A structured methodology for natural deep eutectic solvent selection and formulation for enzymatic reactions

ir. Attila Kovács

Supervisors **prof. dr. ir. Pieter Billen — prof. dr. ir. Iris Cornet — prof. dr. Erik C. Neyts**

Thesis submitted in fulfilment of the requirements for the degree of Doctorate in applied engineering
Faculty of Applied Engineering — Antwerp, 2023

University of Antwerp

University of Antwerp

Faculty of Applied Engineering
Doctorate in applied engineering

A structured methodology for natural deep eutectic solvent selection and formulation for enzymatic reactions

Thesis submitted in fulfilment of the requirements for the degree of
Doctorate in applied engineering
at University of Antwerp

ir. Attila Kovács

Supervisors
prof. dr. ir. Pieter Billen
prof. dr. ir. Iris Cornet
prof. dr. Erik C. Neyts

Antwerp, 2023

Jury

Chairman
prof. dr. Christophe Vande Velde, University of Antwerp, Belgium

Supervisors
prof. dr. ir. Pieter Billen, University of Antwerp, Belgium
prof. dr. ir. Iris Cornet, University of Antwerp, Belgium
prof. dr. Erik C. Neyts, University of Antwerp, Belgium

Members
prof. dr. Remco Tuinier, Eindhoven University of Technology, The Netherlands
prof. dr. ir. Seyed Soheil Mansouri, Technical University of Denmark, Denmark
dr. ir. Els D'Hondt, I-CatS Labs, Belgium
prof dr. ir. Philippe Nimmegeers, University of Antwerp, Belgium

Contact
ir. Attila Kovács
University of Antwerp
Faculty of Applied Engineering
Intelligence in Processes, Advanced Catalysts and Solvents (iPRACS) research group
Groenenborgerlaan 171, 2020 Antwerp, Belgium
M: attila.kovacs@uantwerpen.be

Acknowledgment

I would like to thank my promoters for their support, comments, advice and most of all their trust in me and my project. Pieter Billen, as my main promoter, guided me through this venture and taught me a lot about the nuts and bolts of academia along the way. Iris Cornet always gave honest opinions and feedback on my manuscripts, and with her comments (I hope) both the language and the content of my writings went to the next level. Erik Neyts introduced me to the field of computational chemistry, which gave a new direction not only to my doctoral thesis but also to my postdoctoral aspirations. I would also like to thank Marc Wijnants, who was not formally my supervisor, but who was instrumental in getting this project started.

The next group of people to whom I would like to express my gratitude are the students I had the honor to supervise during my doctoral years: Houssam Tauil, Arthur Verstraelen, Nathalie Jansses, Dries Lauwers, Jarne van Loon, Ayan Berkay, Jonas Cassimon and Miguel Mielants. Their work contributed greatly to the achievements of this thesis and their questions helped me to better understand the subject I was working on. All students helped to consolidate the modeling efforts presented, but I would like to give a special thanks to Ayan who contributed to the machine learning models, Nathalie who did most of the measurements for the enzymatic reactions, and Jonas who developed the chromatography protocol for the biosurfactant deacetylation.

Special thanks to my wife Kinga. In addition to general emotional support, she also provided very specific technical support with the code (both VBA, bash, and Python) that appears in the various chapters. Without her, this thesis would not be in front of you now.

Among my colleagues, I would especially like to thank Yanou Fishel for his help. He proofread many parts of the thesis, often in a surprisingly short time. In addition, his cheerful personality made the PhD experience much more enjoyable.

I would also like to thank Dániel Fózer, one of the most dedicated academics I know. I learned a lot from our conversations about academia, its opportunities and threats. I hope that in time I will have at least some of his enthusiasm.

Just as Isaac Newton wrote in his letter, "If I have seen further [than others], it is by standing on the shoulders of giants."[1], I must also express my gratitude to the entire deep eutectic solvent research community, with three specific groups to highlight. First, the iPRACS research group at the University of Antwerp (and its twin group BioWAVE with their common predecessor BIOGEM), which gave this project a home and a foundation. I'm lucky to have such great colleagues and I would like to give special thanks to the members

[1]For full disclosure, he most likely took this metaphor from Bernard of Chartres. For younger readers, I would refer to the 2011 science fiction film Rise of the Planet of the Apes: "Apes together strong".

(past and present) who contributed their knowledge to this thesis: Lukasz Pazdur, Trang Bui Thi, Sven Janssens, Toon Bouwen, Nikolett Wittner, Eliane Goossens and Veerle Akkermans. Second, I would like to thank the Des.solve research group at Nova University of Lisbon for providing the much needed data for the machine learning models. Personally, I would like to thank Reza Haghbakhsh and Ana Rita C. Duarte for providing not only data but also feedback on the models under development. I would also like to mention the efforts of this group to connect the deep eutectic solvent community by organizing the biennial conference series, International Meetings on Deep Eutectic Systems, starting in 2019. Third, I would like to thank the Department of Technical Thermodynamics at the Technical University of Dortmund, with special thanks to Nicolás Gajardo Parra and Moreno Ascani. Their expertise on PC-SAFT led to the recommended solubility model in this thesis.

I would like to give a special thanks to all my family and friends who took an interest in my research and put up with my attempts to explain it. I have to mention Kitti, Patrik and Honey by name. Although they are laymen in this field, their interesting views on chemistry have pushed my research forward in the most unexpected ways.

Last but not least, I would like to express my gratitude to all the members of my Doctoral Committee and Doctoral Jury, who, together with the administrative staff of the University, made the effort to give this thesis its final shape through their comments, questions and supervision: Sofie Delanoye, Els D'Hondt, Philippe Nimmegeers, Christophe Vande Velde, Remco Tuinier and Seyed Soheil Mansouri.

Contents

Abstract

Natural deep eutectic solvents (NADES) show great promise as media for enzymatic reactions in areas where (bio)compatibility with natural or medicinal products is a must. While in theory they can be tailored to the intended reaction to ensure optimized yields, the knowledge to date is predominantly empirical, with some mechanistic reports providing a fragmented view at best. Therefore, it is not easy to explain experimental observations, let alone make predictions. The aim of this study was to develop a structured, holistic understanding of the effects of NADES media on enzymatic reactions, distinguishing between effects on solubility, solvation, viscosity, inhibition and denaturation.

Experimental and computational chemistry methods were combined to separately study the interactions between enzyme, substrate, and NADES as reaction media. The initial enzyme activity and the final conversion of vinyl laurate transesterification by immobilized *Candida antarctica* lipase were studied experimentally. The direct effect of NADES on the same enzyme was modeled by molecular dynamics simulation. The effect of solubility was studied by both experimental and computational methods. To predict the solubility and viscosity of NADES, data-driven models were developed by combining group contribution and machine learning methods, based on the accumulated experimental knowledge on NADES found in the literature. Finally, the composed relationships and prediction models were applied to the practical example of deacetylation of mannosylerythritol lipids (MELs).

The experimental findings show that the chosen NADES system has a significant effect on both the apparent initial activity and the final conversion. However, in the simulations, the enzyme retains its original structure; moreover, NADES has an additional stabilizing effect on the enzyme. In addition, changes in the molar ratio of the compounds in NADES do not show a significant effect on the stability of the enzyme. These results indicate that the main effect of NADES on the reaction is mainly related to the substrate-solvent interactions (solvation energy) and the viscosity of the system. On the other hand, the experimental results only confirmed the significance of solvation, viscosity did not show a clear correlation with the studied reaction parameters. The machine learning models built for solubility and viscosity gave quantitative predictions of these properties. The accumulated knowledge was used to optimize the yield in the deacetylation reaction of MELs.

The combination of these methods provides fundamental knowledge about the effect of NADES on biocatalysis, but the results are also applicable to other uses of NADES.

Samenvating

Natuurlijke diepe eutectische oplosmiddelen (NADES) zijn veelbelovend als medium voor enzymatische reacties op gebieden waar (bio)compatibiliteit met natuurlijke of medicinale producten een must is. Hoewel zij in theorie kunnen worden afgestemd op de beoogde reactie om een optimaal rendement te garanderen, is de kennis tot dusver overwegend empirisch, waarbij sommige mechanistische rapporten in het beste geval een gefragmenteerd beeld geven. Daarom is het niet eenvoudig experimentele waarnemingen te verklaren, laat staan voorspellingen te doen. Het doel van deze studie was een gestructureerd, holistisch begrip te ontwikkelen van de effecten van NADES-media op enzymatische reacties, waarbij onderscheid wordt gemaakt tussen effecten op oplosbaarheid, solvatie, viscositeit, remming en denaturatie.

Experimentele en computationele scheikundige methoden werden gecombineerd om de interacties tussen enzym, substraat en NADES als reactiemedium afzonderlijk te bestuderen. De initiële enzymactiviteit en de uiteindelijke omzetting van vinyllauraattranssesterificatie door geïmmobiliseerd lipase van Candida antarctica werden experimenteel bestudeerd. Het directe effect van NADES op hetzelfde enzym werd gemodelleerd door moleculaire dynamicasimulatie. Het effect van de oplosbaarheid werd bestudeerd met zowel experimentele als computationele methoden. Om de oplosbaarheid en viscositeit van NADES te voorspellen werden datagestuurde modellen ontwikkeld door een combinatie van groepsbijdrage en machine-learning methoden, gebaseerd op de verzamelde experimentele kennis over NADES uit de literatuur. Tenslotte werden de samengestelde relaties en voorspellingsmodellen toegepast op het praktijkvoorbeeld van deacetylering van mannosylerythritol lipiden (MEL's).

Uit de experimentele bevindingen blijkt dat het gekozen NADES-systeem een significant effect heeft op zowel de schijnbare beginactiviteit als de uiteindelijke omzetting. In de simulaties behoudt het enzym echter zijn oorspronkelijke structuur; bovendien heeft NADES een extra stabiliserend effect op het enzym. Bovendien hebben veranderingen in de molaire verhouding van de verbindingen in NADES geen significant effect op de stabiliteit van het enzym. Deze resultaten wijzen erop dat het belangrijkste effect van NADES op de reactie voornamelijk verband houdt met de substraat-oplossingsinteracties (solvatie-energie) en de viscositeit van het systeem. Anderzijds bevestigden de experimentele resultaten alleen het belang van solvatie, viscositeit vertoonde geen duidelijke correlatie met de bestudeerde reactieparameters. De voor oplosbaarheid en viscositeit gebouwde machine-learning modellen gaven kwantitatieve voorspellingen van deze eigenschappen. De verzamelde kennis werd gebruikt om het rendement van de deacetyleringsreactie van MEL's te optimaliseren.

De combinatie van deze methoden levert fundamentele kennis op over het effect van NADES op biokatalyse, maar de resultaten zijn ook toepasbaar op andere toepassingen van NADES.

Chapter 1

Introduction

"A beginning is the time for taking the most delicate care that the balances are correct."
Frank Herbert, Dune

In the summer of 2018, our research group began to wonder how to utilize hydrophobic deep eutectic solvents for the deacetylation reaction of mannosylerythritol lipids. From this question, a long and winding road led to the proposal of a complete methodology for solvent selection. I'm sure that this elaboration is also reflected in the thesis. For this reason, I provide a preface to each chapter, explaining how its content relates to the larger picture. (You are reading one of these.) Many elements of this thesis have already been published as peer-reviewed articles. So as not to disrupt the flow of thought in them, these papers are included unaltered. However, to make it easier to relate to the overall work, I have also added summaries at the end of each chapter, where I explain how the results of the given article relate to the main research questions and overall goals of the thesis.

This chapter provides a brief overview of NADES, their properties and applications, and existing knowledge gaps to help the reader put the research into context. This is followed by an outline of the research, the questions and hypotheses that will be investigated, along with the objectives, scope, planned deliverables, and their limitations. The chapter concludes with an overview of how the rest of the dissertation will be structured and how previously published results will be incorporated.

A final note on the Introduction: since the full literature review is in the next chapter, some of the methodological choices may not be clear from the brief introduction of NADES presented here. I ask the reader to bear with me until the full review of the topic in Chapter 2.

1.1 Context

Deep eutectic solvents/systems are multicomponent mixtures that exhibit melting significantly below the ideal eutectic temperature [1]. In this thesis this is further restricted towards organic compounds, typically quaternary ammonium salts, and metal chlorides or hydrogen bond donors (HBD) [2]. Although the definition of "how deep is deep" is still debated in the scientific community [3], "deviation from such 'ideal-solution' behavior [1], which creates low-temperature melting materials and provides unexpected liquid environments" is the general consideration [4]. The mixtures of practical interest are those with melting point close to or even below room temperature [5]. For example, one of the most commonly studied systems, the 1 to 2 molar ratio mixture of choline chloride and urea, has a melting point of 17 °C [6]. In comparison, solid choline chloride decomposes at 305 °C [7] and urea melts at 132.7 °C [8]. This decrease in melting point is attributed to the strong interactions between the compounds, which includes the formation of an intermolecular hydrogen bonding network through the whole system [9, 10]. The hydrogen bonding network stabilizes the liquid state of the mixture and facilitates its use as a solvent [11]. Natural deep eutectic solvents (NADES) are the subgroup containing naturally derived chemicals [2] [12], usually quaternary ammonium salts and hydrogen bond donating compounds such as organic acids or polyols [13]. NADES are a green alternative to common organic solvents, due to their more benign properties [14, 15].

DES as solvents have many advantageous properties. Since DES are mostly ionic solutions, they have many properties similar to ionic liquids, but with additional advantages. Compared to commonly used organic solvents, DES are non-flammable, have low vapor pressure and low toxicity, providing a better alternative from both environmental and occupational safety perspectives [16, 17]. DES are "designer solvents", meaning that their required physicochemical properties can in theory be tailored to the task at hand [18]. The interactions between the salt and HBD groups define the behavior of the mixture [1]. By exploiting this behavior, the desired properties of the mixture can be optimized by changing the components and their molar ratio [19]. Compared to ionic liquids, DES are cheaper to produce and less sensitive to impurities, facilitating their use on a large scale [5]. NADES offer additional advantages because they are based on primary metabolites (organic acids, amino acids, amines, and sugars) that can be obtained from the waste streams of other processes [20]. As a result, they have good biodegradability, low production costs and are consistent with circular economy principles [21].

Non-flammability, low volatility, low toxicity and low price make NADES attractive candidates for many fields. Due to their novelty, practical applications are limited to date, including mostly electrochemistry [22, 23] and separation/extraction processes [24]. Nevertheless, treatment of waste streams [25, 26], biomass valorization (e.g., cellulose processing [27, 28, 29] or biodiesel production [30]), gas capture [31, 32], metallurgy [33, 34, 35], medical applications [36, 37, 38, 39] and media for biocatalysis [40, 41, 42, 43] are also considered. Due to their biodegradability and possible biocompatibility, NADES are likely to be applied in

[1] 'Ideal-solution' behavior implies that the change in interaction energies between different compounds and the same compounds is small and the decrease in melting temperature is determined by the entropy of mixing

[2] Although these compounds can be produced from natural sources (e.g., agricultural waste streams), currently most of the compounds come from the petrochemical industry

the pharmaceutical, food and feed sectors, where these properties are highly necessary [44, 4]. In this respect, NADES are suitable media for biocatalytic reactions, which emphasizes their potential in the aforementioned biochemistry-related fields.

Their benign properties make NADES a green alternative to commonly used solvents in biocatalysis. Due to their tailorable properties, NADES may even outperform organic solvents. However, these tailorable properties also pose the greatest challenge to their effective use, as knowledge of their structure-property relationships is severely lacking. Simply put, we do not understand how changes in the compounds used and their molar composition affect the physicochemical properties and behavior of NADES. In biocatalysis, NADES have been shown to influence enzymatic reactions, but the pathway of this influence remains unclear. A **general knowledge gap** in the field is that the influence of NADES on enzymatic reactions is poorly understood and their effect in biocatalysis cannot be easily described. This in turn inhibits their current applicability as tailored reaction media for biocatalysis. The more **specific knowledge gap** is the lack of a generic design methodology for NADES due to the lack of predictive models of their properties. Both experimental and computational studies have been published, but the available information is highly fragmented and does not allow the establishment of a generic predictive model, even if their effect on biocatalysis would be understood. This significantly hinders the application of NADES as designer solvents (even outside of biocatalysis).

1.2 Research outline

The fundamental problem with the application of (NA)DES is that we do not understand the relationship between the composition and the behavior of the solvent well enough to guide its design. In particular, the effect of NADES on enzymatic reactions is poorly described and therefore the effect of a novel eutectic system on an enzymatic reaction is unpredictable. This hinders the task-specific design of NADES for enzymatic reactions, as trial-and-error or design of experiments approaches are time consuming. Therefore, the purpose of this research is to describe the mechanism of how NADES affect the enzymatic reaction and to propose a framework for predicting it in a way that enables the design of task-specific NADES for enzymatic reactions.

The solve this problem, the following **research questions** have to be answered:

Q1 How do NADES influence enzymatic reactions?

Q2 How can these effects be expressed by discrete and quantitative properties?

Q3 How can the effect of NADES on biocatalysis be predicted in a structured way?

Q4 How to minimize the amount of empirical data required for the predictions?

Based on the information available in the literature, the following **hypotheses** are formed:

H1 It is clear from the literature that NADES compounds and other actors in the mixture interact with each other primarily though hydrogen bonds and the intermolecular

hydrogen bonding network present in the media. As a simple model, we consider the interactions between the NADES compounds, between the NADES and the substrates of the reaction, and between the NADES and the enzyme. The latter can be further distinguished based on the nature of the interaction. Stabilization/denaturation of the enzyme is possible as well as changes in the active sites leading to altered enzyme activity. Finally, side reactions with the media itself may occur.

H2 The direct interactions between the various actors in the mixture can be expressed by quantitatively measurable and predictable properties that are independent of each other. The interaction between NADES compounds is characterized by their viscosity. Viscosity also plays a significant role in the behavior of NADES, as mass transfer limitation is considered a major issue in these highly viscous media. The solubility and solvation energies of the substrates characterize the interaction between the media and the reactants. In addition, good solubility is a must for solvents. Finally, changes in enzyme activity can be calculated from reaction experiments (or assays) and changes in enzyme structure can be determined from computer simulations. These three types of interactions are independent, but their separate effect on reaction performance is difficult to determine experimentally because their effects add up. Separate determination requires careful experimental design.

H3 Predicting these properties requires a multi-scale and multi-step modeling approach. While the interactions between the enzyme and the NADES require a molecular-scale model, the bulk properties of the media can be predicted by macroscale or data-driven (quantitative structure-property relationship) approaches. Prediction of the behavior of novel systems is possible without experimental measurements, based only on the structure of the novel system.

H4 Data-driven methods require a comprehensive and structured database of available experimental data (from the literature). In addition, these data are also necessary for the validation of the predictions. By analyzing the experimental data and their structural variety used for the built model, the performance of the model for novel systems can be estimated. For the molecular and macroscale models, parameterization, usually based on experimental data, is also required (e.g. force fields for molecular dynamics simulations).

Investigating these questions and testing the hypotheses will provide a structured methodology for natural deep eutectic solvent selection and formulation for enzymatic reactions. The hypotheses are summarized in Figure 1.1 and in Figure 1.2.

These hypotheses divide the research into two distinct parts: First, describing the hypothesized effects of NADES and determining whether this approach provides sufficient information about the effect of NADES on the enzymatic reaction. Second, to develop predictive models of the descriptive properties of NADES and to determine how much data is needed to accurately predict them.

To achieve these aims a twofold combined methodological framework is formed (Figure 1.3): In the descriptive part, experimental measurements and molecular dynamics simulations are combined to determine how well the proposed properties of NADES explain the changes in

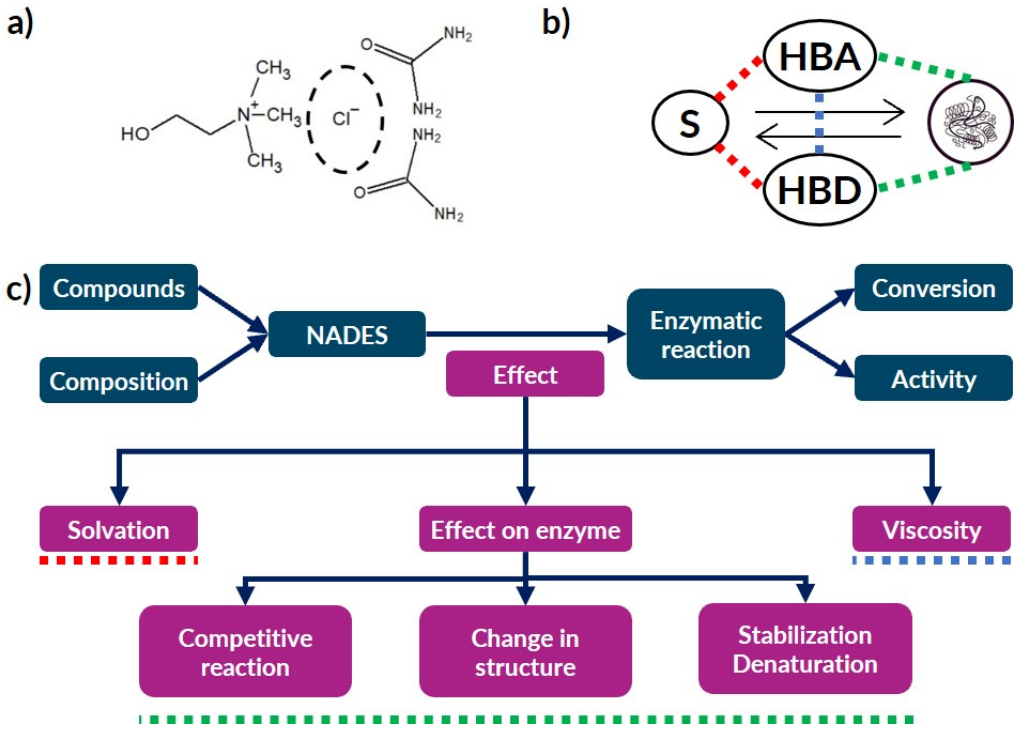


Figure 1.1: Hypotheses 1 and 2: a) The specific behavior of NADES is related to the strong hydrogen bonding in the system (here represented by the choline chloride and urea system). b) The effect of NADES on biocatalysis can be described by the specific interactions between NADES-substrate (S in the figure), NADES-NADES and NADES-enzyme. c) These interactions can be related to substrate solubility/solvation, media viscosity and changes in enzyme structure, respectively.

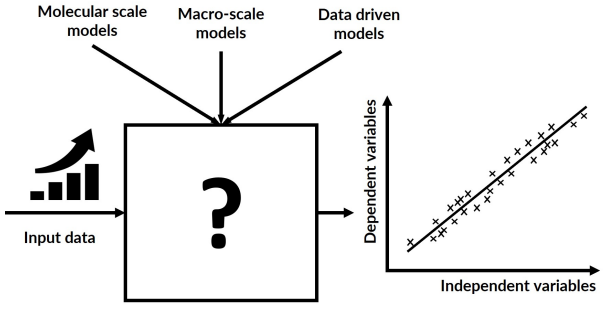


Figure 1.2: Hypotheses 3 and 4: The relevant properties can be predicted by a multiscale model combining molecular, macroscale and data-driven modeling approaches. The currently available application and the amount of additional data required can be determined by analyzing the model accuracy and the database.

biocatalysis (testing H1-H2). The monitored enzymatic reaction parameters are the conversion and the initial reaction activity. These are measured experimentally for the selected biocatalytic reactions. In these experiments, possible side reactions are also investigated. The overall stability/denaturation of the selected enzyme is studied both experimentally and by MD simulations. Structural changes of the enzyme are monitored in these MD simulations. Solubility and viscosity values of NADES and the solutes of interest are determined experimentally.

In the predictive part, data-driven and macroscale models are applied to predict the properties of NADES (i.e., solubility and viscosity) that describe their effect on biocatalysis (testing H3). These methods are based on machine learning with combination of group contribution methods, where predictive models are built based on literature data of NADES. For solubility prediction, machine learning density prediction is combined with PC-SAFT equation-of-state modeling. By analyzing the error of the models based on the amount of data applied and the type of NADES compounds predicted, the amount of additional data needed to increase accuracy can also be determined (testing H4).

Finally, the overall framework will be tested on a practical example to develop a task-specific NADES system for a biocatalytic reaction. This is the enzymatic deacetylation reaction of mannosylerythritol lipids, which is the initiating question of the whole project.

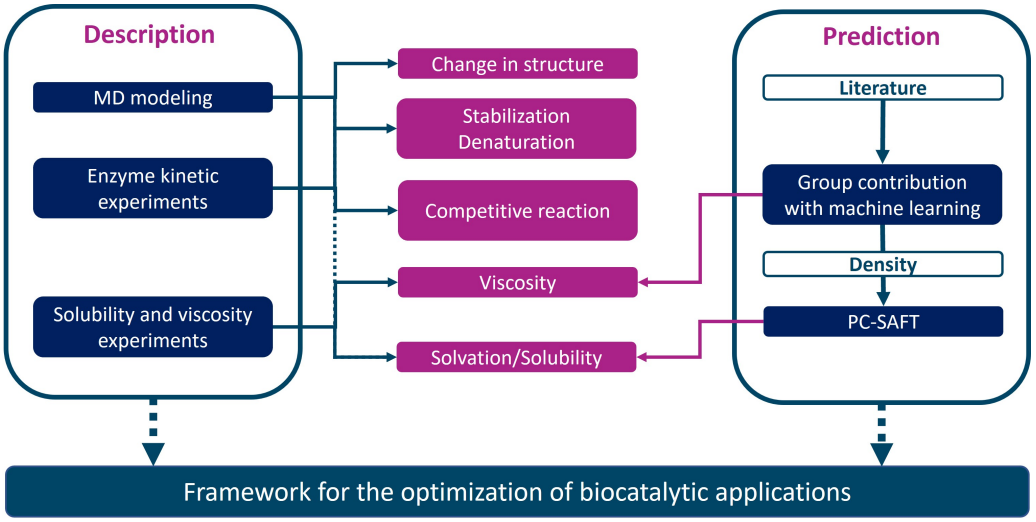


Figure 1.3: Methodological framework of the research project

The number of NADES combinations and possible biocatalytic reactions makes it impossible to give a general description of all possibilities within a single study. Therefore, the **scope of this research** is limited to a few selected systems and serves as a proof of concept. Nevertheless, the insights can later be extended to other systems. In this research only lipase enzymes were considered, more specifically the *Candida antarctica* lipase B. In the experimental part, two enzymatic reactions were studied, the transesterification of vinyl laurate to butyl laurate and the deacetylation of mannosylerythritol lipids. For the experiments and MD simulations, NADES systems containing choline chloride with either urea, ethylene glycol or glycerol were considered. In addition to the composition, the effect of the molar

ratio was also investigated. In the MD simulation, the focus was on simulating the structure of NADES and the enzyme. Therefore, the force field for the simulation was taken from the literature (see Chapter 4). The enzymatic reaction was monitored by calculating the final conversion and the initial reaction activity. For predictive modeling, supervised regression machine learning models were considered. The database for machine learning was based on literature data, the sources used are listed in Chapter 5.

This research provides a theoretical framework for describing and predicting the effect of NADES on enzymatic reactions. The modeled systems described in the scope are used to gather information for the descriptive part and to validate the framework for predicting properties. The analysis of the existing data is an important feature of this research. Since the framework of the proposed prediction methods can be implemented on novel systems, these elements represent a versatile solution. Although this research will not create an off-the-shelf solution for generic NADES application problems, the results are applicable for decision support in real-life NADES applications. However, this research does not attempt to provide a generic solution applicable to any NADES and enzyme with quantitative output, as molecular dynamics simulation is a bottleneck in the recommended framework. The discussed PC-SAFT solubility model (detailed description will follow in Section 2.5.2.2) is well established in the literature, but this work focuses only on the implementation of data-driven methods.

1.3 Thesis outline

The discussion of the research is divided into 7 chapters (the structure is shown in Figure 1.4). After the overview in this chapter, the detailed background of NADES and methods to predict their properties are discussed in detail in Chapter 2. Here, after discussing the general properties and application of NADES, the different modeling techniques of NADES are introduced, together with the related methodological choices of this research. Chapter 3 covers the investigation of enzyme kinetic experiments of vinyl laurate transesterification, which reveals the role of solubility, viscosity and experimental studies on the denaturation of the enzyme, together with the possible side reactions of the studied systems. Chapter 4 discusses the molecular dynamics simulations and the changes in the enzyme structure, together with the experimental measurements of the viscosities of NADES. Chapters 3 and 4 together cover the descriptive part of the study. In the following Chapter 5 the results of the predictive part are described. The first part of Chapter 5 describes the density and viscosity modeling efforts, including a discussion of data sources, capabilities, and limitations of the models developed. The second part covers the combination of the density model with the PC-SAFT method for the prediction of solubility in NADES. In Chapter 6, the developed framework is applied to optimize the deacetylation of mannosylerithritol lipids in NADES. Finally, in Chapter 7, the findings are summarized along with the research lines that were not fully explored and additional considerations that arose during the project.

This thesis is largely organized around previously published or submitted papers and conference presentations. These papers form the core of each chapter. Figure 1.4 shows the structure of the thesis along with the included papers. The already published papers are:

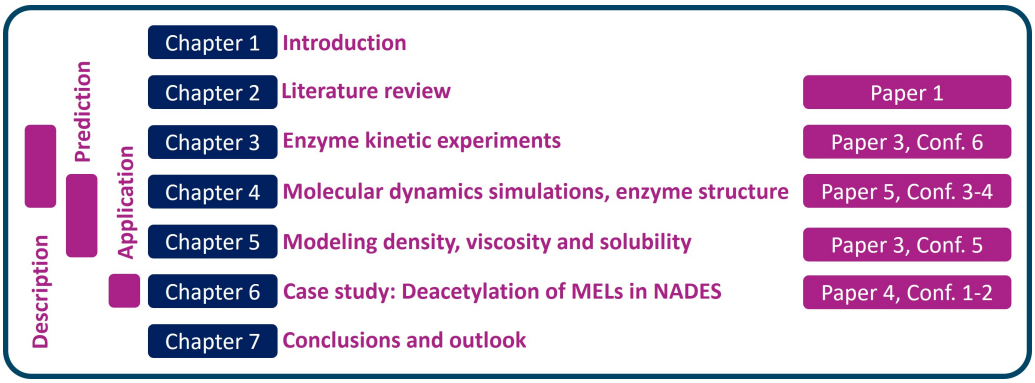


Figure 1.4: Outline of the thesis

Paper 1: Modeling the Physicochemical Properties of Natural Deep Eutectic Solvents - A Review (ChemSusChem 2020, 13, 3789–3804, DOI:10.1002/cssc.202000286). A review of the possible methods for modeling and predicting the properties of DES. This paper gives the main part of the literature review.

Paper 2: Effect of natural deep eutectic solvents of non-eutectic compositions on enzyme stability (Journal of Molecular Liquids 366 (2022) 120180, DOI: 10.1016/j.molliq.2022.120180) This paper discusses the findings of the molecular dynamics simulation and gives the main part of Chapter 4.

The submitted papers are:

Paper 3: Biocatalyzed vinyl laurate transesterification in natural deep eutectic solvents (submitted to Waste and Biomass Valorization). This paper discusses the effect of NADES on the biocatalyzed vinyl laurate transesterification. This research gives the core of Chapter 3.

Paper 4: Deacetylation of mannosylerythritol lipids in hydrophobic natural deep eutectic solvents (submitted to Applied Microbiology and Biotechnology). This paper discusses the application of the developed framework of NADES selection on the example of the deacetylation of mannosylerithritol lipids. This work gives the main part of Chapter 6.

The last paper is still in preparation:

Paper 5: Combining group contribution with machine learning to predict the density of deep eutectic solvents. This paper contains the findings of the predictive modeling of the density of deep eutectic solvents, discussed in Chapter 5.

The content of these publications is used in the following chapters. They are included without change in their content, but with annotations to link their findings to the overall research objectives and questions.

The conference contributions, which are also part of this thesis, are as follows:

Conference 1: Enzymatic Conversion Of Mannosylerythritol Lipids In Natural Deep Eutectic Solvents. 27th European Biomass Conference & Exhibition (2019). In this presentation, the concept of using natural deep eutectic solvents as reaction media for biocatalysis was presented using the deacetylation of mannosylerithritol lipids as an example. The results presented are also part of Chapter 6.

Conference 2: Enzymatic Conversion Of Mannosylerythritol Lipids In Natural Deep Eutectic Solvents. 1st International meeting on Deep Eutectic Systems (2019). This presentation covered the preliminary efforts of biocatalytic deacetylation of mannosylerithritol lipids. The results presented are also part of Chapter 6.

Conference 3: Towards the mechanistic understanding of natural deep eutectic solvents effect on enzyme catalyzed reactions. 1st GREENERING International Conference (2021). This presentation covered the preliminary findings of molecular dynamics modeling of the CALB enzyme in NADES. The results presented are also part of Chapter 4.

Conference 4: Towards the mechanistic understanding of natural deep eutectic solvents effect on enzyme catalyzed reactions. 2nd International meeting on Deep Eutectic Systems (2021). This presentation covered the findings of molecular dynamics modeling of the CALB enzyme in NADES. The results presented are also part of Chapter 4.

Conference 5: Predicting the density of natural deep eutectic solvents by the combination of a group-contribution method and artificial neural networks. 32nd European Symposium on Computer Aided Process Engineering (2022). This presentation covered the preliminary results of the data-driven approach to predict the density of deep eutectic solvents. The results presented are also part of Chapter 5.

