7-2023

Big Data for a Deep Problem: Exploring Natural Deep Eutectic Solvent (NADES) Properties through RDKIT and Data Analytics

Grayson Weavil

Lucas B. Ayres

Carlos D. Garcia

Big Data for a Deep Problem: Exploring Natural Deep Eutectic Solvent (NADES) Properties through RDKIT and Data Analytics

Grayson Weavil[1], Lucas B. Ayres[2], and Carlos D. Garcia[2]
[1]Department of Chemistry, Wake Forest University, 1834 Wake Forest Rd. Winston Salem, NC 27109
[2]Department of Chemistry, Clemson University, 211 S. Palmetto Blvd. Clemson, SC 29634

Introduction

The past decade of scientific research has presented natural deep eutectic solvents (NADES) as an emerging class of supposed eco-friendly solvents with promising potential for use in food, material science, and pharmaceutical industries. NADES are governed by a hydrogen bonding network induced by the heating and stirring of two or more solid compounds to produce a stable liquid with a substantially lower melting point (eutectic point) than either of its respective components. The NADES fabrication process is a complex and intricate process which requires optimization of selective properties such as conductivity, melting point, stability, biodegradability, and viscosity, which are further dependent upon the multifaceted properties (hydrogen bond donor and acceptor count, molecular weight, hydrophobicity, surface area, etc.) of each individual component. Because there is no systematic approach to the curation of NADES due to the unique interplay of variables within each network, insights made about properties of NADES are largely based on empirical work and limited to a small number of formulations with similar features. In order to examine NADES properties on a comprehensive scale, data analytics and Python programming was performed on a NADES database with over 1300 formulations. Through our data analytics approach, we present a "big picture" evaluation of NADES properties to refine the NADES formulation process, reveal commonalities among NADES formulations, and improve future application of NADES.

