First principles based automated kinetic model generation using on-the-fly \textit{ab initio} calculations

Ruben Van de Vijver, Guy B. Marin, Kevin M. Van Geem

Laboratory for Chemical Technology
Introduction

Kinetic models contain many reactions and species
Each species needs **thermodynamic data**
Each reaction needs an accurate **rate coefficient**
This data comes from:
- Experimental measurements
- Theoretic calculations
- Approximation methods

**Problem**: large amount of **data is inaccurate** or is simply **lacking**

**Aim**: develop methodologies to address this data gap by allowing **on-the-fly ab initio calculations** towards accurate thermochemistry and rate coefficients

On-the-fly:
- Without any manual intervention
- As part of the automatic kinetic model generator tool **Genesys**
Molecular representation

Use of **chemoinformatics**:  
Graph representation of species  
No atomic distances, angles or dihedrals
### Need for 3D coordinates

Tabulated (average) bond lengths and angles

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>O</th>
<th>H</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.00</td>
<td>1.25</td>
<td>1.11</td>
<td>1.11</td>
</tr>
<tr>
<td>O</td>
<td>1.21</td>
<td>0.00</td>
<td>2.06</td>
<td>2.06</td>
</tr>
<tr>
<td>H</td>
<td>1.07</td>
<td>1.95</td>
<td>0.00</td>
<td>1.94</td>
</tr>
<tr>
<td>H</td>
<td>1.07</td>
<td>1.95</td>
<td>1.83</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Upper bound** $d_u$

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>O</th>
<th>H</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.00</td>
<td>1.23</td>
<td>1.10</td>
<td>1.09</td>
</tr>
<tr>
<td>O</td>
<td>1.23</td>
<td>0.00</td>
<td>2.00</td>
<td>2.04</td>
</tr>
<tr>
<td>H</td>
<td>1.10</td>
<td>2.00</td>
<td>0.00</td>
<td>1.84</td>
</tr>
<tr>
<td>H</td>
<td>1.09</td>
<td>2.04</td>
<td>1.84</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Lower bound** $d_l$

$$d = d_l + r(d_u - d_l)$$

$$0 < r < 1$$

**Distance geometry** algorithm

1. Start from the **connectivity** of the molecule
2. Build a **distance bounds matrix**
3. Choose a random **distance matrix** which meets the bounds
4. Embed the coordinates based on the distance matrix

**Distance matrix**

$$\sum_{l=i<j}^{N,N} (w_{i,j} (\|x_i - x_j\|^2 - d_{ij}^2))^2$$

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>O</th>
<th>H</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.00</td>
<td>-0.53</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.00</td>
<td>0.67</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>0.00</td>
<td>-0.12</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>0.00</td>
<td>-0.12</td>
<td>0.94</td>
<td></td>
</tr>
</tbody>
</table>
Transition states

Values for bond lengths and bond angles of the atoms that change in connectivity?

Employ a user-defined template

Templates originate from

Vast literature data on transition state geometries

Simple “manual” transition state searches

Same distance geometry algorithm can be used

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
<th></th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>133</td>
<td>135</td>
<td>α</td>
<td>139°</td>
<td>147°</td>
</tr>
<tr>
<td>b</td>
<td>137</td>
<td>140</td>
<td>β</td>
<td>105°</td>
<td>105°</td>
</tr>
<tr>
<td>c</td>
<td>129</td>
<td>132</td>
<td>γ</td>
<td>123°</td>
<td>114°</td>
</tr>
<tr>
<td>d</td>
<td>184</td>
<td>189</td>
<td>δ</td>
<td>102°</td>
<td>100°</td>
</tr>
<tr>
<td>e</td>
<td>139</td>
<td>139</td>
<td>ε</td>
<td>122°</td>
<td>117°</td>
</tr>
<tr>
<td>f</td>
<td>141</td>
<td>142</td>
<td>ζ</td>
<td>104°</td>
<td>99°</td>
</tr>
</tbody>
</table>
Steps in gaussian

Both for **species and transition states**

1. Pre-optimization, fast, low level (PM3)
2. Conformational search (B3LYP/3-21G*)
   a) Open chain species: exhaustive
   b) Ring structures not
   c) Up to 8 rotors (6561 conformers)
3. High level optimization (CBS-QB3)
   With lowest 5 kJ mol⁻¹
4. Hindered rotor calculations
   B3LYP/6-31G(D)
Verifying transition states