Conference 6: Understanding biocatalysis in natural deep eutectic solvents: transesterifications with *Candida antarctica* lipase B in various solvents. 9th International Conference on Engineering for Waste and Biomass Valorisation (2021). This presentation covered the findings on how to describe the effect of NADES on biocatalysis via their effect on the enzyme, their viscosity and the solubility of the substrate. The results presented are also part of Chapter 3.

Chapter 2

Literature review: NADES and their modeling approaches

"You are now about to witness the strength of street knowledge"

NWA

2.1 Preface

The interest in deep eutectic solvents and the number of related publications are still growing year by year. The annual published research on DES has been over a thousand papers in the last three years. Because of that I focus my review on a very selected section of the domain.

I hypothesize that the effect of NADES on enzymatic reactions can be described by certain physicochemical properties, namely the solubility of the substrates, the viscosity of the media, and the structural changes of the enzyme induced by the media. Furthermore, I assume that these properties can be modeled and predicted by the modeling methods available in the literature. This leads to the second related hypothesis, that the existing methods can be combined into a complex model that describes the holistic behavior of NADES in enzymatic reactions. However, to decide if this approach is feasible and to choose which models to include, I first review the NADES modeling methods available in the literature.

To this end, this chapter reviews the available modeling techniques of NADES properties, along with their general characteristics and potential applications. The distinction between micro-, macro- and data-driven levels of modeling is made here, and the review already discusses the available methods in this structure.

This review has already been published in the journal ChemSusChem with the title "Modeling the Physicochemical Properties of Natural Deep Eutectic Solvents" [11]. Co-authors of this paper are Erik C. Neyts, Iris Cornet, Marc Wijnants and Pieter Billen, whose contributions include supervision of the project and review and editing in the writing process. In addition, Pieter Billen acquired the financial support for the project.

2.2 Abstract

Natural deep eutectic solvents (NADES) are mixtures of naturally derived compounds with a significantly decreased melting point due to the specific interactions among the constituents. NADES have benign properties (low volatility, flammability, toxicity, cost) and tailorable physicochemical properties (by altering the type and molar ratio of constituents), hence they are often considered as a green alternative to common organic solvents. Modeling the relation between their composition and properties is crucial though, both for understanding and predicting their behavior. Several efforts were done to this end, yet this review aims at structuring the present knowledge as an outline for future research. First, we reviewed the key properties of NADES and relate them to their structure based on the available experimental data. Second, we reviewed available modeling methods applicable to NADES. At the molecular level, density functional theory and molecular dynamics allow interpreting density differences and vibrational spectra, and computation of interaction energies. Additionally, properties at the level of the bulk media can be explained and predicted by semi-empirical methods based on ab initio methods (COSMO-RS) and equation of state models (PC-SAFT). Finally, methods based on large datasets are discussed; models based on group contribution methods and machine learning. A combination of bulk media and dataset modeling allows qualitative prediction and interpretation of phase equilibria properties on the one hand, and quantitative prediction of melting point, density, viscosity, surface tension and refractive indices on the other hand. In our view, multi-scale modeling, combining the molecular and macro-scale methods, will strongly enhance the predictability of NADES properties and their interaction with solutes, yielding truly tailorable solvents to accommodate (bio)chemical reactions.

2.3 Introduction

Deep eutectic solvents (DES) were first reported in 2001, as novel solvent class based on eutectic systems and potential alternative to ionic liquids, when Abbot et al. observed significant decrease in the melting point of metal chlorides and quaternary ammonium salts [45]. These DES showed many common characteristics with ionic liquids (IL) and their melting point was below room temperature in some cases. Although the term DES was introduced in 2003, no clear agreement was made upon its definition. The formerly published articles set up different requirements for DES: Zhang et al. described DES as the composition of two or more components, which are interacting through a hydrogen bond interaction, forming a eutectic mixture [5]. Francisco et al. connected the lowered melting point to the lowered entropic difference of the phase transition [18]. Smith et al. defined the components as Lewis or Brønsted acids and bases [1]. Paiva et al. only set the requirement of the significantly lowered melting point, compared to the melting point of the individual components [14]. Due to the unclear definition, we introduce only their general characteristics. DES are a mixture of two or more compounds which are associating through hydrogen bonding. The most frequently studied DES are formed by a quaternary ammonium salt and a metal salt or a hydrogen bond donor (HBD) component. The salt components in DES typically have low lattice energy. The components form a hydrogen bonding network, which increases the system's stability and in case of ionic constituents, allows the charge delocalization. The

decreased melting point is the result of these effects. Figure 2.1 illustrates the hydrogen bonding interactions on the most commonly studied DES, the mixture of choline chloride and urea.

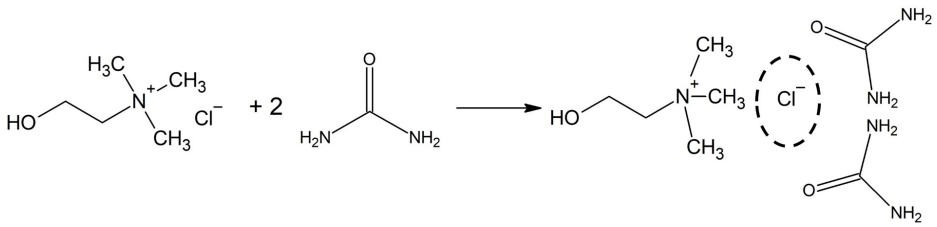


Figure 2.1: Deep eutectic solvent preparation from choline chloride (hydrogen bond acceptor component) and urea (hydrogen bond donor component). The dashed circle represents the complex hydrogen bonding formed between the constituents.

Natural deep eutectic solvents (NADES) were reported in 2011 by Choi et al., as a subclass of DES [13]. NADES are composed of naturally derived components, such as the primary metabolites, carboxylic and amino acids, choline chloride, sugars or urea. Choi et al. discussed the possibility of NADES presence in living organisms as a third liquid phase next to water and lipids, what could rationalize the biosynthesis of poorly water soluble macromolecules in cells that are otherwise aqueous environments, or the survival of organisms in arid environments. Changing common organic solvents to NADES have many benefits, as the latter ones are often non-volatile and non-flammable [14]. Compared to ionic liquids, NADES are cheaper with better biodegradability and lower toxicity. NADES can be manufactured with 100% atom economy and they are less sensitive to impurities than ILs. The potential of NADES was already proven in organic reactions, extraction processes and electrochemistry applications [5]. The most promising field for NADES is the biochemical industry, as NADES not only provide a green medium for enzymatic reactions, but can also increase the efficiency of certain biocatalytic reactions. [41]. The eutectic mixtures seem to modify both the reactivity and stability of enzymes compared to conventional media. The growing interest in NADES is also reflected by the number of academic publications and citations (See Figure 2.2).

The biggest potential of DES and NADES is their application as designer solvents [18]. Designer solvents are systems of which the properties can be tailored according to the application. In DES, the hydrogen bonding network between the components defines the behavior of the given mixture to a large extent. Therefore the desired properties can be achieved by changing the components used and their molar ratio. Currently, the relation between the composition and the properties of NADES is not described adequately and the development of new systems done through trial-and-error. Given the lack of generic design methodology to predict the relevant properties of novel NADES systems, such as viscosity, surface tension or solubility, these eutectics cannot be used as “designer solvents”. The relation between NADES structure, properties and the possible application is conceptually illustrated in Fig. 2.3. Recognizing the importance of the subject various studies investigated the possibility of modeling NADES behavior, both in view of a better understanding of structure-property relations, and in order to predict the properties of these systems. Despite the high number of publications related to this field, to the best of our knowledge, no review structuring the recent advances in the modeling of NADES’ behavior was reported.

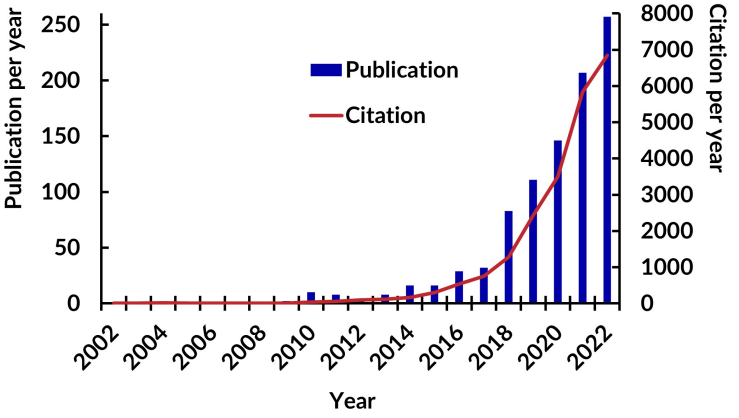


Figure 2.2: Number of yearly publications (right vertical axis), and yearly citations (left vertical axis), using the keywords "natural deep eutectic solvent" OR " natural deep eutectic solvents" OR "NADES" (Source: Web of Science; date of search: 2023.03.14).

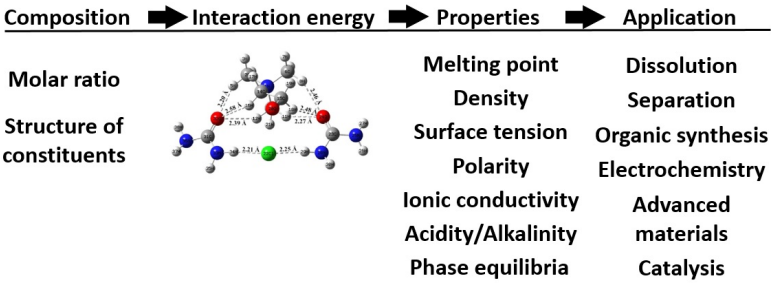


Figure 2.3: Structure-properties-application relation (Interaction energy of choline chloride and urea by Zhu et al.[9]

The aim of this review is therefore to introduce the current state of understanding on the structure-property relations of NADES and the advances in its modeling. This paper gives a general overview of the available methods to researchers and developers, who are new to the field of modeling physicochemical properties. First, we discuss the relevant physicochemical properties of NADES, their relation to the composition and general considerations to tailor the properties of systems that were explored thus far. In the second part, we introduce the available methods to model NADES behavior, starting at the level of molecular interactions, with subsequently the prediction of bulk properties of NADES such as density or viscosity, and finally the analysis of larger datasets with machine learning methods.

2.4 Structure-property relations of NADES

The characteristic properties of NADES (e.g., decreased melting point, high viscosity) are the result of the intermolecular hydrogen bonding system among the components. The

strength of this interaction determines to a large extent the physicochemical properties of the eutectic mixture, which are relevant in their application as solvents (discussed below) and it depends on the structure and ratio of the components in the mixture. To design the properties of NADES, the relation between the properties and composition of the mixture has to be known. For a given application, the tailored medium has to possess good solvation, transport properties and its use should be safe. Nonetheless, the additional effects of NADES on the reaction (e.g., side reactions, inhibition, catalytic effect) have to be considered as well. During development, there are many relevant properties to take into account, such as melting point, density, viscosity, polarity, ionic conductivity, acidity or alkalinity, and surface tension. The solubility of different chemicals in the mixture is for obvious reasons also a key aspect. The water content of the eutectics has a significant effect on all the properties, therefore it will be discussed in more detail, along with the effect of NADES on (bio)catalytic reactions as a major application domain.

2.4.1 Melting point

The application potential of a eutectic system is determined by the eutectic point (the lowest melting point of the system and its associated composition), i.e. the mixture can be used as a solvent only in the temperature ranges where it forms a stable liquid phase. However, the temperature of a specific application determines an allowed composition range for potential use of a given NADES (See Fig. 2.4). Although many NADES were reported in the literature [5, 6, 33, 46, 47, 48, 49], systems with a melting point below room temperature are still scarce. NADES containing amides, carboxylic acids and sugar-derived polyols with organic salts, often have melting points below room temperature.

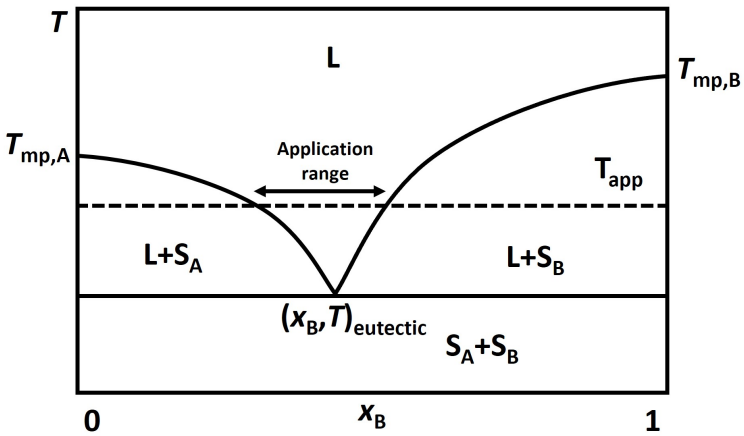


Figure 2.4: T-x phase diagram of simple eutectic behavior: the diagram shows the change of the system's melting point (T) as the function of molar ratio of constituent B (x_B) in a binary mixture of A and B. $T_{mp,A}$, $T_{mp,B}$ are the melting points of the pure A and B constituents, T_{app} is the temperature of the given application, L is the liquid phase, while S_A and S_B are the solid phases of constituent A and B. $(x_B, T)_{\text{eutectic}}$ indicates the eutectic composition and temperature of the mixture.

Zhang et al. comprehensively studied the empirically observed relation between structure and melting point.[5] HDB components with lower molecular weight yields a larger melting point depression. The asymmetry of the cation decreases the melting point (due to smaller lattice energies), while the increasing electron affinity of the anion also has a decreasing effect on melting point (as this yields stronger hydrogen bonds between the two constituents). The addition of a ternary component to the system can also lead to a decrease of the melting point.[48, 49] By comparing the eutectic points of mixtures with organic salts and mono- or dicarboxylic acids, it was found that the complexation of the anion requires two molecular carboxyl groups.[5] These observations are in line with the general assumption that stronger interactions between the components result in a greater decrease of the melting point.

In the NADES of choline chloride and carboxylic acids, the lattice energies of the HBD does not show a correlation with the melting point depression, but the HBD with the lowest molecular weight is associated to the largest melting point depression.[5] From this observation, Abbott et al. one can assume that the lattice energy of the HBD only relates to the HBD-anion interaction, therefore the melting point depression is related to the mixing entropy only.[6] Consequently, in a first approximation the melting point depression is a measure of the entropy change, i.e., the magnitude of the depression relates to the increase of entropy change during mixing.

In this view, we suggest an approach based on the method of Krossing et al. (which was developed to rationalize the low melting points of ionic liquids) to describe the relation between molecular interactions of constituents and the change of the melting point.[50] During the melting (fusion) of the NADES constituents, both the enthalpy (H) and the entropy (S) of the system change. These changes can be described by the change of Gibbs free energy at a phase transition:

$$\Delta_{\text{fus}}G = \Delta_{\text{fus}}H - T \cdot \Delta_{\text{fus}}S \tag{2.1}$$

At the melting temperature, so also at the eutectic point, the ΔG of melting (fusion) is zero:

$$\Delta_{\text{fus}}G^{\text{tot}} = 0 \tag{2.2}$$

At other temperatures, the value of $\Delta_{\text{fus}}G$ indicates, whether the solid ($\Delta_{\text{fus}}G > 0$) or the liquid ($\Delta_{\text{fus}}G < 0$) state is the thermodynamically preferred phase state (See Fig 2.5).

The change of Gibbs free energy during melting can be described as a Born-Fajans-Haber cycle of the fusion, lattice and solvation energies of the components (See Figure 2.6).

By that, the change of Gibbs free energy during melting:

$$\Delta_{\text{fus}}G^{\text{tot}} = \Delta_{\text{fus}}G^{\text{salt}}_{\text{lattice}} + \Delta_{\text{fus}}G^{\text{HBD}}_{\text{lattice}} - \Delta_{\text{fus}}G^{\text{HBD}}_{\text{solvation}} - \Delta_{\text{fus}}G^{\text{Anion}}_{\text{solvation}} - \Delta_{\text{fus}}G^{\text{Cation}}_{\text{solvation}} \tag{2.3}$$

The $\Delta G_{\text{lattice}}$ contains the lattice enthalpy ($\Delta H_{\text{lattice}}$) and the entropy change of forming ionic lattice from infinitely separated gaseous ions ($\Delta S_{\text{lattice}}$):

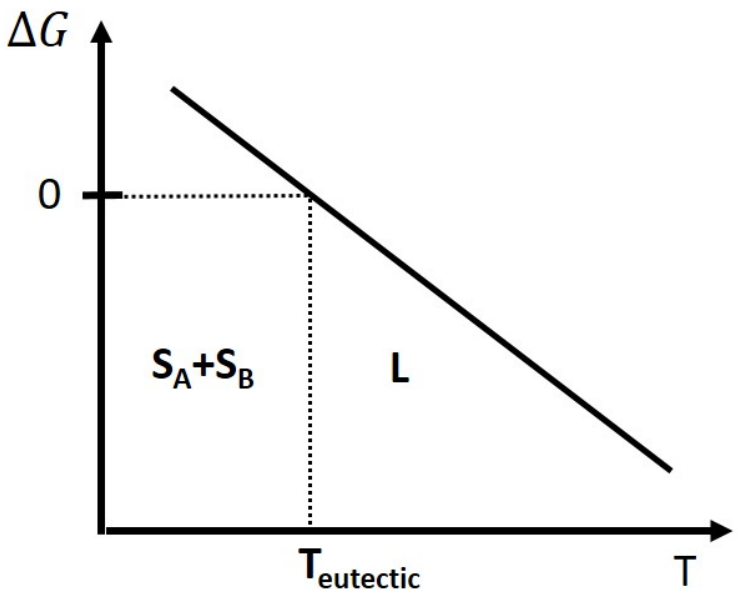


Figure 2.5: The change in Gibbs free energy of melting at different temperatures at eutectic composition: T_{eutectic} is the melting point of the eutectic mixture, L is the liquid phase S_{A} and S_{B} is the solid phase of constituent $\mathbf{A}$ and $\mathbf{B}$.

$$\Delta G_{\text{lattice}} = \Delta H_{\text{lattice}} - T \cdot \Delta S_{\text{lattice}} \tag{2.4}$$

The lattice enthalpy can be described as the function of lattice potential energy and the change in molar volume ($\Delta V_{\text{lattice}}$):

$$\Delta H_{\text{lattice}} = \Delta U_{\text{lattice}} - p \cdot \Delta V_{\text{lattice}} \tag{2.5}$$

We can assume that contribution of molar volume change is smaller for such less ordered, bulky compounds, what also contribute to lower melting points. When these expressions (Eq. 2.3-2.5) are implemented in the equation of the change of Gibbs free energy during melting, and isolating the eutectic temperature, we can identify the effects altering the melting point:

$$\Delta T_{\text{eutectic}} = \left[\frac{\begin{array}{c} \Delta U^{\text{salt}}_{\text{lattice}} + p \cdot \Delta V^{\text{salt}} + \Delta U^{\text{HBD}}_{\text{lattice}} + p \cdot \Delta V^{\text{HBD}} \\ - \Delta H^{\text{HBD}}_{\text{solvation}} - \Delta H^{\text{Anion}}_{\text{solvation}} - \Delta H^{\text{Cation}}_{\text{solvation}} \end{array}}{\Delta S^{\text{salt}}_{\text{lattice}} + \Delta S^{\text{HBD}}_{\text{lattice}} - \Delta S^{\text{HBD}}_{\text{solvation}} - \Delta S^{\text{Anion}}_{\text{solvation}} - \Delta S^{\text{Cation}}_{\text{solvation}}} \right] \tag{2.6}$$

The equation implies that smaller lattice energies, stronger interactions between the constituents and a higher change in entropy during solvation results in a decrease of the melting

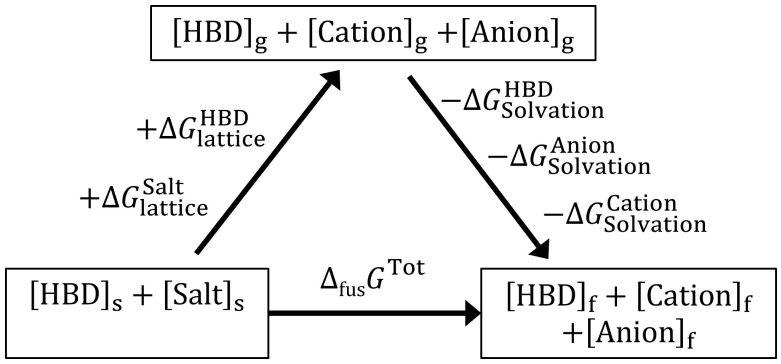


Figure 2.6: Born-Fajans-Haber cycle of NADES melting, lattice and solvation energies.

temperature. This means that weaker crystalline structure, bulky ions and stronger interaction between the constituents yield a lower melting point. These relations are in line with the general observations. However, solving these equations requires extensive modeling to determine the lattice and solvation enthalpies, potential energies and changes in molar volume.[50]

2.4.2 Density, viscosity and surface tension

Most NADES have densities higher than that of water (between 1.1 and 1.4 g/cm) [24] and significantly higher viscosities than common (organic) solvents (e.g., the viscosity of water and toluene is 0.89 mPa·s and 0.56 mPa·s at 25 °C, respectively, while that of the eutectic mixture of choline chloride and urea is 750 mPa·s at 25 °C) [51]. The surface tension of eutectic solvents is also larger than that of most molecular solvents and similar to imidazolium-based ILs [5]. High viscosity entails significant limitations to the mass and energy transfer during chemical reactions, therefore limiting the viscosity of NADES is necessary by selecting smaller constituent molecules with fewer hydrogen bond donating/accepting groups and weaker interaction. However, strong intermolecular forces contribute to low melting points and good solubility, which results in a trade-off between transport and solvation properties.

Density, viscosity and surface tension are mostly discussed together, as they show a similar relation to the intermolecular interaction energy and temperature (See Fig 2.7). Higher interaction energy increases all three properties [47, 5, 52], while higher temperature results in a decrease of these properties [47, 53, 54, 55]. The relation of these properties to the temperature can best be described by non-linear functions, but the effect of the structure of the constituents is only vaguely described. The density showed a quadratic relationship with the temperature [54], while the viscosity-temperature relation follows an Arrhenius equation (see Fig 2.7. a)) [55]. Increasing interaction between the components yields a higher density [5]. Comparing the viscosity of sugar derivatives, ethylene glycol and glycerol based NADES, viscosity shows an increase with the hydrogen bonding ability of these constituents (more hydrogen group donating group in sugars) [5]. Strong hydrogen bonding seems to hinder the mobility in the NADES. The comparison of choline chloride-urea and choline chloride-ethylene glycol mixtures suggests a higher surface tension with the increasing strength of

the hydrogen bonds in the first instance [52]. The disruption of the hydrogen bond network decreases all three values (e. g., the addition of a quaternary component) [47, 5].

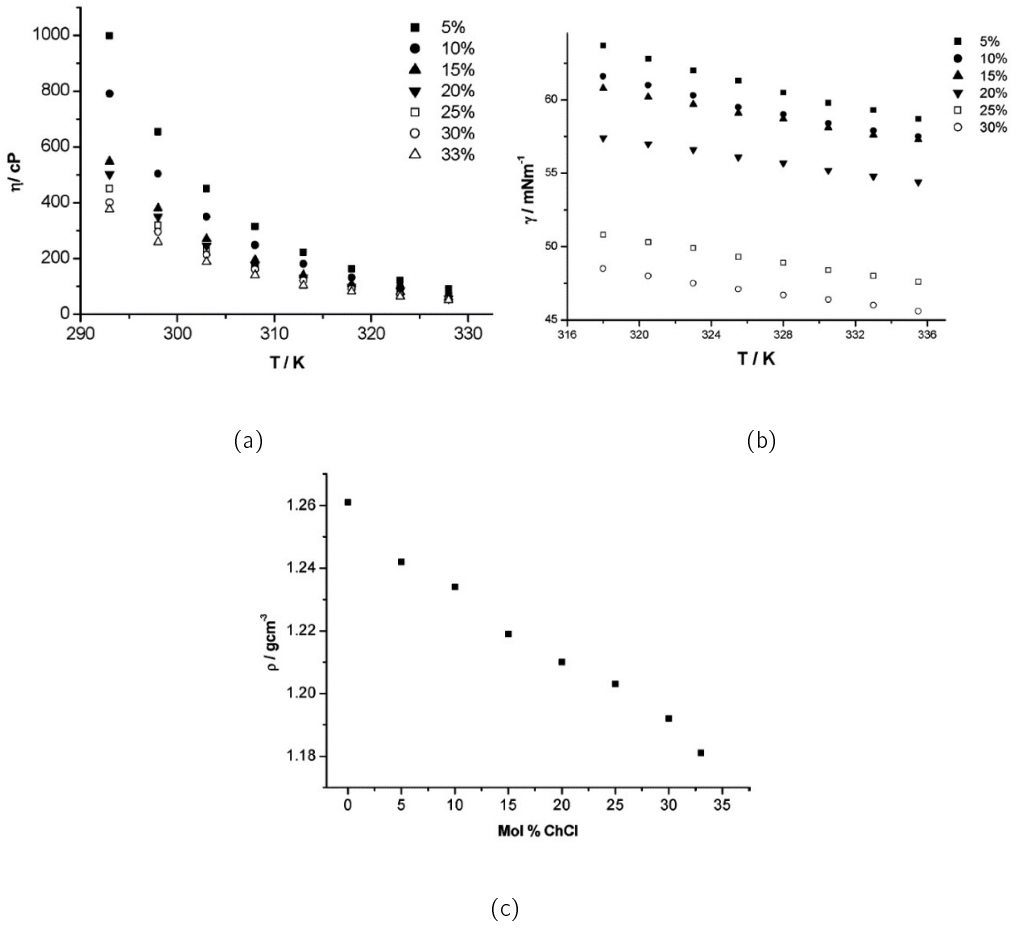


Figure 2.7: Relation between the composition and properties of the NADES of choline chloride - glycerol. (Taken from the work of Abbot et al.[47]) The different markings on a) and b) indicate the choline chloride ratio in the mixture (0.05-0.33 molar ratio). a) Viscosity of the NADES as function of temperature and composition, b) Surface tension of NADES as the function of temperature and composition, c) Density of the NADES as the function of composition.

To rationalize the observed relations, the hole theory was considered [33, 56]. This phenomenological model describes ionic liquids as an ensemble of particles and vacancies with variable size (See Fig 2.8.) [57]. Its application was suggested due to the analogies between NADES and ILs. The average hole size can be calculated from experimental measurements of the surface tension, and can subsequently be used for calculation of other properties of the system. The increase of viscosity is interpreted in the model as a result of a decreased average hole diameter (due to stronger interactions). Abbot et al. studied the design of low viscosity NADES by application of this hole theory, but the model did not prove to be sufficiently accurate [56].

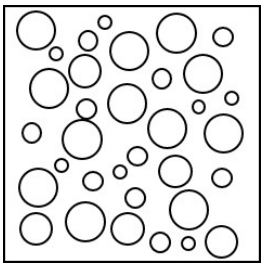


Figure 2.8: The hole theory with randomly located and variable-sized holes in the liquid. The model describes the structure of liquid salts, as they would have "tiny volume elements, varying in size from the subatomic to about six ions, which are empty and constantly fluctuating in size [57]."

2.4.3 Polarity, ionic conductivity and pH

The polarity, ionic conductivity and the acidity/alkalinity of NADES received little attention thus far, although these properties are relevant for many applications in electrochemistry. The polarity of choline chloride and glycerol mixtures shows similar values to $RNH_3^+X^-$, $R_2NH_2^+X^-$ and imidazolium-based ILs and it increases linearly with the salt concentration [47]. Due to their high viscosities, NADES typically show poor conductivity [47]. An increase in temperature (as it decreases the viscosity) increases the conductivity. The molar ratio of constituents also affects the conductivity through its effect on the viscosity. An increased concentration of charge carrying species (salts) has a positive effect on conductivity [48]. The nature of the HBD strongly affects the acidity/alkalinity of NADES. The choline chloride-urea eutectic mixture shows weak alkalinity [58], while sugar derived polyols as HBD yield NADES with neutral pH [49]. The alkalinity makes the absorption of acidic gases, such as CO_2, preferable.

2.4.4 Water content

Many NADES are susceptible to water uptake, as water can take part in the hydrogen bonding network and quaternary ammonium salts are highly hygroscopic. The incorporation of water into the hydrogen bonding system and its altering effect was demonstrated by both NMR spectroscopy [24] and molecular simulations [59]. Consequently the water content affects all of the other properties significantly. $5w/w\%$ of water absorption decreases the melting point of the choline chloride-urea mixture from $25°C$ to $15°C$ [60], while the viscosity decreased by 83% [61]. The absorbed water disrupts the interaction between the components, yielding a decrease in density and viscosity. This sensitivity calls for special care during the characterization of physicochemical properties, as the water intake during the measurements itself yields significant deviation in the results [59]. On the other hand, by controlled addition of water, the properties of NADES can be fine-tuned.

2.4.5 Solubility

Solvation is a complex process, as the solubility and solvation energy depend on many factors. In thermodynamic terms of view, stronger hydrogen bonding among the NADES constituents should decrease the enthalpy of the solvation and therefore increasing the dissolution, but the available information in the literature is limited.

NADES as solvent were studied mainly in three distinctive fields: the dissolution of CO_2, metal oxides and drugs. The dissolution of metal oxides in NADES for the separation and recycling of metals can be a green process in electrochemistry. Abbot et al. demonstrated first the solubility of metal oxides in NADES [33]. The solvation of the metals and metal oxides occurs via their complexation with the NADES molecules, therefore the structure of the applied constituents has a significant effect on the solubility, e.g. metal oxides with more covalent character (such as TiO_2) are poorly soluble [5]. Elevated temperature also increases the solubility. Quantum chemical simulations showed that the hydrogen network forms an open cluster and the solvation accompanied by proton transfer [62].

The application of NADES for CO_2 is intensively studied. The solubility of CO_2 depends on three factors, viz. the partial pressure of CO_2, the temperature and the molar ratio of the NADES components [63]. Higher pressures and lower temperatures increase the solubility of CO_2, while the NADES with eutectic molar composition yield higher CO_2 solubility than NADES with excess of either HBA or HBD constituents. Increasing the water content has a negative effect on the solubility, since water acts as an antisolvent for CO_2 in the eutectic mixture of choline chloride and urea [64].

As the hydrogen bonding network can form strong interactions, even with larger molecules, the solubilization of otherwise poorly soluble drugs might be done in NADES. Morrison et al. reported the increased solubility of poorly water soluble molecules in NADES and NADES/water mixtures [65]. The solubility of rutin also showed improvements in urea, sugar, organic acid and choline chloride based NADES, compared to water [66]. In latter case, the presence of basic sites in the NADES lead to an increased solubility due to the acidic properties of the rutin [67].

2.4.6 Effect on enzymatic reactions

NADES provide not only green media for enzymatic reactions, but they modify the kinetic parameters and yields of reactions compared to conventional media [41]. Strong hydrogen bond donors such as urea were beforehand expected to denature proteins, but enzymes remained stable in the eutectic mixture in earlier reports. The study of Monhemi et al. showed, that the intermolecular hydrogen bonding network decreases the denaturing effect of the individual NADES constituents, by preventing the NADES constituents from diffusion into the protein chain and disrupting its secondary structure [68].

In case of the aminolysis of ethyl valerate with lipase B from *Candida antarctica* in a choline chloride-glycerol mixture, an increase of 13 % in enzyme activity was reported, compared to toluene as solvent [69]. The choline chloride-urea mixture increased the half-life of

horseradish peroxidase enzyme from 50 minutes (in phosphate buffer) up to 350 minutes [70]. The latter observation was related to the stabilization effect of the hydrogen bonding network in NADES.

The reports show that NADES may influence both the conversion and kinetics of enzymatic reactions. The thermodynamic stability of substrates and products is altered by solvation, while the viscosity affects the mass transfer of all the reactants (substrate, product, catalyst). The enzyme/NADES interaction could lead to the stabilization or denaturation of the enzyme, but also change in enzyme's active site, secondary or tertiary structure. These phenomena are the result of the hydrogen bonding interactions between the constituents of the NADES and/or the actors of the enzymatic reaction. Moreover, competing reactions of the NADES constituents with the enzyme and/or the substrates are also possible. However, the available information is highly fragmented and the relationship between these effects and the structure of NADES is still unclear, even on qualitative level. As we will discuss in the later sections, qualitative relations between the NADES structure and its effect on enzymatic reactions can be elucidated by computational methods.

2.5 Modeling the properties of NADES

In the former section we reviewed the relevant physicochemical properties and phenomena of NADES, which have to be taken into account during of the solvent design for new applications. By modeling eutectic mixtures, we can better understand and even predict the behavior of these systems. As mentioned earlier, no comprehensive model of such systems was created yet. However, many different methods are available to describe certain distinct characteristics of given eutectics. While introducing these characteristics, we will proceed from the modeling of intermolecular interactions to the direction of bulk systems. First, the molecular scale models will be discussed; predominantly based on density functional theory and molecular dynamics. Thereafter we discuss macroscale models that may calculate the properties of the bulk material (solubility, density, viscosity). Lastly, we discuss methods based on large amount of experimental data, like group contribution methods and machine learning algorithms. Overall, this section focuses on the theoretical background of the methods, their field of application along with the advantages and drawbacks, case studied of actual applications and future challenges.

2.5.1 Molecular scale modeling

The aim in computational chemistry is to understand the properties of and the interactions in the investigated system and to apply the acquired information in the design of the system. Computational methods permit observation of processes at time and length scales that are not accessible through experimental methods. Additionally, computational methods can complement experimental results, and thereby provide a deeper understanding of the system.

