

Introduction

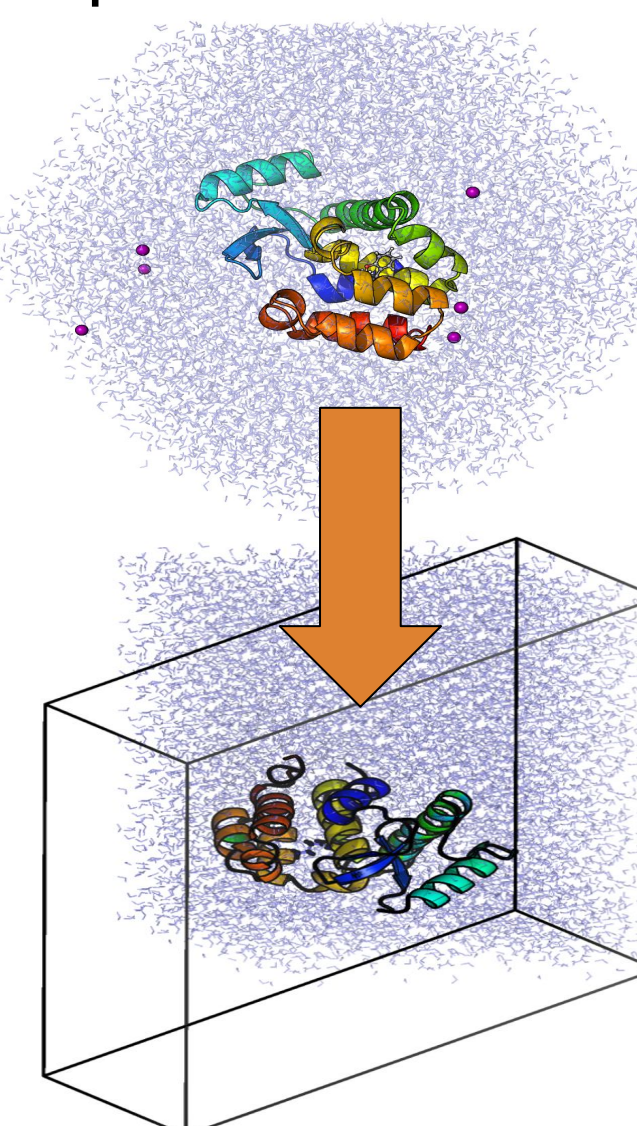
The Genetic Algorithm for Machine Learning (GAML) software has provided a novel method in the creation of Optimized Potentials for Liquid Simulations (OPLS)-based force field (FF) parameters. GAML employs a unique approach that integrates a genetic algorithm (GA) to generate FF parameters. GAML utilizes the Gaussian software to calculate partial atomic charges, which significantly influence the intermolecular interactions and thus behavior of a molecule in a system. Although the Gaussian software provides advanced computational tools for molecular modeling, the cost of the software can be prohibitive. Therefore, we sought to replace its usage in GAML with an open-source alternative that produces indistinguishable results.

