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

Swaroop Venkata Sai Kunapuli

Directrice of the thesis: **Prof. Rachel Schurhammer**

Co-director of the thesis: **Dr. Alain Chaumont**

June 14, 2024

Contents

Résumé

Since their conception Deep Eutectic Solvents (DES) have been a constant subject of interest for researchers working on tailor-made solvents for various applications including catalysis, biomass conversion and drug delivery. The presence of diverse type of constituent molecules that make up these solvents and a spectrum of non-covalent interactions between these constituents help in designing solvents for particular applications by fine-tuning the functional groups of the molecules. The experiments developed to understand the nature of DES are generally accompanied by a computational study to aid in interpretation of the experimental results. In this thesis, we set out to understand the structure and formation of two types of DES using classical molecular dynamics, quantum mechanical calculations and fragmentation simulations of non-covalent clusters. The first part of this work involved parameterization of non-bonding parameters of GAFF v2.11 forcefield for molecular dynamics simulations of Urea : Choline Chloride (2:1 molar ratio) (named Reline) to accurately represent the thermophysical properties (density, diffusion coefficients and viscosity) and the structure of the solvent relative to the experimental results. The quantum mechanical calculations involving DFT and semi-empirical methods were used in the second part of this work for thermochemical calculations of non-covalent clusters of Reline and Evernic/Usinic acid to corroborate the stability results observed in the soft-ionisation mass spectrometric experiments. Finally a major part of this thesis involves adopting and adapting the algorithm for molecular fragmentation simulations (M_3C) to make the computational study of fragmentation of non-covalent in mass spectrometry computationally tractable. Despite the initial computational bottleneck of the fragmentation simulations, with a combination of semi-empirical methods and density functional approximations methods we are able to finalize a workflow that can be employed in calculating the breaking curves of the chemical systems at the scale of supramolecular assemblies in mass spectrometry on a high-performance computing (HPC) cluster.

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 Literature Review

2.1 Parametrization for accurate molecular dynamics simulations

2.2 Thermochemical calculations of Choline Chloride-based systems

2.2.1 Quantification of intermolecular interactions

Include literature on the following:

- NBO Analysis

- QTAIM Analysis

- LED Analysis

2.3 Fragmentation simulations for corroboration of experimental mass-spectrometric data

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

3 Thermochemical calculations of Choline Chloride-based systems

Details of DFT: B3LYP, 6-311++G(d,p), D3
List of methods studied:

- Replacing S with O and O with S in the urea/thiourea most stable geometries calculated previously.

- CREST NCI Sampling and taking the best conformer.

Relplaced S in Thiourea with O to make Urea

System	Energy (Old geometry)	Energy (New geometry)	Energy (CREST geometry)
$[1U+1ChCl+1Cl]^{-1}$	-1475.01399936	-1475.01307918	-1475.01792736
$[1U+2ChCl+1Cl]^{-1}$	-2264.34123298	-2264.33731603	-2264.34844310
$[2U+1ChCl+1Cl]^{-1}$	-1700.39820039	-1700.39980596	-1700.41437298

Table 1: Urea Systems: Comparing energies of new geometries to the old ones and also CREST generated geometries.

Replace O in Urea with S to make Thiourea

System	Energy (old geometry)	Energy (New geometry)	Energy (CREST geometry)
$[1ThioU+1ChCl+1Cl]^{-1}$	-1797.97419795	-1797.97359958	-1797.97877987
$[1ThioU+2ChCl+1Cl]^{-1}$	-2587.29633558	-2587.29763998	-2587.30660929
$[2ThioU+1ChCl+1Cl]^{-1}$	-2346.32084031	-2346.31949541	-2346.33237642

Table 2: Thiourea Systems: Comparing energies of new geometries to the old ones and also CREST generated geometries.

4 Quantification of intermolecular interactions

In this section we will discuss the methods and results of the calculations to quantify the strength of the intermolecular interactions in our choline chloride-based systems.

The Local Energy Decomposition method is implemented in ORCA to decompose the DLPNO-CCSD(T) energies into physically meaningful contributions. Admittedly, the decomposition is to some extent arbitary, but using pair correlation energies for the pairs of localized occupied orbitals help in decomposing the correlation and the reference energy into intra- and interfragment contributions. The

System	Old Geometry	New Geometry	Crest Optimised Geometry
$[1U + 1ChCl + 1Cl]^{-1}$			
$[1ThioU + 1ChCl + 1Cl]^{-1}$			

Table 3: Interaction energies of three geometries of each cluster combination studied

fragments have to be user defined, and in our case these are the urea,thiourea and choline molecules and the chloride ions. We studied the following urea-containing and the corresponding thiourea-containing clusters. For each cluster combinations, we studied three geometries:

4.1 Input structure and parameters for different Calculations

4.1.1 NBO Analysis

GAUSSIAN 16 Package for the NBO population analysis

```
#P B3LYP/6-311++G(d,p) SCF=Tight Pop=NBO
```

4.1.2 QTAIM Analysis

WFN files were generated using GAUSSIAN 16 Package.

```
#P B3LYP/6-311++G(d,p) SCF=Tight output=WFN
...
...
<filename>.wfn
```

4.1.3 Local Energy Decomposition Analysis

ORCA 5.0.4 Package was used for the Local Energy Decomposition (LED) Analysis.

```
! dlpno-ccsd(t) def2-TZVP def2-TZVP/C def2/J rijcosx verytightscf
! TightPNO LED
%mdci DoDIDplot true end
%MaxCore 8000
%pal nprocs 32 end
*xyz -1 1
...
*
```

Resolution of identity (RI) method with Chain of Spheres (COS) approximation (RIJCOSX) implemented in the ORCA package was used to speed up the calculations, along with DLPNO-CCSD(T) method and auxiliary basis set def2-TZVP/C. The calculations requested data for DIDplot which will be presented in this report.

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.

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

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

Table 5: Single-point energy of the most stable conformer from *part0* calculations of CENSO

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 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.

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

Table 6: Energy of the most stable conformer from *part2* calculations of CENSO

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

Table 7: Various approaches for optimization with ORCA/CENSO and the corresponding computation times on the 32-core Intel Cascadelake machine.