and  $E_a > 0$  the inverse Laplace transform can be easily evaluated to give

$$k(E) = A \frac{\rho(E - E_a)}{\rho(E)} \quad E > E_a$$

We can also determine an expression when  $n > 0$  and  $E_a > 0$  using a convolution integral:

$$k(E) = A \frac{\phi(E - E_a)}{\rho(E)} \quad E > E_a$$

$$\phi(E) = \mathcal{L}^{-1}[T^n Q(\beta)] = \frac{1}{k_B^n \Gamma(n)} \int_0^E (E - x)^{n-1} \rho(x) dx$$

Finally, for cases where  $n < 0$  and/or  $E_a < 0$  we obtain a rough estimate by lumping these contributions into the preexponential at the temperature we are working at. By redoing this at each temperature being considered we minimize the error introduced, at the expense of not being able to identify a single  $k(E)$ .

### RRKM Theory

RRKM theory – named for Rice, Ramsperger, Kassel, and Marcus – is a microcanonical transition state theory. Like canonical transition state theory, detailed information about the transition state and reactants are required, e.g. from a quantum chemistry calculation. If such information is available, then the microcanonical rate coefficient can be evaluated via the equation

$$k(E, J) = \frac{N^\ddagger(E, J)}{h \rho(E, J)}$$

where  $N^\ddagger(E, J)$  is the sum of states of the transition state,  $\rho(E, J)$  is the density of states of the reactant, and  $h$  is the Planck constant. Both the transition state and the reactants have been referenced to the same zero of energy. The sum of states is related to the density of states via

$$N(E, J) = \int_0^E \rho(x, J) dx$$

The angular momentum quantum number dependence can be removed via

$$k(E) = \sum_J (2J + 1) k(E, J)$$

## The Full Master Equation

The governing equation for the population distributions  $p_i(E, J, t)$  of each isomer  $i$  and the reactant concentrations  $y_{nA}(t)$  and  $y_{nB}(t)$  combines the collision and reaction models to give a linear integro-differential equation:

$$\begin{aligned} \frac{d}{dt} p_i(E, J, t) &= \omega_i(T, P) \sum_{J'} \int_0^\infty P_i(E, J, E', J') p_i(E', J', t) dE' - \omega_i(T, P) p_i(E, J, t) \\ &+ \sum_{j \neq i}^{N_{\text{isom}}} k_{ij}(E, J) p_j(E, J, t) - \sum_{j \neq i}^{N_{\text{isom}}} k_{ji}(E, J) p_i(E, J, t) \\ &+ \sum_{n=1}^{N_{\text{reac}}} y_{nA}(t) y_{nB}(t) f_{in}(E, J) b_n(E, J, t) - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} g_{ni}(E, J) p_i(E, J, t) \\ \frac{d}{dt} y_{nA}(t) &= \frac{d}{dt} y_{nB}(t) = \sum_{i=1}^{N_{\text{isom}}} \int_0^\infty g_{ni}(E, J) p_i(E, J, t) dE \\ &- \sum_{i=1}^{N_{\text{isom}}} y_{nA}(t) y_{nB}(t) \int_0^\infty f_{in}(E, J) b_n(E, J, t) dE \end{aligned}$$

A summary of the variables is given below:

Variable	Meaning
$p_i(E, J, t)$	Population distribution of isomer $i$
$y_{nA}(t)$	Total population of species $A_n$ in reactant channel $n$
$\omega_i(T, P)$	Collision frequency of isomer $i$
$P_i(E, J, E', J')$	Collisional transfer probability from $(E', J')$ to $(E, J)$ for isomer $i$
$k_{ij}(E, J)$	Microcanonical rate coefficient for isomerization from isomer $j$ to isomer $i$
$f_{im}(E, J)$	Microcanonical rate coefficient for association from reactant channel $m$ to isomer $i$
$g_{nj}(E, J)$	Microcanonical rate coefficient for dissociation from isomer $j$ to reactant or product channel $n$
$b_n(E, J, t)$	Boltzmann distribution for reactant channel $n$
$N_{\text{isom}}$	Total number of isomers
$N_{\text{reac}}$	Total number of reactant channels
$N_{\text{prod}}$	Total number of product channels

The above is called the two-dimensional master equation because it contains two dimensions: total energy  $E$  and total angular momentum quantum number  $J$ . In the first equation (for isomers), the first pair of terms correspond to collision, the second pair to isomerization, and the final pair to association/dissociation. Similarly, in the second equation above (for reactant channels), the pair of terms refer to dissociation/association.