In both density functional theory (DFT) and classical molecular dynamics (MD), the model system is described by the interaction between the atoms (or particles) of the system, from which the distance, spatial distribution, strength of the interaction between the constituents,

etc., can be determined accurately. However, the two methods are based on a different theoretical framework. DFT is a first principles method where the interaction energy is expressed as a functional of the electron density. DFT does therefore not require empirical parameters, although it does require the application of a functional which is only an approximation to the "true" functional (e.g., the functional of exchange and correlation energies among the electrons are not known exactly, therefore approximations are used). DFT typically has a high computational intensity, which limits the system size (up to few hundred atoms). In classical MD on the other hand, the atoms interact through a (semi-)empirical force field or interaction potential. While this allows the computation of much bigger systems (up to $10^7 - 10^{11}$ atoms), it does not allow the calculation of electronic or magnetic properties such as band gaps or magnetic moments.

Molecular scale methods are useful for the rationalization of intermolecular interactions in NADES, which are assumed to be the main reason of their outstanding properties. As DFT and MD can obtain information at a resolution and length scale currently inaccessible to experiments, they can help the interpretation of chemical reactions in NADES, e.g. enzymatic catalysis or electrochemical treatment of surfaces. These methods may also aid in the rationalization of macroscale NADES properties. However, the direct interpretation of the results at the macroscale is limited by the size and length of the simulation, i.e., investigated attribute does not reach its equilibrium within a feasible calculation time and the correlation length of some attributes could exceed the size limitation of the model. For example, in MD simulations of NADES the calculated physicochemical properties, such as density and surface tension showed good agreement with experimental data, but the calculated transport properties had significant error. DFT calculations are also the basis of the COSMO-RS continuum model, which will be discussed later.

2.5.1.1 Density functional theory

DFT simulations of NADES describe the conformation of the components and the strength of interaction between them. The evaluation of site-site distances identifies the hydrogen bonding interactions among the molecules. In case of hydrogen bonding the distances become smaller than the sum of the van der Waals radius of the constituents. The simulation describes the electronic behavior of the components. The electrostatic potential analysis (ESP) determines the spatial distribution of electrostatic potential, which provides information about the electrostatic interactions in the system. Meanwhile, reduced density gradient analysis (RDG) is also an useful tool to describe the non-covalent interactions in the system. From DFT results, vibrational spectra can be calculated. The vibrational spectra of NADES are typically difficult to evaluate due to the complex interactions, hence the computed spectra facilitate their interpretation.

The disadvantage of the DFT methods is the absence of London dispersion forces in the model. NADES often contain alkyl side chains and aromatic moieties, therefore dispersion forces have a significant contribution to non-covalent interaction energies. Also, the dispersion forces may significantly influence ionic interactions. This issue can be handled by the application of dispersion corrected functionals. Although this correction does not have a significant effect on the vibrational spectra (dispersion forces can be detected only in the far infrared region), in structure optimization it results in a significant difference. The

application of DFT simulation for gas phase molecular structures results in a good agreement between the calculated spectra and experimental results. However, the simulation of NADES' physicochemical properties requires the simulation in liquid phase.

In the available literature two main application domains of DFT are distinguished; the description of molecular structures and intermolecular interactions between the NADES constituents on the one hand, and the rationalization of NADES applications by computational simulations on the other hand. Garcia et al. combined the DFT calculations with the topological analysis of the NADES electronic density [71]. By this, they set up a relation between the melting temperature and structure of the hydrogen bonding network of the NADES. The structure was described by the AIM (Atoms in Molecules) approach (Method for the topological analysis on the electronic density of the system). Others applied charge decomposition analysis (method for determination of direction and extent of charge transfer between the constituents) to correlate the strength of interactions in the system to the melting temperatures [72]. The DFT simulation of Stefanovic et al. revealed fundamentally different hydrogen bond network structures with different constituents which rationalized the significantly different melting point changes to some extent [73]. Zhu et al. used the DFT simulations to calculate the vibrational spectra of the NADES and identified the peaks of the experimental spectra [9]. Rimsza et al. studied the application of NADES in surface etching of copper [62]. The DFT simulation described the ionic character of the urea during the interaction with elemental and ionic copper. That was done by comparing the different binding energies between neutral and anionic form of urea and elemental copper and copper oxide. In another study, the simulation explained the electrochemical deposition process of magnesium metal in the mixture of choline chloride and magnesium chloride hexahydrate [10]. The most likely reactions near the cathode were determined by identifying the cationic species in the system. The CO_2 capturing ability of DES was also studied by DFT calculations: the hydrogen bonding interaction in the system was calculated before and after the addition of CO_2 to the system (new NADES-CO_2 interactions) [74].

In the future, DFT simulation will remain a key method for the investigation of molecular interactions, charge transfer and thermodynamic changes associated with the formation of NADES. Such simulations will be also useful to rationalize novel applications of NADES. The experience of simulations of IL could be a useful starting point for these studies, e.g., the process of cellulose dissolution in NADES is assumed to be similar to the dissolution in IL, which is already described in the literature. However, the size and time limitations of the method require its combined application with experimental methods or simulation of the extended system by e.g. molecular dynamics.

2.5.1.2 Molecular dynamics

MD considers either atoms or groups of atoms as the basic particles of the system; simulations are done by integrating the relevant equations of motion through discretization of time. New positions, velocities and accelerations are obtained from atomic forces which are obtained as the negative gradient of some empirically derived force field. These force fields are typically built on some functional form determined by the type of system to be studied, and then parameterized for the specific system. MD methods can simulate larger systems and are able to calculate physical chemical parameters, like density or viscosity.

MD is also applicable to study the structural characteristics of the system. This latter is done by calculating the radial distribution function (RDF) and spatial distribution function (SDF). The first describes the density function of the distance between selected particles. The second is the visualization of the spatial distribution of the constituent particles around a selected molecule. With these methods the hydrogen bonding interaction of the system can be studied. As the solvation dynamics are too fast to be monitored by experimental methods, MD simulations constitute a solid complementary method for such studies. The simulation of larger systems allows applying MD for the investigation of the microstructural properties of enzymes in NADES and rationalize experimental results.

Although MD can simulate bigger systems, the accuracy of the method depends on the applied force field and its proper parametrization. As this method doesn't say anything about the electrons and their interaction in the system, MD is not applicable for the determination of electrical properties (therefore the calculation of vibrational spectrum) or the molecular structure optimization, what is required for the determination of the equilibrium structure (unlike DFT). Therefore MD calculations are often amended by DFT calculations to find the optimal molecular structure.

Similarly to DFT, the reported applications of MD mainly consider the rationalization of NADES behavior and their practical applications (e.g., CO_2 absorption), but the simulation of macromolecular system and selection of proper force field for the calculations is also discussed. The investigation of applied force field parameters showed the significant effect of applied charge schemes regarding the calculated properties and intermolecular interactions of the system [75]. The right method for the determination of charge assignment is vital for accurate simulation. Doherty et al. developed new force field potentials for choline chloride based DES (non-polarizable force field, OPLS (Optimized Potentials for Liquid Simulations) -DES) [76]. The simulations with this force field gave good agreement with the experimental results of density, viscosity, heat capacity and surface tension. However, the quantitative simulation of self-diffusion coefficients proved to be a challenge with the presented potentials. Sun et al. simulated the structural characteristics of choline chloride-urea system and determined the relation between the molar composition and interactions energies of the system [77]. The simulation also revealed a long range ordered structure among the ionic compounds of the system and yielded information about the strength and lifetime of hydrogen bonding interactions between specific sites. Another study used MD to calculate density, heat capacity and self-diffusion coefficient of the simulated system [78]. Although MD is not able to reproduce vibrational spectra, the structural and hydrogen bond analysis of the system was applied to rationalize the experimental spectrum. Das et al. combined MD simulations with steady state fluorescence emission measurement to study the relaxation dynamics, spatial and dynamic heterogeneity aspects of eutectic solvents [79]. The simulation of particle displacement during the relaxation of the system proved its homogeneity. Monhemi et al. studied the microstructural properties of a macromolecule, i.e. *Candida antarctica* lipase B in NADES [68]. The simulations described the enzyme-NADES interactions and the potential diffusion of the constituents inside the macromolecule. The comparison of these processes in water to an aqueous solution of the NADES constituents on the one hand and to the pure NADES on the other hand rationalized the stabilizing effect of NADES on macromolecules. Ullah et al. studied NADES for CO_2 capturing along with their physical chemical properties by MD simulations [74]. The method described the hydrogen bonding network in the system and its change during the absorption of CO_2. The

affinity between the solvent and CO_2 was determined in the gas-liquid interfacial region to rationalize the absorption and migration of the CO_2 molecules toward the bulk fluid region.

As NADES can be used as green solvent in the field of food, feed and pharma industries, it is imperative to better understand their interaction with proteins and polymers. This also covers the application of NADES to host enzymatic reactions. As MD methods are appropriate for the modeling such systems, they may be vital in the development of these novel applications.

2.5.2 Macroscale modeling

We refer to macroscale modeling as all methods aiming to model NADES properties in the bulk rather via the direct intermolecular interactions among the constituents. The targeted parameters include their physicochemical properties as density, viscosity, which gives some overlapping with the earlier discussed methods, but also solvation energies and thermodynamic equilibrium properties of liquids properties, which are relevant to practical applications. Although such macroscale models still incorporate theoretical considerations, they mostly rely on input data from empirical observations or on the results of other calculation methods.

2.5.2.1 COSMO-RS

Conductor-like screening model for a real solvent (COSMO-RS) is a method to describe the thermodynamic properties of pure compounds and mixtures of compounds based on the unimolecular quantum chemical calculations of constituents. COSMO-RS combines quantum chemical calculations with statistical thermodynamic approaches to overcome the limitations of dielectric continuum solvation models (in these models the solvent is modeled as a polarizable continuum to decrease the computational intensity). Klamt gave a comprehensive overview of the model,[80] therefore we only give a short description about its working principle (See Fig. 2.9.). The model uses the output of quantum chemical calculations such as the charge density surface of the molecule (σ-surface). The model transforms this into discrete surface segments with an area and screening density charge. Here the contribution of hydrogen bonding, electrostatic misfit (the deviation of the electrostatic interaction energy from the idealized contact of same charges with different polarities) and van der Waals interactions are taken into account. Next, the model calculates the sigma profile: the screening density charge distribution of the molecule surface. The sigma profile of the complete system is the sum of the individual sigma profiles weighted with the molar ratio of the compounds. The chemical potential is calculated by an iterative function of the sigma profile of the system. This chemical potential is the basis for the computation of other thermodynamic properties. The model calculates the thermodynamic properties of liquid mixtures including solubility, partition coefficient and the liquid-liquid equilibria.

Since the method does not need experimental data of the studied system, it is appropriate for preliminary screening and rationalization of structural characteristics (i.e., the probability distribution of charge density can be used for the interpretation of molecular interactions). The drawback of this technique is the unknown composition and dissociation state of the investigated ionic compounds (when such are present), which has to be determined separately.

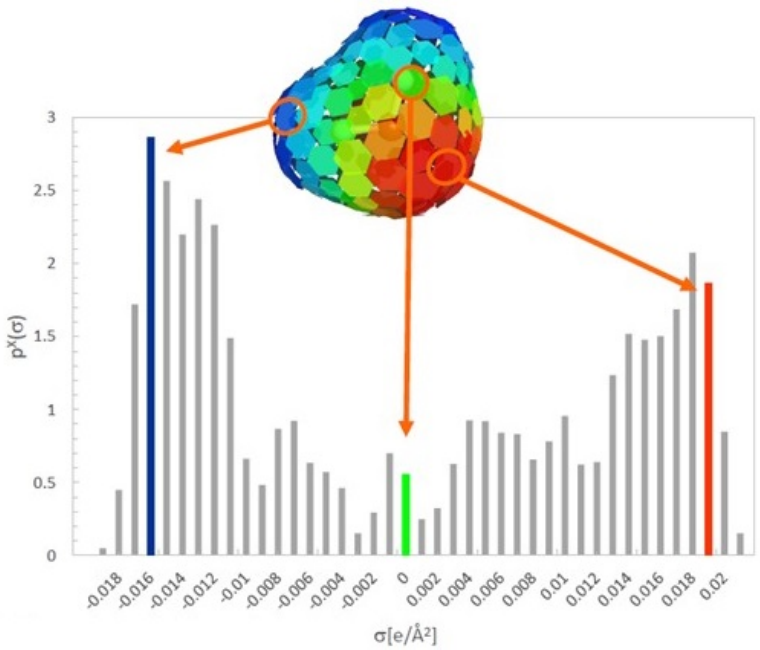


Figure 2.9: Determination of sigma profile of water based on the results of DFT simulation. (Taken from COSMO-RS theory, The Basics.[81])

Although the simplicity of the model's algorithm makes its application easy, the necessity of preliminary quantum chemical calculations could be a difficulty for novel compounds.

COSMO-RS was primarily applied in screening of DES for applications where the liquid-liquid equilibria are relevant (i.e. liquid-liquid separation, chromatography). Mulyono et al. investigated the liquid-liquid extraction of BTEX aromatics from n-octane by using DES [82]. They used COSMO-RS to predict the ternary liquid-liquid equilibria diagram of the studied systems, where the method showed only qualitative agreement with the experimental results. Also the calculated sigma profiles were used to describe the interactions between the DES and the solute compounds. Another research screened DES for use in extractive denitrification of diesel [83]. The activity coefficient of nitrogen compounds in DES was predicted and it was used for the calculation of selectivity, capacity and performance index of the DES. Based on the results, they also set up assumptions regarding the relations of the strength of interaction to the structure of the compounds. However, due to the lack of experimental validation of these results, it is possible that the screened liquid eutectic systems do not exist in reality. Gouveia et al. also tested DES for the separation of aromatic and aliphatic hydrocarbons via liquid-liquid extraction [84]. They used COSMO-RS to predict the phase behavior and tie lines of the ternary mixtures with small relative error. The method was also able to describe the trend of distribution ratio and selectivity of different DES. A different research group used the COSMO-RS to predict the interaction mechanism between the constituents of DES [85]. They compared the sigma profiles of the pure constituents and the formed DES. They also found good agreement between the predicted and experimental density. Bezold et al. assessed the application of COSMO-RS in model for the calculation of thermodynamic properties of DES in liquid-liquid chromatography [86].

They found qualitative agreement between the activity coefficients, liquid-liquid equilibria data and partition coefficients taken from experimental measurements and literature data. They found the overall prediction quality to be sufficient for pre-screening procedure in solvent system selection. Jeliński et al. used the COSMO-RS methodology for the screening of DES for the solvation of rutin by predicting the solubility [67]. Their applied model also took into consideration the possible ionic and neutral form of the HBA. They used their findings to describe relation between the structural properties of the NADES and the solubility values. Finally, COSMO-RS was also used as a screening tool to predict the thermodynamic properties of sugar based, ternary DES [87]. The comparison of predicted and measured eutectic point showed good agreement.

Due to its capabilities, COSMO-RS will remain a valuable method for pre-screening tasks in the future for applications related to the liquid equilibria properties of NADES and their solutes. However, the experimental validation of the results cannot be omitted, as the predictions thus far often yielded only qualitative agreement with the actual results. This also raises the question, to which extent can the results of the calculations (without experimental validation) be used in the investigation of the general structure-property relation of NADES.

2.5.2.2 PC-SAFT

Perturbed-Chain Statistical Association Fluid Theory (PC-SAFT) is an advanced equation of state model. PC-SAFT is a thermodynamic model for phase equilibrium calculations: it describes the relation between energy, volume, pressure, temperature and composition in the fluid region. It is also an association model, which means it also describes the effects of hydrogen bonding between the compounds of the system. To calculate the total energy of the hydrogen bonding, the model uses a perturbation theory: they approximate the solution of a complex problem with the solution of a simpler problem, then they use perturbation parts to incorporate the differences. The detailed theoretical background of the method can be found elsewhere [88], here we give only a short description of the main considerations. In this model molecules of a reference fluid are constituted by the addition of equal sized spherical segments into a hard chain (See Fig. 2.10). The total interaction energy is expressed as the sum energy of the ideal gas system and the residual energy of the interactions (See Eq. 2.7.). The residual energy contains the contribution of the interaction in the hard-chain reference, the dispersion interactions among the chains and the specific site-site hydrogen bonding interactions in the system (See Eq. 2.8). The equation is usually expressed in Helmholtz free energy as most thermodynamic properties can be obtained by its differentiation.

$$a^{total} = a^{ideal} + a^{residual} \tag{2.7}$$

$$a^{residual} = a^{hard-chain} + a^{dispersion} + a^{association} \tag{2.8}$$

Each compound in the model requires characterization by five pure component parameters: the number of segments, the diameter of segments, energy of the segment, volume of association and energy of association. Empirical binary interaction parameters can be also added to the model. This increases the accuracy of the model by incorporating the interactions

among the constituent and solutes. These parameters have to be determined experimentally by regression on the vapour pressure and liquid density data of the pure components or by group contribution methods (if available).

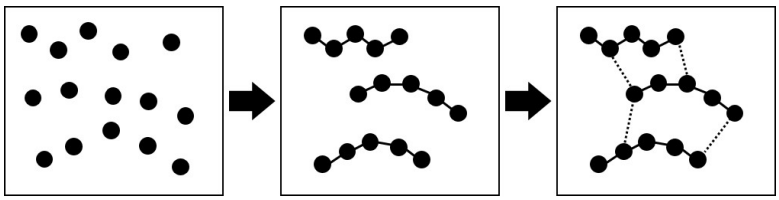


Figure 2.10: Schematic representation of PC-SAFT theory: the equal-sized spherical segments (a) connected by covalent bonds, forming the hard-chain reference fluid. (b) The specific interactions among the chains are described by the perturbation parts. (c)

The accurate prediction of the thermodynamic properties is the main advantage of the method. The calculated behavior of vapor-liquid, liquid-liquid and solid-liquid phase equilibria show quantitative agreement with the experimental results in many cases. Therefore solubility, separation properties, gas absorption (e.g., CO_2) and the melting point can be predicted by this method. More complex properties can be determined by coupling PC-SAFT with additional models, where the calculated thermodynamic properties are the input for further steps (see work of Haghbaksh et al. in the next paragraph). On the other hand, the determination of the pure component parameters can be difficult. Their determination based on experimental dataset requires a significant amount of extra work. The pure component parameters of NADES raises another question: the mixture of the compounds can be described as one pseudo-pure component or individually. The second approach yields a more general model, but the determination of individual parameters is not always feasible. In case of NADES, where solid constituents are applied, the determination of vapor pressure and liquid density data is not feasible. To overcome this issue, the parameters can be determined from the aqueous solutions of the constituents, however that is not feasible in case of hydrophobic constituents.

Similar to COSMO-RS, PC-SAFT can also be used for solubility calculations. Verevkin et al. applied PC-SAFT to calculate the limiting activity coefficients of aliphatic and aromatic organic compounds in DES [89]. The comparison of calculated and experimental values showed agreement at the order of magnitude. Moreover, the predicted temperature dependency followed the trend of experimental results. Gas absorption capabilities can be investigated by the determination of liquid-vapor equilibria. Zubier et al. applied a PC-SAFT model to investigate CO_2 capture in DES [90]. They determined the phase behavior of the eutectic systems with CO_2. They applied both of the above-mentioned strategies for the determination of pure component parameters. Although both strategies yielded accurate predictions on phase behavior, the individual-component approach yielded more versatile parameters: with pseudo-pure approach not just the pure component parameters were DES-specific (as they were determined for given DES), but also the binary interaction parameters became ratio specific. An additional research group investigated the CO_2 solubility of DES with constituents, that were hydrophobic and solid at room temperature, therefore they could only use pseudo-pure parameters for their calculations [91]. Still, the calculations predicted

CO_2 solubility with reasonable accuracy, even without the application of binary interaction parameters. PC-SAFT is also useful for the determination of the melting point in eutectic systems. Pontes et al. applied PC-SAFT to describe the solid-liquid phase diagrams of 15 quaternary ammonium and fatty acid based DES [92]. Based on the experimental results they connected the increasing chain length with increasing molecular interaction and lower melting point as consequence. They incorporated that effect in the binary interaction parameter of the thermodynamic modeling. The predicted melting point showed good agreement with the experimental data (around 7K average absolute deviation). As we mentioned earlier the PC-SAFT can be coupled with additional models. Haghbaksh et al. combined PC-SAFT and Cubic plus Association models with free volume theory to create a predictive viscosity model for DES. The equation of state models provided the density of the system, which is a required parameter for the viscosity calculations by the free volume theory [93]. The viscosity calculation with PC-SAFT yielded accurate predictions (with 2.7% average deviation). In their second work, they combined the same equation of state models with the frictional theory to build an additional viscosity model [93]. Here PC-SAFT was used to determine the repulsive and attractive pressure values, which are the input parameters of the frictional theory. The final model showed reasonable predictive accuracy with 4.4% average deviation.

PC-SAFT has a future potential in the determination of thermodynamic properties of the NADES systems. As more and more data become available about eutectic systems, the pure component parameter determination will become easier, making the method application more straightforward. The coupling of the equation with additional models on both input and output side is still a relatively unexploited area, where the initial results are promising.

2.5.3 Group contribution and machine learning methods

The methods we discuss in this section focus on the description of the structure-property relationship in DES systems. It is common in the group contribution and machine learning methods, that the aim is not the general understanding anymore, but rather the quantitative prediction of the systems properties. To achieve this, these methods use large datasets to describe the relation between the structural properties and the behavior of the investigated systems. Rather than theoretical considerations, these methods work with empirical models relying on heavy parametrization. In case of machine learning black box models are often used, were the model does not yield minimal information about the rationale of the described relation. The quantitative agreement between the predicted and experimental data can be achieved, however the amount of required data is often the bottleneck of such methods.

2.5.3.1 Group contribution methods

Group contribution methods calculate the properties of chemical compounds based on their structure. In principle the method assigns contributions of the groups to a given property, that describes how the chemical compound is built. The property of the whole compound is the added contributions of the consisting groups. By that, large number of different compounds can be characterized with the small amount of information on the group con-

tributions. The application of such contribution method reduces the required data and computational intensity of the calculations. Still, the determination of the group contributions require a large experimental dataset (e.g., Hansen et al. applied around 1200 data points for their software to predict the Hansen solubility parameters [94]). One common example is the Lydersen-Joback-Reid method, where the thermodynamic properties of the chemicals is determined based on their structural composition.

The main advantage of group contribution methods is the simplicity of the calculations, which only require the structure as input. The method can yield quantitative accuracy, multiple physicochemical properties were determined with an average error lower than 5.0%. On the other hand quality of the estimation depends on the limitations of the method itself and the structural domain covered. The development of the contribution dataset is also labor intensive, but available system can be enhanced subsequently.

For the determination of solvation properties, Hansen solubility parameters are usually determined through group contribution methods (e.g., Hoftyzer and van Krevelen method). However, their implementation for NADES due to the special hydrogen bonding interactions and the presence of ionic compounds is not straightforward. Still, Lee et al. used the group contribution methods of van Krevelen of Hansen solubility parameters to predict the melting point decrease of DES [95]. They take into account the difference in polar and hydrogen bonding parameters of the HBD component to compose the regression model of melting point depression of choline chloride based DES. The regression model gave 0.738 as coefficient of determination. The other frequently used method for DES is the earlier mentioned Lydersen-Joback-Reid group contribution, usually coupled with other calculations where the group contribution gives the input values. Shahbaz et al. developed a predictive model for the density of DES [96]. They used the Modified Lydersen-Joback-Reid method for the determination of the critical properties of molecules. They combined the results with Lee-Kesler mixing rules and the modified Rackett equation to calculate the liquid density. The error of the predicted densities was 1.9%. In their other study, they used the same method for the evaluation of additional DES and to compare the method to artificial neural network based estimations [53]. The group contribution based method yielded a 2.03% error in that case. Shahbaz et al. also used the group contribution method of Knotts et al. for the prediction of parachor values of DES [97]. The parachor value links the surface tension, density and structure of the compound and they used the group contribution method for the determination of surface tension and density. Additionally, they combined the results with the Othmer equation to describe the temperature dependency of the surface tension of the investigated DES. The experimental and predicted surface tension and density values showed good agreement with 6.4% and 1.61% average error, respectively. The predicted temperature dependency of the surface tension also gave good correlation with 2.57% average error. The method of Wildman and Crippen for the prediction of refractive indices was the third method investigated for the prediction of DES behavior [98]. They calculated the molar refraction values, and then they used the Lorentz-Lorenz equation for the determination of refractive indices and the density. The error of the refractive indices were 0.56%, while it was 1.43% in the case of density calculations. Mjalli et al. compared the Modified Lydersen-Joback-Reid group contribution and Eötvos method for the calculation of critical temperatures of DES systems and the application of these for the prediction of density and surface tension values of the DESs. For the latter, they combined the Rackett and Guggenheim empirical equations. As the Eötvos and Guggenheim methods based on

the experimental values of density and surface tension, it gave better results than group contribution method, especially in the higher temperature range. Finally, Mirza et al. used the combination of modified Lydersen-Joback-Reid group contribution method with the Lee-Kesler mixing rules to determine the critical properties of a large number of DES [99]. For validation, they used the Rackett-equation to calculate densities and compare with experimental values. The average deviation of the density values was 4.9%. However, the DESs containing aromatic groups had a bigger deviation, probably due to the stronger interaction forces in the system.

The main issue with the currently used group contribution methods is that they do not include the strong secondary interactions (hydrogen bonding) and often do not include ionic compounds, therefore they are not applicable to many deep eutectic systems. With the increase of available data on eutectics, group contribution methods directly for NADES can be developed. These specific systems would be useful for DES–based application development, due to their simplicity and predictive accuracy.

2.5.3.2 Machine learning methods

Machine learning is the process of computer systems performing specific tasks based on patterns and inference instead of explicit instructions. In the description of structure-property relations this usually means a supervised learning task: the machine is a mathematical function, an algorithm that maps the relation between the output variable and the possible input variables. Training datasets with known input-output pairs are used for the development. After this step, the trained algorithm can be evaluated and preferably used for the determination of the unknown output values of novel cases. In case of DES the main method used is artificial neural networks. A neural network (see Fig. 2.11.) is made up of neurons. Neurons are some non-linear functions (e.g., sigmoid, binary, hyperbolic tangent) that sums up the values of the different input signals (from other neurons or input variables). The neurons are connected to each other and mainly aggregated into layers. The inputs are the signals from the input variables or the output of other neurons. The connections between neurons have weights adjusting the strength of the input signal. The learning process is based on the adjustment of these weights during the training.

If proper datasets are available machine learning methods can yield good quantitative accuracy, even in the case of complex relations. On the other hand, the quality and the size of the available dataset is vital for accurate models. During the training process, the applied training data will determine to what extent of structural variability will be covered and what precision is possible. By applying a model for vastly different compounds than it was trained for, it becomes less accurate. The intention of machine learning is primarily the prediction of properties. Therefore the final model is often hard to interpret regarding the underlying structural relations (black box model).

Despite its potential, published studies on machine learning with NADES are scarce. Shahbaz et al. used machine-learning methods to predict the density of different DES system based on their composition and the temperature [53]. They used a feed forward backpropagation neural network and compared its efficiency to the earlier discussed group contribution method. The method had a 0.14% error and resulted in better predictions especially at higher

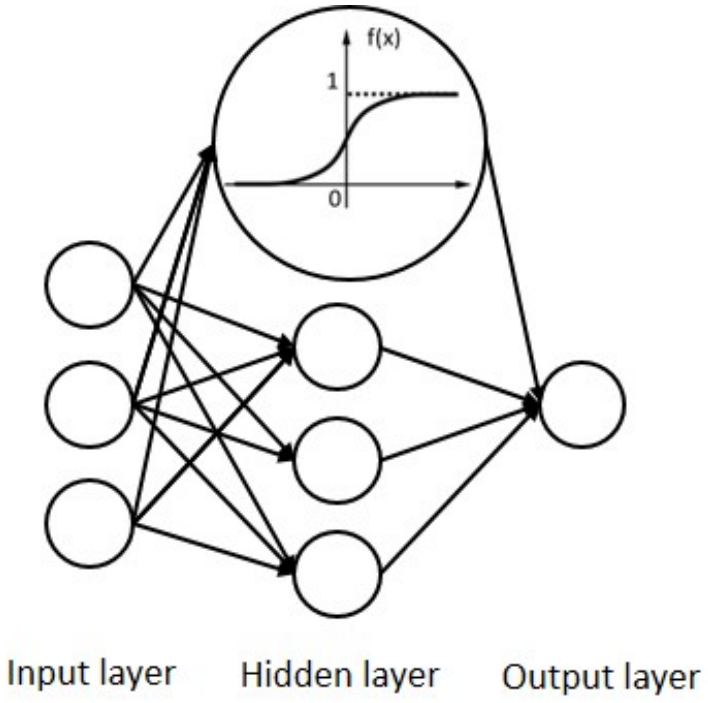


Figure 2.11: Artificial neural network with one hidden layer, three input variables and one outcome, The four nodes in the hidden layer use sigmoid functions for signal processing, one of the commonly applied non-linear function in neural networks.

temperatures compared to group contribution methods. Bagh et al. used a similar feed forward backpropagation neural network to predict the electrical conductivity of DES based on the temperature and the molar composition [100]. The predicted values showed good correlation with the experimental results with a regression coefficient of 0.9988. Adeyemi et al. used both feed forward backpropagation neural networks and bagging neural networks to predict the density and conductivity of multiple amine based DES systems [101]. The bagging method yielded a significant increase in the prediction quality of these properties with $2.799 \cdot 10^{-7}$ and $5.820 \cdot 10^{-4}$ values of normalized mean square of errors in case of density and conductivity, respectively.

Similarly to group contribution methods, as more and more data become available on DES, more complex machine learning methods will be possible for modeling different properties of such systems. On the other hand, good quality data is imperative for proper model development. In our experience, composing high-quality datasets based on literature is challenging to date.

2.6 Summary and outlook

Compared to common organic solvents and ionic liquids, the application of NADES may offer an economic and ecological alternative due to the cheap constituents, efficient synthesis, low toxicity and volatility. The possibility to design NADES properties through the structure of the constituents and their molar ratio in the mixture means both potential and challenge. Their application as designer solvent requires a good understanding of the structure-property relationships of these systems.

The earlier studies on the physical chemical properties of NADES showed that these are strongly related to the strength of hydrogen bonding in the system. Therefore changes of properties such as density, viscosity, surface tension or phase equilibria can be qualitatively described by the changes of the hydrogen bonding's strength. The strength of the hydrogen bonding changes with the structure of the components and their molar ratio. In the first case the number of hydrogen bond donor and acceptor sites on the molecule are determinative, but also the size and the electron affinity has effect as other intermolecular interactions are significant. The molar ratio is relevant with respect to the number of hydrogen bonds between the constituents: in the eutectic composition the ratio facilitates the maximal hydrogen bonding interaction between the constituents, leading to maximal interaction energy. This results minimal melting point, but also maximal viscosity in the eutectic point. However, the quantitative description of the structure-property relationship requires in depth modeling of these phenomena.

We discussed the models of DES in three separate sections. The molecular scale covers models based on density functional theories and molecular dynamics based methods, where the primary goal is to describe the steric configuration of the molecules in the eutectic system and the actual interaction energies between the constituents. These methods help rationalizing the behavior of DES and also the feasibility of biocatalytic reactions in strongly hygroscopic environment. In macroscale modeling the aim is to describe the properties of the bulk solution, based on semi-empirical methods instead of the actual intermolecular background. The two main methods were COSMO-RS and PC-SAFT. Both are applicable for the determination of equilibria phase properties. The third section discussed empirical models based on a large amount of data. Group contribution and machine learning models can yield quantitative precision, but their application requires large amount of experimental data.

In the following figure (Fig. 2.12.) we compared the different aspects of the investigated methods. While the molecular and macroscale methods can aid the interpretation of the already experienced phenomena, the macroscale, group contribution and machine learning models can predict the outcome of novel cases. While the macroscale methods yield mostly qualitative results, with group contribution and machine learning methods one can acquire quantitative predictions. The published studies determined the vibrational spectra, density, viscosity and diffusion properties with the aid of molecular scale properties. With macroscale models the phase equilibria properties like liquid-liquid, solid-liquid and vapor-liquid equilibria and solubility were calculated. Group contribution and machine learning methods predicted the density, viscosity, surface tension, electric conductivity and refractive indices different NADES systems.

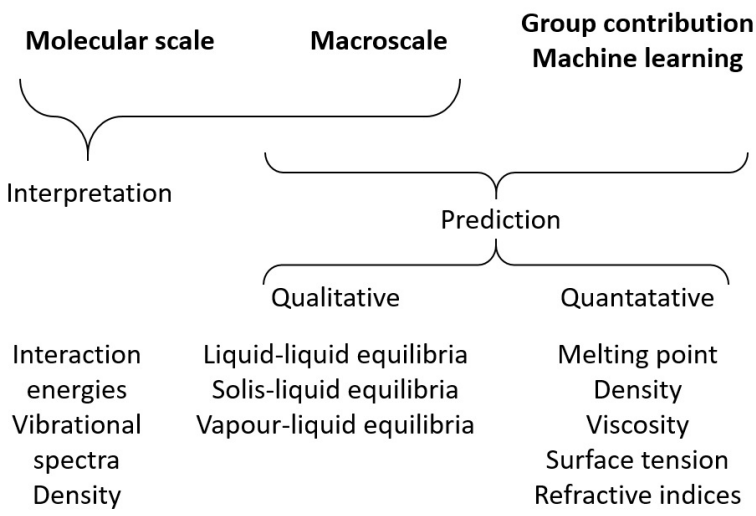


Figure 2.12: Different methods and their capabilities for modeling the properties of NADES.

