VT-2005 Sigma Profile Database

A detailed tutorial for generating sigma profiles using Accelrys’ Materials Studio v3.2 software package

Eric Mullins, Y.A. Liu, Richard Oldland, Shu Wang, Stanley I. Sandler, Chau-Chyun Chen, Michael Zwolak, and Kevin Seavey

Honeywell Center of Excellence in Computer-Aided Design
SINOPEC/FPCC/AspenTech Center of Excellence in Process Systems Engineering

Department of Chemical Engineering, Virginia Tech
Blacksburg, VA 24060

www.design.che.vt.edu/VT-2005.html
Objectives

• An Introduction to Sigma Profiles, P(\(\sigma\))
• Using Accelrys’ Materials Studio (MS)
  – Detailed procedure for creating a molecule
  – Detailed procedures for optimizing molecular geometry and performing COSMO calculations
• Overview of the Sigma Profile averaging program
Sigma Profiles at a Glance

- Sigma Profiles, $P(\sigma)$, depict the surface charge density distribution over the entire molecule.
- Graphically: Area vs. Charge/Area
- Profiles are available with our database.
- Each molecule’s profile is unique.
- Profiles are sensitive to conformation
COSMO-based models used sigma profiles to predict physical properties.

- Solubility (SLE)
- Vapor-Liquid (VLE)
- Partition coefficients ($K_{ow}$)
- pKa
The VT-2005 sigma profile database currently contains 1266 compounds.

- Compounds include alkanes, alkenes, alkynes, alcohols, amines, acids, aldehydes, aromatics, epoxides, esters, ethers, fluorocarbons, chlorocarbons, ketones and others.
- Compounds consist of the following atoms: N, C, O, H, F, Cl, P, S, I, and Br.
Using Accelrys’ Materials Studio (MS), we will construct our model molecule.

- We will use propane for our example.
- Molecules can be drawn manually or downloaded from other online sources as *.mol files and imported into MS.
  - NIST database
    http://webbook.nist.gov/chemistry/
  - Scifinder™
    http://www.cas.org/SCIFINDER/SCHOLAR/index.html
  - National Library of Medicine
We begin by creating a new project in MS, and then opening a new molecule window.

- Select a new “3D Atomistic Document.”
- Right click on the “3D Atomistic Document” in the project window and select “Rename.”
- Type “Propane” and hit Enter.

www.design.che.vt.edu/VT-2005.html
We display the molecule with the “Ball and Stick” option.

- Right click in the molecular window and select “Display Style.”
- Select “Ball and Stick” in the display style window and then close the window.
Use the sketch tool to draw propane (C₃H₈)

- Select the “Sketch Atom” tool from the toolbar.
- With the drop-down menu, select the Carbon atom.
Draw the three carbon atoms, but do not draw the hydrogen atoms for now.

- Left click once to place the first atom. Move your cursor and click again; it will automatically create a second atom connected with a single bond.
- Draw a three carbon backbone.
- Hitting “Esc” will stop adding atoms.
- In the instance of double bonds, simply clicking on the single bond will change its bond type.
The program automatically adds the required hydrogen for each molecule with the “Adjust Hydrogen” tool.

- Click the “Adjust Hydrogen” icon from the toolbar to add the hydrogen atoms.
Highlight the entire molecule and change the properties filter to “molecule.”

- Changing the filter option to molecule shows various properties like the mass, number of atoms, and the chemical formula.
- Quickly verify the chemical formula is correct.
Now we use the “Clean” tool to correct the bond lengths and bond angles.

- Left click on the “Clean” tool icon on the toolbar.
- The “Clean” tool does a rough geometry optimization and adjusts the bond lengths and angles to the correct values for each atom and bond type.
- We are now finished constructing our example molecule, Propane.
Now we move onto setting up the Geometry Optimization calculation.