We can also simplify the above to the one-dimensional form, which only has  $E$  as a dimension:

$$\begin{aligned} \frac{d}{dt} p_i(E, t) &= \omega_i(T, P) \int_0^\infty P_i(E, E') p_i(E', t) dE' - \omega_i(T, P) p_i(E, t) \\ &+ \sum_{j \neq i}^{N_{\text{isom}}} k_{ij}(E) p_j(E, t) - \sum_{j \neq i}^{N_{\text{isom}}} k_{ji}(E) p_i(E, t) \\ &+ \sum_{n=1}^{N_{\text{reac}}} y_{nA}(t) y_{nB}(t) f_{in}(E) b_n(E, t) - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} g_{ni}(E) p_i(E, t) \\ \frac{d}{dt} y_{nA}(t) &= \frac{d}{dt} y_{nB}(t) = \sum_{i=1}^{N_{\text{isom}}} \int_0^\infty g_{ni}(E) p_i(E, t) dE \\ &- \sum_{i=1}^{N_{\text{isom}}} y_{nA}(t) y_{nB}(t) \int_0^\infty f_{in}(E) b_n(E, t) dE \end{aligned}$$

The equations as given are nonlinear, both due to the presence of the bimolecular reactants and because both  $\omega_i$  and  $P_i(E, E')$  depend on the composition, which is changing with time. The rate coefficients can be derived from considering the pseudo-first-order situation where  $y_{nA}(t) \ll y_{nB}(t)$ , and all  $y(t)$  are negligible compared to the bath gas M. From these assumptions the changes in  $\omega_i$ ,  $P_i(E, E')$ , and all  $y_{nB}$  can be neglected, which yields a linear equation system.

## The Energy-Grained Master Equation

Except for the simplest of unimolecular reaction networks, both the one-dimensional and two-dimensional master equation must be solved numerically. To do this we must discretize and truncate the energy domain into a finite number of discrete bins called *grains*. This converts the linear integro-differential equation into a system of first-order ordinary differential equations:

$$\frac{d}{dt} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ y_{1A} \\ y_{2A} \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{K}_{12} & \dots & \mathbf{F}_{11} \mathbf{b}_1 y_{1B} & \mathbf{F}_{12} \mathbf{b}_2 y_{2B} & \dots \\ \mathbf{K}_{21} & \mathbf{M}_2 & \dots & \mathbf{F}_{21} \mathbf{b}_1 y_{1B} & \mathbf{F}_{22} \mathbf{b}_2 y_{2B} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ (\mathbf{g}_{11})^T & (\mathbf{g}_{12})^T & \dots & h_1 & 0 & \dots \\ (\mathbf{g}_{21})^T & (\mathbf{g}_{22})^T & \dots & 0 & h_2 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ y_{1A} \\ y_{2A} \\ \vdots \end{bmatrix}$$

The diagonal matrices  $\mathbf{K}_{ij}$  and  $\mathbf{F}_{in}$  and the vector  $\mathbf{g}_{ni}$  contain the microcanonical rate coefficients for isomerization, association, and dissociation, respectively:

$$(\mathbf{K}_{ij})_{rs} = \begin{cases} \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} k_{ij}(E) dE & r = s \\ 0 & r \neq s \end{cases}$$

$$(\mathbf{F}_{in})_{rs} = \begin{cases} \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} f_{in}(E) dE & r = s \\ 0 & r \neq s \end{cases}$$

$$(\mathbf{g}_{ni})_r = \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} g_{ni}(E) dE$$