2.7 Linking results to hypotheses

In this chapter, I have reviewed the available methods for modeling NADES properties and behavior. The available methods are divided into molecular level, macro level and data driven approaches. Molecular level models describe the direct interactions between the different actors in the system, they can e.g., determine the interaction energies, the vibrational spectra or the density of the system. The methods discussed here are density functional theory and molecular dynamics simulations (ab initio and classical). Macroscale models consist of PC-SAFT and COSMO-RS modeling. These methods are suitable to determine the physicochemical properties of the bulk media and also the dissolution in NADES. Data-driven methods, such as group contribution and machine learning modeling, also predict the physicochemical properties of the media, but they are based on large amounts of experimental data instead of semi-empirical relationships.

Based on the reviewed models, it is possible to describe and predict the properties of NADES that I am interested in (**Q3** and **H3**). This review confirms that all models necessary for a holistic modeling approach are available. For the physicochemical properties density and viscosity, I choose machine learning modeling. The results of the density predictions will be used in combination with PC-SAFT modeling to determine the solubility of the substrates. To study the effect of NADES on the enzyme structure, I choose classical molecular dynamics simulation.

However, the review also shows many gaps in the currently available models. Although many different approaches are discussed in the literature, holistic methods that aim at predicting complex effects of NADES are still missing. With this research I try to bridge this gap. Most of the current machine learning approaches are still too rudimentary in their features (input variables) and have insufficient accuracy to generically predict the properties of novel NADES systems. This problem can be addressed in the future by building larger databases of

NADES. This will happen as the field develops, but because it is a labor-intensive process, I do not include screening of large numbers of novel systems in this work. Instead, in Chapter 5, I present an analysis of the currently available data to provide an overview of the current capabilities in terms of modeling accuracy and structural application domain. Molecular dynamics simulations require an accurate force field to produce reliable results. The reviewed works have shown that proper force fields are rarely available for NADES and they strongly influence the final results. Force field optimization is not a straightforward process, and obtaining them for novel NADES systems can be a bottleneck in modeling efforts.

Chapter 3

Biocatalytic transesterification of vinyl laurate in NADES

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka" but "That's funny..."

Isaac Asimov

3.1 Preface

Many studies describe enzymatic reactions in NADES, and they usually include some comparison of reaction performance (e.g., final conversion or initial reaction rate) in relation to the NADES used. Rarely, however, is a relationship established between the properties of the NADES and the reaction performance that would abstract from the specific compounds in the given NADES. Therefore, before trying to predict the presumably relevant properties, I check if changes in these properties can really explain the changes in the reaction performance.

I hypothesize that the NADES influence the enzymatic reaction through the strong secondary interactions between substrate, media and enzyme. Furthermore, I hypothesize that these interactions can be described by measurable and predictable properties, namely solubility (substrate-NADES interaction), viscosity (NADES-NADES interaction). I try to interpret changes in reaction performance and select the optimal solvent based on these properties.

In this chapter I discuss the results of the enzymatic transesterification of vinyl laurate in different NADES, where I investigate the effect of the NADES through the above mentioned properties. To do this, I measure the viscosity of the NADES, the solubility of the substrate, the conversion over time, and monitor side reactions. From this, I will determine how the changes in reaction performance are related to the investigated properties and what is the relative weight of these properties in the effect of NADES on the reaction.

The content of this chapter has also been submitted to the journal Waste and Biomass Valorization with the title "Biocatalyzed vinyl laurate transesterification in natural deep

eutectic solvents". The co-authors of this paper are Nathalie Janssens, Miguel Mielants, Erik Neyts, Iris Cornet and Pieter Billen. Nathalie Janssens and Miguel Mielants contributed to the development of the methodology, the formal analysis of the results and the writing of the original draft. Erik Neyts, Iris Cornet, and Pieter Billen contributed to the supervision of the project and to the review and editing in the writing process. In addition, Pieter Billen acquired financial support for the project.

3.2 Abstract

Natural deep eutectic solvents (NADES) represent a green alternative as reaction media, offering more benign properties compared to conventional organic solvents. Their application is being considered in the treatment of waste streams and biomass (e.g., cellulose processing), and as media for biocatalysis. To efficiently design NADES for biocatalysis, a better understanding of their effect on these reactions is needed. We hypothesize that this effect can be described by separately considering (1) the solvent interactions with the substrates, (2) the solvent viscosities and (3) the enzyme stability in NADES.

We investigated the effect of substrate solvation, the effect of NADES on enzyme stability and the effect of NADES' viscosity on mass transfer limitation of the substrate. To this end, we monitored the conversion over time of the transesterification of vinyl laurate with 1-butanol by the lipase enzyme *Candida antarctica* B in NADES of different compounds and molar ratios. The studied NADES are composed of choline chloride with urea, ethylene glycol and glycerol. The effect of solubility and enzyme incubation on the reaction rate was also investigated in a selection of abovementioned systems. Oversaturated solutions were used to study the dissolution of the substrate in NADES and to determine the phase in which the reaction takes place.

The catalytic reaction takes place in most of the NADES and, despite their higher viscosities, the initial reaction rate is often higher than in the reference n-hexane system. We found no correlation between viscosity and reaction rate, implying that viscosity has only a marginal effect on the reaction. The reaction proceeded in the time frame of 2-24 hours, showing that the enzyme retains at least some of its activity after the first 2 hours of reaction. The reactions in the saturated CCEG12 system indicate that the reaction proceeds well in the NADES phase. Enzyme incubation in CCEG12 prior to the reaction resulted in a reduced reaction rate which may be due either to an almost complete inactivation of the enzyme or to the slow dissolution and high viscosity of NADES, or to both phenomena.

Many of the studied NADES systems are suitable media for the transesterification reaction of vinyl laurate, and the proposed framework of solvation-viscosity-enzyme stabilization proves suitable for describing the effect of NADES on the enzymatic reaction. The plateau we experienced in the conversion of CCEG systems and the slow reaction rate of the incubated enzyme in CCEG12 system are most likely the combined effect of enzyme denaturation, limited solubility of the substrate, and limited access to the enzyme due to high viscosity.

3.3 Novelty statement

In this study, we report the use of NADES as reaction media in biocatalysis. To our knowledge, the systematic use of non-eutectic NADES has not been reported in the literature. Furthermore, this is the first time that a mechanistic framework is reported that proposes to describe the effect of NADES on enzymatic reactions via the media-media (viscosity), media-substrate (solvation), and media-enzyme (changes in enzyme structure) interactions. The experimental results are discussed within this framework.

3.4 Introduction

Natural deep eutectic solvents (NADES) are a novel class of reaction media that offer more benign properties than conventional organic solvents [14]. Deep eutectic solvents (DES) consist of a mixture of organic compounds, typically quaternary ammonium salts with metal chlorides or hydrogen bond donors (HBD) [5]. The term deep eutectic refers to the significantly decreased melting point of the mixture, compared to the pure compounds [33]. Although the definition of deep is still debated in the scientific community [102], mixtures of practical interest have a melting point close to or even below room temperature [5]. For example, the typical "case study" of DES, i.e., the mixture of 1 to 2 molar ratio of choline chloride and urea has a melting point of 17 °C, while solid choline chloride decomposes at 305 °C and urea melts at 132.7 °C [6]. This eutectic behavior is attributed to the strong secondary interactions between the compounds, in particular the formation of an intermolecular hydrogen bonding network throughout the system [9, 10]. This network stabilizes the mixture in a liquid state [11], allowing the use of DES as a solvent. NADES are a subset of DES that contain naturally derived compounds, usually quaternary ammonium salts and hydrogen bond donating components such as organic acids or polyols [13].

Since NADES are mostly ionic solutions, they have similar physicochemical properties to ionic liquids, but offer additional advantages. Compared to volatile organic compounds, NADES are often non-flammable, have low vapor pressure and low toxicity, making them a better choice from both an environmental and occupational safety perspective [16, 17]. In addition, NADES are considered "designer solvents", i.e. their physicochemical properties can be tailored according to the task. The interactions between the salt and HBD groups define the behavior of the given mixture [1]. Therefore, in principle, the desired properties can be achieved by appropriate selection of the components used and their molar ratio [19]. Compared to ionic liquids, DES can be prepared with 100% atom economy and are less susceptible to impurities, which facilitates their large-scale use [5]. Containing organic acids, amino acids, amines and sugars, NADES offer biodegradability, low production cost and they are in line with the principles of circular economy [20].

Although non-flammability, low volatility, low toxicity, and low price make NADES attractive candidates for many fields, practical applications to date include mainly electrochemistry [23] and separation/extraction processes [24]. Treatment of waste streams and biomass (e.g., cellulose processing [103], biodiesel [104], and fatty acid ester production [105]) have also been tested on a laboratory scale. Due to their biodegradability and possible biocompatibility, NADES are likely to be applied in the pharmaceutical, food and feed sectors where these properties are preferred [106]. In this respect, NADES are also suitable for certain enzyme-catalyzed reactions [107, 108], underlining their future potential for the aforementioned biochemistry-intensive industries.

The first biocatalytic reaction in NADES was described by Gorke et al. in 2008 [109] and many other studies followed. Xu et al. published a review of the enzymatic reactions studied in 2017 [41]. However, the varying effectiveness of NADES as media in these reactions is rarely discussed. The available literature indicates that the intermolecular H-bonding network of NADES is also the main way how they affect biocatalytic reactions [110, 111, 112]. First, the intermolecular network helps protein structures to remain stable in NADES, although these media often contain strong denaturing agents such as urea. A computational study

of NADES showed that the intermolecular network prevents the diffusion of the denaturing compound inside the enzyme structure [68]. Thus, the intramolecular interactions of the protein remain intact and the enzyme remains active. In addition, NADES provide extra stability to the enzyme: the half-life of horse radish peroxidase increased to 350 minutes in choline chloride-urea mixture compared to 50 minutes in phosphate buffer [70]. Simulations suggest that this is due to strong secondary interactions between the eutectic constituents and surface residues of the enzyme, which provide additional stability to the enzyme structure [110, 111]. NADES, as media, also influence the kinetics of the enzymatic reaction and this effect can be either positive or negative. In the aminolysis of ethyl valerate by immobilized *Candida antarctica* lipase B (iCALB), the mixture of choline chloride-urea and choline chloride-glycerol outperformed ionic liquids in terms of final conversion and showed initial reaction rates similar to toluene [109]. The reaction in a mixture of choline chloride and acetamide yielded only one third of the conversion compared to toluene. Similar differences in conversion rate were found in the case of lipase-catalyzed transesterification [113] and hydrolysis reactions [114].

The combination of their designer nature and biocompatibility makes NADES a green alternative for biocatalytic media that may outperform conventional solvents due to their tailorable properties. However, the tailorable properties of NADES also pose the greatest challenge to their effective use, as knowledge of their structure-property relationships is severely lacking. Simply put, we don't understand how changes in the compounds used and their molar composition affect the physicochemical properties and behavior of NADES. NADES have been shown to influence enzymatic reactions, but the exact mechanism behind this influence remains unclear.

The observed differences in performance may be due to several phenomena. Differences in the solvation energies of substrates and products can shift the reaction equilibrium and also change the activation energy of the reaction. The high viscosity of NADES can lead to mass transfer limitations. This is even more problematic in the case of immobilized enzymes, where the reactants must diffuse into the porous carrier beads. As mentioned above, the media can have a stabilizing or denaturing effect on the enzyme itself, but side reactions with the NADES compounds are also possible.

For biocatalytic applications of NADES, lipases are the most intensively studied group. Firstly, they tolerate a wide range of temperatures and many different solvents, making them applicable to many synthetic and hydrolytic reactions [41, 107, 108, 109]. Secondly, lipases are the most widely used enzymes in the food, pharmaceutical and cosmetic industries, as well as in the synthesis of surfactants and the production of biodiesel [41, 104, 105, 107, 108].

To determine the activity of lipase enzymes, Goujard et al. developed a method based on the transesterification of vinyl laurate using UV-Vis spectrophotometry [115]. Later, the same reaction was used by Durand et al. to study the initial reaction rate in NADES [113]. With their method, they determined the feasibility of lipase-catalyzed reactions in NADES and compared the performance of different NADES systems and common organic solvents. This approach is a fast and effective way to determine enzyme activity by measuring the conversion of vinyl laurate over time, compared to more time-consuming and expensive chromatography methods [113].

At present, the designer property of NADES as solvents for biocatalysis is inaccessible be-

cause we do not fully understand how the NADES affect the reaction. In this research, we present a framework that mechanistically describes the effects of NADES on the enzymatic reaction, which can then be related to the structure-property relationships of the NADES. To achieve this, we need to know how to break down the complex effect that NADES have on enzymatic reactions into elements that can be discussed separately. Since NADES exert their specific behavior mainly through their strong secondary interactions, we hypothesize that the separate discussion of media-substrate, media-media and media-enzyme interactions would be expedient. This means the description of the solvation of the reactants, the viscosity of the NADES systems and the direct effect of the NADES on the enzyme structure. This breakdown should provide enough information to understand the performance differences between two or more NADES systems or even in comparison with volatile organic compounds. To test our hypothesis, we studied the transesterification of vinyl laurate in various NADES with different compositions and molar ratios. To investigate the effect of dissolution during the reaction, we also used supersaturated NADES solvents. We monitored the initial reaction rate, the final conversion and possible side reactions. We determined the viscosity of the media to study its effect on the reaction. We also studied the long-term stability of the enzyme by incubating the lipase for 24 hours before starting the reaction, and we studied the solubility of vinyl laurate in selected NADES systems. We compared the initial reaction rate and final conversion in NADES with n-hexane, a volatile organic solvent, as a reference system.

3.5 Materials and Methods

3.5.1 Materials

For the synthesis of NADES systems choline chloride (≥99%, Across Organics), ethylene glycol (≥99%, Fisher Scientific) and glycerol (≥99%, Fisher Scientific) were used. For the water content measurement Hydranal-composite 5 (Honeywell Fluka) was used. For the enzymatic reaction and the related chromatographic measurements vinyl laurate (≥99.0%, Sigma-Aldrich), 1-butanol (SLR, Fisher Scientific), butyl laurate (≥99%, Fisher Scientific) and methyl laurate (≥98.0%, Sigma-Aldrich) were used. For the sample preparation and as reference solvent n-hexane (≥99%, HPLC grade, Chem-Lab) was used. For the catalysis Immozyme CALB-T2-150XL immobilized lipase enzyme (Chiralvision) and concentrated hydrochloric acid (37%, a.r., Chem-Lab) to prepare the stopping reagent were used.

3.5.2 Preparation and measurement of NADES

The salt and HBD compounds were weighted together according to the molar ratio of the given NADES (made in batches of 200 grams). Table 3.1 shows the different NADES with their composition and molar ratio (along with the abbreviations used for the different NADES systems). CCUR samples with HBD ratio higher than 1:2 were not included in the research, as these mixtures were not stable liquids at 60 °C. Since the physicochemical properties of NADES are sensitive to their water content, the water uptake of the samples was controlled and prevented. To protect the mixtures from atmospheric moisture, they

were prepared and stored in sealed flasks. The mixtures were heated to 80 °C and stirred with a magnetic stirrer bar at 600 RPM for 2 hours, resulting in a colorless transparent liquid. The water content of each mixture was monitored by Karl-Fischer titration (Mettler Toledo V30 Volumetric KF Titrator) and only samples with less than 1 % water content by weight were used for the reaction. The viscosities of the prepared NADES were determined using a HAAKE RotoVisco 1 Controlled Rate rotational viscometer. For the viscosity measurements a "PP60 Ti" 60 mm titanium sensor plate was used with a constant shear stress of 1 Pa applied for 60 seconds.

Table 3.1: Prepared NADES systems. In the article we use the abbreviations that are defined in column 'NADES'.

NADES	HBA	HBD	Molar ratio
CCEG12	Choline chloride	Ethylene glycol	1:2
CCEG13	Choline chloride	Ethylene glycol	1:3
CCEG14	Choline chloride	Ethylene glycol	1:4
CCGLY12	Choline chloride	Glycerol	1:2
CCGLY13	Choline chloride	Glycerol	1:3
CCGLY14	Choline chloride	Glycerol	1:4
CCUR12	Choline chloride	Urea	1:2

3.5.3 Transesterification reaction

The transesterification reaction (see Figure 3.1) was performed in 40 ml sealed glass vials in a glycerol bath at 60 °C while stirring the mixture continuously at 300 rpm. The composition of the reaction mixture was 5.00 ml NADES, 270 μl vinyl laurate (1.0 mmol), 550 μl 1-butanol (6.0 mmol), and 5±0.5 mg iCALB enzyme. We grounded and sieved the immobilized enzyme beads, using the size range of 75 μm to 355 μm for the reaction, as described by Durand et al. [113]. The starting mixture contained the solvent, vinyl laurate and 1-butanol. The solvents were oversaturated for vinyl laurate by used a concentration of approximately 45 mg/ml. For comparison, Durand et al. used about 9 mg/ml for similar NADES systems [113]. After 5 minutes of temperature equilibration, the reaction was initiated by adding the enzyme to the mixture. Due to the possible heterogeneity of the samples, the entire reaction mixture was processed at each reaction time studied, rather than successive samples of the same system. Therefore, each sampling of the reaction was a new single-point experiment, as this resulted in higher repeatability. To stop the transesterification at the selected time, 5.00 ml of 1M HCl was added to the reaction mixture and the vials were removed from the glycerol bath.

To extract the vinyl laurate, 20.00 ml of hexane was added in a single step and the mixture was vortexed. The hexane formed a separate phase which was filtered through a 0.45 μm PTFE syringe filter. Reactions were carried out at 0, 5, 10, 15, 20, 30, 40, 60, 90 and 120 minutes and 24 hours. The various reactions were always performed in triplicate to monitor the standard error and repeatability of the measurements.

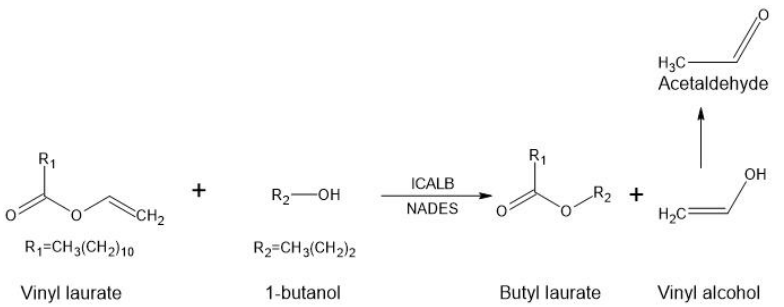


Figure 3.1: Transesterification of vinyl laurate by iCALB with 1-butanol. The side product vinyl alcohol tautomerizes into acetaldehyde and leaves the liquid phase.

3.5.4 Analysis of reaction

For the analysis of the vinyl laurate content of the sample we used the UV-Vis spectroscopy method described by Goujard et al. [115] and Durand et al. [113] In this method we follow the depletion of the vinyl laurate in the sample by measuring the absorption of the vinyl group at 200 nm. The absorption-concentration calibration curve for vinyl laurate in n-hexane was determined in the 20-100 μM range with an uncertainty of ± 3 μM for the concentration of the sample (See Figure A.1). The n-hexane extract was sampled and diluted 500 times in two steps and filtered with a 0.45 μm PTFE syringe filter. All absorbances were measured using quartz cuvettes (VWR European; Cat. Number 634-6018; 10 mm) with a Thermo Scientific Genesys 10S UV-Vis spectrophotometer. We calculated the conversion as the difference between the measured vinyl laurate concentration after the reaction and the calculated initial concentration. To take into account the small amount of vinyl laurate left in the polar phase, we also did the extraction steps on samples without reaction (zero time measurement) and we adjusted the conversion in each NADES medium to zero, based on the concentration difference determined in this step.

The products and substrates were identified by gas chromatography (Shimadzu GC-2010 gas chromatograph) using a non-polar column (Zebron ZB-5ms; length: 30 m; inner diameter: 0,25 mm; film thickness: 0,25 μm) coupled with a mass spectrometer (Shimadzu GCMS-QP2010S gas chromatograph mass spectrometer) using helium as the carrier gas. The oven temperature profile is shown in the supporting information (See Figure A.2). The n-hexane extracts were diluted 20 times (so that each component was below 0.1 mass percent). Measurements and analysis were performed using GCMSsolution software (Shimadzu) and the NIST 11 Mass Spectral Library.

In addition, quantitative analysis of both vinyl laurate and butyl laurate was performed on an Agilent 6890N Network Gas Chromatograph with a flame ionization detector using a column with low polarity (DuraBond DB-5HT; length: 30 m; diameter: 0,320 mm; film thickness: 0,10 μm) with helium as the carrier gas. Methyl laurate was used as an internal standard. Samples were diluted 20 times (so that each component is below 0.1 mass percent). The oven temperature profile used is given in the supporting information (See Figure A.3).

3.6 Results

3.6.1 Water content and viscosity of NADES

Since the presence of water significantly alters the physicochemical properties of NADES (e.g., viscosity and solubility) and may also induce undesirable side reactions, we controlled the water content in the investigated systems. The water content was less than 0.5 weight% in every system tested (see Table 3.2). According to Yadav et al. [61], this water content causes a decrease of about 9 % in the viscosity of the CCUR system at 60 °C, which we considered an acceptable uncertainty for our research objectives.

Table 3.2 and Figure 3.2 show that the viscosity of NADES is highly dependent on the compounds used and their molar ratio. Our measurements at 60 °C show that CCUR12 has the highest viscosity, followed by the CCGLY systems with slightly lower values and finally the CCEG systems with significantly lower viscosities. In the CCEG systems the viscosity also decreases with increasing HBD ratio. This is caused by the increasing molar ratio of ethylene glycol and the weakening of the intermolecular interactions between the different compounds due to the deviation from the eutectic composition (1:2). The CCGLY systems do not show this trend, possibly due to the higher viscosity of glycerol. As mentioned above, the CCUR systems with higher urea content were unstable at 60 °C and were excluded from further investigation.

Based on these results, we expect the reaction to be significantly faster in the CCEG systems because they are less viscous and impose less mass transfer limitations on the reaction components. In addition, the molar composition should have a similar effect, although the induced difference in viscosity is smaller.

Table 3.2: Water content and viscosity of NADES at 60 °C

NADES	Viscosity [mPa · s]		Water content [%]	
	Mean	SD	Mean	SD
CCEG12	12.188	0.090	0.488	0.007
CCEG13	9.322	0.024	0.486	0.006
CCEG14	7.605	0.025	0.151	0.003
CCGLY12	52.451	0.600	0.428	0.003
CCGLY13	50.923	0.300	0.414	0.006
CCGLY14	54.200	0.150	0.424	0.006
CCUR12	70.106	0.160	0.434	0.011

3.6.2 Observations on the reaction setup

The addition of vinyl laurate to the NADES medium makes the mixture opaque, as the medium, substrate and enzyme form an apparently heterogeneous mixture under magnetic stirring. By adding aqueous hydrochloric acid and n-hexane as a stopping reagent and extraction solvent, the mixture separates into two phases, a polar and an apolar phase. The

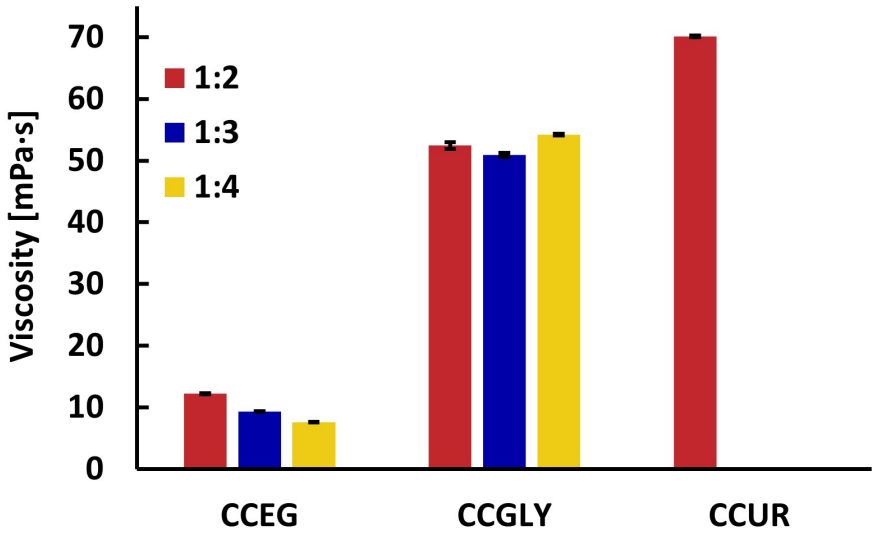


Figure 3.2: Comparison of the viscosity of different NADES systems. The error bars show the standard error of three repeated measurements (See Table 3.1 for the applied abbreviations).

immobilized enzyme beads remain between the two phases. Analysis of the apolar phase shows that the polar NADES compounds remain completely in the aqueous phase. On the other hand, the recovery of vinyl laurate was not complete; a small part of it remained in the polar phase. Therefore, the calculated conversions in different NADES were adjusted based on the measured concentrations in different NADES at zero time measurements.

When the mixing was stopped after two hours, all samples showed the separation of vinyl laurate and NADES, confirming the oversaturation of the samples (as intended). The low solubility was expected considering the difference in polarity between vinyl laurate and the NADES used.

3.6.3 Conversion over time

In Figures 3.3, 3.4 and 3.5 we summarize the conversion of vinyl laurate over time based on the UV-VIS measurement. Based on the amount of enzyme in the sample and the amount of substrate converted in the linear range of the reaction (the first 5-15 minutes), we estimate the enzyme activity. The results are plotted in Figure 3.6 as μmoles of substrate converted per mg of enzyme per minute by averaging the values of the three parallel reactions.

The reference results of the conversion in n-hexane are also plotted in Figures 3.3, 3.4 and 3.5 and these follow the expected trend. For the first 20 minutes (i.e., the initial reaction phase in n-hexane) the conversion is linear. Thereafter, the conversion gradually converges to a plateau at about 90 % conversion, as the mixture is depleted of the substrate after two hours.

Due to the higher viscosity of NADES and their stabilizing effect on the enzyme structure through secondary interactions, one might expect the initial phase of the reaction to be slower than in the reference solvents, but stable for a longer period of time. However, we observe the opposite trend in our NADES systems. The initial phase is much faster and after the first 30 minutes of reaction, it reaches a higher conversion than the reference n-hexane system. The only exception is the CCUR12 system, where only a small (about 10%) conversion was observed. This is in contradiction with previous reports where the same enzyme in the CCUR12 system had a good reactivity even for the same reaction [109, 113]. On the other hand, in CCEG systems the initial phase is followed by a sudden plateau, at a much lower final conversion than in n-hexane, while CCGLY systems reach an almost complete conversion after one hour. The plateau at low conversions in CCUR and CCEG systems may indicate the early denaturation of the enzyme.

The comparison of NADES with different compounds in Figures 3.2 and 3.6 shows no clear correlation between the viscosity and the reaction rate of the biocatalysis. The reaction showed similar initial reaction rate in CCGLY and CCEG systems, both exceeding the activity in n-hexane, while the viscosity difference between the two NADES is fivefold. In addition, the NADES system of urea, where the viscosity is 35 % higher than in CCEG, the reaction barely proceeded. The obtained initial reaction rates for different molar ratios of the NADES compounds are also surprising. In the CCEG systems, the reaction rate decreases with decreasing viscosity (increasing ratio of HBD compound), while we expected the opposite relationship. In the case of the CCGLY systems, there is no clear trend between the solvent viscosity and the initial reaction rate. These results indicate that viscosity plays a lesser role in the effect of NADES on biocatalysis than originally thought. Furthermore, contrary to our expectations, viscosity does not seem to be the bottleneck in the selection of NADES for (enzymatic) applications, as no mass transfer limitation was observed.

It is possible that the reaction rate decreases sharply at higher conversions, because the undissolved vinyl laurate reacts first (in the more accessible separate phase) and the reaction is not controlled by the viscosity of the NADES until the undissolved vinyl laurate is depleted. This would also mean that the reaction provides little information about the interactions between the substrate and the NADES medium in the initial phase, when the reaction is only influenced by the enzyme activity. Previous work found that iCALB enzyme retains 45 % and 95 % of its activity in CCGLY12 and CCUR12 systems, respectively, after 5 hours [113], arguing against complete inactivation of the enzyme in the time range of our study. However, they also followed the reaction with only a single point measurement, and therefore obtained less information about the evolution of the reaction over time.

To determine the final conversion of the transesterification, we measured the conversion of vinyl laurate in the different NADES after 24 hours (see Figure 3.7). These measurements show that after the apparent plateau at 2 hours (Figures 3.3, 3.4 and 3.5), still a significant amount of additional vinyl laurate reacts. The final conversion reached about 80-90% in each NADES system tested, except for the CCUR12 system, where the conversion actually remained unchanged after the first two hours. Our prior measurements showed no conversion of vinyl laurate in the absence of the enzyme. This indicates that in the CCEG and CCGLY systems the enzyme retains some of its activity after the first 2 hours of the reaction. However, in the 2-24 hour time frame, the decrease in the enzyme activity and the depletion of vinyl laurate in the system also play a significant role in the decrease of the reaction rate.

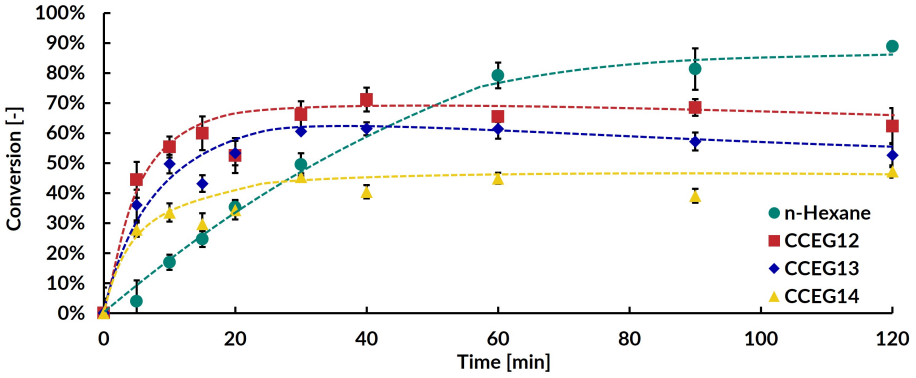


Figure 3.3: Conversion of vinyl laurate over time in CCEG based systems. The error bars show the standard error of the three repeated measurements. The dashed curves are a visual guide to the progress of the reaction.

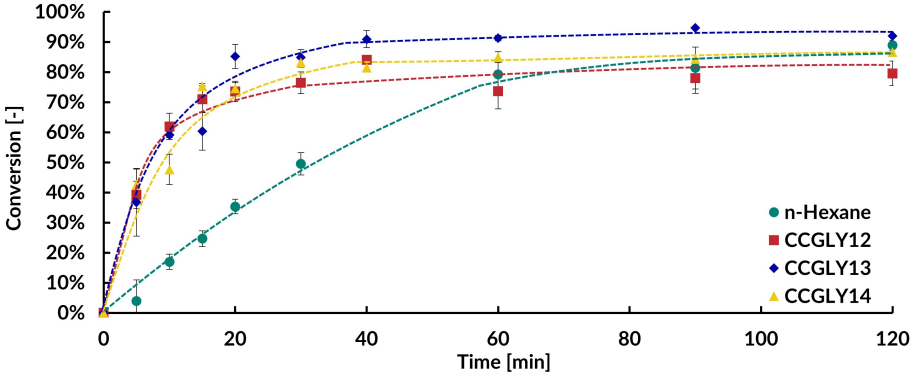


Figure 3.4: Conversion of vinyl laurate over time in CCGLY based systems. The error bars show the standard error of the three repeated measurements. The dashed curves are a visual guide to the progress of the reaction.

3.6.4 Side reactions

Up to this point, conversions have been calculated based on vinyl laurate depletion followed by UV spectroscopy. GC-FID and GC-MS measurements were also included to identify and quantitatively measure the products and possible side reactions (see Figure 3.8). These measurements reveal a significant side reaction between the ethylene glycol and the vinyl laurate in CCEG systems, resulting in 2-hydroxyethyl laurate, which accounts for about half of the observed vinyl laurate conversion. The relative amount of side product increases slightly with increasing molar ratio of the ethylene glycol, indicating that the more abundant HBD compound is less affected by the intermolecular hydrogen bonding network and is more likely to react. In addition, the proportion of butyl laurate in the product is constant over the different measurement times, indicating that the side reaction proceeds in parallel with the main reaction and does not take place after the main reaction has finished.

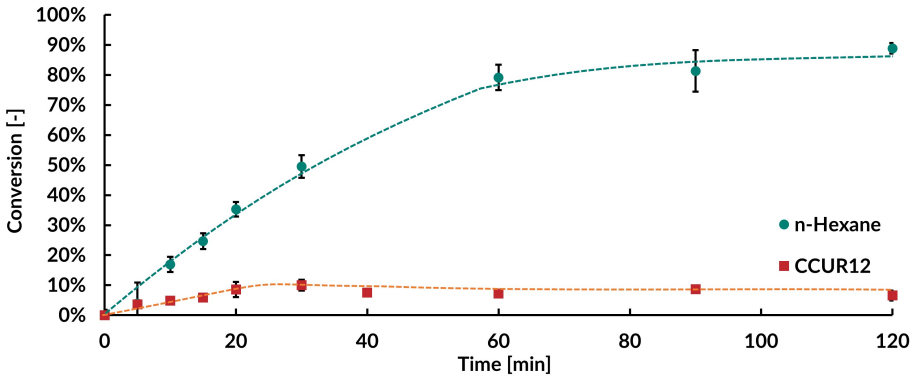


Figure 3.5: Conversion of vinyl laurate over time in CCUR based systems. The error bars show the standard error of the three repeated measurements. The dashed curves are a visual guide to the progress of the reaction.