Methods

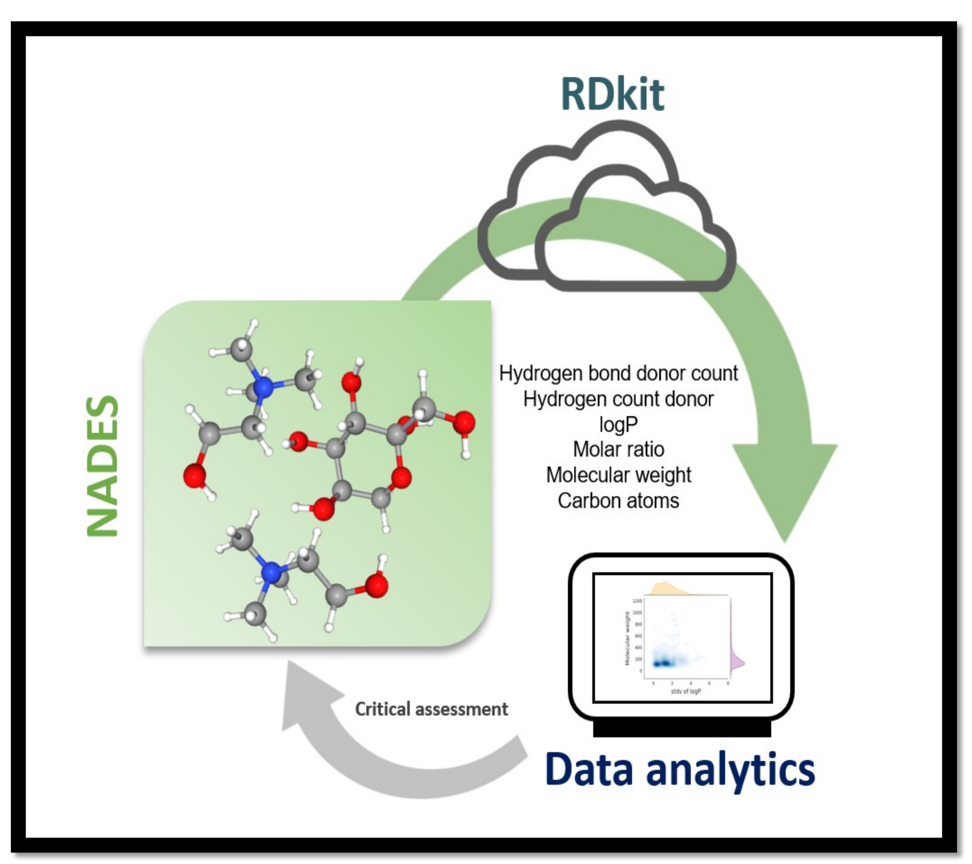

Figure 1: Experimental Scheme

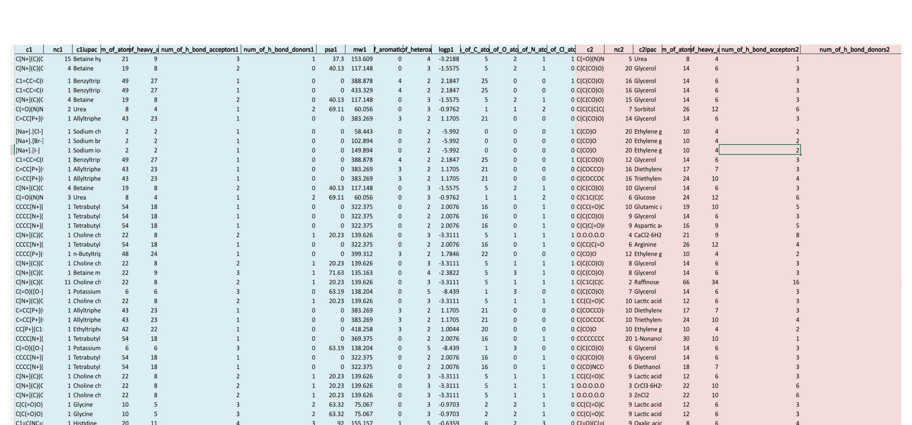

Figure 2: Augmented Database Sample using RDKIT

To obtain a comprehensive view of NADES, we selected an initial dataset of over 1300 NADES formulations published in Omar and Sadeghi's DES review. The database was expanded to include a variety of selected chemical descriptors (polar surface area, molecular weight, number of hydrogen bond donating and accepting compounds, number of aromatic rings, logP, and number of C, H, N and O atoms) by accessing RDKIT, a chemical informatics toolkit. The original database compounds were converted into canonical Simplified Molecular-Input Line-Entry System notation (SMILES), which was input into RDKIT to produce a master csv file with over 1300 NADES formulations, containing over 85,000 data entries. Data analytics was performed using Python and Seaborn visualization plotting to gain insights regarding the relationships among the variables of NADES components.

