Tutorial For Modified Seminario Method

Alice E. A. Allen,[†] Michael C. Payne,[†] and Daniel J. Cole[*,‡]

TCM Group, Cavendish Laboratory, 19 JJ Thomson Ave, Cambridge CB3 0HE, United Kingdom, and School of Chemistry, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom

E-mail: daniel.cole@ncl.ac.uk

[*]To whom correspondence should be addressed
[†]University of Cambridge
[‡]Newcastle University

1 Gaussian Calculation

Parametrizing a molecule using the modified Seminario program involves a number of steps. The first of which is obtaining the QM Hessian matrix of the molecule. To do this a frequency calculation must be carried out using Gaussian 09.[1] An example of the input file for Gaussian is given in the folder 'Benzene' in the file 'lig.com'. Further information about Gaussian frequency calculations is given at *gaussian.com/freq/*. It is important to ensure that the calculation is fully converged and this can be checked in the '.log' file, in the table stating whether forces and displacements have converged.

The output of a Gaussian frequency calculation is a .'chk' file, this can be converted into a 'fchk' file using the Gaussian 09 utility: *$g09root/g09/formchk lig.chk lig.fchk*

Additionally, if you are planning to use a 'Zmat.z' file it is important that the order of atoms in the Gaussian input file is the same as that in the 'Zmat.z' file.

2 Computing the Parameters

2.1 Input Files

There are two input files that are required. One file is the Gaussian '.log' file and the other is the Gaussian '.fchk' file. These files must be called 'lig.fchk' and 'lig.log' to be used by the program.

An optional z-matrix can also be used as an input with the name 'Zmat.z'. A z-matrix can be produced using the LigParGen server found at *www.jorgensenresearch.com/ligpargen*.[2–4] This informs the program of the OPLS atom types and therefore allows for averaging over

bond/angle classes. The 'Benzene' folder contains example input files for a benzene molecule.

2.2 Running the Program

The program is written in python and the main function is 'modified_Seminario_method.py'. There are three arguments that need to be passed to the function: the input file folder (containing the 'lig.fchk' and 'lig.log' files) , the output file folder and the vibrational scaling factor:

modified_Seminario_method(inputfolder, outputfolder, scalingfactor)

The *inputfolder* and *outputfolder* are strings and the *scalingfactor* is a double. The vibrational scaling factor is a constant that accounts for anharmocity and inaccuracies in the level of theory used, it can be set to one if a scaling factor is not wanted. A list of vibrational scaling constants is given at *cccbdb.nist.gov/vibscalejust.asp*.

For the benzene example that is supplied you can try running the program with the following line: *python modified_Seminario_method.py '../Benzene/' '../Benzene/' 0.957*. It should take a few seconds to run on a standard desktop (molecules with linear groups will take slightly longer).

2.3 Output Files

Depending on whether a z-matrix is supplied determines whether there are four or seven output files produced. A log file named *MSM_log* is always produced.

2.3.1 No Z-matrix Supplied

The first file produced is *input_ coords.xyz* and contains the structure that was submitted to Gaussian. The two other files produced are *Modified_ Seminario_ Bonds* and *Modified_ Seminario_ Angle*. They contain a list of all the bond and angle parameters found. The force constant (kcal/mol/Å^2 or kcal/mol/rad^2) is stated first followed by the bond length(Å) or equilibrium angle(°) . The bonds and angles are described by a name such as *C2-C1-C3*, with the letter giving the element involved and the numbering identifying the specific atom (the order the numbers follow is the same as the order the atoms occur in the .xyz file).

2.3.2 Z-matrix Supplied

Again, the files *input_ coords.xyz*, *Modified_ Seminario_ Bonds* and *Modified_ Seminario_ Angle* are produced. However the bonds and angles are now given a name based on the atom types such as *CA-CA-HA*, the angle is also identified by numbers at the end of the line.

Two additional files, *Average_ Modified_ Seminario_ Bonds* and *Average_ Modified_ Seminario_ Angle*, are produced. These average the force constants and bond length/equilibrium angle over the bond/angle class. The final file produced is *Modified_ Scaled_ Seminario.sb*, this contains the bond and angle parameters with OPLS atom types.

Examples are given in the *./Benzene/* folder.

References

(1) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnen-

berg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision E.01. 2009; Gaussian Inc. Wallingford CT.

(2) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331–W336.

(3) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864–3870.

(4) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665–6670.