The results indicate that the good final conversion of the vinyl laurate does not always lead to the desired product. In addition, NADES with different molar ratios are suitable media for enzymatic reactions, but we have to consider potential side reactions during application.

3.6.5 Solubility of vinyl laurate and stability of enzyme

To get a better insight into the solvation effect, we determined the solubility of vinyl laurate in NADES CCEG12, since in this system the conversion reached a plateau around 60 % conversion after the fast initial phase (see Figure 3.3). We measured a solubility of 11.6 mg/ml of vinyl laurate in CCEG12, while the previously applied concentration would have been 45.3 mg/ml. This means that only 25 % of the vinyl laurate was dissolved. To see how the initial phase of the reaction changes when the substrate is completely dissolved, we also ran the reaction with saturated samples (i.e. 11.6 mg/ml). The results are compared with the oversaturated CCEG12 system (see Figure 3.9). The initial phase of the reaction does not change: the initial reaction rate is only slightly lower and the reaction reaches almost complete conversion after 5 minutes. These results indicate that the main reason for the slowdown of the reaction is not the depletion of the undissolved vinyl laurate in the separate phase and the mass transfer limitation in the saturated solvent. The complete reaction for the saturated solute indicates the importance of the solvation effect in the NADES behavior. However, the plateau is not consistent with the ratio of the dissolved to not dissolved vinyl laurate (i.e., the reaction does not slow down at the potential depletion of the undissolved vinyl laurate or the NADES phase). Based on these results, the reaction proceeds well in the NADES phase, but it is not clear whether the undissolved vinyl laurate also reacts directly or the reaction proceeds exclusively via the NADES.

The long-term effect of NADES on the enzyme was also tested. We compared the behavior of the enzyme in n-hexane and in CCEG12 after 24 hours of enzyme incubation at 60 °C in these solvents. After incubation, the enzyme was used for the first 15 minutes of the reaction. In Figure 3.10 we compare the conversion over time results of the n-hexane and CCEG12 systems. While in n-hexane the reaction has maintained its original rate, yielding

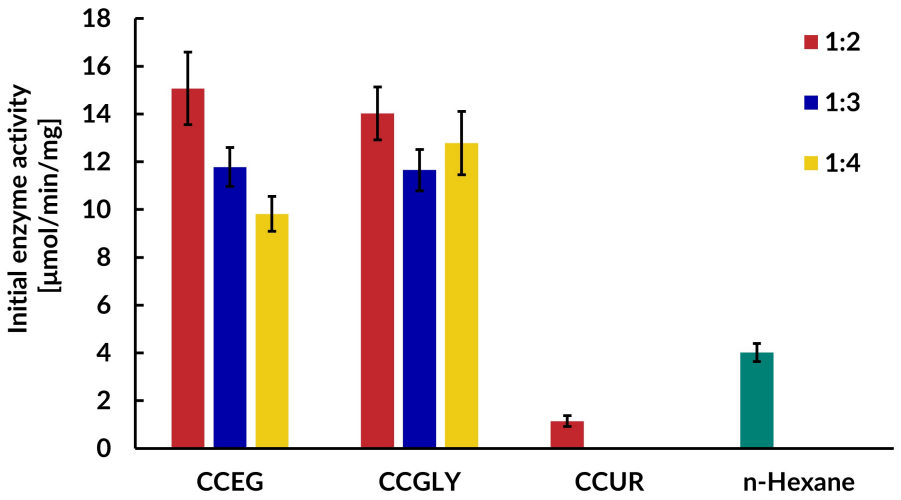


Figure 3.6: Initial reaction rate in different NADES systems based on the linear range (first 5-15 minutes) of the reaction. The error bars show the standard error of the three repeated measurements.

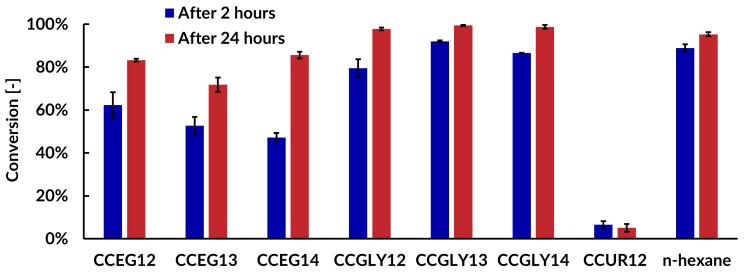


Figure 3.7: Comparison of 2 and 24 hours reaction conversions. The error bars show the standard error of the three repeated measurements.

a similar 20 % conversion after 15 minutes, in the CCEG12 system the reaction rate has dropped by a tenth, reaching only 5 % conversion compared to the original 60 % (cf. Figures 3.3 and 3.10).

While this demonstrates the long-term stability of the enzyme in the reference solvent, the results of the CCEG12 system are less conclusive. Although there is still a conversion in the NADES, the initial reaction rate is much lower than in the original setup. The decreased reaction rate could indicate the almost complete denaturation of the enzyme by the medium. However, in this setup the reaction is started by adding the substrate to the system, so the results are not directly comparable to the original reaction setup, where the enzyme addition started the reaction. This can also be interpreted as the higher viscosity of the media and the resulting slower dissolution of the substrate causing the lower reaction rate, even with partial inactivation of the enzyme. In connection with the solubility experiments, this would also indicate that the reaction takes place mainly in the NADES phase. Nevertheless, the present experimental setup is not fit to differentiate the weight of the different effects in

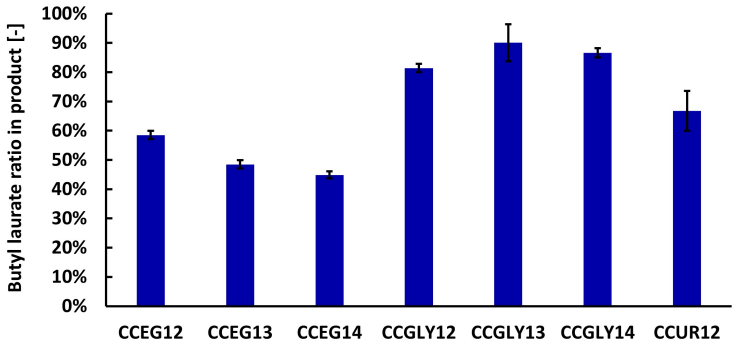


Figure 3.8: Ratio of butyl laurate to total product, showing the side product reactions in different solvents systems. All measurements are averaged over the first 120 minutes of the reaction. The error bars show the standard error of the repeated measurements.

the final measurable conversion rate.

3.7 Discussion

We found that the conversion over time varied to a large extent between the different NADES (see Figure 3.3, 3.4 and 3.5). With the exception of CCUR12, we measured higher initial enzyme activities in NADES than in n-hexane (see Figure 3.6). All reactions reached a plateau after two hours, which can be related to the combined effect of substrate depletion, enzyme inactivation and mass transfer limitation. The reactions show additional conversion in the 2-24 hour period (see Figure 3.7), indicating that the enzyme retains part of its activity in NADES after these two hours. Based on previous experimental and computational findings [113, 111], the retained stability of the enzyme in NADES can be expected. In all NADES except CCUR12, the final conversion is greater than 70% after 24 hours. In the case of CCGLY systems, it is over 95%. This indicates that certain NADES are ideal candidates for the enzymatic reactions studied. The high conversions in eutectic NADES are in agreement with the previous results of Gorke et al. [109] and Durand et al. [113]. However, the suitability of non-eutectic compositions has not been previously reported. Changing the molar ratio has only a small effect on the initial rate and the final conversion of the reaction, providing another factor to tune the properties of the solvent without hindering the reaction. The comparison of NADES viscosities and conversion over time (cf. Figure 3.2 and 3.6) shows that the viscosities vary over a wide range, but they show no correlation with the observed reaction rates. This suggests that the high viscosities of NADES and potential mass transfer limitations have less of a role in the conversion rates than originally thought and previously discussed in the literature. Additional analysis revealed a significant side reaction in the CCEG systems, resulting in 2-hydroxyethyl laurate as side product. Solubility tests in the CCEG12 system showed that the reaction is fast in the NADES media, but were inconclusive as to where the reaction is more likely to occur when a separate vinyl laurate phase is present. Incubation of the enzyme in the solvent prior to the reaction showed that the initial reaction rate remained essentially unchanged in n-hexane,

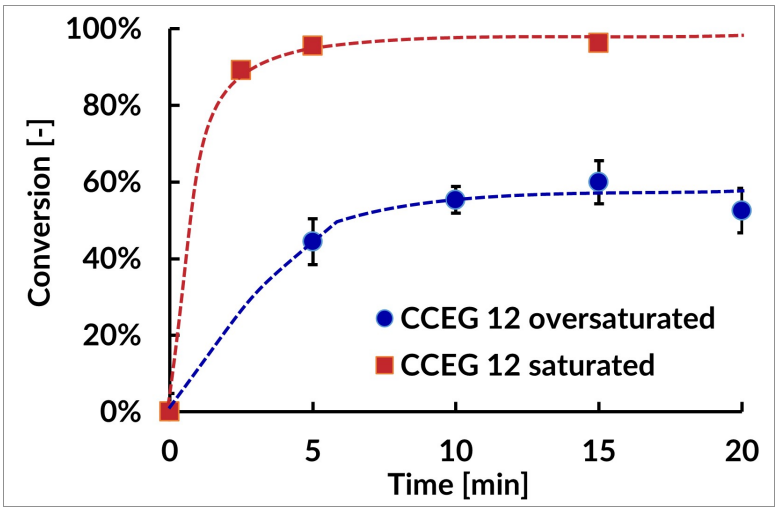


Figure 3.9: Conversion over time in saturated vs. oversaturated CCEG12 solvent systems. The error bars show the standard error of the three repeated measurements. The dashed curves are a visual guide to the progress of the reaction.

but it showed a large decrease in the case of CCEG12. However, these results could indicate both the inactivation of the enzyme in NADES and the importance of the solvation of the substrate in NADES prior to the effective reaction. Separation of the latter two will require a new experimental setup.

Our initial hypothesis was that the effect of NADES on the enzymatic reaction could be described by the solubility of the reactants, the viscosity of the NADES, and the activity of the enzyme in the given system. Based on the conversion over time (see Figure 3.3, 3.4 and 3.5), the different viscosities of the NADES have no clear effect on the reaction rate in the early phase of the reaction. Since the initial phase show no correlation with the NADES viscosity; the conversion reached a plateau relatively quickly after the initial phase; and we used the vinyl laurate in excess of its solubility limit; our first theory was that the reaction rate is governed by the reaction of the undissolved substrate, which forms a separate phase and is more accessible to the enzyme than the substrate dissolved in a viscous medium. The second, slower stage of the reaction begins when the separate, more reactive vinyl laurate phase is depleted. This stage would be governed by the mass transfer limitation of the substrate and the continuously decreasing enzyme activity.

However, this theory was disproved by the solubility experiments. In the saturated CCEG12 media, almost full conversion was reached after 5 minutes of reaction, which is a comparable reaction rate to thee original, supersaturated setup (see Figure 3.9). This proves that the more viscous media does not limit the reaction rate. However, this experiment does not reveal whether the reaction via the NADES media is more favorable than the direct reaction in the pure vinyl laurate phase. Since more substrate was reacted in the supersaturated setup than was present in the saturated CCEG12 sample, either the reaction occurred in parallel in the pure vinyl laurate and the NADES or the NADES was replenished by the dissolution of substrate from the pure vinyl laurate. In addition, the ratio of the dissolved to undissolved vinyl laurate does not match the position of the plateau in the conversion over

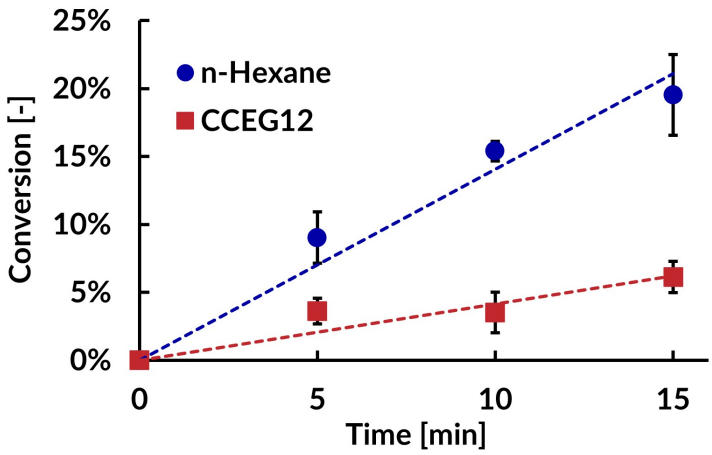


Figure 3.10: Effect of enzyme incubation in n-hexane and CCEG12 on initial rate enzyme activity. The error bars show the standard error of the three repeated measurements.

time. In both cases, the reaction was faster than in the n-hexane reference, emphasizing the positive effect of NADES on the reaction and the lack of hindering effect of viscosity. Where this additional rate comes from is currently unclear. Since the computational studies did not reveal any structural changes in the enzyme itself that would indicate increased activity [112], the most reasonable assumption is that a solvation layer is formed around the enzyme by the NADES or that the interaction energies between the NADES and the substrate make the catalytic reaction more favorable than in the conventional organic solvent.

The other potential reason for the slowing of the reaction is the continuously decreasing activity of the enzyme. Again, the 24 hour measurements still showed a significant conversion in the 2-24 hour range. To gain more insight, we incubated the enzyme for 24 hours at 60 °C in n-hexane and CCEG12 and then ran the reaction with the original amount of substrate. However, due to the prior incubation, the reaction setup was changed. While originally, the reaction was started by adding the enzyme to the equilibrated solvent-substrate system, here the reaction was started by adding the substrate to the incubated enzyme. In both cases, the co-substrate alcohol was added to the solvent prior to the reaction. In the case of n-hexane, due to the good solubility of the substrate and the low viscosity of the organic solvent, this did not make much difference, and accordingly we measured similar results to the original setup, indicating very similar enzyme activities after 24 hours, compared to the original setup. In the case of the CCEG12, the high viscosity and the limited solubility could also hinder the reaction, so the severely hindered reaction rate could not be clearly linked to the enzyme inactivation. Moreover, the almost complete denaturation would contradict to previous experimental results. It is more likely, that we experienced here the combined effect of enzyme denaturation, limited solubility of the substrate and limited access to the enzyme due to the high viscosity. However, at this time we do not have conclusive evidence of either enzyme inactivation or reaction limitation due to dissolution. Further investigation of these effects will require a revised experimental setup.

The significance of the side reaction is identified. Of the NADES tested, the CCEG systems

show significant side reactions with the HBD compound forming a side product with the substrate (see Figure 3.8). This is probably related to the interaction energies between the NADES compounds, as the increasing HBD ratio (less bound ethylene glycol molecules) results in a higher amount of side product.

Our results show that the effects included in the original hypothesis (solvation, viscosity, enzyme stability and side reactions) can describe the effect of NADES on the biocatalytic reaction. However, the original weights of the elements are different. There are three competing phenomena in the system that change the conversion rate over time: the original activity of the enzyme in the given reaction and solvent, the transport rate of the substrate, and the denaturation of the enzyme. Our results indicate that in this transesterification reaction and with the tested NADES, the limiting factor is not the viscosity induced mass transfer limitation, but the solubility of the substrate. At the same time, the initial reaction rates indicate a much higher apparent affinity of the enzyme in NADES than in n-hexane. The interaction to which this is related requires further investigation. The enzyme itself is shown to be stable in each system, in agreement with previous simulations [68, 111, 110] and experimental results [109, 113, 41]. However, with the applied experimental setup, we could not prove the long-term stability of the enzyme in NADES. However, the formation of the side product is an issue in certain systems. Interestingly, an earlier report on the same reaction did not mention similar side reactions at lower concentrations [113].

The distinction of the solvation/enzyme activity effects requires further investigation in the future, especially since previous computational studies have found a significant role of medium-substrate interactions in the outcome of the reaction [116, 62]. From a practical point of view, solubility should be the primary consideration in application development. In our reaction, the bottleneck is getting the substrate into the solvent and to the enzyme, not the reaction itself. In light of these findings, fed-batch systems, for example, would not be optimal. A two-step process could also be considered: first solubilization of the substrate, then the reaction. Further studies are needed to better understand the role of solvation. Here, molecular dynamics simulations of solvation free energies could give a better insight [117], especially regarding the role of hydrogen bonding [118].

3.8 Conclusions

In this work, we discuss the performance of NADES as a solvent for biocatalysis based on the solubility of the substrate, the viscosity of the media and the activity of the enzyme. For this purpose, we studied the transesterification reaction of vinyl laurate with 1-butanol by the immobilized lipase enzyme *Candida antarctica* B in different oversaturated NADES media. The NADES consisted of choline chloride with either urea, ethylene glycol or glycerol; in eutectic composition and with increased HBD ratio. We experimentally followed the conversion over time using gas chromatography and UV spectrometry. From these data, we determined the initial reaction rate and the final conversion achieved after 24 hours. In addition, we compared the reaction rate in saturated and oversaturated solutions and the effect of 24 hours of enzyme incubation in selected solvents.

With the exception of the urea-based system, good conversions were measured in the NADES

systems. The initial reaction rate and final conversion were comparable and in many cases higher than in n-hexane, which was used as a reference solvent. However, in CCEG systems the reaction reached a plateau before full conversion was achieved. Our results showed no correlation between the viscosity and the reaction rate, indicating that viscosity has little effect on the reaction and that mass transfer limitation is not an issue. In addition, our results indicate that NADES with increased HBD ratio are also suitable media for enzymatic reactions. Side reactions were observed in the ethylene glycol based systems, with an increasing amount at higher hydrogen bond donor ratios. The reaction proceeded in the time frame of 2-24 hours, indicating that the enzyme retains some of its activity over the long term. The reactions in the saturated CCEG12 system indicate that the reaction proceeds well in the NADES phase, but it is not clear whether the undissolved vinyl laurate also reacts directly or the reaction proceeds exclusively via the NADES. Incubation of the enzyme in NADES for 24 hours prior to the reaction showed that the enzyme retains almost all of its activity in n-hexane. The results in CCEG12 were inconclusive, as the initial reaction rate decreased to one tenth of the original setup. However, this could be the result of either the almost complete inactivation of the enzyme or the limitation of the reaction due to the slow dissolution of the substrate and high viscosity of NADES.

It is more likely that the plateau we experienced in the conversion of CCEG systems and the slow reaction rate of the incubated enzyme in CCEG12 system are the combined effect of enzyme denaturation, limited solubility of the substrate, and limited access to the enzyme due to high viscosity. However, the experimental setup presented in this study is not suitable for a more in-depth study and separation of the underlying phenomenon. Further investigation of these effects is needed.

Our results suggest that the most important effect of NADES is related to substrate-media interactions. Molecular dynamics simulations of the solvation energies of the solutes in NADES are a promising research direction to describe these interactions.

3.9 Linking results to hypotheses

In this chapter, I studied the proposed solubility-viscosity enzyme activity framework for describing the effect of NADES on biocatalysis. The case study for this work was the enzymatic transesterification of vinyl laurate by 1-butanol and iCALB enzyme. To determine the relationship between NADES properties and reaction performance, we measured the viscosity of the media, the solubility of the substrate, and the conversion of the substrate over time during the reaction. From the latter, we also determined the initial reaction rate and the final conversion. The tested NADES with different compounds and molar ratios are all effective media for the reaction. This demonstrates that the molar ratio is an additional variable to tune the properties of NADES. Although mass transfer limitation was expected to be a problem with highly viscous NADES, the reaction rate did not show any correlation with the viscosity of the media. On the other hand, the dissolution of the substrate in the NADES seems to be the most important factor. The substrate dissolved in the NADES reacts rapidly and then the reaction slows down as it is sustained only by the additional substrate dissolved from the separate phase. The slow dissolution of the additional vinyl laurate is the critical step and bottleneck of the reaction. The enzyme retains some of its activity for the 24 hours of the reaction, but due to the combined effects of solvation, viscosity, and changes in enzyme structure, we have not been able to determine experimentally how much of its activity is lost. In addition, side reactions are observed with certain NADES, and their magnitude increases with increasing molar ratio of the HBD compound. This calls for caution when using non-eutectic compositions, as the hydrogen bond network responsible for the inertness of certain compounds is weaker in these.

In this chapter I investigated the hypotheses that the effect of NADES can be described by the specific interactions between NADES-substrate, NADES-NADES and NADES-enzyme; and that the effect of NADES on enzymatic reactions can be related to substrate solubility/solvation, media viscosity and changes in enzyme structure. The results presented here partially support these hypotheses. While viscosity shows no clear relationship to reaction rate, solubility plays the major role in determining reaction rate. We could not determine the role of enzyme activity, as it cannot be separately determined in the experiments.

In this study, the experimental design was not suitable for separately examining the different effects of NADES on activity. In the future, a better designed experimental setup would be necessary. Alternatively, as discussed in the next chapter, molecular modeling can describe the direct NADES-enzyme interactions, which also helps to interpret the results seen here. It is now clear that solubility plays the major role in the effect of NADES on the reaction. In the future, NADES with a wider range of polarity should be used to gain more insight into this aspect. The role of viscosity seems to be less than we expected, since viscosity did not show any correlation with the reaction rate. It is possible that mass transfer limitation will not be a problem in practical applications, but this needs to be tested on a wider range of reactions.

Chapter 4

Effect of NADES with non-eutectic composition on enzyme stability

"Seed is sown, I'm chippin' in
Roll the bones, I'm chippin' in
Embed the code, I'm chippin' in
Mayhem flows"

Samurai

4.1 Preface

In the previous chapter, I concluded that it is difficult to study the effect of NADES on the structure of the enzyme separately. However, understanding this effect is important because potential interactions between the NADES and the enzyme may determine the stability of the enzyme in the media. In addition, these interactions may also induce changes in the active sites, which may lead to increased activity of the enzyme.

In this chapter, I further test the hypothesis that the effects of NADES on enzymatic reactions can be related to substrate solubility/solvation, media viscosity, and changes in enzyme structure, focusing on the role of direct enzyme-NADES interaction. In addition, I begin to test my third hypothesis, namely that the relevant properties can be predicted by a multiscale model involving the combination of molecular, macroscale, and data-driven modeling approaches. Obviously, I focus here on the possibility of implementing molecular dynamics simulation in the proposed multi-scale model to predict the effect of NADES on the enzyme.

In this chapter I discuss the results of the molecular dynamics simulation of a single *Candida antarctica* lipase B enzyme in NADES with different compounds and molar compositions. The computational cost of this method is high, so I limit the simulation to the initial phase, the first 200 ns of the system. I monitor how the overall stability of the enzyme changes in the different media, together with the potential stabilizing effect of the NADES compounds in the form of strong secondary interactions with the surface residues of the enzyme. Finally,

I look for structural changes in the active site of the enzyme and whether the NADES compounds diffuse into the enzyme structure and form hydrogen bonds with the residues there.

The content of this chapter is also published in the Journal of Molecular Liquids with the title "Effect of natural deep eutectic solvents of non-eutectic compositions on enzyme stability" [112]. Co-authors of this paper are Maksudbek Yusupov, Iris Cornet, Pieter Billen and Erik C. Neyts. Maksudbek Yusupov contributed to the development of the methodology, the formal analysis of the results, and the writing of the original draft. Iris Cornet, Pieter Billen, and Erik C. Neyts contributed to project supervision and review and editing during the writing process. In addition, Pieter Billen obtained financial support for the project.

4.2 Abstract

Natural deep eutectic solvents (NADES) represent a green alternative to common organic solvents in the biochemical industry due to their benign behavior and tailorable properties, in particular as media for enzymatic reactions. However, to fully exploit their potential in enzymatic reactions, there is a need for a more fundamental understanding of how these neoteric solvents influence the course of these reactions. Thus, the aim of this study is to investigate the influence of NADES with various molar compositions on the stability and structure of enzymes, applying molecular dynamics simulations. This can help to better understand the effect of individual compounds of NADES, in addition to eutectic mixtures. More specifically, we simulate the behavior of *Candida antarctica* lipase B (CALB) enzyme in NADES composed of choline chloride with either urea, ethylene glycol or glycerol. Hereto, we monitor the NADES microstructure, the general stability of the enzyme and changes in the structure of its active sites and surface residues. Our simulations show that none of the studied NADES systems significantly disrupt the microstructure of the solvent or the stability of the CALB enzyme within the time scales of the simulations. The enzyme preserves its initial structure, size and intra-chain hydrogen bonds in all investigated compositions and, for the first time reported, also in NADES with increased hydrogen bond donating compound ratios. As the main novelty, our results indicate that, in addition to the composition, the molar ratio can be an additional variable to fine-tune the physicochemical properties of NADES without altering the enzyme characteristics. These findings could facilitate the development and application of task-tailored NADES media for biocatalytic processes.

4.3 Introduction

Deep eutectic solvents (DES) were described first by Abbot et al. in 2001 [45]. DES are mostly ionic solvents, composed of Lewis and Brønsted acids and bases [1]. The strong interaction between the compounds forms a eutectic mixture with a significantly lower melting point than each constituting compound individually. In many cases the melting point is near or even below room temperature, which makes their application as solvent possible.

Natural deep eutectic solvents (NADES) are a subgroup of DES, composed of naturally derived (or derivable) compounds. However, there is no clear consensus on the definition of NADES. Generally, they are mixtures of two or more naturally derived organic compounds, that form a eutectic mixture through strong secondary interactions between the actors. These interactions form an intermolecular hydrogen bonding network, accompanied by charge delocalization in the system. These two processes stabilize the liquid state of the mixture at lower temperatures, resulting in a decreased melting point [1]. Typically, NADES are formed between quaternary ammonium salts (e.g., choline chloride) and hydrogen bond donating (or hydrogen bond donor (HBD)) compounds, like urea, polyols or organic acids [5]. NADES were first reported in 2011, as a subgroup of regular DES, containing metabolic products [13]. They are nonvolatile, nonflammable, biodegradable and less toxic than conventional organic solvents. Compared to ionic liquids, NADES are less sensitive to impurities, cheaper to manufacture and less toxic to the environment. NADES are considered as designer solvents: their behavior (e.g., melting point, viscosity, interaction with

substrates) can vary in a wide range based on their composition. These beneficial properties make NADES good candidates as green reaction media in many applications. Despite their novelty, NADES have already found applications in electrochemical processes, extractions, biomass pretreatment and as organic synthesis media [5].

NADES provide biocompatibility, biodegradation with limited or no toxicity, which is often required in biocatalytic processes. Thus, biocatalysis is one of the potential applications of NADES, e.g., in cosmetics, food, feed and pharmaceuticals [14]. Efforts were already made to test the applicability of NADES in biocatalysis: the first enzymatic reaction was described by Gorke et al. [109] and a comprehensive review of the advances in this field was made by Xu et al. in 2017 [41]. These investigations showed that NADES have a large effect on the reaction rate and final conversion. For example, in the work of Gorke et al. [109], in the aminolysis of ethyl valerate by immobilized *Candida antarctica* lipase B (iCALB), the mixture of choline chloride-urea and choline chloride-glycerol outperformed ionic liquids regarding the final conversion rate and showed similar initial reaction rate as in toluene. Meanwhile, the reaction in the mixtures of choline chloride and acetamide yielded only one third of the conversion as compared to that in toluene [109]. Differences in conversion rates and initial reaction rates were also found in the case of vinyl laurate transesterification using iCALB [113], which is studied in our research as well.

These differences can be related to multiple phenomena, which are associated with the strong secondary interactions in NADES. Firstly, such interactions play a role in the stability of the enzyme. The intermolecular network prevents the diffusion of the denaturing NADES compounds (like urea) inside the enzyme structure, which in turn leaves the intramolecular interactions of the protein intact and the enzyme active [68]. At the same time, the solvent constituents bind on the enzyme surface with strong secondary interactions, providing additional stability to the enzyme's structure at elevated temperatures and during long reactions [109]. For example, the half life of horseradish peroxidase increased to 350 minutes in choline chloride-urea mixture compared to 50 minutes in phosphate buffer solution [70]. Secondly, the NADES can also form strong interactions with the substrates, which can affect the activation energy and the equilibrium of the reaction. Thirdly, the strong interaction between the NADES compounds results in a high viscosity, which is typical for NADES and may lower mass transfer. These effects can explain the differences observed in biocatalytic reactions. However, the contribution of each effect to the final performance of the enzymatic reactions is still unclear. To effectively design novel NADES and to find the best solvent system for a given enzymatic application, this relation has to be understood.

Many studies related to this challenge have been published. These often include the use of computational methods to describe the interactions and other properties of NADES [11]. In this respect, molecular dynamics (MD) simulation is a commonly used method [77, 76, 79, 74]. Classical MD simulations provide information about the structural characteristics of NADES, such as the spatial distribution of specific groups or hydrogen bonding in the system, but also about physicochemical properties like density and viscosity. For instance, Sun et al. investigated the structural characteristics of choline chloride and urea-based NADES [77]. They determined the relationship between the molar composition and interaction energies, but also revealed an extensive ordered structure among the compounds. Das et al. studied the relaxation dynamics, spatial and dynamic heterogeneity of certain NADES systems [79]. Ullah et al. simulated the changes in the hydrogen bonding network upon the absorption of

carbon dioxide [74]. The advantage of the MD method in the study of biocatalytic systems is that it can handle relatively large systems compared to other computational chemistry methods, like density functional theory [68, 110, 111]. This enables the investigation of solvent-enzyme interactions by MD, which plays a central role in activity of the enzyme and final conversion of the reaction. Yet, the literature on this subject (i.e., solvent-enzyme interactions) is still very limited. The first study on this was conducted by Monhemi et al. [68], where they investigated the diffusion of the NADES compound molecules into a lipase enzyme structure and the interactions between the eutectic media and the different sites of the enzyme. Nian et al. studied which mechanism keeps the enzyme activated in NADES [110]. They found that the medium has no effect on the binding pockets, but it is possible that NADES increase the nucleophilic properties of the substrate, which promotes the reaction [110]. Shehata et al. studied the effect of an increasing water content in NADES on solvated thermoalkalophilic lipases and found that the mobility of the enzyme residues increase together with the hydration level of NADES [111].

To achieve reliable results, MD simulations require accurate force fields, which can be hard to acquire for NADES [75, 76]. This is due to the novelty of the systems and the unusual, strong secondary interaction between the compounds, as well as the required level of accuracy. Garcia et al. found that different charge assignment methods significantly affect the accuracy of the simulation [75]. Consequently, an optimization step before the application of an existing force field is often necessary, including the proper validation of the modifications. Doherty and Acevedo optimized the general OPLS-AA (Optimized Potentials for Liquid Simulations - All Atom) force field for a set of choline chloride-based DES, which resulted in good agreement with the experimental results of density, viscosity, heat capacity and surface tension, but the calculated self-diffusivity coefficient in the 298-328 K temperature range yielded errors up to 31% [76].

Hitherto, numerous studies investigated the effect of NADES on biocatalysis employing both experimental work and computer simulations. Still, our understanding is fragmented on how NADES affect the enzyme's structure, stability and activity. In addition, when the design of NADES is discussed, the emphasis is always on the use of different compounds in eutectic ratio. Deviation from the eutectic molar ratio, however, would be an equally important aspect to study for two reasons. Firstly, considering the temperature of a given application, there is a range of molar compositions in which the NADES remain liquid and hence a usable medium. Consequently, the molar composition could be an additional parameter to fine-tune NADES properties, similar to the water content that is a more thoroughly discussed topic [64, 61]. Secondly, by studying NADES with various molar ratios, we could acquire better understanding of the effect of individual compounds on the mixture and additionally on the enzyme itself. So far, research on NADES behavior in relation to composition only covered the water content of the mixture [119, 120, 121, 122].

Thus, in this study we investigate how the stability of the enzyme changes in NADES consisting of non-eutectic compositions and whether these solvent systems are applicable media for biocatalysis. Based on the available knowledge, we expect that the enzyme remains stable in eutectic and close-to-eutectic NADES. However, when increasing the HBD compound ratio, more intense interactions between the enzyme and solvent compounds are expected, which eventually leads to the disruption of the enzyme structure and the loss of its stability. In our study, we simulate CALB enzyme in NADES with eutectic compositions and increased

HBD compound ratio to investigate this effect. As reference systems, we also perform additional simulations with water, n-hexane (as benchmark solvents where the enzyme remains stable) and a concentrated aqueous urea solution to model the effect of a HBD compound without the hydrogen bond donating network of NADES. We monitor the microstructure of the NADES, the general stability of the enzyme and the changes in the structure of its active sites and surface residues.

4.4 Simulation details

We performed MD simulations using model systems consisting of the CALB enzyme embedded in NADES compositions. Specifically, we used three NADES structures that contain either urea, ethylene glycol or glycerol as HBD compounds and choline chloride as the quaternary ammonium salt (i.e., eutectic salt). Moreover, each NADES composition contained the eutectic salt and HBD compound in ratios of 1:2, 1:3 and 1:4. Thus, we employed nine model structures for CALB+NADES systems. We also simulated the reference solvent structures as benchmark systems. Our reference solvents were water, n-hexane (as experimentally applied solvents for lipase enzymes) and a 8 M concentrated aqueous solution of urea as a model of free HBD compound (without the effect of the intermolecular hydrogen bonding network of NADES). For the simulations, we have chosen 333 K temperature. Firstly, because the studied CALB enzyme is applied at this temperature in practice and we are interested in the behavior of the NADES and the enzyme at this temperature. Secondly, at elevated temperature, the equilibration of the model system is faster and thus the computational costs are decreased, without inducing fundamental changes in the structure of NADES or the enzyme. Every model system was simulated twice (including the benchmark systems) with different initial velocities of the molecules to check the consistency of the results. The simulated model systems are summarized in Figure 4.1.