Objective

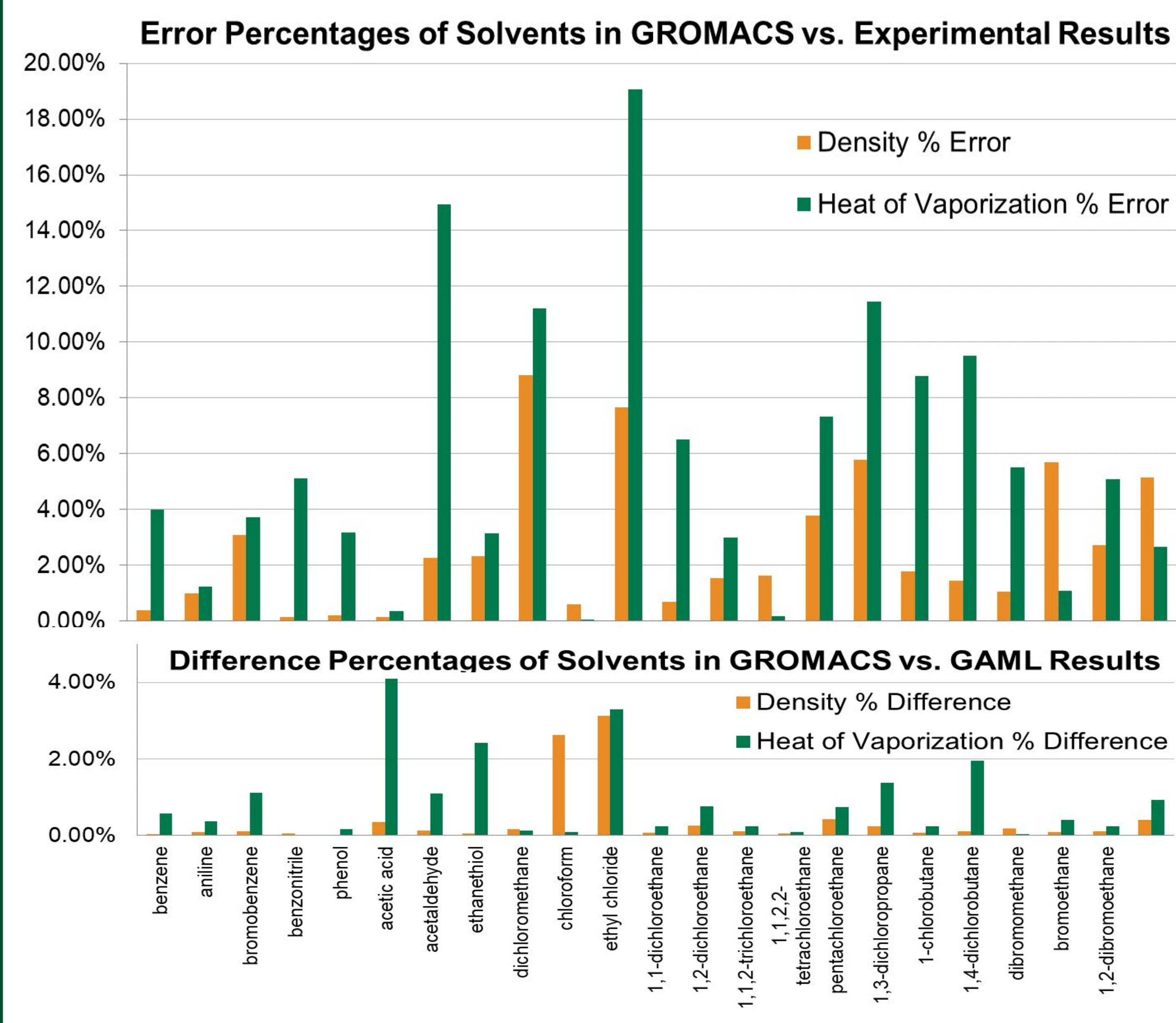
Our project incorporates free and open-source software such as Psi4 to replace the commercial Gaussian 09 program in GAML. By implementing this open-source suite of ab initio quantum chemistry programs, we can calculate RESP (Restrained Electrostatic Potential) partial charges when simulating molecular properties. This is an important step in facilitating GAML for the development of force field parameters for ionic liquids and deep eutectic solvents, which can provide alternatives to environmentally hazardous processes.

GROMACS: Molecular Dynamics

GROMACS, a versatile package for performing molecular dynamics simulations, was employed to validate previously developed force field (FF) parameters generated using GAML and was compared with experimental data through the following steps:



