Towards the understanding of the structure and formation of Deep Eutectic Solvents using classical molecular dynamics and fragmentation simulations

Forcefield parametrization and benchmarking the density functional approximation methods.

Author: **Swaroop Kunapuli**
Directrice: **Prof. Rachel Schurhammer**
Co-director: **Dr. Alain Chaumont**

May 5, 2024

Contents

1 Introduction

For the last few decades, a new class of solvents callled Deep Eutectic Solvents (DES) have garnered much attention from researchers interested in finding solvents tailor-made for specific applications. DES are a mixture of two or more components that form a eutectic mixture with a melting point lower than that of the individual components. The invention of DES lead to the studies of wide-range of applications as their nature and tailoring help in significant processes. The modeling of DES as helps as they explain the experimental observations and help in the design of DES. In this thesis, the DES systems considered are:

1. Urea - Choline Chloride (1:2 molar ratio) colloquially called Reline.
2. Evernic Acid - Usinic Acid NADES found in the lichen *EverniaPrunastri*

And the following modeling strategies/tools were utilised for this purpose:

1. Molecular Dynamics
2. Fragmentation Simulations

2 Résumé

Since the last CSI, the following projects were undertaken and certain protocols and work flows were developed based on the benchmarking of certain semi-empirical and density functional approximation methods. The following projects were undertaken:

1. **Developing the algorithm for fragmentation simulations of non-covalent clusters**
2. Studying the stability and water retention properties of the proposed NADES systems in lichen EverniaPrunastri.

3 Current status and information for the preparation of CSI

This section summarizes the data (results and discussion) of the work done in the academic year 2023-2024, to be presented in the CSI-2024. The summary is divided into two categories: results and discussion that is already available and the work that is currently being done.

3.1 Results and discussion of the work (academic year 2023-2024)

The following is the list of distinct projects that were undertaken in the academic year 2023-2024:

1. **Developing the algorithm for fragmentation simulations of non-covalent clusters**
2. Studying the stability and water retention properties of the proposed NADES systems in lichen EverniaPrunastri.
3. Benchmarking the workflow for Metadynamics simulations at semi-empirical and DFT levels of theory.
4. Going beyond the thermochemical calculations, Local Energy Decomposition Analysis (LEDA) as an attempt for explaining the stability of certain non-covalent clusters observed in the experiment.

Set	Energy (Hartree)	Relative Energy (kcal/mol)	Time (min)	No. conformers
1	-134.71150		26	11
2	-134.70620		40	12
3	-134.70223		39	20

Table 1: Energy of the most stable conformer after optimization at GFN2 level of theory on the MSM (16-core) machine

The following part of this project starts with the section 4 on literature review for the above listed works. The section **??** explains our approach to the fragmentation simulations and attempts to solve the computational bottleneck for these simulations. The section 5 gives some results on benchmarking on different QM methods that are planned to be employed in the fragmentation simulations. The section 6 gives the results of the benchmarking of different density functional approximations for the geometry optimization calculations. The section 7 gives the results of the LEDA calculations for the non-covalent clusters. The section **??** introduces the lichen system *Evernia prunasitri* and the experimental MS data to understand their water-retention properties via possible formation of NADES. The results of the thermochemical calculations of different combinations of clusters are presented here.

4 Literature Review

4.1 Parametrization for accurate molecular dynamics simulations

4.2 Fragmentation simulations for corraboration of experimental mass-spectrometric data

4.3 Thermochemical calculations of the lichen *Evernia prunastri* using Semi-empirical and DFT methods

5 Metadynamics xTB/CREST workflow benchmarking

5.1 Benchamarking at Semiemperical level with CREST

Three different initial coordinates were generated using packmol for the cluster combination $[3Ch+3U+4Cl]^{-1}$. workflow: The initial coordinates were used as input for the NCI algorithm of CREST, that performs iMTD sampling. The energy of the conformers was calculated at GFN2-GFNFF composite method level to preserve the topology of the system. The conformers that fall in the energy range of 12 kcal/mol from the lowest energy conformer were selected and further optimized at GFN2 level of theory.

5.2 Benchamarking at DFT level with CENSO

The ensemble of conformers generated from the previous step were used as input for two types of calculations in CENSO:

1. Single point energy calculations of each conformer at B97-3c level of theory with def2-SV(P) basis set (called *part0* calculations in CENSO).

2. Geometry Optimization of all conformers at r2scan-3c level of theory (called *part2* calculations in CENSO).

It was verified that the threshold energy value (the free energy value above which the conformer is ignored by CENSO) for optimization calculations was set very high so as to not ignore any conformers in the optimization calculations.

The results above show that the choice of initial geometry does have an effect on the energies of the

Initial ensemble	Energy (Hartree)	Relative Energy (kcal/mol)
1	-3500.3072605	
2	-3500.3211375	
3	-3500.3016773	

Table 2: Single-point energy of the most stable conformer from *part0* calculations of CENSO

Initial ensemble	Energy (Hartree)	Relative Energy (kcal/mol)
1	-3503.1954934	
2	-3503.2027487	
3	-3503.1849456	

Table 3: Energy of the most stable conformer from *part2* calculations of CENSO

most stable structures but at the level of complexity that comes with these number of combinations, this workflow gives pretty reasonable results.

6 Benchmarking density functional approximations

List of available methods and codes for speeding up the optimization calculations:

1. ORCA (DFA and composite methods)
2. CENSO (with ORCA as the QM calculation driver and xTB as the semi-empirical method; parameters already benchmarked by Grimme's lab)
3. ML-based methods (AIQM1 method in MLatom 3)

Eventhough ORCA has its own optimization algorithm, even with approximations included the calculations are long. xTB gives a way to do optimization on a fixed number of cycles while calculating single-point energies with ORCA. This is what CENSO does but can also be done with xTB and ORCA alone.

7 Local Energy Decomposition

7.1

Systems Studied:

1. $[1Ch + 1U + 2Cl]^{-1}$ $(m/z = 235)$
2. $[1Ch + 2U + 2Cl]^{-1}$ $(m/z = 295)$
3. $[1Ch + 3U + 2Cl]^{-1}$ $(m/z = 355)$

Directory	Orca settings	xTB settings	time(min)
20			
19			
18			
17			

Table 4: Various approaches for optimization with ORCA/CENSO and the corresponding computation times on the 32-core Intel Cascadelake machine.

4. $[2Ch + 0U + 3Cl]^{-1}$ $(m/z = 313)$