MD simulations and analyses were carried out in GROMACS (v2018.01) [123, 124, 125]. The initial model systems were created using PACKMOL [126]. Visualizations were made in Visual Molecular Dynamics (VMD) [127]. The structure of the simulated enzyme CALB was taken from the Protein Data Bank (PDB ID: 1TCA [128, 129]). For the enzyme, n-hexane and aqueous urea systems, the OPLS-AA force field with Simple Point Charge water model (SPC) was used [130]. The topology of urea was taken from [131]. The force field parameters of n-hexane and urea were generated by LigParGen, a web-based application, which provides OPLS-AA force field parameters for organic molecules [132]. For the NADES compounds, the modified OPLS-AA force field by Doherty and Acevedo was used [76].

The CALB+NADES systems were prepared in the following way: The CALB enzyme was placed in a cubic box with dimensions $10\times10\times10$ nm^3 and periodic boundary conditions. Subsequently, the box was filled with 1500 solvent molecules (e.g., CCUR12 that contain 500 choline cations, 500 chloride anions and 1000 urea molecules) surrounding the enzyme. The systems were then energy minimized for maximum 10^5 steps applying the steepest descent algorithm. Subsequently, the equilibration of model systems was carried out in two steps: 2 ns simulation was performed first applying the NVT ensemble (number of molecules, volume and temperature are constant during the simulation), followed by 10 ns simulation in the NPT ensemble (number of molecules, pressure and temperature are constant). The

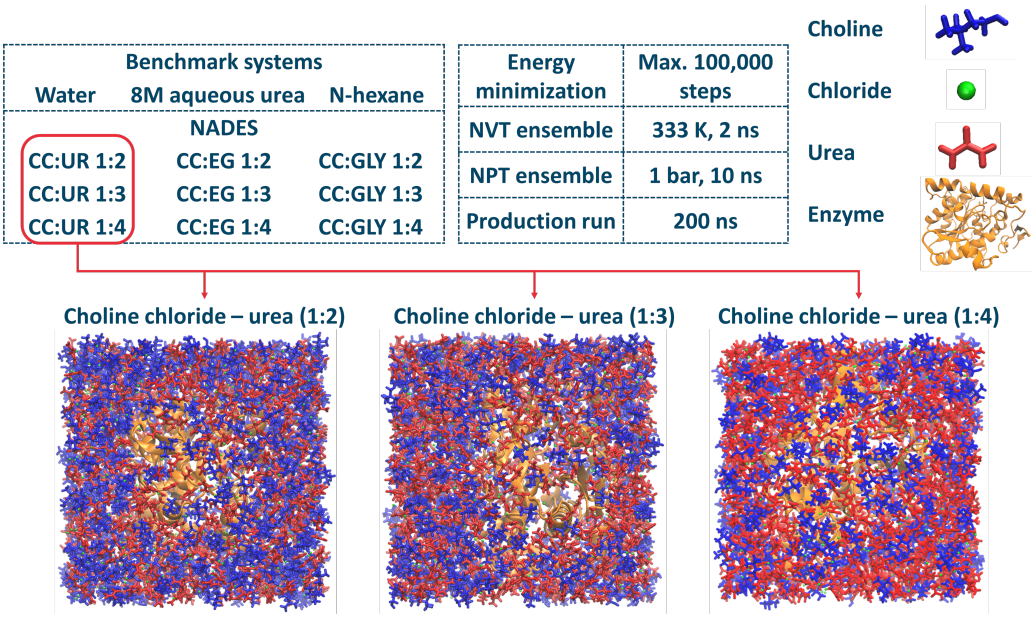


Figure 4.1: Simulated solvent systems and simulation setup. Here, the applied abbreviations of NADES systems used throughout the article are summarized as well: first the composition (choline chloride (CC), urea (UR), ethylene glycol (EG), glycerol (GLY)) and then the molar ratio of the compounds (ration of quaternary ammonium salt to HBD 1 to 2 (12), 1 to 3 (13) or 1 to 4 (14)).

production run was then performed for 200 ns employing the NPT ensemble again. The simulations were carried out at 333 K and 1 bar, applying the thermostat of Bussi et al. [133] with a coupling constant of 0.1 ps and isotropic Parinello-Rahman barostat [134] with a compressibility and coupling constant of $4.5 \cdot 10^{-5}$ bar^{-1} and 2.0 ps, respectively. During the equilibration the time step of 2 fs was used, whereas during the production run the time step was set to 1 fs. Note that this decrease in time step was necessary to prevent unacceptably large forces that would result in a failure of the integrator. A 1.6 nm cutoff was used for the electrostatic and Van der Waals interactions. The long range electrostatics were treated by Particle-Mesh Ewald summation [135]. To prevent the deformation of the enzyme residues before the production run (i.e., during the equilibration), restraints were applied with a force constant of 10^3 kJ/mol $\cdot$ nm^2 on the residues.

Since the optimized OPLS-AA force field was developed for eutectic NADES, we validated its use for non-eutectic compositions by performing additional simulations of NADES without the enzyme and compared the calculated density values with experimental results. The same simulation approach as mentioned above was used here, the only difference was the length of the runs: the NVT, NPT equilibration and production run were 1, 10 and 50 ns, respectively. As the model systems did not contain enzyme in these cases, a shorter simulation was sufficient to reach equilibrium, thus we used the same simulation time as Doherty and Acevedo [76].

To analyze the changes in the NADES microstructure and in the enzyme's overall stability, we

calculated the radial distribution functions (RDFs) between certain atoms and compounds, the root mean square deviation (RMSD) of the Cα atoms of the enzyme, the radius of gyration of the enzyme (R_g), the root mean square fluctuation (RMSF) values of the enzyme residues and the intra-main chain hydrogen bonds in the enzyme [136]. To investigate the stability of the catalytic triad residues of the enzyme (i.e., serine 105 (SER105), aspartate 187 (ASP187) and histidine 224 (HIS224), see below) and the structural changes of specific sites, we calculated pairwise distances between the atoms of the active site residues, the solvent accessible surface area (SASA), the ratio of secondary structures in the enzyme, the pairwise distance of residues in certain alpha-helices and the RDF of NADES compounds and certain surface residues. As most of the calculated values have high deviations, we always display 10 ns moving averages on our graphs to keep the changes observable.

The RDF is a measure of the probability of finding a particle at a distance of r away from a given reference particle, relative to that for an ideal gas.) RDF is defined by Equation 4.1:

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2} \tag{4.1}$$

where $\langle \rho_B(r) \rangle$ the particle density of type B at a distance r around particles A, and $\langle \rho_B \rangle_{local}$ the particle density of type B averaged over all spheres around particle A with radius r_{max}.

The RMSD is the deviation of certain atomic positions in the protein compared to a reference structure by least square fitting. RMSD is calculated from Equation 4.2:

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \left\| \mathbf{r}_i(t_1) - \mathbf{r}_i(t_2) \right\|^2 \right]^{1/2} \tag{4.2}$$

where $M = \sum_{n=1}^{\infty} m_i$ and $\mathbf{r}_i(t_1)$ is the position of atom i at time t [136].

This gives a measure of similarity between protein structures and quantifies the possible unfolding of the protein in time. A steady increase of the RMSD values indicates the disruption of the original structure. It thus gives information about the (possible) denaturation of the protein and its mechanism. We calculated the RMSD values of the Cα atoms of the enzyme by utilizing the *gmx rms* tool of GROMACS [136]. R_g describes the dimensions of the folded protein. R_g is calculated from Equation 4.3:

$$R_g = \left(\frac{\sum_i \left\| \mathbf{r}_i \right\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \tag{4.3}$$

where m_i is the mass of atom i and $\mathbf{r}_i$ the position of atom i with respect to the center of the molecule [136].

An increase in R_g indicates the unfolding of the enzyme and, by that, its denaturation. We used the *gmx gyrate* tool of GROMACS to calculate the R_g values in our simulations [136]. RMSF is the average deviation of particle positions over time (with respect to the initial reference position). RMSF is calculated from Equation 4.4:

$$RMSF = \frac{1}{N} \sum_{i=1}^{N} \left| \mathbf{x}^{(i)}(t) - \mathbf{x}^{(i)}(0) \right|^2 \tag{4.4}$$

where N is the number of particles to be averaged, vector $\mathbf{x}^{(i)}(0) = \mathbf{x}^{(i)}{}_0$ is the reference position on the i-th particle and vector $\mathbf{x}^{(i)}(t)$ is the position of the i-th particle at time t [137].

In our simulation, the calculation was done on the Cα atoms of each residue and the reference point was the initial position of the simulation. RMSF is an indicator of macromolecular flexibility. We used the *gmx rmsf* tool of GROMACS to identify the flexible regions of the enzyme in our simulations [136]. We considered a geometric criterion to identify hydrogen bonds between the residues, viz. a distance between acceptor and donor smaller than 3.5 Å and the angle between hydrogen-donor and donor-acceptor smaller than 30° [123]. We calculated the pairwise distances by *gmx distance* to determine structural changes in the catalytic triad residues and between the residues of α-helices [136].

To identify possible changes in the structure of the surface residues of the enzyme, we calculated the SASA as a function of time by *gmx sasa* tool of GROMACS [138]. To investigate how the solvents affect the α-helices and β-sheets in the enzyme structure, we used the DSSP program to calculate the changes in the secondary structure of the enzyme over time [139, 140]. The DSSP program was run via the *gmx do_dssp* tool of GROMACS.

4.5 Experimental details

Because experimental data are not available on non-eutectic compositions, we acquired these as well. Thus, for the synthesis of NADES, we used choline chloride (Acros Organics, 99%), ethylene glycol (Fischer Scientific, $\geq$ 99%), glycerol (Fischer Scientific, $\geq$ 99%) and urea (Fischer Scientific, 99.5%). The mixtures were prepared by stirring (600 RPM) the samples at 80 °C for at least 2 h until they became homogeneous. The water content of the NADES was determined by Karl Fischer titration with a Mettler Toledo V30 Compact Volumetric KF Titrator. The density measurements were performed on samples with water content below 0.5 m%, with Quantachrome Ultrapyc 1200e Automatic Gas Pycnometer with a 10 cm^3 sample cell with helium gas at 60 °C. The NADES samples were stored in a vacuum desiccator with silica gel to prevent water uptake.

4.6 Results and discussion

4.6.1 Validation of the force field

Before the analysis of the simulation results, we checked the reliability of the applied force field for the simulation of NADES systems used in this study. We compared our calculated densities with our own experimental measurements and data from the literature. Table 4.1 summarizes the experimentally determined density values (measured by Yadav et al. [61] and Leron et al. [141] and extended by our results) and simulation results (by Doherty and Acevedo [76] and our results). Unfortunately, literature data were only available for eutectic compositions for the temperature used in this study (i.e., 333 K).

System	Results taken from literature				Results of this study			
	Experiment [61, 141]		Simulation [76]		Experiment		Simulation	
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
CCUR12	1.177	0.000	1.129	NA	1.221	0.003	1.128	0.003
CCUR13	NA	NA	NA	NA	NA	NA	1.135	0.003
CCUR14	NA	NA	NA	NA	NA	NA	1.139	0.003
CCEG12	1.097	0.000	1.107	NA	1.140	0.008	1.082	0.003
CCEG13	NA	NA	NA	NA	1.130	0.003	1.081	0.003
CCEG14	NA	NA	NA	NA	1.129	0.008	1.081	0.003
CCGLY12	1.172	0.000	1.173	NA	1.221	0.003	1.172	0.003
CCGLY13	NA	NA	NA	NA	1.228	0.004	1.188	0.003
CCGLY14	NA	NA	NA	NA	1.240	0.005	1.198	0.003

Table 4.1: Densities [g/cm^3] of NADES systems calculated from experiments and simulations. The literature experimental density results are taken from the work of Yadav et al. for CCUR12 [61] and Leron et al. for CCEG12 and CCGLY12 [141]. The literature simulation results are from the work of Doherty and Acevedo [76]. Experimental results are missing for CCUR13 and CCUR14 systems, because these were unstable at 60 °C.

The differences between our calculated and experimentally obtained densities (both ours and literature) were below 5%. In comparison, the difference between the experimental values and simulation results of Doherty and Acevedo were below 2%, whereas our measured densities were higher than those found in literature by 4% in every case. This systematic difference is believed to be related to the applied method. An error due to water uptake is improbable as this would decrease the measured density. Nevertheless, when comparing our experimental and simulation results, we observed similar trends. In case of CCEG, the density decreased with increasing molar ratio of HBD, whereas it increased by increasing the ratio of HBD in cases of CCUR and CCGLY (see Table 4.1). These findings are in line with the work of Abbot et al., where they measured density as function of composition at room temperature [56].

As density is a convoluted result of many factors, good agreement in densities alone does not validate the force field. Because the development and full validation of the force field are out of the scope of this research, we consider the density check sufficient at this point. Based on the good agreement between the measured and simulated densities, we consider the applied force field appropriate to simulate the non-eutectic NADES of our interest.

However, thorough validation in future work would include the comparison of viscosity, heat capacity, enthalpy of vaporization and self-diffusivity coefficients in addition to densities. These properties, at this point, were only calculated for the eutectic compositions [76]. On the other hand, the feasibility of force field refining was proven by Spittle et al., who recently fitted and validated the AMBER force field for CCGLY systems with various molar ratios [142]. Similarly, Zhang et al. reported on the refinement of the GAFF force field for the simulation of CCEG system with different molar compositions [143].

In Table 4.1, measurements are missing on non-eutectic CCUR (namely CCUR13 and CCUR14) samples, as these systems were already unstable upon preparation. This highlights the limitation of this approach: the simulation does not provide information about the stability of the NADES system itself. This finding emphasizes the necessity of thorough

validation of simulation results.

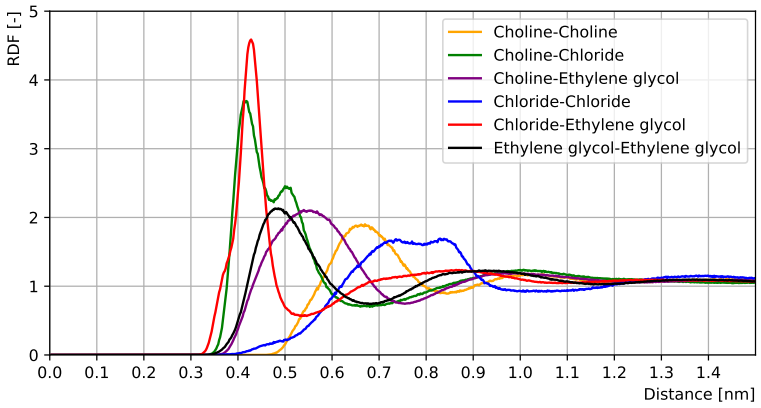
4.6.2 Microstructure of NADES

To understand how changes in molar composition affect the biocatalysis, the first thing to consider are the changes in the microstructure of NADES. Disruption of the interactions in the solvent system can also lead to a change in the behavior of the solutes.

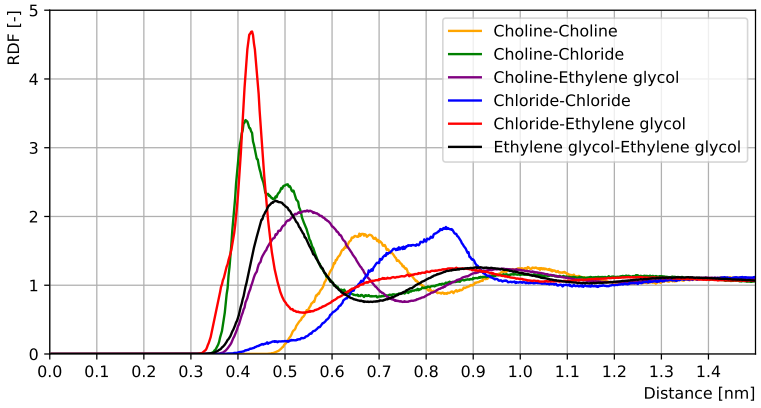
Thus, we first compared the center-of-mass (COM) radial distribution functions (RDF) of the different NADES systems. Figure 4.2a depicts the COM-RDF of the CCEG12 system. There are strong and narrow peaks between the choline and chloride and between chloride and ethylene glycol (HBD). In contrast, the peaks corresponding to interactions between the same compounds with each other are wider and less intense. The interactions between choline and ethylene glycol are similar in strength and width to the interactions between the same compounds with each other. The weakest interaction is between chloride anions. The obtained results indicate that the original structure of the pure compounds is largely lost. Instead, in the NADES the two compounds (choline & ethylene glycol) interact through the chloride anion and this interaction is highly stable (the distribution of their distance is narrow, see Figure 4.2a). Meanwhile, choline and HBD (ethylene glycol) molecules interact with themselves and each other in a more widely distributed way.

To evaluate the effect of the molar ratio on the NADES microstructure, we compared the distance and intensity of interactions again by their RDFs. Figure 4.2b and 4.2c shows the COM-RDF peaks of CCEG system in different molar ratios (i.e., CCEG13 and CCEG14). The position of the peaks did not change significantly with the molar ratio. However, the intensity of the peaks slightly changed with the composition (cf. Figure 4.2, a, b and c). With increasing HBD ratio the interaction between choline-choline, choline-chloride and chloride-chloride decreased, while every interaction involving HBD slightly increased. This is reasonable: by increasing the concentration of HBD compound, all other constituents interact more likely with it. The fact that the distances remained the same, but only the intensity changed, indicates that the nature of interactions in NADES does not change intrinsically in the studied composition range. The only change in the RDFs shape with the molar ratio was the chloride-chloride interaction. The double peak in the chloride-chloride RDF is related to their complex interaction with the choline and ethylene glycol molecules. The choline-chloride RDF already has two distinctive peaks and chloride anions simultaneously interact with the HBD and choline compounds. Shifting the molar ratio results in a more dominant interaction with the HBD compounds, which relates to the more distant peaks between the anions. Additionally, the decreased relative concentration of the ionic compounds can also contribute to the slightly longer average distances. These results are in line with earlier findings of Ferreira et al. [119] and Kaur et al. [121] on the effect of water on NADES: a NADES can retain its structure even in mixtures with high water content (up to 35%). It is reasonable that moderate changes in molar composition cause less disruption in the NADES' microstructure than the formation of a ternary system with water.

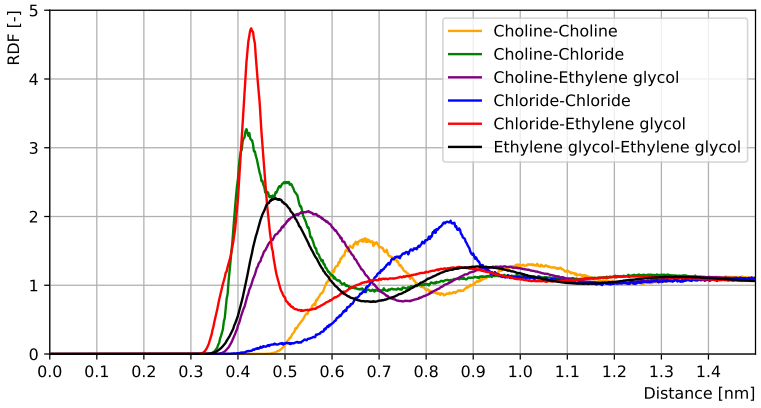
The other two compositions (i.e., CCUR12 and CCGLY12) gave similar results (Figure A.4, A.5): the strongest interactions were formed between choline-chloride and HBD-chloride.



(a)



(b)



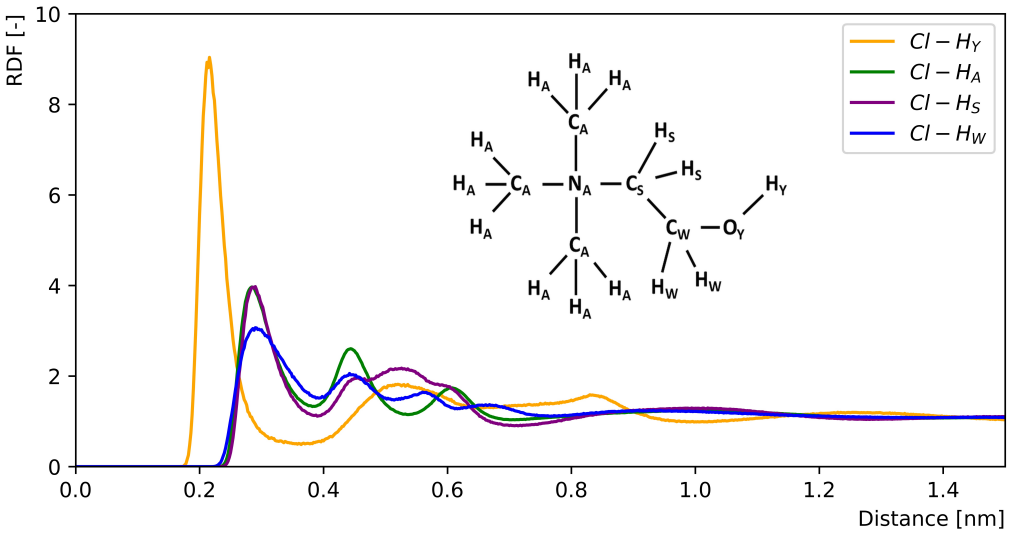
(c)

Figure 4.2: Center-of-mass radial distribution function of (a) CCEG12, (b) CCEG13 and (c) CCEG14 systems.

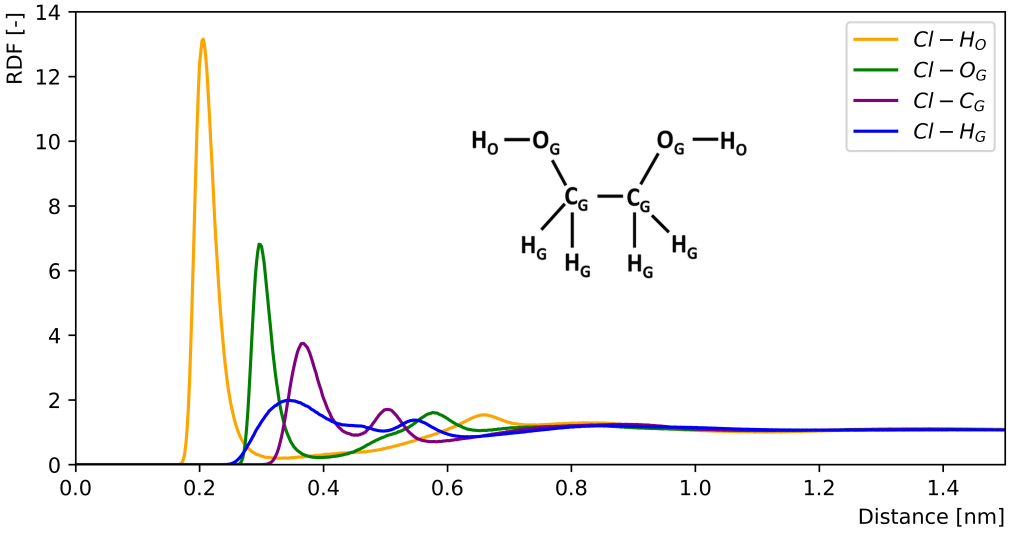
This is followed by the wider peaks of choline-choline, HBD-HBD and HBD-choline interactions. The only unexpected result is that in the intensity of the glycerol systems the HBD-chloride peak remained relatively small and appeared at longer distance (Figure A.5a). The possible causes could be the larger molecular volume of glycerol and its three hydroxyl groups. Due to its larger volume, the neighboring molecules are located farther. This and the three positions for hydrogen bonding result in a more flexible interaction and wider distribution. Our findings are in line with the results of Monhemi et al. [68] and Nian et al. [110], who found similar distributions in choline chloride-urea and choline chloride-glycerol NADES. Thus, the results of the RDF showed that increasing the molar ratio of the HBD compound leads to a stronger interaction of all other constituents with it and weaker interaction among the HBD molecules. However, the nature of the interaction does not change fundamentally with the increase of HBD compound (cf. Figure 4.2 and A.5). As the CCUR systems with increased HBD compound ratios were not stable in experimental work, they will not be discussed.

To study the possible changes in the nature of the interactions between the constituents, we calculated the atom-atom RDF of choline and HBD compounds relative to chloride anions. Figure 4.3 shows the atom-atom RDF of the choline chloride-ethylene glycol 1:2 system. The RDFs with lowest distances in the RDF plots of the two constituents are the hydroxyl group of the choline and the two hydroxyl groups of the ethylene glycol. This indicates that the compounds interact with the chloride anion through these functional groups. The nature of the interacting groups and the length of the interactions indicate that the compounds form hydrogen bonds. The observations are similar in the other two NADES systems: The urea interacts with chloride through the hydrogens of the amine groups (see Figure A.6), meanwhile, the glycerol interacts through all three of its hydroxyl groups (see Figure A.8a and A.8d). The similar intensity of the interaction of the three hydroxyl groups of glycerol is also in line with the wider distribution of its COM-RDF. These results indicate that all of the examined NADES form strong hydrogen bonding interactions between the constituents as the primary mode of interaction. These findings are in agreement with earlier results of Monhemi et al. [68] and Shehata et al. [111] who described similar interactions between urea and choline. In addition, we studied the direct interactions between the choline cation and the HBD compound by calculating the atom-atom RDF between the hydroxyl group of the choline and the HBD functional groups of the HDB compounds (see Figure A.9). The results show peaks in the distance range of hydrogen bonding interaction in case of CCUR (Figure A.9a) and CCEG (Figure A.9b, A.9c and A.9d) systems. However, these peaks are smaller and at slightly longer distances than the RDF of the chloride anion. In the CCGLY systems, ((Figure A.9e, A.9f and A.9g)) no distinctive peak was observed. Furthermore, the intensity and location of the peaks do not show a relation to the molar ratio of the NADES. These results indicate that the primary interaction occurs via the anion.

The effect of molar composition was similar to the COM-RDF results: the molar ratio did not affect the position of peaks in the atom-atom RDF (cf. Figure 4.3, A.7 and A.8). On the other hand, the intensity of the peaks changed slightly with the molar composition. However, in this case the different NADES follow different trends. While in CCGLY systems (Figure A.8) the intensities of choline peaks are increasing with increasing HBD ratio, in CCEG (Figure 4.3 and A.7) systems the choline peaks are decreasing with increasing HBD ratio. At the same time, the HBD is not affected by the altered molar ratio. By increasing the concentration of HBD, a decrease in the choline-chloride interactions would be expected.



(a)



(b)

Figure 4.3: Atom-atom RDF of (a) choline cation and (b) ethylene glycol relative to chloride anion in CCEG12 system.

The results of CCEG systems are in line with earlier studies, where upon addition of water, the interaction between different compounds was decreasing [111]. It is possible that the CCGLY system is more sensitive to HBD surplus as the extra hydroxyl group and addition of HBD promotes the interaction between choline and chloride instead of interaction between the different compounds. Moreover, Weng and Toner found that the addition of water to the CCGLY12 system decreased the number of interactions through the chloride anion, but the NADES preserved its unique properties in a wide range of hydration levels [144]. Again, their findings support the conservation of the core NADES behavior in a wide composition range.

Overall, the simulation did not indicate significant changes due to changes in molar composition in the microstructure or the nature of interactions in the studied NADES. There are small deviations in the interaction intensity among the constituents, but the distributions of constituents are comparable to each other and the nature of the interactions remained the same in every investigated system.

4.6.3 Enzyme stability

4.6.3.1 Root mean square deviation of atomic positions and radius of gyration

To evaluate the stability of the simulated enzyme, we first describe how much its structure changes as a function of time (See the graphical representation of the enzyme structure in the CCEG12 system in A.10). To describe this change, we calculated the root mean square deviation (RMSD) of the Cα atoms of the enzyme and the radius of gyration (R_g) of the enzyme as a function of time. First, we evaluated the RMSD and R_g of the enzyme in eutectic NADES (i.e., the NADES systems with molar ratio 1:2) and reference solvents. Figure 4.4 displays the smoothed RMSD and R_g values (10 ns moving averages as mentioned in simulation details) of the above mentioned systems.

As shown in Figure 4.4, all RMSD values of the C_α atoms of the enzyme in the eutectic NADES remain lower than those in water. Except for CCUR12 and 8 M aqueous urea, the RMSD reached a stable value before the end of the simulation. Although the CCUR12 system did not show stability, the final value of RMSD is still below the equilibrium value in water. In contrast, RMSD in 8 M aqueous urea quickly surpasses the values in water and steadily increases until the end of simulation. The R_g displays similar trends: the compactness of the enzyme did not change significantly in most of the simulated systems. The NADES R_g values remain below the values obtained in water. The two exceptions are again the simulations of 8 M aqueous urea and CCUR12, where the final value of R_g is still close to the value in water, but the increase is steady. However, our results do not show complete denaturation in the aqueous urea system during the 200 ns simulation. Still, these results are in line with earlier experimental results where the enzyme remained stable in water and the simulated eutectic NADES, while 8M aqueous urea caused denaturation [109].

At the same time, n-hexane showed unexpected results: the RMSD reached an equilibrium value that was the highest of all studied systems. The compactness of the enzyme remained the lowest among all. A possible reason for this may be the non-polar nature of the solvent.

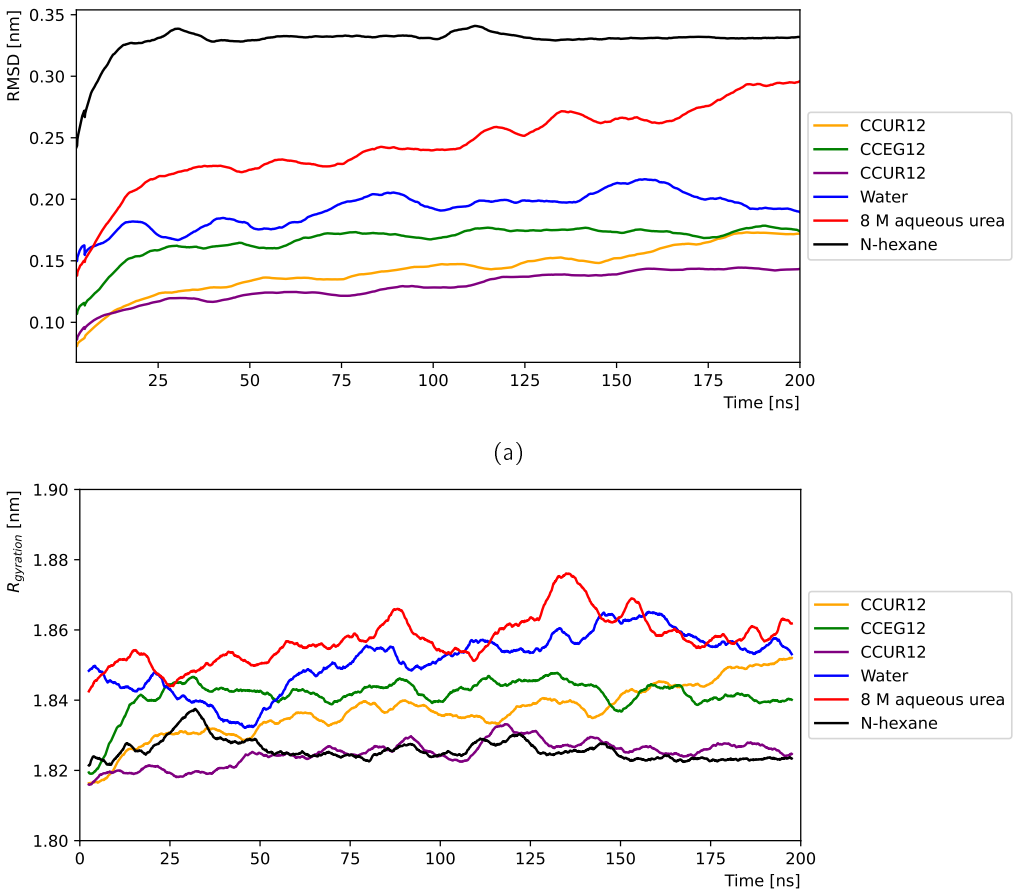


Figure 4.4: (a) RMSD of Cα atoms and (b) radii of gyration in CALB during 200 ns simulation: comparison of eutectic NADES and reference solvent systems.

