NIST
Computational Chemistry Comparison and Benchmark Database

A website which compares computed and experimental ideal-gas thermochemical properties
http://srdata.nist.gov/cccbdb

Russell D. Johnson III
Computational Chemistry Group, NIST
http://www.nist.gov/compchem
A few of the online databases through NIST Standard Reference Data Division

http://www.nist.gov/srd/online.htm

NIST Chemistry WebBook
Fundamental Physical Constants
PDB: The Protein Data Bank
Enzyme Catalyzed Reactions
NIST Ceramics WebBook
## NIST Chemistry WebBook


<table>
<thead>
<tr>
<th>Thermodynamic Data</th>
<th>Other Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas phase</td>
<td>Gas phase IR spectrum</td>
</tr>
<tr>
<td>Phase change</td>
<td>Mass spectrum</td>
</tr>
<tr>
<td>Condensed phase</td>
<td>UV/Vis spectrum</td>
</tr>
<tr>
<td>Reaction</td>
<td>Vibrational and/or electronic spectrum</td>
</tr>
<tr>
<td>Ion energetics</td>
<td>Constants of Diatomic Molecules</td>
</tr>
<tr>
<td>Ion cluster</td>
<td>Henry's Law</td>
</tr>
<tr>
<td></td>
<td>Thermophysical Properties of Fluid Systems</td>
</tr>
</tbody>
</table>

NIST Computational Chemistry Comparison and Benchmark Database  [http://srdata.nist.gov/ccccbdb](http://srdata.nist.gov/ccccbdb)
Background - Data needs for modeling chemical processes

• Modeling is becoming more important.
  – Faster computers.
• Models need data
  – (which is often unavailable).

Need methods to estimate data!
Background –
Estimation methods

• When experimental data is lacking, estimation methods are used, but the accuracy of the methods is unknown.
  – Without knowledge of the accuracy, their value in process simulation is diminished.

• Estimation schemes can be validated by comparing with accurate experimental measurements for prototypical systems.
Background - Computational Chemistry

- Computational chemistry is becoming a valued estimation method for thermochemical properties (enthalpies of formation, etc.), but the accuracy of the estimates depends on the computational methods being used. Typically higher accuracy requires more "expensive" computational methods.

Cost vs. Accuracy
Background - Industrial needs

• The 1996 ACS symposium on Computational Thermochemistry revealed needs:
  – A database of critically evaluated experimental thermochemistry.
  – This database needs about 500 species for the testing and comparison of computational methods.
NIST CCCBDB

- WWW interface.
- Benchmark thermochemical data, both experimental and computational.
- Tools for comparing experimental and computational data.
NIST CCCBDB

• Set of 615 gas-phase species.
  – No more than 6 heavy (non-hydrogen)atoms.
  – No more than 20 total atoms.
  – No atoms with atomic number > 17 (Chlorine).
  – No ions or noble gases.
  – Uncertainty of Enthalpy of formation < 10kJ/mol.
Description

The Computational Chemistry Comparison and Benchmark Database (CCCBDB) allows process modelers to analyze the accuracy of well-defined computational methods for estimating thermochemical properties.
Error Bars

• Estimation methods need error bars.
  – Using the CCCBDB error distributions can be obtained for selected methods applied to selected classes of molecules or reactions.
  – The process modeler can then choose the least expensive computational chemistry estimation method that provides the desired accuracy for the system under investigation.
  – Error bars can be assigned to the estimates, which in turn allows reliable analysis of design parameters in process modeling.
CCCBDB Contents

• Experimental data
• Computational data
• Comparison methods
• Costs ( timings )
• Auxiliary information
  – Description of computational techniques
  – Primer on Statistical mechanics
  – Description of energy units
  – Why the vibrational frequencies are scaled
  – Designing isodesmic reactions
Experimental Data in CCCBDB

- Enthalpy of formation ($\Delta_f H$) at 0K and 298.15K
- Entropy (S) at 298.15K
  - Mass
  - Vibrational frequencies
  - Geometries
  - Rotational Constants
  - Excited and/or degenerate electronic states
- Heat Capacities
- Dipole Moments
Computational Data in CCCBDB

- **Energies**
  - Electronic
  - Vibrational
  - Torsional barriers (internal rotation)

- **Geometries**
  - Cartesians

- **Derived Values**
  - Vibrational zero-point energies
  - Atomization Enthalpies
  - Entropies
  - Products of moments of inertia
# Computations in CCCBDB