The matrices  $\mathbf{M}_i$  represent the collisional transfer probabilities minus the rates of reactive loss to other isomers and to reactants and products:

$$(\mathbf{M}_i)_{rs} = \begin{cases} \omega_i [P_i(E_r, E_r) - 1] - \sum_{j \neq i}^{N_{\text{isom}}} k_{ij}(E_r) - \sum_{n=1}^{N_{\text{reac}} + N_{\text{prod}}} g_{ni}(E_r) & r = s \\ \omega_i P_i(E_r, E_s) & r \neq s \end{cases}$$

The scalars  $h_n$  are simply the total rate coefficient for loss of reactant channel  $n$  due to chemical reactions:

$$h_n = - \sum_{i=1}^{N_{\text{isom}}} \sum_{r=1}^{N_{\text{grains}}} y_{nB} f_{in}(E_r) b_n(E_r)$$

## Further Reading

The interested reader is referred to any of a variety of other sources for alternative presentations, of which an illustrative sampling is given here [Gilbert1990] [Baer1996] [Holbrook1996] [Forst2003] [Pilling2003].

### 3.2.3 Methods for Determining Phenomenological Rate Coefficients

Solving the energy-grained master equation is often prohibitively expensive for chemical reaction mechanisms of even modest size. Instead, we seek to reduce the master equation matrix down to a set of phenomenological rate coefficients

$k(T, P)$ . In particular, we need to replace the isomer population distributions  $p_i(E, t)$  with the corresponding time-dependent total isomer populations  $x_i(t)$ .

Three methods of varying rigor, computational cost, and robustness will be discussed in the upcoming sections. The modified strong collision (MSC) method is the fastest and most robust, but utilizes the least realistic approximations. The reservoir state (RS) method uses better approximations, which leads to increased accuracy, but requires more computational effort. Finally, the chemically-significant eigenvalues (CSE) method is the most theoretically sound, but is very computationally expensive and not very robust. Your choice of method will depend on the particular balance between expense, robustness, and rigor that is required for your intended application.

## A Common Formalism

All of the methods discussed here can be expressed in terms of a common formalism. Each method seeks to express the population distribution vector  $p_i(E, t)$  for each unimolecular isomer  $i$  as a linear combination of the total populations  $x_j(t)$  and  $y_{mA}(t)y_{mB}$  of unimolecular isomers  $A_j$  and reactant channels  $A_m + B_m$ :

$$p_i(E, t) = \sum_{j=1}^{N_{\text{isom}}} x_j(t)u_{ij}(E) + \sum_{m=1}^{N_{\text{reac}}} y_{mA}(t)y_{mB}v_{im}(E)$$

The function  $u_{ij}(E)$  represents the portion of the population distribution of unimolecular isomer  $i$  at energy  $E$  that tracks the population of isomer  $j$ . In the modified strong collision and reservoir state methods, this is because the energy levels of isomer  $i$  are in pseudo-steady-state relationships with isomer  $j$ . The interpretation is a bit different for the chemically-significant eigenvalues method, but the form of the equations is the same. Similarly, the function  $v_{im}(E)$  represents the population distribution of unimolecular isomer  $i$  at energy  $E$  that tracks the population of reactant channel  $m$ . Both functions  $u_{ij}(E)$  and  $v_{im}(E)$  are functions of energy only, and not of time.