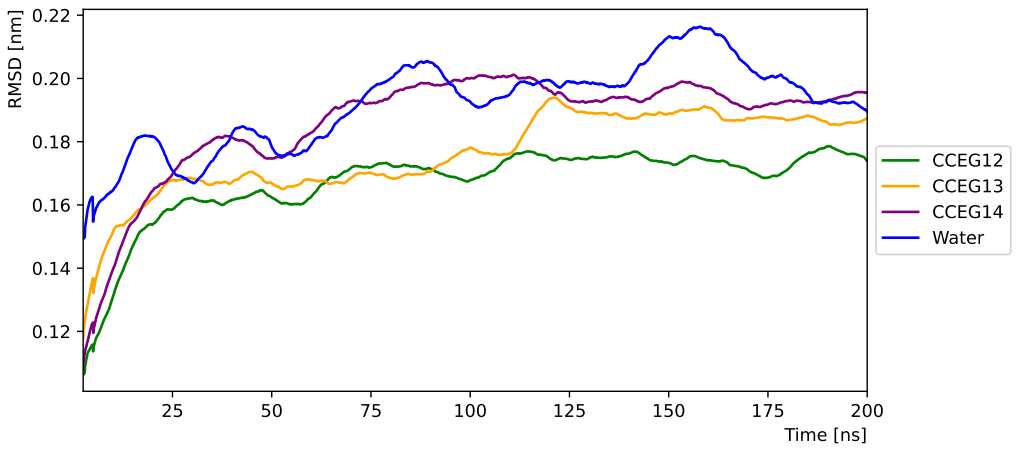
The small and apolar hexane molecules do not form strong interactions with each other or the enzyme residues and therefore they can increase the observed fluctuation of certain residues (hence the high RMSD values), without disrupting the interaction between the enzyme residues and by that the overall structure of the enzyme (viz. low R_g values). Overall, none of the studied systems show clear denaturation, although, the 8 M aqueous urea and CCUR12 systems show decreasing structural organization during the simulation.

These results are generally in line with earlier studies: simulations of eutectic CCEG [110] and CCUR [68] systems showed agreement with our results in RMSD values. The calculated RMSD values showed concurrence in many simulations of different force fields, namely OPLS-AA [110], GROMOS96 [68], CHARM36-AA [111] and optimized OPLS-AA (used in this work). Discrepancies (greater than 10% in RMSD values) appeared only in concentrated aqueous solutions and at high simulation temperatures (373 K) [68, 111]. Results were especially divergent for 8 M aqueous urea. While Monhemi et al. [68] measured a high RMSD of 1.2 nm at 300 K after 50 ns, in the simulation of Shehata et al. [111] it was 0.2 nm at 373 K after 300 ns. The recent study of Monhemi on the comparison of force fields showed that the united atom force field exhibits a higher deviation of the protein structure from the native state, suggesting that it requires more time for equilibration or a higher temperature to reduce simulation time [145]. An earlier study attributed the similar 0.5 nm RMSD value to the denaturation of the barnase enzyme in aqueous urea, which is a 50% relative increase compared to the initial structure [146]. However, the RMSD reached these extreme values (absolute or relative) in none of our simulations.

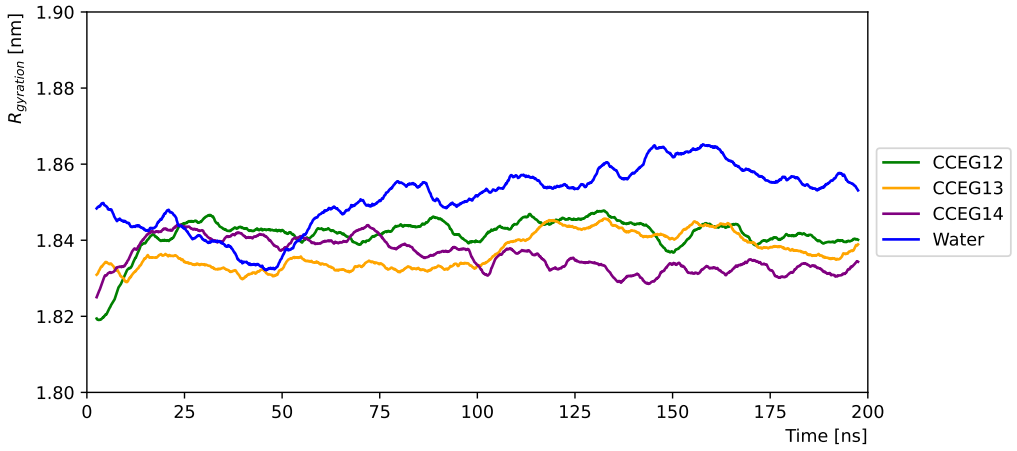
To study the effect of an increasing HBD ratio in our system, we also compared the RMSD and R_g values in the non-eutectic NADES systems (i.e., CCEG13/14, CCUR13/14 and CCGLY13/14). Results of the RMSD and R_g values of CCEG systems with different HBD ratios as a function of time are shown in Figure 4.5.

Our initial expectation was that the enzyme structure disrupts more heavily with the increasing ratio of HBD. With increasing HBD ratio, we expected more interaction between the HBD and the enzyme, which would result in stronger alterations in the enzyme and consequently higher RMSD and R_g values. Interestingly, this happened in none of the studied systems. As is obvious from Figure 4.5, there is a minimal increase in RMSD with the increase of HBD ratio in CCEG systems, but all three systems are still in the stable region. The difference is even less clear for the R_g values and the fluctuations within a single simulation are larger than the differences between various systems. We found similar results in CCGLY systems (see Figure A.11). We conclude that based on these findings, the increase of HBD compounds in the NADES mixture does not result in stronger interactions with the enzyme nor in disruptive effects on its structure, at least in case of the studied NADES and composition ranges.

Overall, we did not find evidence for denaturation of the enzyme in our simulations. The enzyme showed stability, not just in NADES with various compositions, but also in increasing HBD ratios, as the enzyme overall structure was not negatively affected.



(a)



(b)

Figure 4.5: (a) RMSD of Cα atoms and (b) radii of gyration in CALB during 200 ns simulation: comparison of choline chloride-ethylene glycol (CCEG) systems with different molar ratios.

4.6.3.2 Root mean square fluctuations

The flexible domains (with high RMSF value) are the most likely to change their structure into an unfolded state. These are therefore the most probable starting points of the enzyme denaturation. We determined the flexible regions of CALB enzyme and how the flexibility changes with different NADES.

Based on earlier studies we expected two main regions with high flexibility:

- The α5 helix (residues 142-146) and the turns next to it (residues 132-141 and 147-151)
- The α10 helix (residues 268-287) and the turn before it (residues 243-267)

These regions are suspected to form a 'lid' in the enzyme structure and thus control access to the active site. Earlier MD studies also revealed the increased fluctuations near the N and C termini [68]. The residue fluctuations in the enzyme are shown in Figure 4.6.

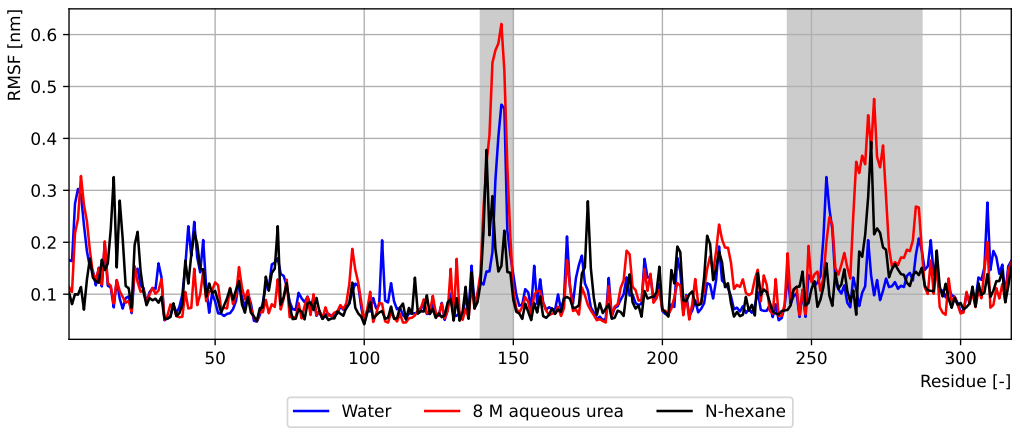
Based on our simulations of the reference solvents, the most mobile regions were indeed the α5 and α10 helices and their neighboring residues. The increase in fluctuations were visible in both regions when the solvent was changed from water to 8 M aqueous urea (Figure 4.6a). However, compared to the results of Monhemi et al. [68], the disruptive effect of urea was smaller in our simulations, possibly due to the lower temperature used. Comparison of the results of water and eutectic NADES showed that no new flexible region appeared and, in general, fluctuations were smaller in NADES (Figure 4.6b). Similarly, the molar ratio of NADES compounds had little effect. Although in the region of the α5 helix there was a small increase in the RMSF values with increasing molar ratio, the overall effect is negligible and still below the fluctuations measured in water (Figure 4.6c). In CCGLY systems the increasing HBD ratio had a similarly small effect with little fluctuation increase mostly in the α5 helix, but without clear trend (see Figure A.12). However, the fluctuations in all molar ratios were below the values of water.

Furthermore, the results of Monhemi et al. [68] and Nian et al. [110] are in line with our findings (regarding NADES of CCUR12 and CCGLY12). The only exception is in the 8 M aqueous urea, where the fluctuations in the flexible helix sections were about twice compared to our simulation (possibly due to the higher simulation temperature, i.e., 373 K) [68].

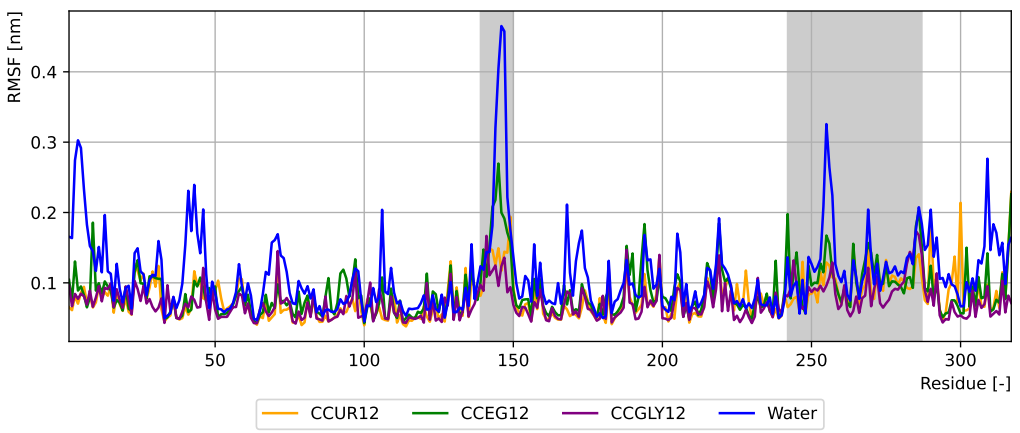
In summary, NADES show a dampening effect on the fluctuation of enzyme residues. The composition and molar ratio have some effect on the measured RMSF values, but the fluctuations in the flexible regions remained below the values measured in water.

4.6.3.3 Intramolecular hydrogen bonds in the enzyme

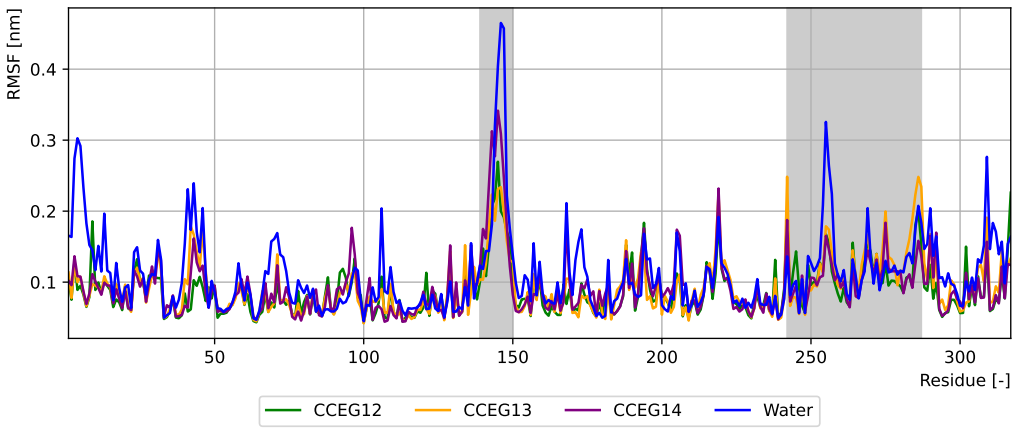
The enzyme's secondary and tertiary structure is stabilized by hydrogen bonds between the residues of the enzyme. Their breakdown would result in the deterioration of the enzyme structure. We therefore monitored the number of intra-main chain hydrogen bonds in the enzyme during our simulations. We expect that in systems where a HBD compound is


RMSF [nm]
0.6
0.5
0.4
0.3
0.2
0.1
50
100
150
200
250
300
Residue [-]
Water
8 M aqueous urea
N-hexane

(a)


RMSF [nm]
0.4
0.3
0.2
0.1
50
100
150
200
250
300
Residue [-]
CCUR12
CCEG12
CCGLY12
Water

(b)



(c)

Figure 4.6: Root mean square fluctuation (RMSF) of enzyme residues: (a) in reference solvents of water, 8 M aqueous urea and n-hexane, (b) in eutectic NADES of CCUR12, CCEG12, CCGLY12 and water and (c) in three CCEG systems with increasing HBD ratios of 1:2, 1:3 and 1:4. The grey area highlights the mobile region of the enzyme: the α5 helix and its neighboring region (residue 132-151) and the α10 helix and its neighboring region (residue 243-287).

abundant (like 8 M aqueous urea and NADES with increased HBD ratio) the number of intra-main chain hydrogen bonds will significantly decrease as new interactions are formed with the more potent HBD compound. Earlier results support these expectations, where in eutectic NADES hydrogen bonds in the enzyme remained stable, while in 8 M aqueous urea the number of hydrogen bonds between enzyme residues decreased rapidly [68].

Figure 4.7 shows the number of intramolecular hydrogen bonds in the enzyme as a function of time in the solvent for various compositions and molar ratios. Figure 4.7a shows that the number of hydrogen bonds in the enzyme reaches a stable value within our simulation in all reference solvents and eutectic NADES. The only unstable value, which kept decreasing, was in 8 M aqueous urea. However, the decrease during our simulation was still relatively small (around 10% of the initial hydrogen bonds in 200 ns) compared to earlier results (Monhemi et al. described a nearly 50% loss in 50 ns [68]). The enzyme preserved its hydrogen bonds best in eutectic NADES systems. In water and n-hexane the enzyme had a similar amount of intermolecular hydrogen bonds, which were slightly lower than those in eutectic NADES. However, this is an interesting result, taking into account the apolar nature of n-hexane. In the latter case, the apolar solvent is unable to form strong secondary interactions with the enzyme residues. Consequently, displacement by the solvent is not the cause of the lower number of intra-main chain hydrogen bonds in the enzyme. Figure 4.7b displays the differences in CCEG systems with increasing HBD ratio. As expected, an increased ratio of HBD results in a small decrease in the number of intra-main chain hydrogen bonds in the enzyme, but even with this decrease the value remains higher than in water. The differences between CCGLY systems are even less pronounced (see Figure A.13).

The qualitative results are also in line with earlier studies. In their simulation, Monhemi et al. measured the number of intra-main chain hydrogen bonds in CCUR12 and 8 M aqueous urea systems [68]. In CCUR12 this number decreased from the initial 125-145 range to 110-130 after 50 ns. In case of 8 M aquatic urea, the number decreased to 50-70 [68]. This decrease relates possibly to the higher simulation temperature (373 K) used. Nian et al. simulated CCGLY12 and betaine-xylitol systems at 333 K for 30 ns. In their simulation the number of hydrogen bonds between NADES components remained stable, which suggests no additional hydrogen bonding formed between the NADES and the enzyme [110]. Shehata et al. also found that the number of hydrogen bonding interactions between solvent and enzyme remained stable during their 300 ns simulation at 373 K (in CCUR12 system with various water content) [111].

Our findings indicate that NADES have a less disruptive effect on the enzyme's intra-main chain hydrogen bonding than water or n-hexane. The increase of HBD ratio had a small effect, where an excess in HBD is more likely to interact with the enzyme. However, the effect of increased HBD ratio does not surpass the disruptive effect of the simulated reference solvents.

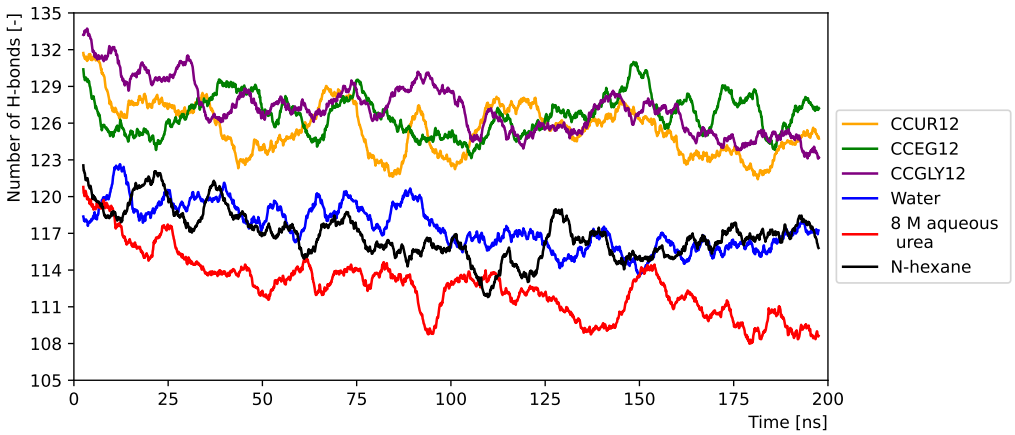
4.6.4 Changes in the enzyme structure

Based on the results so far, the general enzyme stability does not decrease significantly. The enzyme remains stable in every simulated NADES system in the timeframes studied, including the ones with increased HBD ratio. The RMSD, R_g, RMSF and the number of hydrogen bonds do not indicate the denaturation of the enzyme. However, it is possible that NADES affect the active site of the enzyme. In that case the enzyme keeps its stability, but its activity might change. It is also possible that strong secondary interactions of NADES with the enzyme surface help to stabilize its structure at higher temperatures and in longer reactions. (To study these possible effects, we investigated these specific interactions.)

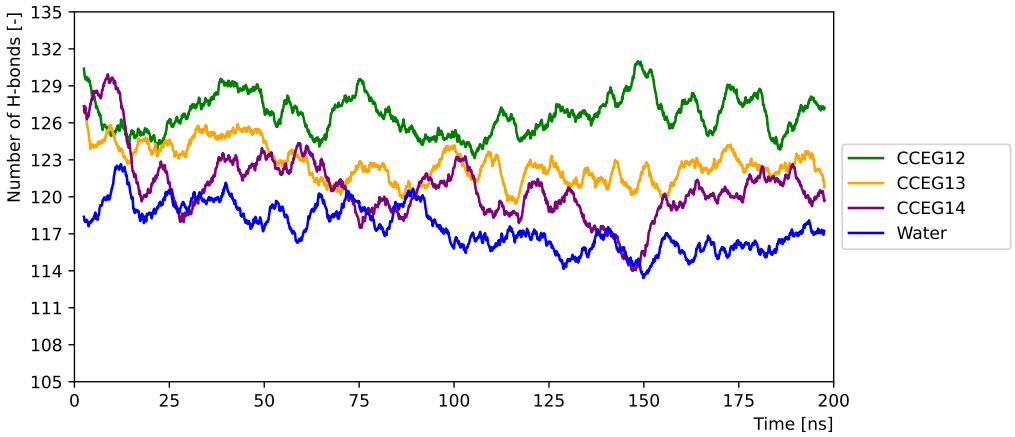
4.6.4.1 Stability of the catalytic triad

The enzyme catalytic site includes three residues: serine 105 (SER105), aspartate 187 (ASP187) and histidine 224 (HIS224), i.e. the catalytic triad. In the enzymatic reaction, the serine acts as a nucleophile and attacks the carbonyl carbon of the ester. The histidine promotes this attack by accepting a proton from serine. Strong interactions between SER105-HIS224 and ASP187-HIS224 are important to stabilize the transition state of the reaction [147]. If the distance between these sites increases due to interactions with the solvent system, they cannot promote the catalytic process and the enzyme loses its activity. We therefore calculated two distances during our simulations, i.e., between the OH group of aspartate and NH group of histidine, and the OH group of serine and NH group of histidine.

Based on earlier results, we expect that in the presence of a large amount of free HBD compound (either in systems with increased HBD ratio or in 8 M aqueous urea) the catalytic triad suffers structural changes. Figure 4.8 shows the distance between the two catalytic triad pairs as a function of time in solvent systems with different compositions and molar ratio.

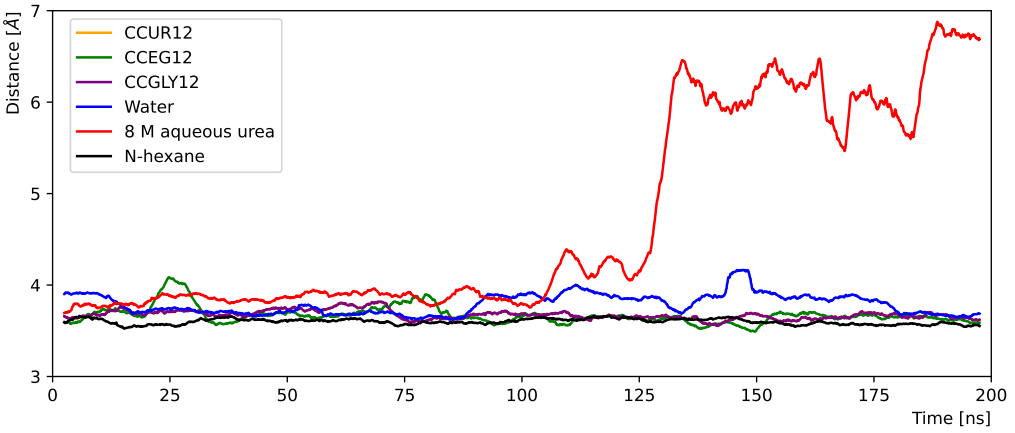


(a)

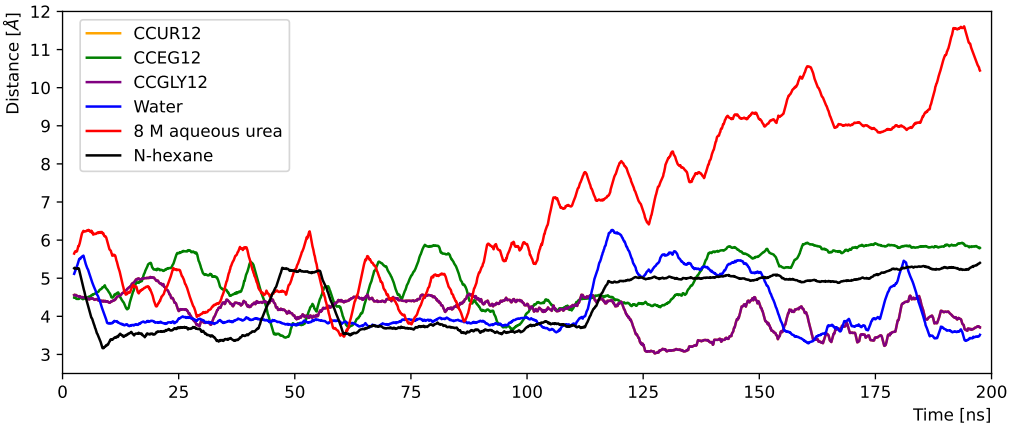


(b)

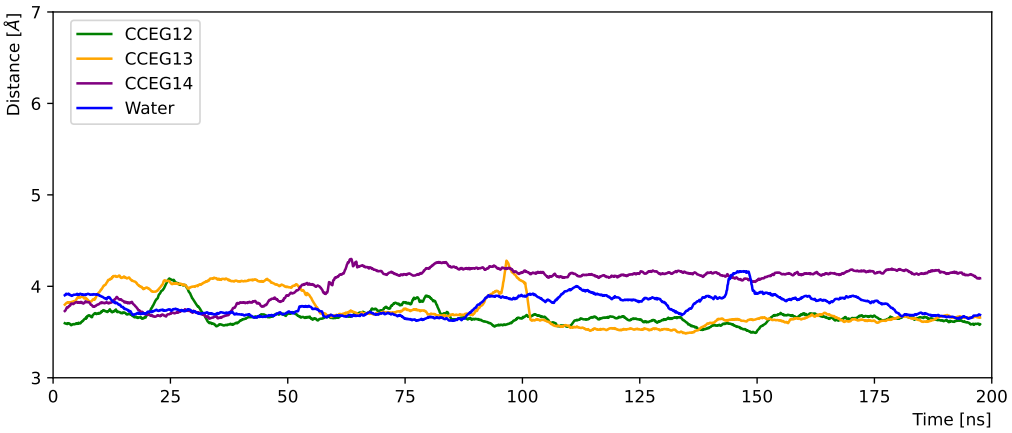
Figure 4.7: (a) Number of intramolecular hydrogen bonds in the enzyme as function of time in eutectic NADES and reference solvents and (b) number of intramolecular hydrogen bonds in the enzyme as function of time in CCEG with increasing HBD ratio.



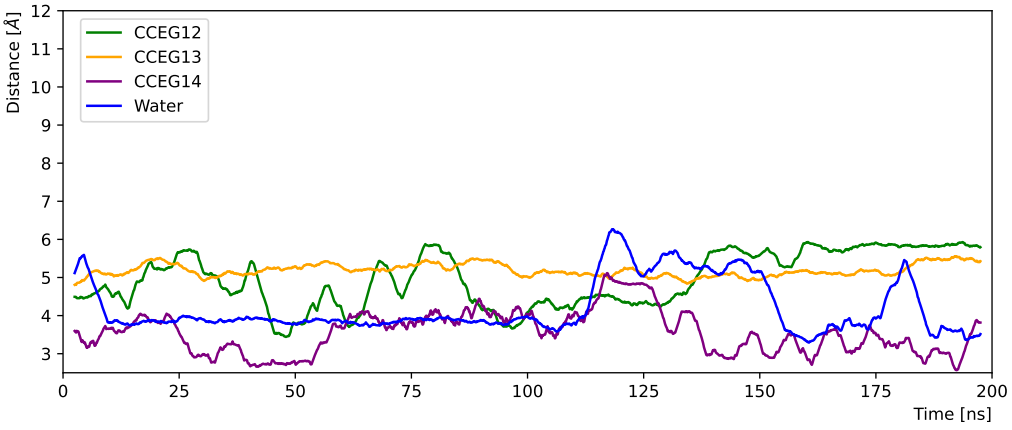
(a) ASP187-HIS224



(b) SER105-HIS224



(c) ASP187-HIS224



(d) SER105-HIS224

Figure 4.8: Distance between the catalytic triad residues in the enzyme as function of time in eutectic NADES and reference solvents: (a) OH group of ASP187 and NH group of HIS224 and (b) OH group of SER105 and NH group of HIS224. In CCEG with increasing HBD ratio: (c) OH group of ASP187 and NH group of HIS224 and (d) OH group of SER105 and NH group of HIS224.

The comparison of reference systems and eutectic NADES (Figure 4.8a,4.8b)) shows that, with the exception of 8 M aqueous urea, the distance remains stable between the residue pairs. In case of ASP187-HIS224, the distance was within the 3.5-4 Å range with around 0.5 Å fluctuation during the simulation. In case of SER105-HIS224, the distance was larger and showed much higher deviation, but still fluctuated around the initial value. The differences between solvents were higher in this case and varied between 3-6 Å. The fluctuation within a single solvent system was also greater, reaching up to 2 Å. 8 M aqueous urea was the only system where residue distances increased significantly and rapidly. Here, the distances between both residue pairs had a quick increase between 100 and 120 ns. This structural change most probably diminished the enzyme's ability to catalyze transesterification. On the other hand, the increase of HBD ratio in the NADES did not affect the distance between residue pairs neither in CCEG (Figure 4.8b) nor in CCGLY (Figure A.14)systems. The fluctuation and the mean value remained very close to the results for water.

Our results are in line with earlier findings of Monhemi et al. who measured 4.3 and 3.0 Å, for SER105-HIS224 and ASP187-HIS224 in CCUR12, respectively [68]. However, they measured 17.7 and 16.6 Å in 8 M aqueous urea, respectively, which are much greater values than we measured, i.e., in a shorter simulation time and at a lower temperature. This might also be related to the different force field and simulation setup they employed. Shehata et al. did simulations on a different, thermoalkalophilic enzyme, but with the same residues in the catalytic triad [111]. They measured slightly higher absolute values for the residue pair distances, but the trends were similar to our results: while in NADES the distances remained stable, in 8 M aqueous urea the distance reached a higher value. In addition, they studied NADES system with different water content, but similar to our results they did not find significant differences due to changes in the compound ratios.

The study of the catalytic triad revealed the first process clearly responsible for the loss of enzyme activity in 8 M aqueous urea. Although denaturation does not fully occur in the simulated time frame, the disruption of the catalytic triad is clear. However, a similar effect did not occur in NADES of any composition or molar ratio.

4.6.4.2 Stability of helices and surface residues

Apart from the catalytic triad, earlier studies identified two structural elements where the solvent-enzyme interactions can play a significant role. Specifically, Uppenberg et al. found that the α5 and α10 helices form a channel which provides access to the active site for external molecules [128]. Moreover, Monhemi et al. identified a number of enzyme residues (GLN23, THR267, GLU269, LYS271, LYS308, THR316, PRO317) which are situated on the surface of the enzyme and form strong secondary interactions with the NADES compounds [68]. This provides additional stability to the enzyme and also affects the distribution of the solvent molecules around the enzyme.

To investigate the effect of NADES on the channel formed by the helices, we calculated the distance between them (namely, between GLY142 and LEU278 as well as ALA146 and ALA287 residues which are part of these two helices) as a function of time. An increase of these distances indicates the opening of the 'lid', which increases the accessibility to the active site. Additionally, we determined the changes in the secondary structure of the enzyme

during the simulation. This way, we monitor if the changes in the distances between the helical residues are related only to the opening of the 'lid' or to a more fundamental change in these structural elements. Lastly, we calculated the SASA to explore possible correlation between the changes in the secondary structure and the accessibility of the enzyme by the solvent.

If the solvent has a disruptive effect on the helices, we would expect a decrease in the number of residues forming helices and an increased fluctuation in the pair distances. These two changes could also induce changes in the SASA.

Figure 4.9 shows the observed changes in the enzyme's secondary structure and SASA in the eutectic NADES and reference systems. The number of residues that participate in helices decreased in every simulated system (Figure 4.9a). This decrease was greater in hexane and 8 M aqueous urea, while in eutectic NADES it was similar to water. There is no clear trend between the different NADES compositions, just as there is also no clear trend between the different molar ratios (Figure A.15). The number of residues participating in β-sheet structures remained stable during the simulation in every system, although the fluctuation is slightly larger in hexane (Figure A.15b and A.15d). Again, there was no clear trend in NADES with different molar ratios (Figure A.15).

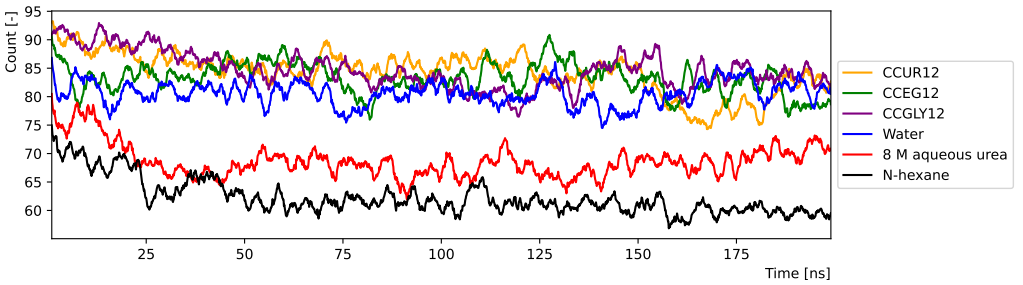
In Figure 4.9b the distance between the residues in α5 and α10 helices showed great fluctuation. The distance between the residues decreased in most of the systems and it was more significant in non-eutectic compositions (Figure A.16). However, in hexane and CCGLY12 the distance remained stable (Figure 4.9b). Again, 8 M aqueous urea showed the largest fluctuations.

The results of SASA (Figure 4.9c) did show a little increase (below 10%) in every system. While in CCUR12 and CCEG12 the SASA increases during the entire simulation, in other systems it reached a plateau before the end. Interestingly, while 8 M aqueous urea had the largest increase in SASA, the smallest measured value was found in hexane. There was no clear trend related to different molar ratios, but the differences were small (Figure A.17).

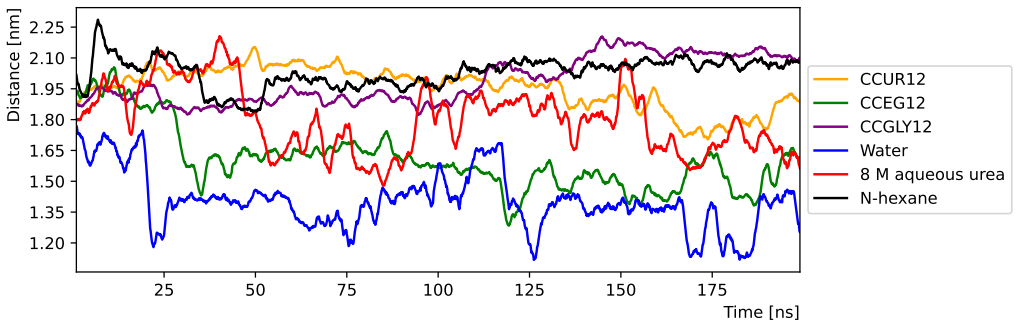
These results indicate that NADES affect the secondary structure of the enzyme and the channel-forming helices, which leads to a change in the SASA of the enzyme. However, these changes are also present in the reference systems and the effect of NADES is comparable with water. None of the applied compounds and the molar ratios in NADES showed a clear trend or significant differences in the measured values. Nevertheless, the free HBD compounds in the 8 M aqueous urea and also the apolar hexane had a more notable effect on these metrics.