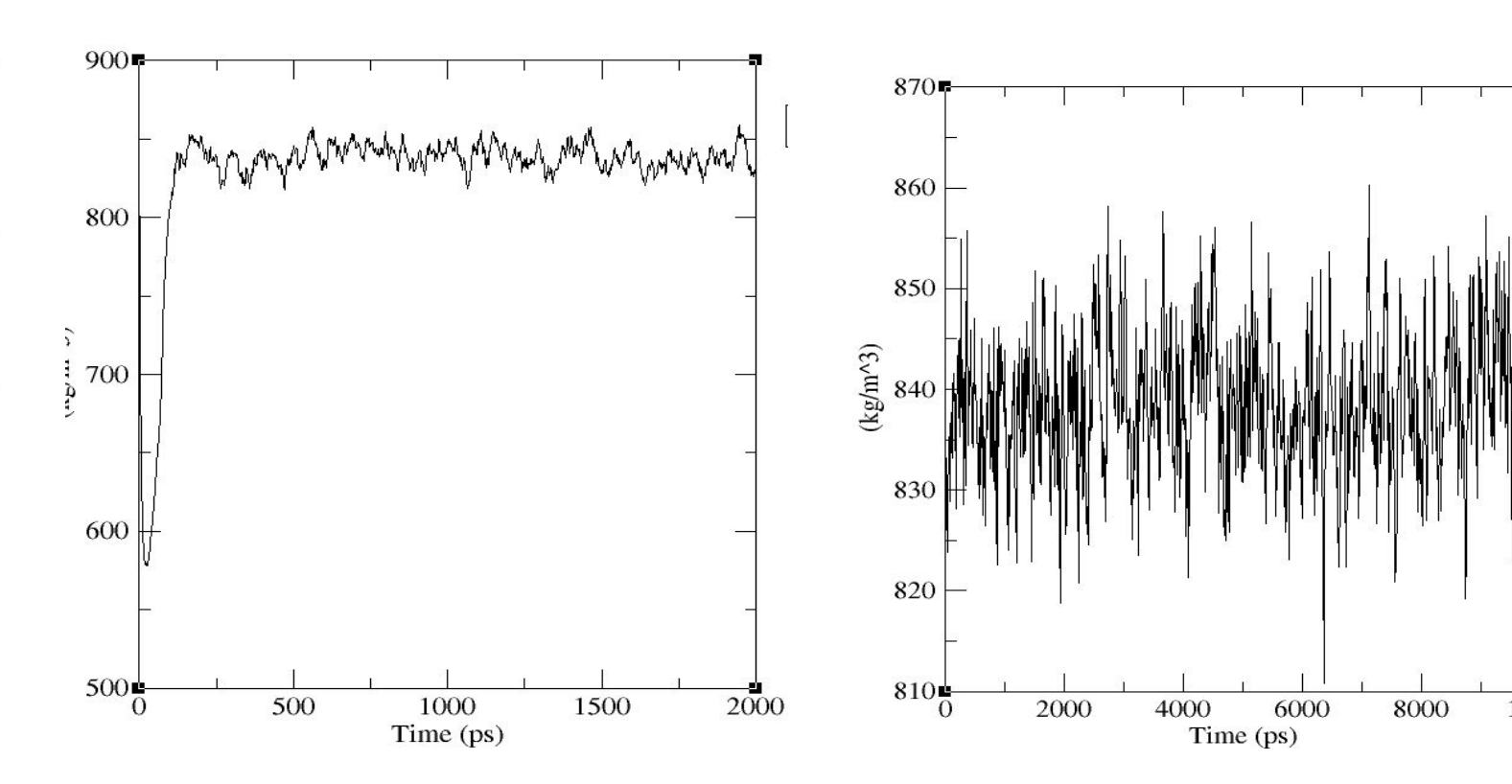
- Define simulation box and solvate:** Using the free PACKMOL software, we created a 3D box that packs as many atoms as needed, excluding their van der Waals repulsive interactions. This configuration can be customized via the number of ions and box size depending on each solvent. A PDB file is created.
- Establish parameters:** Using the "editconf" module, we changed the box dimensions and utilized the PDB file to generate a GRO (molecular structure in Gromos87 format) file. Using a text editor, the units were changed from nanometers (ns) to angstroms (Å). For GROMACS to read this information, we loaded the unbonded and bonded parameters from the ITP files through the TOP file.
- Energy Minimization:** Before beginning dynamics, we relaxed the structure to ensure it had no steric clashes or inappropriate geometry, by finding the lowest local energy value. We used PME summation for long-range interactions and set a 3.8 Å cutoff range for short-range electrostatics to generate binary files.
- NPT Equilibration:** To equilibrate the solvent, the system must reach the proper density with volume allowed to change. The compound's heavy atoms (anything excluding hydrogen) undergo position restraining forces. The pressure is stabilized using an NPT (constant number of particles and temperature) ensemble via the Berendsen pressure coupling method.
- Production:** With a well-equilibrated system, we released the position restraints and ran production MD to collect data and compute properties at the target conditions.
- Analysis:** The FF parameters used in GAML are, by default, validated against densities and heats of vaporization, so we reproduced these two sets of data as criteria for evaluation and compared them to the experimental results.