After discretizing the energy domain, the above becomes

$$\mathbf{p}_i(t) = \sum_{j=1}^{N_{\text{isom}}} x_j(t)\mathbf{u}_{ij} + \sum_{m=1}^{N_{\text{reac}}} y_{mA}(t)y_{mB}\mathbf{v}_{im}$$

The phenomenological rate coefficients can be constructed from the energy-grained master equation matrix and the vectors  $\mathbf{u}_{ij}$  and  $\mathbf{v}_{im}$ :

$$\begin{aligned} k_{ij}(T, P) &= \sum_{s=1}^{N_{\text{grains}}} (\mathbf{M}_i \mathbf{u}_{ij})_s + \sum_{\ell \neq i}^{N_{\text{isom}}} \sum_{s=1}^{N_{\text{grains}}} (\mathbf{K}_{i\ell} \mathbf{u}_{\ell j})_s \\ k_{im}(T, P) &= \sum_{s=1}^{N_{\text{grains}}} (\mathbf{M}_i \mathbf{v}_{im})_s + \sum_{\ell \neq i}^{N_{\text{isom}}} \sum_{s=1}^{N_{\text{grains}}} (\mathbf{K}_{i\ell} \mathbf{v}_{\ell m})_s + \sum_{s=1}^{N_{\text{grains}}} (\mathbf{F}_{im} \mathbf{b}_m)_s \\ k_{nj}(T, P) &= \sum_{\ell=1}^{N_{\text{isom}}} \mathbf{g}_{n\ell} \cdot \mathbf{u}_{\ell j} \\ k_{nm}(T, P) &= \sum_{\ell=1}^{N_{\text{isom}}} \mathbf{g}_{n\ell} \cdot \mathbf{v}_{\ell m} \end{aligned}$$

Above, the indices  $i$  and  $j$  represent unimolecular isomers of the initial adduct,  $m$  represents bimolecular reactants,  $n$  represents bimolecular reactants and products, and  $s$  represents an energy grain. Thus, the rate coefficients above are for isomerization, association, dissociation, and bimolecular reactions, respectively.

The output from each of the three methods is a set of phenomenological rate coefficients  $k(T, P)$  and the vectors  $\mathbf{u}_{ij}$  and  $\mathbf{v}_{im}$  which can be used to construct the approximate population distribution predicted by that method.

### The Modified Strong Collision Method

### The Reservoir State Method

### The Chemically-Significant Eigenvalues Method

- `genindex`
- `modindex`
- `search`



## BIBLIOGRAPHY

- [GRIMech3.0] Gregory P. Smith, David M. Golden, Michael Frenklach, Nigel W. Moriarty, Boris Eiteneer, Mikhail Goldenberg, C. Thomas Bowman, Ronald K. Hanson, Soonho Song, William C. Gardiner, Jr., Vitali V. Lissianski, and Zhiwei Qin [http://www.me.berkeley.edu/gri\\_mech/](http://www.me.berkeley.edu/gri_mech/)
- [RDKit] RDKit: Open-source cheminformatics; <http://www.rdkit.org>
