Chemistry by numbers
or: How I learned to stop worrying
and love computational chemistry.

James C. Womack

University of Bristol

04 July 2012
Outline

▶ Computational chemistry.
▶ Method development.
▶ My own research.

Photo: “Red Onion Slice” by photobunny (Flickr)
What is computational chemistry?

Using computers to study chemistry!

- Many different approaches, e.g. informatics, simulation.

**Theoretical chemistry:** “the subfield [of chemistry] where mathematical methods are combined with fundamental laws of physics to study processes of chemical relevance” [1].

**Quantum chemistry:**

Electrons  Nuclei  Constants  Environment  \[ \Psi \]  Thermochemistry  Spectral properties  Reactivity

![Diagram]

Why use computational chemistry?

In place of lab work:

▸ Eliminate the unprofitable or irrelevant.
  ▸ e.g. virtual screening of drug candidates [1].

▸ Some systems **inaccessible to experiment**.
  ▸ e.g. astrochemistry [2], radioactive decay [3].

---

Why use computational chemistry?
To enhance understanding and gain new insights:

- Inspire and inform new research.
- Can “play” with a system easily.
- Facilitate rational design of approaches and products.
  - e.g. drug design (raltegravir) [1]

![Raltegravir structure](image)

- e.g. catalyst design (Mannich) [2]

![Mannich reaction](image)


Method development: faster, better science

Motivations:
- Faster calculations.
- Greater accuracy.
- New systems and situations.
- Improved practicality / accuracy compromise.

... more science per Watt!

Image: BlueCrystal supercomputer, University of Bristol.
Method development: example

Accurate enzyme reaction barriers with QM/MM methods [1].

Image: Hen egg-white lysozyme in solvent (Mike Limb, University of Bristol).
Method development: example

Modelling the QM region:

▶ Semi-empirical methods, error \( \approx 10 \text{ kcal mol}^{-1} \).
▶ DFT underestimates barrier heights (several kcal mol\(^{-1}\)).
▶ Higher level methods accurate but too expensive.
▶ Local methods allow high level calculations at lower expense.

<table>
<thead>
<tr>
<th>Method</th>
<th>CM</th>
<th>PHBH</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFT</td>
<td>10.2</td>
<td>8.4</td>
</tr>
<tr>
<td>Local CCSD(T0)</td>
<td>13.1</td>
<td>13.3</td>
</tr>
<tr>
<td>Experiment</td>
<td>12.7</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Averaged activation enthalpies (300K), kcal mol\(^{-1}\)

Improved accuracy allows...

“quantitative studies of reaction mechanisms in enzymes”.

Table adapted from [1]. CM = chorismate mutase, PHBH = \textit{para}-hydroxybenzoate hydroxylase.
My research: electron correlation

Quantum chemistry: find $E$ by solving $\hat{H}\Psi = E\Psi$

Electron-electron correlation is important to describe chemistry!
- Bond dissociation.
- Dispersion interactions.

...important for accurate $E$.

Conventional wavefunction, $\Psi$, based methods:
- Electron coordinates are independent.
  - Inefficient for describing correlation.
- Poor scaling of cost with error in correlation energy.
  - Reducing the error by a factor of 10 requires a 10000-fold increase in computer time [1].

My research: explicitly-correlated methods

Explicitly correlated methods:

- Electron coordinates are not independent.
- Electron correlation “built in” to wavefunction.
- Improved scaling of cost with error in correlation energy.

BUT: Many-electron integrals arise:

- All methods require 1- and 2-electron integrals.
- Many-electron integrals are costly and numerous.
My research: approximating integrals

Current widely used method: resolution of the identity (RI) [1]

\[
\langle ijm | f_{12} f_{23} | mkl \rangle \approx \sum_p \langle ij | f_{12} | mp \rangle \langle pm | f_{23} | kl \rangle
\]

\[O(N^6), \text{3} \text{occ} \] [2]

Alternative method: density fitted orbital pairs [3]

\[
(im | f_{12} | jk | f_{23} | ml) \approx \sum_{A,B,C} D^i_A D^j_B D^m_C \langle A | f_{12} | B | f_{23} | C \rangle
\]

\[O(N^5), \text{2} \text{occ} \] [2]


My research: implementation

▶ Derive many-electron integrals without RIs.
▶ Derive approximate density-fitted forms of the integrals.
▶ Write code to generate the integrals in Molpro [1].
▶ Run calculations using the new integral approximation.
▶ Compare against existing methods.

Acknowledgements

I would like to thank the following people and organizations:

▶ Prof Fred Manby
▶ Prof Neil Allan
▶ Prof Adrian Mulholland
▶ Dr Natalie Fey
▶ Colleagues at the Centre for Computational Chemistry.
▶ SCI College of Scholars
▶ EPSRC
▶ University of Bristol
# Appendix: Activation enthalpies (300K) (kcal mol\(^{-1}\))

<table>
<thead>
<tr>
<th>Method</th>
<th>CM</th>
<th>PHBH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hartree-Fock</td>
<td>28.3 (2.1)</td>
<td>36.7 (2.6)</td>
</tr>
<tr>
<td>B3LYP</td>
<td>10.2 (1.8)</td>
<td>8.4 (1.4)</td>
</tr>
<tr>
<td>LMP2</td>
<td>9.5 (1.0)</td>
<td>10.7 (1.2)</td>
</tr>
<tr>
<td>LCCSD(T0)</td>
<td>13.1 (1.1)</td>
<td>13.3 (1.5)</td>
</tr>
<tr>
<td>Experiment</td>
<td>12.7</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Activation enthalpies from averages of energy differences from single-point QM/MM calculations for the reactant complex and the TS on different adiabatic pathways. aug-cc-pVTZ basis used on oxygen and cc-pVTZ on all other atoms; point-charge representation of the MM environment was included in the QM calculations.

---

CM = chorismate mutase, PHBH = *para*-hydroxybenzoate hydroxylase
Table adapted from slide provided by A. J. Mulholland, University of Bristol
Appendix: Computational science at Bristol

- BlueCrystal:
  - 3360 2.6 GHz x86 processor cores.
  - Some GPGPU and large memory nodes.
  - >600 users across the university.

- Uses for chemistry:
  - Docking simulations.
  - Molecular dynamics.
  - Climate modelling.
  - Quantum chemistry.
  - Reaction dynamics.

BlueCrystal information and image: Dr Ian Stewart, Director of Advanced Computing, University of Bristol.