Data Analysis & Results

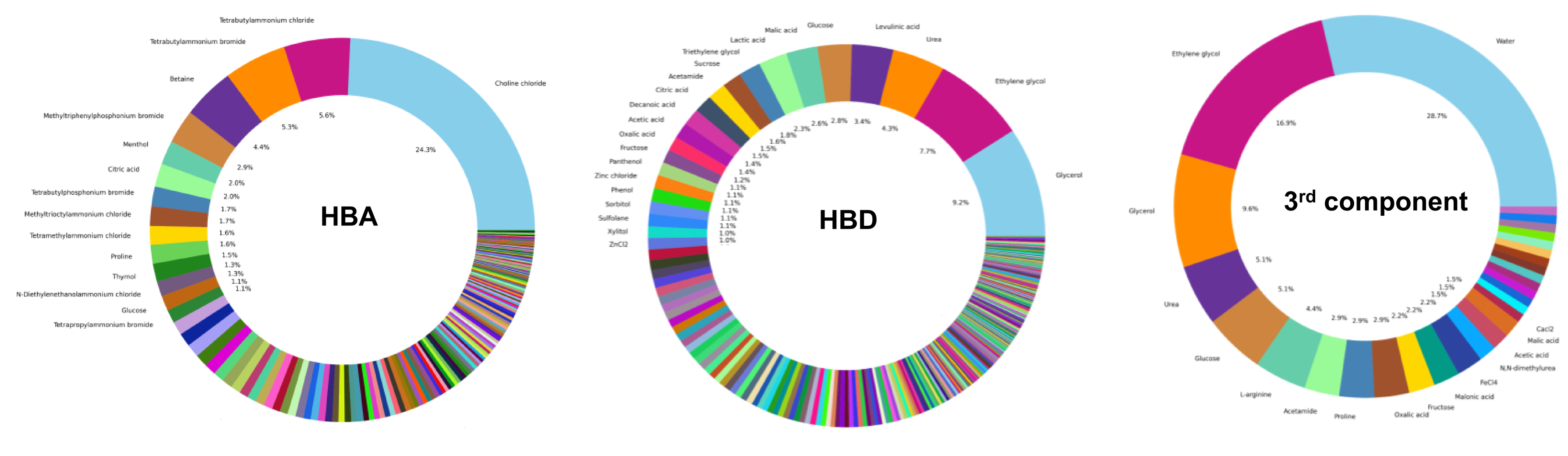

Figure 3. Relative percentage of hydrogen bond acceptor group (HBA), hydrogen bond donor group (HBD) and 3rd component in NADES formulations. Relevant compounds: **HBA;** Choline chloride, Tetrabutylammonium chloride, Tetrabutylammonium chloride, Betaine, Methyltriphenylphosphonium bromide, Menthol, Citric acid, and Tetrabutylphosphonium bromide – **HBD;** Glycerol, Ethylene glycol, urea, Levulinic acid, glucose, malic acid, lactic acid, triethylene glycol, sucrose, acetamide, citric acid, decanoic acid, and acetic acid – **3rd component;** Water, ethylene glycol, glycerol, urea, glucose, L-arginine, acetamide, proline, oxalic acid, and fructose.

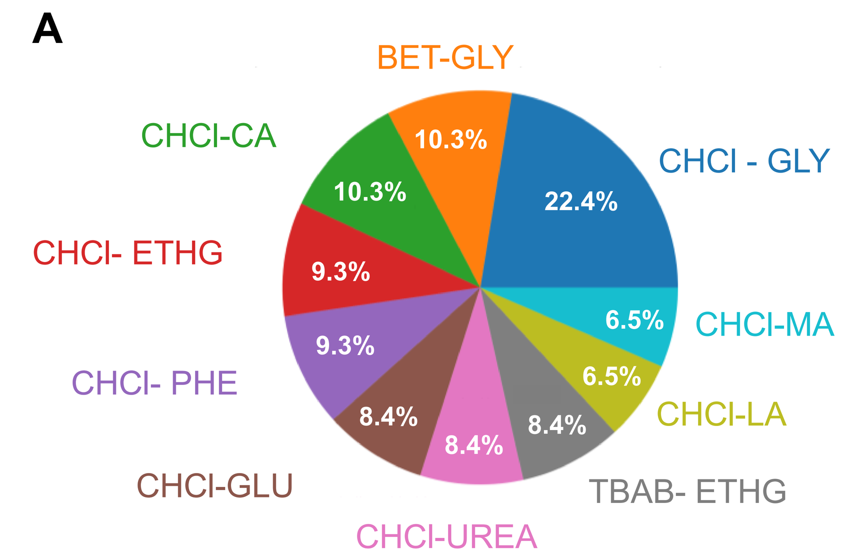

Figure 4. Top ten binary NADES formulations

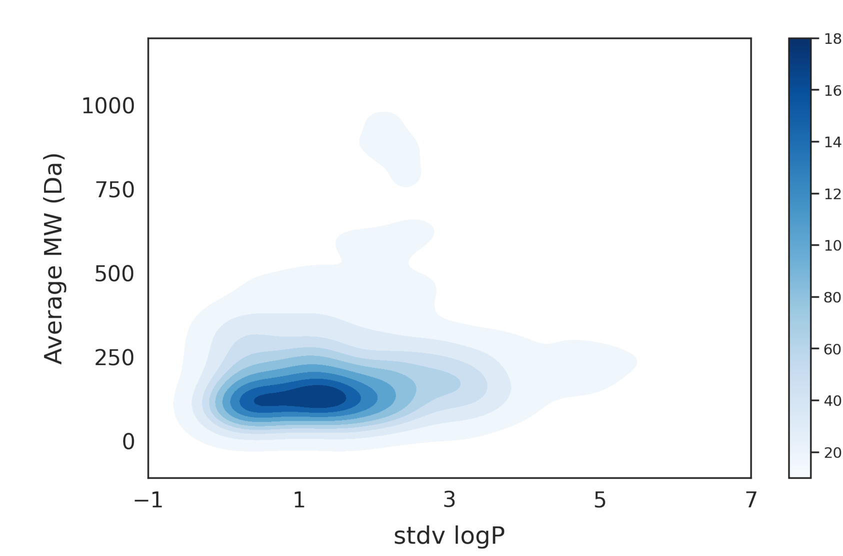

Figure 5. Heatmap of average molecular weight and standard deviation of logP of NADES