- [Chang2000] A.Y. Chang, J.W. Bozzelli, and A. M. Dean. “Kinetic Analysis of Complex Chemical Activation and Unimolecular Dissociation Reactions using QRRK Theory and the Modified Strong Collision Approximation.” *Z. Phys. Chem.* **214** (11), p. 1533-1568 (2000).
- [Green2007] N.J.B. Green and Z.A. Bhatti. “Steady-State Master Equation Methods.” *Phys. Chem. Chem. Phys.* **9**, p. 4275-4290 (2007).
- [Benson] Benson, Sidney William. “Thermochemical kinetics.” (1976)
- [Lay] Lay, T.; Bozzelli, J.; Dean, A.; Ritter, E. *J. Phys. Chem.* 1995, **99**, 14514-14527
- [Allinger] Allinger, N. L., & Lii, J.-H. (2008). MM4(2008) and MM4(2003).
- [Patchkovskii] Patchkovskii, S. (2003). SYMMETRY, <http://www.cobalt.chem.ucalgary.ca/ps/symmetry/>.
- [Yu] “Estimation method for the thermochemical properties of polycyclic aromatic molecules” (Ph.D), Joanna Yu, M.I.T (2005)
- [RDKit] Landrum, G. (2012). RDKit, <http://rdkit.org>.
- [Vitha2006] M. Vitha and P.W. Carr. “The chemical interpretation and practice of linear solvation energy relationships in chromatography.” *J. Chromatogr. A.* **1126(1-2)**, p. 143-194 (2006).
- [Abraham1999] M.H. Abraham et al. “Correlation and estimation of gas-chloroform and water-chloroform partition coefficients by a linear free energy relationship method.” *J. Pharm. Sci.* **88(7)**, p. 670-679 (1999).
- [Jalan2010] A. Jalan et al. “Predicting solvation energies for kinetic modeling.” *Annu. Rep. Prog. Chem., Sect. C* **106**, p. 211-258 (2010).
- [Poole2009] C.F. Poole et al. “Determination of solute descriptors by chromatographic methods.” *Anal. Chim. Acta* **652(1-2)** p. 32-53 (2009).
- [Platts1999] J. Platts and D. Butina. “Estimation of molecular linear free energy relation descriptors using a group contribution approach.” *J. Chem. Inf. Comput. Sci.* **39**, p. 835-845 (1999).
- [Mintz2007] C. Mintz et al. “Enthalpy of solvation correlations for gaseous solutes dissolved in water and in 1-octanol based on the Abraham model.” *J. Chem. Inf. Model.* **47(1)**, p. 115-121 (2007).
- [Mintz2007a] C. Mintz et al. “Enthalpy of solvation corrections for gaseous solutes dissolved in benzene and in alkane solvents based on the Abraham model.” *QSAR Comb. Sci.* **26(8)**, p. 881-888 (2007).
- [Mintz2007b] C. Mintz et al. “Enthalpy of solvation correlations for gaseous solutes dissolved in toluene and carbon tetrachloride based on the Abraham model.” *J. Sol. Chem.* **36(8)**, p. 947-966 (2007).