• Select the “DMol3 calculation” from the “Modules” Menu or use the icon on the toolbar.
• The calculation dialog box will appear.
Select the task option in the DMol3 calculation window.

- Geometry optimization fixed the atomic coordinates for the molecule by minimizing the total energy of the molecule.
Adjust the Geometry Optimization tolerance to “Fine.”

- This can be done in the Setup or Electronic tab.
- This is the accuracy of the Hamiltonian matrix element convergence.
- This specifies the accuracy to which the SCF (Self-Consistent Field) equations are converged. The “Fine” setting is recommended by the software documentation for highly accurate geometry optimizations. It represents a convergence of $10^{-6}$. 
Select DNP as the Basis Set in the Electronic tab.

- DNP = Double Numerical basis with Polarization functions, i.e., functions with angular momentum one higher than the highest occupied orbital in the free atom.
- According to the software documentation, “minimal basis sets are generally inadequate for anything but qualitative results, while DNP sets are the most reliable.” The DNP option instructs the program to ignore extraneous functions that “eliminate” certain atomic orbitals.
- When Quality it set to “Fine,” DNP is the default basis set.
Select GGA (local correlation) and VWN-BP (gradient-corrected functional) for the Functional option in the Setup tab.

- GGA = Generalized Gradient Approximation.
Select the “More” button in the Job Control tab.

- Check the “Retain server files” box. This will save the files to the hard drive in the “jobs” directory.
Click the “Files” button at the bottom of the DMol3 calculation dialog box.

• Click the “Save Files” button. This will create the input files necessary for the calculation. They are visible in the Project window.
Open “propane.input” from the Project window.

- Insert the code “Basis_version v4.0.0” into the input file.
- This statement instructs the program to use version 4.0.0 DNP basis set instead of the default v3.5 DNP basis set. The software documentation recommends v4.0.0 for COSMO calculations.
Leave the “Spin Unrestricted” box unchecked in the Setup tab.

- Checking the “Spin Unrestricted” is necessary when working with radicals, charged molecules, and organo-metallics.
- Although it is possible to run an organic molecule with the “Spin Unrestricted” setting, it will be much slower computationally.
- Checking the “Symmetry” box in the setup tab can also apply to certain molecules, but not all.
Click the “Run Files” button to begin the DMol3 Geometry Optimization.

- The geometry optimization predicts the energy level of the molecule in the ideal gas phase.
- This calculation is the longest step of the procedure and will require ~75% of the time to produce a sigma profile.
The Project window will show the files created and tracks the progress of the calculation.

- The *.xsd file is the optimized geometry. All further calculations should use this file.
- The *.outmol file will show the ideal gas phase energy. There are additional files that do not show up in the Project window but are stored in the “Jobs” folder. (C:\Program Files\Accelrys\MS Modeling 3.2\Gateway\root_default\dsd\jobs)
We have now completed the Geometry Optimization and will proceed to the COSMO calculation.

- Begin by opening the geometry optimization output, *vt-0003.xsd*, and open another DMol3 calculation dialog box.
- Select the task, “Energy.”
- Leave all other options the same as the geometry optimization.
Click the “Files” button in the calculation window.

- Click the “Save Files” button to create the input files.
Open the propane.input file and add the COSMO calculation keywords.

- The COSMO keywords tell the program which parameters to use in the COSMO calculation including atomic radii, basis set version, etc.
Now we review the parameters and settings in the COSMO keywords.

- This statement turns the COSMO calculation on.
- Again, we will use basis set version 4.0.0 instead of the default v3.5.
- The first number is the atomic number and the second term is the fitted atomic radii. Klamt (1998) fitted these parameters to several sets of data.
Now we will run the COSMO calculation.

- Click on the “Run Files” button from the calculation window to begin the energy calculation.
- Upon completion, the Project window will contain a *.cosmo and *.outmol file. There are additional files stored in the “Jobs” folder.
The *.cosmo file contains much useful information.