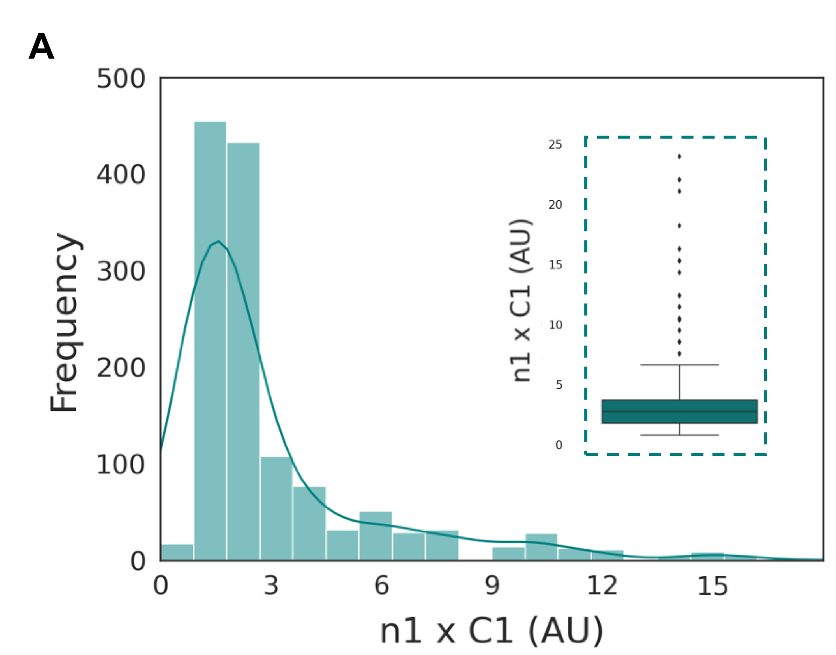

Figure 6. Frequency distribution of the hydrogen bond acceptor count times its number of mol

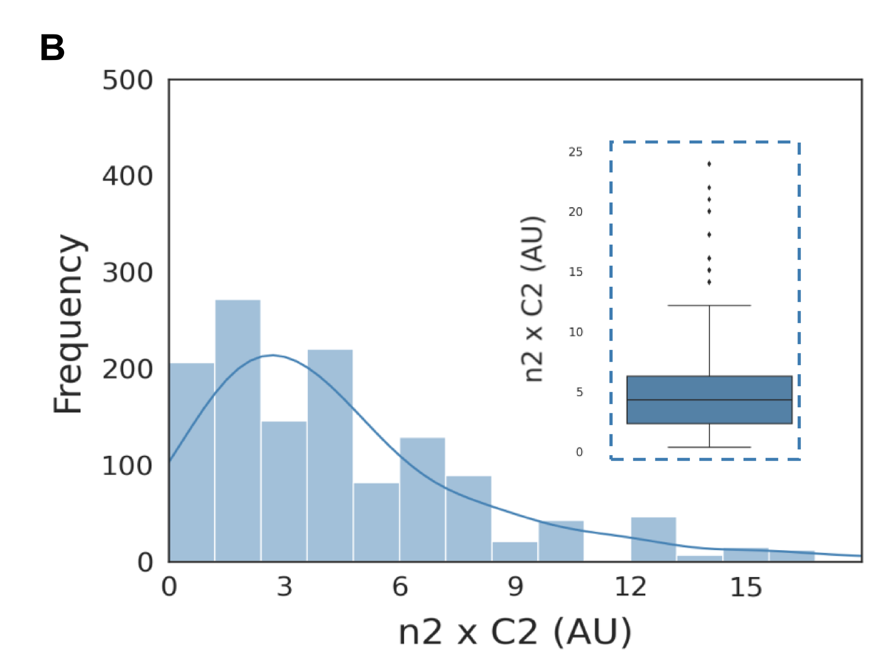

Figure 7. Frequency distribution of hydrogen bond donor count times its number of mol