<table>
<thead>
<tr>
<th>Methods</th>
<th>Basis Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM1, PM3</td>
<td>3-21G, 3-21G*</td>
</tr>
<tr>
<td>HF, ROHF</td>
<td>6-31G, 6-31G*</td>
</tr>
<tr>
<td>MP2(full), MP2(frozen core), MP4</td>
<td>6-311G*</td>
</tr>
<tr>
<td>BLYP, B3LYP, B3PW91, MPW1PW91</td>
<td>6-31G**</td>
</tr>
<tr>
<td>CID, CISD</td>
<td>6-31+G**</td>
</tr>
<tr>
<td>QCISD, QCISD(T)</td>
<td>cc-pVDZ</td>
</tr>
<tr>
<td>CCD, CCSD, CCSD(T)</td>
<td>cc-pVTZ</td>
</tr>
<tr>
<td>G1, G2, G2MP2, CBS-Q</td>
<td>ECPs</td>
</tr>
</tbody>
</table>

NIST Computational Chemistry Comparison and Benchmark Database  [http://srdata.nist.gov/ccccdb](http://srdata.nist.gov/ccccdb)
Comparing Apples and Oranges

- Experiment - measure a reaction, such as combustion:
  \[ \text{CH}_4 + 2\text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O} \]

- Ab initio calculations provide an energy for separating a molecule into nuclei and electrons:
  \[ \text{CH}_4 \rightarrow \text{C}^{6+} + 4\text{p} + 10\text{e}^- \]

- Must define reactions to compare experiment and theory.
  - Atomization, although a poor choice, is often used.
  - A reaction which keeps the kinds of bonds conserved leads to the canceling of systematic errors.
Comparison Methods

- Atomization energies
- User-specified reaction enthalpies
- Compare isomer energies
- Compare entropies
- Compare vibrational frequencies
  - Calculate a least-squares best scaling factor
- Compare geometries
- Semi-empirical methods yield enthalpies of formation directly
Costs (time)

Optimization Times for CH3OH (1 cycle Cs symmetry)

Time (sec) vs Basis Functions

- HF
- B3LYP
- MP2
- MP4
- CCD
- CCSD
- QCISD
- CCSD(T)
Questions that arise from the data

• Why is *that* number for *that* molecule so far off?
  – Error in experimental value
  – Error in data manipulation (e.g. typos)
  – Computational method inadequate for that species
    • Multiconfiguration problems
    • Insufficient basis set
Sicklist database

- Database of molecules and particular computational methods which produce erroneous results.
  - [http://srdata.nist.gov/sicklist](http://srdata.nist.gov/sicklist)
- Open for contributions and comments
CCCDBDB Progress

• What we have:
  ▪ 24 methods, 16 basis sets
  ▪ Comparison methods
  ▪ Input and Output files

• What we plan on having:
  – Empirical group additivity methods
  – Transition States, NMR shifts, charges
  – Combination methods
    • basis set extrapolation
    • user-supplied combinations, similar to G2
  – Tutorials: Educate the users on how to use the data!
The user chooses the property of interest.
The user chooses the molecule or molecules.
The CCCBDB returns the comparison with experiment.
The user can choose finer details (more information) for a particular method/basis set.
Guts

Almost any browser

internet

Windows NT Server (IIS4)

Active Server Pages (ASP)

Microsoft ACCESS database

NIST Computational Chemistry Comparison and Benchmark Database  http://srdata.nist.gov/cccdb
## Enthalpy of formation of Glyoxal

### Reaction Energy differences (kJ/mol)

<table>
<thead>
<tr>
<th>Reactants</th>
<th>Products</th>
<th>Reaction Energy</th>
<th>HF/6-31G*</th>
<th>B3LYP/6-31G*</th>
<th>MP2/6-31G*</th>
<th>G2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₂H₂O₂</td>
<td>2C+2H+2O</td>
<td>2,579.7 ± 1.0</td>
<td>-904</td>
<td>-14</td>
<td>-113</td>
<td>10</td>
</tr>
<tr>
<td>C₂H₂O₂</td>
<td>2 HCO</td>
<td>296.0 ± 7.1</td>
<td>-64</td>
<td>-7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>C₂H₂O₂</td>
<td>2 H₂CO</td>
<td>-5.4 ± 1.1</td>
<td>-16</td>
<td>1</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>C₂H₂O₂ + H₂</td>
<td>2 H₂CO</td>
<td>59.8 ± 1.2</td>
<td>-2</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>C₂H₂O₂ + CH₄ + C₂H₆</td>
<td>2 H₂CO</td>
<td>59.8 ± 1.2</td>
<td>-2</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
Computational Chemistry Comparison and Benchmark Database
http://srdata.nist.gov/cccbdb

• Purpose:
  – Compare calculated and experimental ideal-gas thermochemical data.
  – Provide benchmark data for evaluating theoretical methods.

Contents:
  Enthalpy of formation, entropies, geometries, vibrational frequencies
  615 species
  >50,000 calculations
  >16 comparison options
    Enthalpy of reaction
    Entropy of reaction
    Geometry comparisons

Activity:
  >4000 web pages served/month