**Visualization** of normal mode corresponding to imaginary frequency

Intrinsic reaction coordinate (IRC) calculations

In Genesys: in two fold

1. User-supplied **template**
2. Normal mode analysis

```
\begin{align*}
A & \longrightarrow B & C \\
2.31 & \quad & 1.35 \\
2.33 & \quad & 1.35
\end{align*}
```

Normal mode in Cartesian coordinates

Translate to **internal coordinates** via the Wilson B Matrix

Contribution of the length of forming/breaking bonds to the normal mode
1D Hindered rotor calculations

Approximation: all internal rotations can be **uncoupled**

**Relaxed** scans by changing one dihedral angles

Optionally **semi-relaxed** by fixing other dihedrals or bond lengths

Selecting a **vibrational frequency** corresponding to the internal rotation

Normal mode analysis

![Energy plot](image)

Translate to **internal coordinates**
via the Wilson B Matrix

Contribution of the dihedral angle to the normal mode

Success rate $\approx 90\%$
Standard enthalpy of formation

278 molecules
Up to 12 heavy atoms
H, C, N, O and S

<table>
<thead>
<tr>
<th></th>
<th>Without BACs</th>
<th>With BACs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAD</td>
<td>7.64</td>
<td>2.28</td>
</tr>
</tbody>
</table>

Including bond additive corrections (BACs)
Entropy and heat capacity

\[ S = S_{AI} - R \ln(\sigma_{\text{ROT}}) \]

\[ \sigma_{\text{ROT, ext}} = 2 \]

\[ \sigma_{\text{ROT, int}} = 3 \]

\[ \text{Success rate } > 90\% \]
High pressure limit rate coefficients

\[ k(T) = n_e \kappa(T) \frac{k_B T (V_m)^{-\Delta n^*}}{h} \exp \left( \frac{\Delta S^*}{k_B} \right) \exp \left( -\frac{\Delta H^*}{k_B T} \right) \]

<table>
<thead>
<tr>
<th></th>
<th>Inter</th>
<th>Intra</th>
<th>Inter</th>
<th>Intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>135</td>
<td>135</td>
<td>α</td>
<td>110°</td>
</tr>
<tr>
<td>b</td>
<td>231</td>
<td>231</td>
<td></td>
<td>100°</td>
</tr>
</tbody>
</table>
High pressure limit rate coefficients

\[ \begin{align*}
\text{Inter} & \quad \text{Intra} \\
\alpha & \quad 179° \quad 153° \\
\alpha & \quad 197° \\
\end{align*} \]

\[ \begin{align*}
\kappa(300K) & \quad 37.1 \quad 25.1 \\
\kappa(1000K) & \quad 1.29 \quad 1.31 \\
\end{align*} \]
Rate coefficients

{Graphs showing rate coefficients for different reactions at various temperatures.}