- This file contains the volume of the cavity around the molecule in the theoretical conducting medium, and it is used in the COSMO-RS/COSMO-SAC models.

- It also contains the condensed phase energy, the number of surface segments, and their charge.
We edit the *.cosmo file to make it compatible with the sigma averaging FORTRAN program.

- Record the number of surface segments.
- Delete all information above the last table, which contains the segment charges and coordinates.
- Then save as a *.txt text file.

![Segment information table]

```
Segment information:

<table>
<thead>
<tr>
<th>n</th>
<th>atom</th>
<th>position (x, y, z) [au]</th>
<th>charge</th>
<th>area</th>
<th>charge/area</th>
<th>potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-2.20885 -4.13856 -4.13856</td>
<td>0.00119</td>
<td>0.23228</td>
<td>0.00514</td>
<td>-0.01720</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-2.20885 -4.13856 -4.13856</td>
<td>0.00126</td>
<td>0.37185</td>
<td>0.00339</td>
<td>-0.01946</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-2.29018 -4.45988 -4.45988</td>
<td>0.00132</td>
<td>0.37185</td>
<td>0.00339</td>
<td>-0.01946</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-3.00000 -3.00000 -3.00000</td>
<td>0.00167</td>
<td>0.41810</td>
<td>0.00399</td>
<td>-0.0265</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-3.42656 -3.92340 -3.92340</td>
<td>0.00269</td>
<td>0.38582</td>
<td>0.00488</td>
<td>-0.01430</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-3.92656 -3.92340 -3.92340</td>
<td>0.00315</td>
<td>0.27874</td>
<td>0.00400</td>
<td>-0.01295</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>-4.39184 -4.10783 -4.10783</td>
<td>0.00385</td>
<td>0.38582</td>
<td>0.00488</td>
<td>-0.01430</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>-3.42656 -4.12797 -4.42797</td>
<td>0.00414</td>
<td>0.27874</td>
<td>0.00503</td>
<td>-0.01480</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>-1.35218 -3.17181 -2.63333</td>
<td>0.00440</td>
<td>0.41810</td>
<td>0.00336</td>
<td>-0.01395</td>
</tr>
</tbody>
</table>
```
Our sigma averaging program uses the modified *.cosmo file to calculate the sigma profile.

- We average the surface charge densities from the *.cosmo file to find an effective surface charge density on a standard surface segment.

\[
\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp \left( - \frac{d_{mn}^2}{r_n^2 + r_{av}^2} \right)}{\sum_n \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp \left( - \frac{d_{mn}^2}{r_n^2 + r_{av}^2} \right)}
\]
The averaging program creates a new text file, the sigma profile.

- The output file is “‘Chemical name’ Sigma-profile.txt” is located in the C:\Profiles\ directory. This directory must be created manually before running the averaging program.
- The output can be input directly into the COSMO-SAC program or can be plotted graphically.
We run the FORTRAN program to average the surface charges over a standardized bonding site.

- Run sigmaprofile.exe.
- Follow the prompts and enter:
  - New COSMO output (c:/propane.txt)
  - Chemical Name (Propane)
  - The Number of Surface Segments
The sigma averaging program output is text, but can be easily converted into graphical form.

Sigma values, $(e/A^2)$ [X-axis]

P(σ) values, $(A^2)$ [Y-axis]
Using Microsoft™ Excel, we display the sigma profile below.
Acknowledgments

• We want to thank Alliant Techsystems (particularly Anthony Miano, Vice President), AspenTech (particularly Dustin MacNeil, Director of Worldwide University Programs, Larry Evans, Board Chairman, and Chau-Chyun Chen), China Petroleum and Chemical Corporation (particularly Wilfred Wang, President) and Honeywell Specialty Materials and Honeywell International Foundation for supporting our educational programs in computer-aided design and process systems engineering at Virginia Tech. Michael Zwolak gratefully appreciates support from an NSF Graduate Fellowship.
References


References