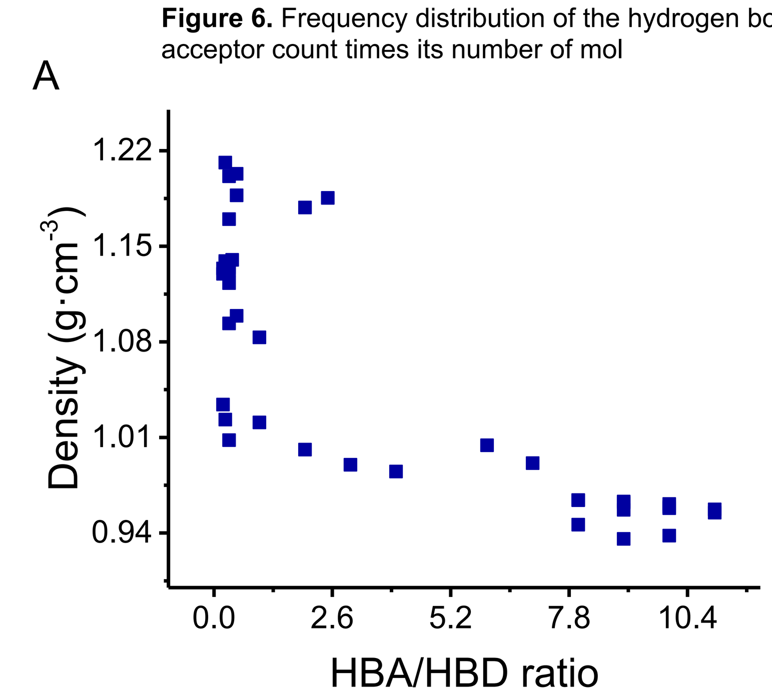

Figure 8. Dependence of density of NADES as a function of HBA/HBD ratio at 30°C

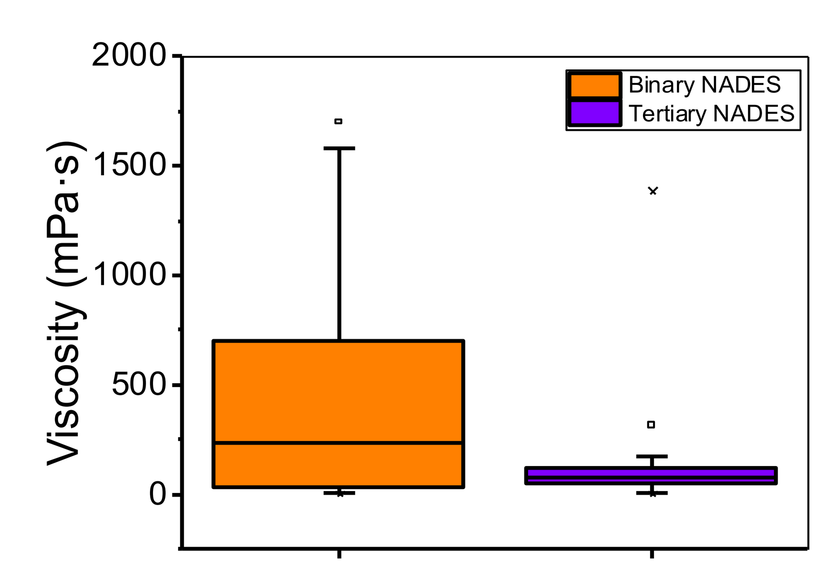

Figure 9. Side by side viscosity boxplot of binary and tertiary formulations

Conclusions

Our data analysis provides a comprehensive evaluation of the general trends of NADES reported in the last ten years. Our work reveals that choline chloride is a cornerstone component of NADES formulations, as it is the most common hydrogen bond acceptor, commonly paired with glycerol, ethylene, glycol, and glucose. Furthermore, a small molecular weight and compatibility of the NADES components (logP) are two crucial factors to consider when formulating a NADES with optimal conditions. Additionally, according to our analysis, NADES are generally formulated with slightly more hydrogen bond donating groups than accepting groups and the addition of a third component to a NADES is correlated with lowering the viscosity of the system. Overall, our study integrates data analytics to provide insights into the general characteristics of NADES and unveils prevalent trends regarding NADES formulations.

Future Work

This analysis provides a guide to researchers who want to optimize NADES conditions and addresses the obstacles associated with new formulations. Researchers may use this analysis to select components that produce greener, more reliable NADES for future applications in chemistry, specifically for use in the pharmaceutical industry. Considering that high viscosity is a prominent obstacle in the NADES fabrication process, future work may investigate the role of the third component on lowering viscosity and further refine which compounds are most effective at lowering the viscosity of NADES without disrupting the hydrogen bonding system.

Acknowledgements

- National Science Foundation Grant Award CHE-2050042 - IMPRESS Chemistry Research Experience for Undergraduates (REU)
- Clemson University Department of Chemistry
- Thank you to the Garcia group for their mentorship, encouragement, and support throughout the past ten weeks at Clemson.

References

- K.A. Omar, R. Sadeghi, Database of deep eutectic solvents and their physical properties: A review, J. Mol. Liq. 384 (2023) 121899.
- Lucas B. Ayres, Grayson Weavil, Mays Alhoubani, Barbara G. S. Guinati, and Carlos D. Garcia, Big Data for a Deep Problem: Understanding the Formation of NADES Through Comprehensive Chemical Analysis and RDKit. *J. Mol. Liq.* **2023**