1. Reaction a: SH + CH₃ → + SH
2. Reaction b: OH → + H₂O
3. Reaction c: + H₂
4. Reaction d: + O₂
## Rate coefficients

<table>
<thead>
<tr>
<th>Reaction</th>
<th>A (s, m, mol)</th>
<th>n</th>
<th>E (kJ, mol)</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{O} + \text{CH}_3 \rightarrow \text{H}_2 \text{O} )</td>
<td>6.6 (10^{-2} )</td>
<td>2.37</td>
<td>23.0</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>5.7 (10^{10} )</td>
<td>0.97</td>
<td>123.3  </td>
<td> </td>
</tr>
<tr>
<td>( \text{H}_2 \text{O} + \text{CH}_3 \rightarrow \text{H}_2 \text{O} )</td>
<td>4.0 (10^{7} )</td>
<td>0.70</td>
<td>29.1</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>8.0 (10^{11} )</td>
<td>0.77</td>
<td>124.3  </td>
<td> </td>
</tr>
<tr>
<td>( \text{O} + \text{O} \rightarrow \text{O}_2 )</td>
<td>9.1 (10^{-9} )</td>
<td>4.48</td>
<td>32.0</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>3.1 (10^{-10} )</td>
<td>4.83</td>
<td>48.3  </td>
<td> </td>
</tr>
<tr>
<td>( \text{O} + \text{H}_2 \rightarrow \text{H}_2 \text{O} )</td>
<td>2.6 (10^{2} )</td>
<td>2.56</td>
<td>64.4</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>6.0 (10^{1} )</td>
<td>2.40</td>
<td>11.7  </td>
<td> </td>
</tr>
<tr>
<td>( \text{H}_2 \text{O} + \text{CH}_3 \rightarrow \text{H}_2 \text{O} )</td>
<td>5.6 (10^{-1} )</td>
<td>1.90</td>
<td>50.6</td>
<td> </td>
</tr>
<tr>
<td></td>
<td>1.1 (10^{-3} )</td>
<td>2.64</td>
<td>50.5  </td>
<td> </td>
</tr>
<tr>
<td>( \text{H}_2 \text{O} + \text{H}_2 \rightarrow \text{H}_2 \text{O} )</td>
<td>8.2 (10^{4} )</td>
<td>2.50</td>
<td>268.</td>
<td> </td>
</tr>
<tr>
<td></td>
<td>1.5 (10^{-10} )</td>
<td>4.52</td>
<td>224.1  </td>
<td> </td>
</tr>
<tr>
<td>( \text{H}_2 \text{O} + \text{H}_2 \rightarrow \text{H}_2 \text{O} )</td>
<td>8.7 (10^{7} )</td>
<td>1.46</td>
<td>228.5</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.3 (10^{-4} )</td>
<td>2.70</td>
<td>125.0  </td>
<td> </td>
</tr>
<tr>
<td>( \text{H}_2 \text{O} + \text{H}_2 \rightarrow \text{H}_2 \text{O} )</td>
<td>6.5 (10^{3} )</td>
<td>2.29</td>
<td>170.4</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>1.8 (10^{-11} )</td>
<td>4.31</td>
<td>148.8  </td>
<td> </td>
</tr>
</tbody>
</table>
Concluding remarks

**Summary:**
Fully automated ab initio calculations, **minimal manual interventions**
Thorough conformational search and 1D hindered rotor approximation
Ideal gas statistical **thermodynamics**
High-pressure limit **rate coefficients**
Satisfactory comparison to experimental and theoretical data

**Future work:**
Conformational analysis when many (>8) torsional modes exist
More robust rotational frequency selection and rotational symmetry perception
Acknowledgments

• The Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).
• The Research Board of Ghent University.
• The European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 290793.
• The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government – department EWI.
LABORATORY FOR CHEMICAL TECHNOLOGY

Technologiepark 914, 9052 Ghent, Belgium

E  ruben.vandevijver@ugent.be
T  +32 9 264 56 75

https://www.lct.ugent.be
Additional remarks on distance geometry

Distance bounds matrix contains all conformers

\[
\begin{array}{ccc}
O & O & H & H \\
O & 0.00 & 1.21 & 0.97 & 1.79 \\
O & 1.19 & 0.00 & 1.79 & 0.97 \\
H & 0.95 & 1.75 & 0.00 & 2.61 \\
H & 1.75 & 0.95 & 1.82 & 0.00 \\
\end{array}
\]

Refinement of the coordinates

\[\phi = 16.2^\circ\]
Ideal gas statistical thermodynamics

\[ q = q_{\text{trans}} \cdot q_{\text{rot}} \cdot q_{\text{vib}} \cdot q_{\text{elec}} \]

\[ q_{\text{vib}} = \prod_{i=1}^{3N-6-n_{\text{rot}}} q_{\text{vib},i} = \prod_{i=1}^{3N-6-n_{\text{rot}}} \frac{e^{-\frac{\hbar \nu_i}{2k_B T}}}{1 - e^{-\frac{-\hbar \nu_i}{k_B T}}} \]

\[ \frac{1}{I_{\text{red}}} = \frac{1}{I_{\text{top},1}} + \frac{1}{I_{\text{top},2}} \]

\[ I_{\text{top},1} = \sum_{k=1}^{N_1} m_k d_{A,k}^2 \]

\[ q_{\text{rot},i} = \frac{1}{\sigma_i} \sum_{k=1}^{m} g_k e^{-\epsilon_{ki}/k_B T} \]

\[ S = S_{AI} - R \ln (\sigma_{\text{ROT}}) \]

Success rate > 90%
Writing of **input files**

Choosing appropriate computational resources, keywords and options

Submitting the calculations to a **scheduler**

Retrieving and interpreting the **output files**

In case of an **error**

- Gaussian error: adjust keywords and options, optionally freeze coordinates
  - Number of optimization steps, coordinate system, grid size, etc.
- Computer error: adjust computational resources
  - Wall time, memory and number of cores