Criteria for Evaluation

- Density:** Using the GROMACS command "gmx energy -f npt.edr -o density.xvg", density is directly generated from a box of 500 liquid molecules.
- Heat of Vaporization (Hvap):** Using the same command, Potential Energy (PE) is extracted from a box of 1 gas molecule and another box of 500 liquid molecules and then added to the product of standard room temperature and gas constant.

$$\Delta H_{vap} = PE_{Gas} - PE_{Liquid} + RT$$



Cloud Computation

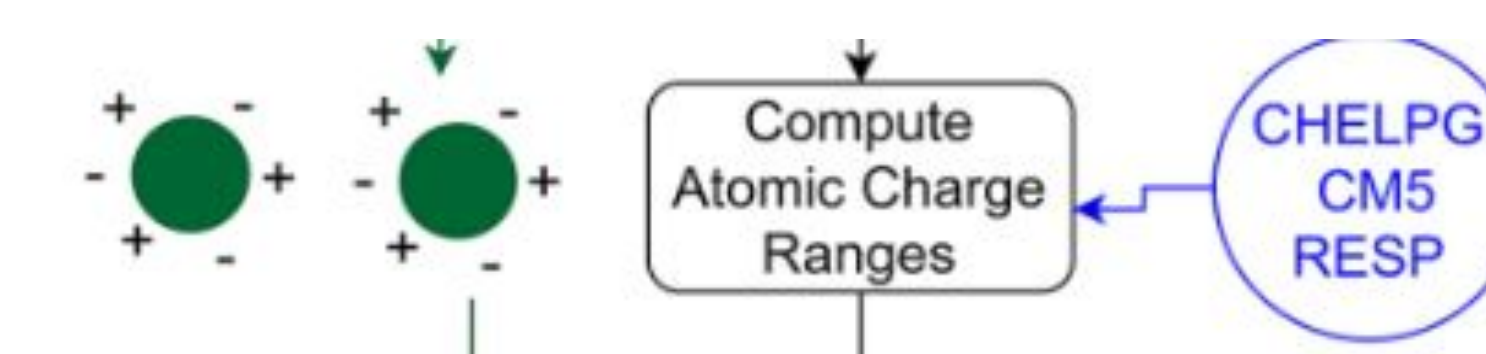
By combining the free computational resources of Google Collab with the capabilities of the OpenMM MD toolkit, the "Making it Rain" Jupyter Notebook was used to test the reproduction of the GROMACS derived values as an alternative method. Reasonably accurate results were found for chloroform and ethanethiol, but inconsistent values were obtained for the other conventional solvents. Given the differences in the MD packages, the environmental options and setup will need to be evaluated in OpenMM to improve the accuracy.

Percent Error and Difference for Density of Solvents in "Making it Rain" vs. GAML