- [Mintz2007c] C. Mintz et al. “Enthalpy of solvation correlations for gaseous solutes dissolved indimethyl sulfoxide and propylene carbonate based on the Abraham model.” *Thermochim. Acta* **459(1-2)**, p. 17-25 (2007).
- [Mintz2007d] C. Mintz et al. “Enthalpy of solvation correlations for gaseous solutes dissolved inchloroform and 1,2-dichloroethane based on the Abraham model.” *Fluid Phase Equilib.* **258(2)**, p. 191-198 (2007).
- [Mintz2008] C. Mintz et al. “Enthalpy of solvation correlations for gaseous solutes dissolved inlinear alkanes (C5-C16) based on the Abraham model.” *QSAR Comb. Sci.* **27(2)**, p. 179-186 (2008).
- [Mintz2008a] C. Mintz et al. “Enthalpy of solvation correlations for gaseous solutes dissolved inalcohol solvents based on the Abraham model.” *QSAR Comb. Sci.* **27(5)**, p. 627-635 (2008).
- [Mintz2009] C. Mintz et al. “Enthalpy of solvation correlations for organic solutes and gasesdissolved in acetonitrile and acetone.” *Thermochim. Acta* **484(1-2)**, p. 65-69 (2009).
- [Rice1985] S.A. Rice. “Diffusion-limited reactions”. In *Comprehensive Chemical Kinetics*, EditorsC.H. Bamford, C.F.H. Tipper and R.G. Compton. **25**, (1985).
- [Gao2016] C. W. Gao, J. W. Allen, W. H. Green, R. H. West, “Reaction Mechanism Generator: automatic construction of chemical kinetic mechanisms.” *Computer Physics Communications* (2016).
- [Susnow1997] R. G. Susnow, A. M. Dean, W. H. Green, P. K. Peczak, and L. J. Broadbelt. “Rate-Based Construction of Kinetic Models for Complex Systems.” *J. Phys. Chem. A* **101**, p. 3731-3740 (1997).
- [Wong2003] B. M. Wong, D. M. Matheu, and W. H. Green. *J. Phys. Chem. A* **107**, p. 6206-6211 (2003). doi:10.1021/jp034165g
- [Lindemann1922] F. A. Lindemann. *Trans. Faraday Soc.* **17**, p. 598-606 (1922).
- [Hinshelwood1926] C. N. Hinshelwood. *Proc. Royal Soc. A* **17**, p. 230-233 (1926). JSTOR:94593
- [Rice1927] O. K. Rice and H. C. Ramsperger. *J. Am. Chem. Soc.* **49**, p. 1617-1629 (1927). doi:10.1021/ja01406a001
- [Kassel1928] L. S. Kassel. *J. Phys. Chem.* **32**, p. 1065-1079 (1928). doi:10.1021/j150289a011
- [Marcus1951] R. A. Marcus and O. K. Rice. *J. Phys. Coll. Chem.* **55**, p. 894-908 (1951). doi:10.1021/j150489a013
- [Siegert1949] A. J. F. Siegert. *Phys. Rev.* **76**, p. 1708-1714 (1949). doi:10.1103/PhysRev.76.1708
- [Bartholomay1958] A. F. Bartholomay. *Bull. Math. Biophys.* **20**, p. 175-190 (1958). doi:10.1007/BF02478297
- [Montroll1958] E. W. Montroll and K. E. Shuler. *Adv. Chem. Phys.* **1**, p. 361-399 (1958).
- [Krieger1960] I. M. Krieger and P. J. Gans. *J. Chem. Phys.* **32**, p. 247-250 (1960). doi:10.1063/1.1700909
- [Gans1960] P. J. Gans. *J. Chem. Phys.* **33**, p. 691-694 (1960). doi:10.1063/1.1731239
- [Widom1959] B. Widom. *J. Chem. Phys.* **31**, p. 1387-1394 (1959). doi:10.1063/1.1730604
- [Keck1965] J. Keck and G. Carrier. *J. Chem. Phys.* **43**, p. 2284-2298 (1965). doi:10.1063/1.1697125
- [Troel1967] J. Troe and H. Gg. Wagner. *Ber. Bunsenges. Phys. Chem.* **71**, p. 937 (1967). doi:10.1002/bbpc.19670710904
- [Troel1973] J. Troe. *Ber. Bunsenges. Phys. Chem.* **77**, p. 665 (1973). doi:10.1002/bbpc.19730770903
- [Tardy1966] D. C. Tardy and B. S. Rabinovitch. *J. Chem. Phys.* **45**, p. 3720-3730 (1966). doi:10.1063/1.1727392
- [Gear1971] C. W. Gear. *Commun. ACM* **14**, p. 176-179 (1971). doi:10.1145/362566.362571
- [Beyer1973] T. Beyer and D. F. Swinehart. *Commun. ACM* **16**, p. 379 (1973). doi:10.1145/362248.362275
- [Stein1973] S. E. Stein and B. S. Rabinovitch. *J. Chem. Phys.* **58**, p. 2438-2444 (1973). doi:10.1063/1.1679522
- [Astholz1979] D. C. Astholz, J. Troe, and W. Wieters. *J. Chem. Phys.* **70**, p. 5107-5116 (1979). doi:10.1063/1.437352
- [Gilbert1990] R. G. Gilbert and S. C. Smith. *Theory of Unimolecular and Recombination Reactions*. Blackwell Sci. (1990).

- [Baer1996] T. Baer and W. L. Hase. *Unimolecular Reaction Dynamics*. Oxford University Press (1996).
- [Holbrook1996] K. A. Holbrook, M. J. Pilling, and S. H. Robertson. *Unimolecular Reactions*. Second Edition. John Wiley and Sons (1996).
- [Forst2003] W. Forst. *Unimolecular Reactions: A Concise Introduction*. Cambridge University Press (2003).
- [Pilling2003] M. J. Pilling and S. H. Robertson. *Annu. Rev. Phys. Chem.* **54**, p. 245-275 (2003).  
[doi:10.1146/annurev.physchem.54.011002.103822](https://doi.org/10.1146/annurev.physchem.54.011002.103822)