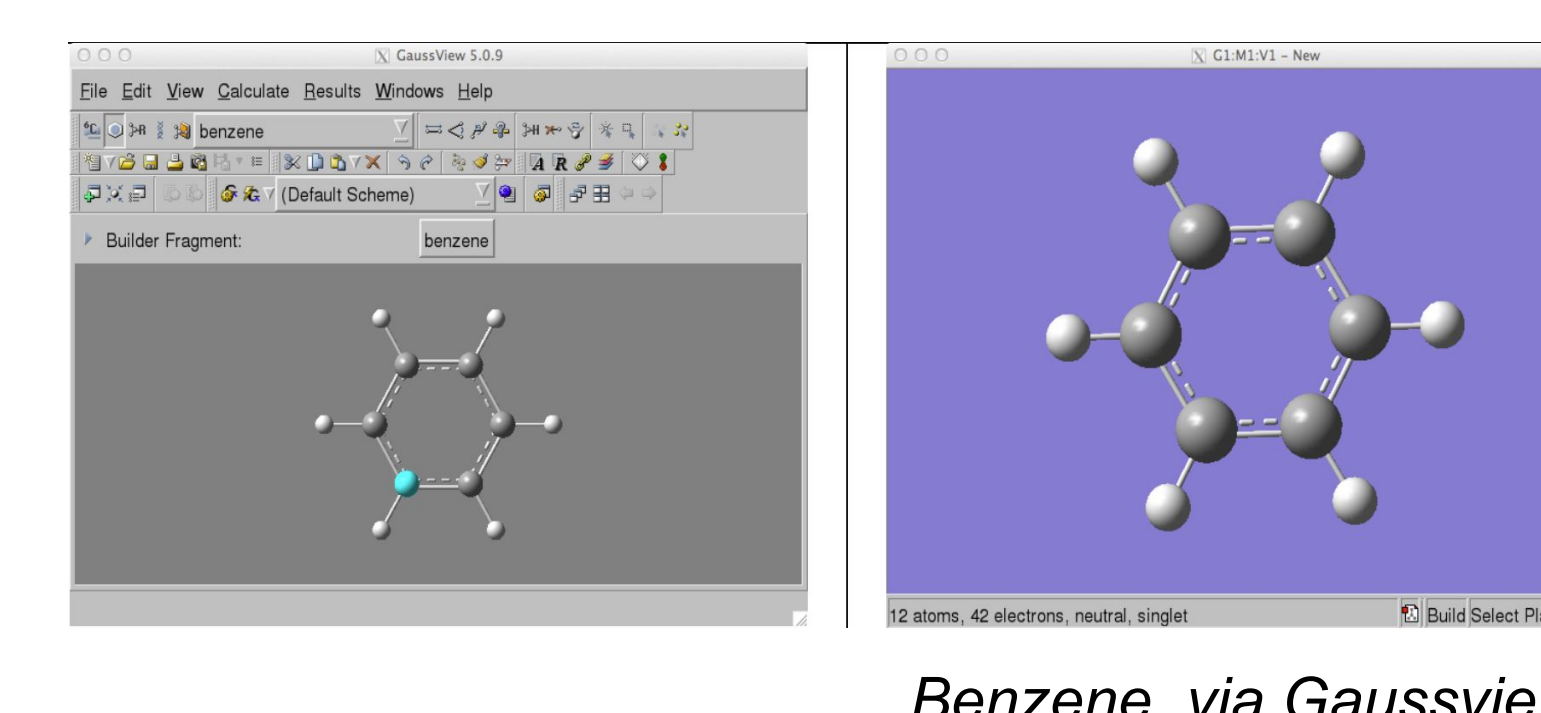
Software: Gaussian09 vs. Psi4

Our results for a sample of solvents showed that partial charges calculated using Psi4 were in excellent agreement with those obtained using Gaussian, demonstrating the consistency of the open-source Psi4 software. By incorporating our custom script into the GAML software, it becomes fully usable to those without a Gaussian 09 license and can be easily migrated onto an accessible powerful cloud-computational environment. Our script utilized Psi4 to calculate partial charges on a molecule using the CHELPG scheme. The following steps were used:



Solvent	G09	PSI4
benzene	-0.089	-0.089
chloroform	0.22	0.22
ethanethiol	-0.06	-0.06
acetaldehyde	-0.51	-0.51
ethyl chloride	-0.15	-0.15
1-chlorobutane	0.03	0.03
dichloromethane	0.15	0.15
1,1-dichloroethane	0.27	0.27
1,2-dichloroethane	0.04	0.04
1,4-dichlorobutane	-0.3	-0.3
1,3-dichloropropane	0.35	0.35

Representative Atomic Charges (e), CHELPG HF/6-31G using Gaussian09 and Psi4 software



- Load PDB File and read in atomic coordinates from the molecule of interest.
- Configure Psi4 by setting memory and output file.
- Set Computation Options using the basis set (6-31G), SCF (Self-Consistent Field) equations (PK), and energy convergence criterion (1e-8).
- Geometry Optimization and SCF Calculation of the ground state wavefunction using SCF method. This step provides the electronic structure of the molecule.
- Compute Electrostatic Potential.
- CHELPG Fitting (CHarges from ELectrostatic Potentials using a Grid-based) method calculates the atomic partial charges by fitting the electrostatic potential to the potential of point charges located at the atomic positions.
- The output charges are calculated and returned via a TXT file.

Future Prospects

This research highlights the progress made implementing Psi4 into the GAML package as an alternative to Gaussian and in the validation of an alternative MD package (OpenMM versus GROMACS). The results suggests a promising future for GAML, where all software components are open-source and simulations could be performed for free on powerful cloud-based computers. A simple-to-use, computationally automated, and scientifically rigorous method that allows the development of FF parameters for unique solvents environments could dramatically enhance the study of chemical reactivity in solution.

Acknowledgments

We want to thank our mentor Dr. Orlando Acevedo, the P.I for the REU Dr. Burton Rosenberg, Dr. Kun Yue, and Dr. Xiang Zhong. This material is based upon work supported by the National Science Foundation under Grant No. CNS-1949972 and CHE-2102038. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.