Lastly, we investigated the possible interactions between the surface residues and NADES compounds. Figure 4.10 shows the RMSF values of earlier identified surface residues by Monhemi et al. [68]. In every NADES the fluctuations of surface residues were equal or below the values measured in water with the exception of PRO317. Neither the change of applied compounds nor molar ratios showed clear trends (cf. Figure 4.10 and A.18).

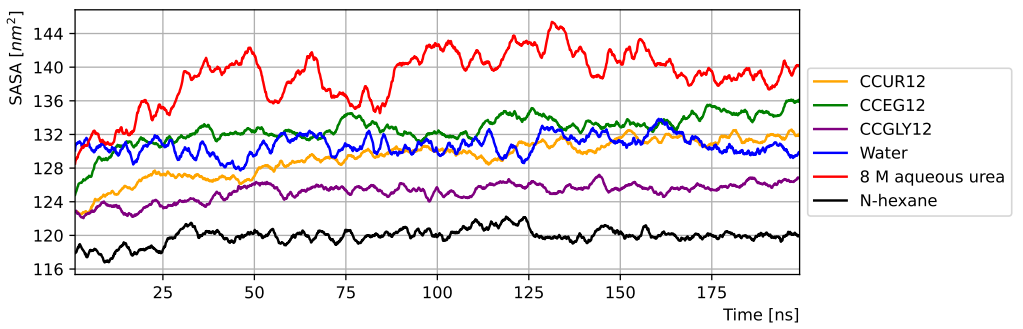
We also calculated the COM RDF between selected residues and the NADES compounds (cf. Figure A.19, A.20 and A.21). These measurements indicate that the hydrogen bonding interactions are also present in non-eutectic compositions. However, the increased amount



(a)



(b)



(c)

Figure 4.9: NADES effect on secondary structure and solvent accessibility: (a) number of residues in α-helix structures, (b) distance between α5 and α10 helices based on residues GLY142 and LEU278 and (c) solvent accessible surface area.

of HBD compounds has more intense and broader peaks in these cases. This indicates that with their increased concentration they have a larger role in enzyme-solvent interactions.

However, the COM RDFs of enzyme-HBD compound showed peaks around 1 nm distance which indicates that the HBD compounds can diffuse into the enzyme structure and the interaction is not limited to the surface residues (cf. Figure A.22a, A.22b and A.22c). This contradicts with earlier results of Monhemi et al. [68], who found peaks of urea at larger distances in CCUR systems. In contrast, Nian et al. [110] described similar diffusion of glycerol into the enzyme structure in CCGLY solvent. Interestingly, the distances of the HBD peaks did not change with the applied compounds, not even with 8 M aqueous urea, but they became shorter with the increase of the HBD ratio. This latter finding is probably related to the weaker interaction of the intermolecular hydrogen bonding network with the excess HBD molecules. This supports the concept that NADES do prevent the free diffusion of their compounds via the intermolecular hydrogen bonding network.

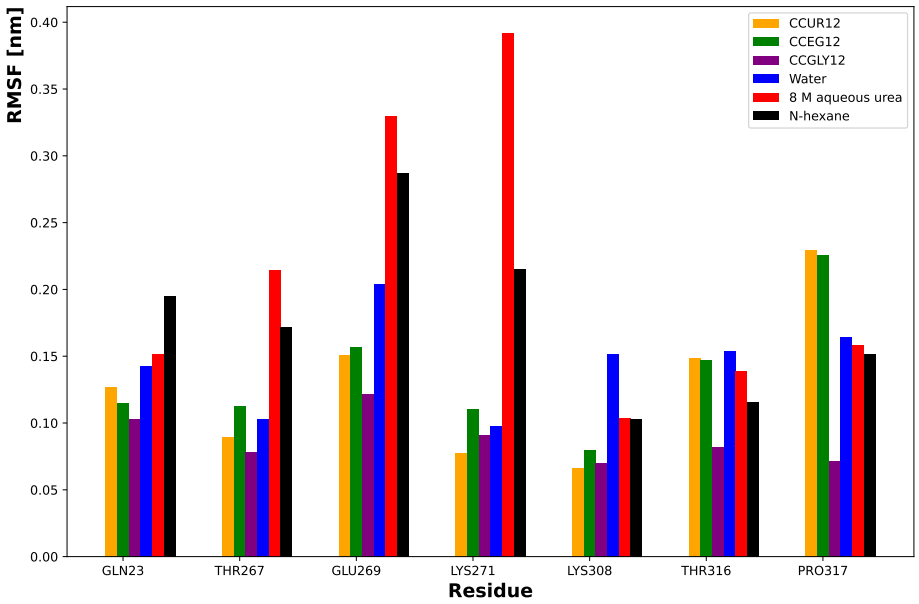
Overall, our results indicate that NADES have some disruptive effect on the secondary structure of the enzyme and this leads to a small increase in the solvent accessible surface. However, the magnitude of this effect is very similar in different compositions and molar ratios, thus the general behavior of NADES does not change with the molar ratio in the time span of our simulations. Moreover, our results confirm the presence of strong interactions between the surface residues and NADES compounds, which might be the cause of the experimentally described thermal stability of enzymes in NADES [113, 70]. However, we also observed the diffusion of the HBD compounds into the enzyme structure, therefore the interactions are not limited solely to the surface of the enzyme. Based on the enzyme-HBD RDF, the increase of the HBD compound ratio facilitated further diffusion of the HBD compounds into the enzyme. These two findings indicate that the intermolecular hydrogen bonding network prevents the diffusion of NADES compounds into the enzyme to a limited extent only.

4.7 Conclusions

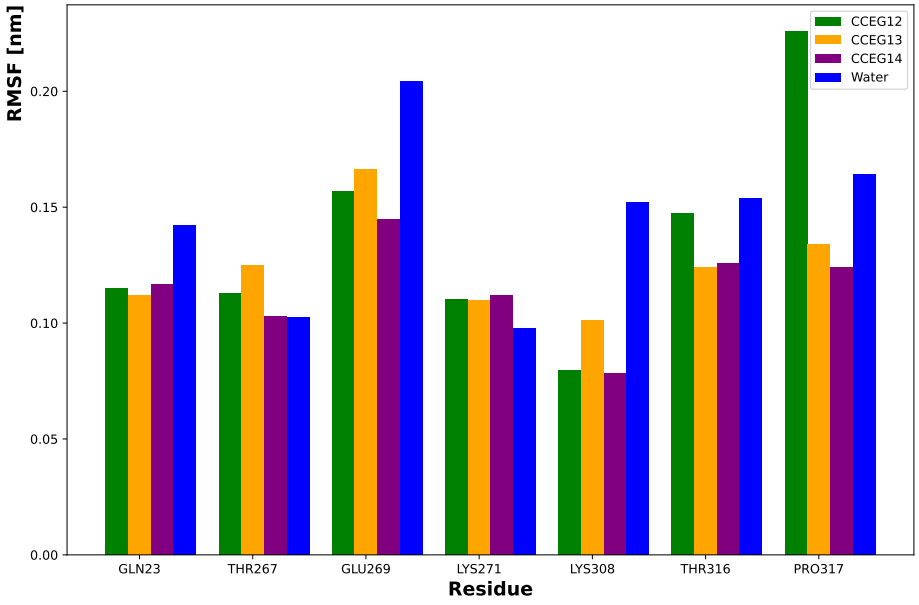
We identified a wide range of molar compositions of NADES where the microstructure of NADES did not change significantly. Our MD simulation showed that the NADES did not alter the stability nor the active site of the lipase enzyme. The composition range, where NADES preserve their characteristic behavior, while the physicochemical properties are slightly modified, provides an additional dimension to fine-tune their properties.

However, these results do not explain the changes in long term thermal stability and initial reaction activity of enzymes in NADES, observed in earlier experiments [109, 41, 113, 70]. Our current results suggest that the changes in reaction rate are related to solvent-substrate interactions rather than to enzyme-solvent interactions. To the best of our knowledge this the first study which discusses the behavior of NADES with non-eutectic compositions in biocatalysis.

Future research in this field are the experimental validation of simulation results on enzyme structure by Raman optical activity spectroscopy and the extension of simulation efforts towards the substrate-solvent interactions and solvation energies.



(a)



(b)

Figure 4.10: (a) RMSF of surface residues of the enzyme in NADES and reference systems. (b) RMSF of surface residues in CCEG system with increasing HBD ratio and in water as reference system.

4.8 Linking results to hypotheses

In this chapter, I studied the molecular dynamics simulation of the CALB enzyme in different NADES. I monitored the stability of the whole enzyme, the surface interactions between the enzyme and the NADES compounds, the structural changes of the active site of the enzyme, and the potential diffusion of the NADES compounds inside the three-dimensional enzyme structure. I have simulated here the same NADES as discussed in the experimental chapter (Chapter 3). The enzyme showed stability in all tested NADES systems. Surface stabilization was observed by hydrogen bonding between surface residues and NADES compounds. However, no changes in the active site were found, nor did the NADES compounds diffuse inside the enzyme. Therefore, it is unlikely that the effect of NADES on biocatalysis is related to changes in enzyme activity. In addition, these results forecast the long-term stability of the enzyme in NADES. Although the time frame of the simulation cannot prove the long-term stability, the short-term simulation results on the enzyme stability are consistent with the experimental results (see Chapter 3). Finally, the similar results of NADES with different molar compositions indicate that the physicochemical properties of NADES can be tuned over a wide range without negative effects on the enzyme.

In this chapter, I tested the hypothesis that the effect of NADES on enzymatic reactions can be related to substrate solubility/solvation, media viscosity, and changes in enzyme structure. More specifically, I investigated whether the NADES induce changes in the enzyme structure. In this context, I also investigated how molecular dynamics simulation can fit into the holistic multiscale model to describe the effect of NADES on enzymatic reactions. My results indicate that the NADES have a stabilizing effect on the enzyme. However, structural changes that would alter the initial activity of the enzyme were not observed. Based on the simulation and the experimental results, the activity change does not play a role in the suspected ternary effect of NADES on the enzyme (i.e. stabilization/denaturation, activity change and side reactions), at least in the studied reaction. Molecular dynamics simulation is a suitable modeling technique for describing the direct effect of NADES on the enzyme, provided that an appropriate force field and computational resources are available. However, the effect of NADES on the enzyme activity is small and not related to the high initial rates of the reaction measured experimentally.

Molecular dynamics simulation is a useful tool, but suitable force fields for NADES are scarce. This may change in the future as more work is done in this area. Otherwise, the meticulous task of force field optimization has to be included in the simulation work. This, together with the already high computational requirements, makes the cost-benefit ratio of the method questionable, especially in a screening process. In consideration, MD should be implemented in a multiscale model in a consecutive step, preceded by a screening step in which only selected systems are simulated. Alternatively, experimental validation could replace the modeling step. So far, the number of tested/simulated enzymes in NADES is limited, therefore it is possible that certain enzymes are more sensitive and NADES have a more significant effect on them. More diverse research is needed in this area in the future. In addition, due to computational costs, only free enzymes can be simulated, whereas in practice immobilized enzymes are usually used. As these are more resistant to structural changes, the results of the simulations cannot be directly applied. Finally, the fact that some NADES were not stable in reality, while the simulation showed no problems, emphasizes

the importance of experimental validation. In fact, validation should be extended to other physicochemical properties in future work.

Chapter 5

Modeling density, viscosity and solubility in NADES

"Modified to outperform
Acceleration the new norm"

Die Krupps

5.1 Preface

In the review of modeling methods in Chapter 2, I pointed out that there are many approaches to describe certain properties of NADES, but little effort has been made to integrate them into a holistic system. For this purpose, data-driven methods seem to be the most useful because of their general and predictive applicability (a model should give information about many new systems with few or no experimental measurements). Although the application of data-driven and machine-learning based methods to NADES is present in the literature, the lack of sufficiently large datasets has prevented their straightforward application. (The development of a model on the density values of 18 DES is not exactly data-driven [97].) In recent years, however, the amount of information on these systems seems to have reached a breakthrough threshold, even if it is a highly fragmented and methodologically incoherent pile of data.

This chapter discusses two of the central questions of the thesis, how to predict the behavior of NADES in a structured way and how to minimize the necessary empirical data. I hypothesize that the complex behavior of NADES can be described by a multiscale model. For the bulk properties of the media, data-driven, machine learning-based methods are a good fit. In the previous chapters, I concluded that viscosity and solubility play a significant role in the behavior of NADES, so I aim to predict these properties here. I hypothesize that sufficient data are already available in the literature. This will be tested by analyzing the learning curves and error distribution of the final models.

In this chapter, I discuss the results of combining machine learning methods with group contribution to develop two models for predicting the density and viscosity of DES systems

with different structures, molar ratios, and temperatures, based on general descriptive structural parameters, namely the number of functional groups in the compounds. The data for building these models are collected from the literature, which will limit the model capabilities, but in the range of structural variety described, the models can be applied to novel systems. Through analysis of the data set, I discuss these limitations. Commonly used machine learning algorithms are compared and fine-tuned to find the best candidates. The data requirements of the different models and their limitations are discussed by analyzing the learning curve and the errors of the final models. This chapter provides practical models of density and viscosity of DES, but the detailed analysis of the model also explains the current limitations and future potential of this modeling approach.

Based on the contents of this chapter, a manuscript entitled "Combining group contribution with machine learning to predict DES properties" has been written. The co-authors of this manuscript are Berkay Ayan, Philippe Nimmegeers, Pieter Billen. Berkay Ayan composed the initial data set and the initial density model. Philippe Nimmegeers and Pieter Billen contributed to the supervision of the project and to the review and editing of the manuscript. In addition, Pieter Billen acquired the financial support for the project.

5.2 Density prediction

5.2.1 Introduction

Deep eutectic solvents (DES) are a novel group of solvents with great potential to replace volatile organic compounds in many chemical applications. DES are mixtures of two or more compounds that exhibit a large decrease in melting point due to strong secondary interactions between the compounds [1]. This decrease results in a liquid mixture at or near room temperature, which facilitates their use in place of conventional organic solvents. Compared to the latter, the advantages of DES are non-volatility, non-flammability, very low toxicity and good biodegradability (sometimes even biocompatibility) [14]. DES are generally formed by hydrogen bond accepting (HBA, usually quaternary ammonium or phosphonium salts) and hydrogen bond donating (HBD, e.g. amines, carboxylic acids or polyols) compounds [5]. Based on their composition, Smith et al. classified DES into four types [1]. Since then, however, new types of DES have been described [148]. In addition, the discovery of natural deep eutectic solvents (NADES), which are formed by primary metabolites, added even more variety to DES compositions [13].

The behavior of the mixture is primarily determined by the complex intermolecular hydrogen bonding network between the compounds. Furthermore, this intermolecular network is determined by the compounds applied, making DES designer solvents. This means that the properties of DES can be tailored or fine-tuned by the variety of compounds and their ratios to meet the requirements of a particular application. In addition, the variety of compounds available allows for optimization across a wide range of physicochemical properties. These properties make DES promising candidates for many applications. Although DES are a novel class of solvents, practical applications already exist in electrochemistry [149], separation processes [150], biodiesel production [151], and biochemistry [152], to name a few.

As mentioned earlier, DES are designer solvents, meaning that their properties can be tailored to specific applications by changing their composition. However, the large number of possible compounds and possible combinations makes the development of DES applications by trial and error or experimental screening impractical. Therefore, modeling the relevant physicochemical properties of NADES is important, not only to better understand their behavior and structure-property relationships, but also to predict the properties of novel DES systems.

In our previous work, we gave a comprehensive overview of the available modeling methodologies of DES [11]. Here we focus on group contribution methods and machine learning approaches for predicting the densities of DES. The general advantage of these approaches is that they do not require much computing power and are therefore much faster than e.g. molecular dynamics simulations. In addition, prior experimental information is not always required for this type of modeling. This allows for high-throughput screening prior to the start of time-consuming and costly experiments.

The evolution of the density modeling approach based on group contribution and machine learning methods can be summarized in three major steps. The first works are based on the corresponding state principle and the Lydersen-Joback-Reid group contribution method [153, 154, 99] to calculate the critical properties of the pure compounds in the NADES. Then Lee-Kesler mixing rules are used to determine the critical properties of the DES mixture [155]. Finally, the density of the system is calculated using the Rackett equation [156]. Shahbaz et al. have made the most significant contribution to this approach [96, 53, 97, 98].

The next step was the application of genetic algorithms [157] to optimize the density estimates. This was done either directly on the output variable (property of interest) or by parameter optimization in a semi-empirical equation of the property. On this topic, Haghbakhsh et al. published a series of papers [158, 159, 160]. The recent publications shift from the optimization of existing equation-of-state models to machine learning based models and the use of more complex independent variables, such as functional group contribution parameters. This includes both the extension of existing group contribution methods with additional parameters, as well as changes to purely chemical structural descriptors taken from other quantitative structure-property relationship solutions. However, the change is gradual, earlier applied genetic algorithms are still present [161], while multiple linear regression [162], neural networks [96] and other more complex algorithms are also described [163].

Shahbaz et al. reported the first implementation of the combination of the Lydersen-Joback-Reid (LJR) group contribution method with the Lee-Kesler (LK) mixing rules and the Rackett equation [96]. They modeled 9 DES in the temperature range of 298.15-368.1 K to describe the temperature dependence of the density and achieved an average relative percentage error (ARPE) of 1.9 %. In their model, the precision was dependent on the DES compounds used and the error increased with the absolute value of the density. Nevertheless, they achieved an average absolute relative percentage error (ARPE) of 1.9 %. In their next work, they considered the combination of LJR and LK methods with the Othmer equation to predict the surface tension of DES [97]. On the data set of nine DES in this study, the Othmer method gave an average absolute deviation (AARD) of 2.57 %. An alternative approach was the use of Wilderman and Grippen's atomic contribution method for calculating molar refractive index in combination with the Lorentz-Lorenz equation [98]. Thus, they

established a relationship between molar refraction, refractive index and density to determine the latter. In their work, they included 24 DES based on ammonium and phosphonium salts. The prediction of the refractive index gave 0.56 % ARPA and 0.9822 coefficient of determination. The prediction of the density gave 1.43 % and 0.9768 coefficient of determination, but the method systematically underestimated the density.

Similarly, Mjalli et al. used the modified Rackett equation, optimized for ionic liquids, to predict the molar volume as a function of composition and temperature [164]. They introduced the optimization by genetic algorithm, which they applied to the compressibility factor in the Rackett equation, using the molar volumes as the target parameter. The optimization was performed on a set of 86 points from 12 DES, and they tested their results on a set of 91 points from 6 DES. Compared to the LJR-LK method, their approach reduced the average relative deviation (ARD) from 0.625 % to 0.211 %. However, they found that the assumption of the Rackett equation (as the reduced temperature tends to zero, $ln(V/V_C)$ equals the compressibility factor Z_R) is not fulfilled in DES. Furthermore, the temperature-molar volume relationship was linear, indicating that DES behaves as a single compound.

In a different approach, Hou et al. combined the LJR group contribution method with the bonding group interaction contribution method [165] to account for the effect of strong intermolecular interactions typical of DES [161]. They used two data sets, one with 3648 points on 645 binary DES and one with 174 points on 36 ternary DES. The model achieved an average absolute relative deviation (AARD) of 1.56 % on the binary DES test set and 2.29 % on the ternary set.

Another approach to optimize the prediction of densities has been the application of genetic algorithms [157]. In their first work, Haghbakhsh et al. combined the LJR-LK method with a genetic algorithm to estimate the density of DES as a function of temperature and composition [158]. They used the critical properties and the accentric factor from the group contribution method as input variables, and the density was the output value of the genetic algorithm. With this approach, they obtained 3.87 % AARD on a data set of 1239 data points and 149 DES. In their subsequent work, they used a similar approach for predicting surface tension, where they added 6 customizable parameters to the surface tension model of Escobedo and Mansoori and optimized the value of these parameters by genetic algorithm [159]. This model achieved 13.9 % AARD based on a data set of 553 points from 112 different DES. In their later work, Haghbakhsh et al. used the genetic algorithm to optimize the group weights in the group contribution and atomistic contribution methods for calculating the density, refractive index, heat capacity, sound velocity, and surface tension of various DES [160]. The size of the database varied between 398 and 1239 data points depending on the property, and the AARD of the different models varied between 0.37 and 9.33 %. Compared to the general models, both the group contribution and the atomistic models outperformed all previous models. In general, the group contribution model was more accurate, while the atomistic model was easier to compose due to the smaller number of parameters. The authors believe that predictions based on group contributions are better because they embed the real chemical structure of the compounds.

More recently, interest has shifted to solutions based on machine learning. The potential of this approach is well demonstrated in the work of Afzal et al. [166]. They trained a deep neural network based on a series of MD simulations of small organic molecules to determine

their density. In their work, they ran 10^5 MD simulations of small organic molecules and trained a deep neural network on them to predict their density. They applied the trained model to 1.5 million additional generated data points, demonstrating the applicability of the model for hyperscreening. They also analyzed the efficiency of the neural network by analyzing the learning curve of the algorithm and determined the data requirements of the task.

Shahbaz et al. were the first to apply a machine learning approach to DES [53]. They used a feed-forward back-propagation artificial neural network (ANN) to predict the density of DES with the molar fractions of the applied compounds and temperature as input parameters. The data set contained 270 instances from the combination of 18 DES and 15 temperatures. ANN yielded a coefficient of determination of 0.9995, while the LJR-LS Rackett method yielded 0.9748 on the same dataset. In addition, the ANN was more accurate at higher temperatures, but the accuracy depended on the groups used. Similarly, Adeyemi et al. compared the accuracy of the LJR-LK Rackett method to ANN in predicting density and conductivity [101]. Their data set contained 105 data points of 3 amine-based compounds with 3 molar ratios in the temperature range of 293.15-353.15 K. They combined feed-forward back-propagation multilayer ANN with bagging [1] to achieve higher accuracy, and their best model reached $2.799 \cdot 10^{-7}$ normalized mean square error. In addition, the method proved to be more accurate at higher temperatures than the group contribution method.

In comparison, Halder et al. applied a more general approach, using structural descriptors from Dragon software as input variables with a sequential forward selection based multiple linear model [162]. As a significant novelty, they discarded the random train/test split method and composed the two data sets based on omitting certain compounds or mixtures from the training set. This prevents the model from being trained on all possible compounds in the entire dataset, making the evaluation of predictive ability more realistic. They created several individual models and then combined them using consensus learning. The best single model yielded 2.589 % AARD. Similarly, Khajeh et al. implemented multiple linear regression but added a neuro-fuzzy interface system to predict DES density [163]. Their density dataset contained 2005 points and they generated more than 1500 molecular descriptors in Dragon software from which they selected features by modified particle swarm optimization. The final model gave 0.993 coefficient of determination and 1.56 % AARD.

To date, the combination of group contribution parameters with machine learning methods has not been demonstrated. However, the use of group contribution parameters as input variables would provide a more theoretical basis for the model than the often very abstract structural descriptors described in previous publications [163, 162]. To prove the feasibility of such an approach, three aspects need to be investigated:

- The functional groups as input parameters provide sufficient flexibility to establish a large domain of application.
- Sufficient data is available to train the machine learning model.
- The developed method is at least comparable in accuracy to other approaches.

[1] bootstrap aggregation: an ensemble learning method commonly used to reduce variance within a noisy dataset by randomly selecting samples with replacement into training sets and aggregating the results of trained models

To investigate these aspects, we built a database based on the available experimental data on binary DES densities in the literature. We developed two machine learning based approaches for density prediction. The input variables were generated from the number of functional groups in the two DES compounds and the dataset was divided into a training and a test set. After initial testing, two models were selected, random forest and gradient boosting regression ensembles. The accuracy of the developed models was compared with previously published density prediction models. The application domain was investigated by statistical analysis of prediction errors. The sufficiency of the data set was investigated by analyzing the learning curve of our models by varying the train/test ratio.

5.2.2 Machine learning workflow

Model Development Workflow: The model development workflow is summarized in Figure 5.1. The work is divided into four main steps: The first step is to assemble the data set by collecting experimental data, generating the input variables, and cleaning the data set of missing data points and outliers. (In machine learning projects, generating and selecting the right input variables is often a central step. In our case, however, we decided in advance to use the functional groups, so there was no need for extensive feature engineering and selection). The second step is to explore preliminary machine learning models to find good candidates that adequately describe the underlying relationship between the input and output variables. The third step is to optimize the promising candidate models by tuning the hyperparameters of the models to achieve the highest accuracy without overfitting the model. The last step is to test the model on the test set and further analyze the model behavior.

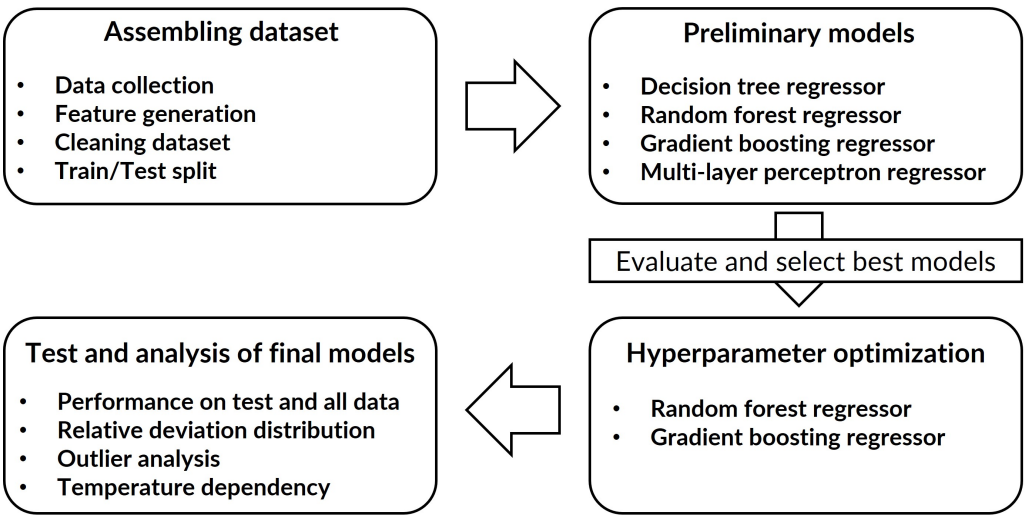


Figure 5.1: Flowchart of the machine learning workflow.

Assets of the model development: The code of the model is written in Python 3 language. Pandas package is used for the database. Machine learning models are from scikit-learn package. For calculation of chemical properties and functional groups RDKit is used. Chemical Identifier Resolver is used to calculate the SMILES of the chemicals.

Composing the dataset: To compile the initial dataset, we reviewed the previously reported experimental measurements of DES densities [160]. In our dataset, we included the HBA and HBD compounds, their ratio, temperature, and measured density. The dataset contains measurements only at atmospheric pressure, but most of the NADES have been measured in a wide range of temperatures.

To calculate the input variables (aka features), we first used the Chemical Identifier Resolver to calculate the SMILES code of the chemicals from their names. Then we used RDKit to calculate the molecular weight, the number of hydrogen bond accepting and donating sites in both HBD and HBA compounds, and the number of different functional groups in the molecules. The functional groups included are listed in Figure 5.2. After calculating

the number of functional groups, the missing functional groups as descriptors were removed from the data set. We also removed the outlier data points that had a standard score of $z \geq 3.0$.

Non-aromatic groups	Aromatic groups
CH3-	CH3-
-OH	-OH
-COOH	-COOH
-NH2	-NH2
-CHO	-CHO
-COO- (ester group)	-COO- (ester group)
-COHN- (amide group)	-COHN- (amide group)
-O- (ether group)	-O- (ether group)

Figure 5.2: Functional groups used as input variables. Functional groups present in aromatic and non-aromatic structures are differentiated.

After cleaning the data set, it was divided into a training and a test set. Due to the small number of instances, the training ratio was 0.9. The instances were randomly distributed. In this way, certain compounds may be present in both the training and test sets. This can lead to an overestimation of the accuracy of the final model, as the model learns about every possible instance (i.e. data snooping). The small number of included DES makes random selection necessary, but using the functional groups as input variables instead of descriptors as the whole structure mitigates this problem.

Evaluating Preliminary Models: Model development was started by selecting a few candidate models and evaluating their initial performance prior to any optimization. For this evaluation, 10-fold cross-validation was used. In this setup, the training data set was randomly divided into 10 equal-sized groups, the model was trained on 9 of them, and then evaluated by comparing the predictions and experimental densities in the last group. This process was then repeated for all groups. Based on this, the mean and standard deviation of the metrics were calculated. The following metrics were used for the evaluation:

- **Root Mean Squared Error (RMSE):** The RMSE is the aggregated measure of the error in the predictions of all data points, and thus a good measure of accuracy to compare the predictive power of different models for a given data set. The equation 5.1 shows the calculation of the RMSE:

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}\left(y_i - p_i\right)^2} \tag{5.1}$$

  where y_i is the experimental density of the ith instance, p_i is the predicted density of the ith instance and N is the number of instances.

- **Average Absolute Relative Deviation (AARD):** AARD is also a measure of the prediction accuracy of a model, but because it expresses the relative error, different

results from different data sets are comparable and thus give a better sense of the magnitude of the error. The calculation of AARD is described in the equation 5.2:

$$AARD = \frac{100\%}{N}\Sigma_{i=1}^{N}\left|\frac{y_i - p_i}{y_i}\right| \tag{5.2}$$

where y_i is the experimental density of the ith instance, p_i is the predicted density of the ith instance and N is the number of instances.

- **Coefficient of Determination (R^2):** R^2 measures how much of the variance of the target is explained by the predictor variable. It is a dimensionless property, so performance on different data sets is comparable. The equation 5.3 describes the calculation of R^2:

$$R^2 = 1 - \frac{\Sigma_{i=1}^{N}(y_i - p_i)^2}{\Sigma_{i=1}^{N}(y_i - \hat{y})^2} \tag{5.3}$$

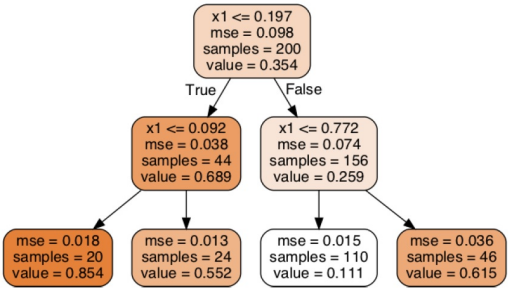
where y_i is the experimental density of the ith instance, p_i is the predicted density of the ith instance, $\hat{y}$ is the average value of experimental density and N is the number of instances.

Four commonly used machine learning models were selected for preliminary modeling:

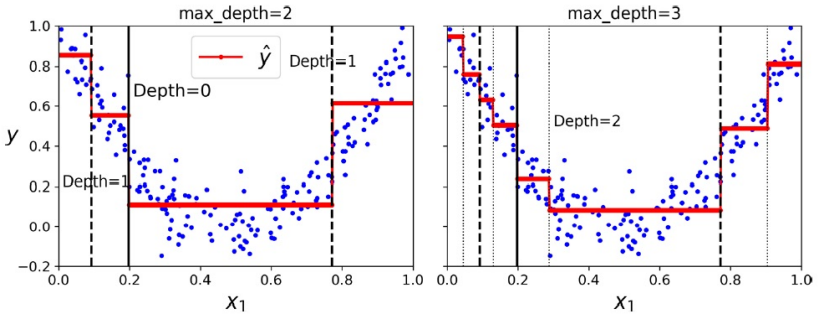
- **Decision Tree Regressor:** Decision tree is a nonparametric supervised learning method in which the model predicts the value of a target variable by learning simple decision rules inferred from data features. (See a representation of decision tree regressors and their predictions in Figure 5.3.) Decision trees are easy to interpret and visualize, require little data preparation, and have low computational cost. Disadvantages of decision trees are that they tend to overfit, so regularization (called pruning in this case) is often necessary. In addition, since the predictions are not smooth in the case of regression, they tend to extrapolate poorly.

- **Random Forest Regressor:** Random forest is an ensemble method based on decision trees. In ensemble methods, several weak learners (e.g., individual decision trees) are combined to jointly describe a complex problem. The group of learners is called an ensemble. By aggregating the predictions of a group of predictors, the combined result is often better than the best individual predictor. The forest is random because each tree is built on a random sample of the test set and may use a random subset of the available features. This randomness reduces the variance of the forest estimator, whereas individual decision trees typically have high variance and tend to overfit. Random forests reduce variance by combining different trees, which in practice often results in a significantly better model. This is a significant advantage, even at the cost of significantly increased computational complexity.

- **Gradient Boosting Regressor:** Gradient boosting regressor is a boosting ensemble method based on fitting the residual errors made by the individual predictors. A boosting method is an ensemble method where the learners are trained sequentially,



(a) A decision tree for regression



(b) Predictions of two Decision Tree regression models

Figure 5.3: Illustrating the working principle of decision tree regressor from the work of Geron [167].

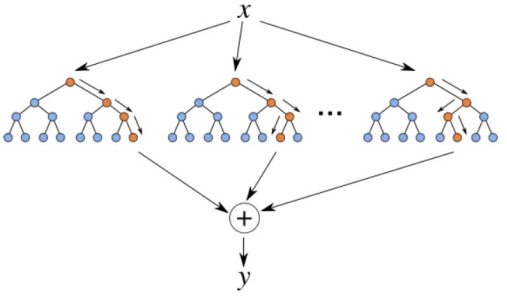


Figure 5.4: Illustrating the working principle of a random forest. Each tree in the random forest ensemble is built from a sample drawn from the training set. Each tree makes its own prediction then the average of those prediction are taken.

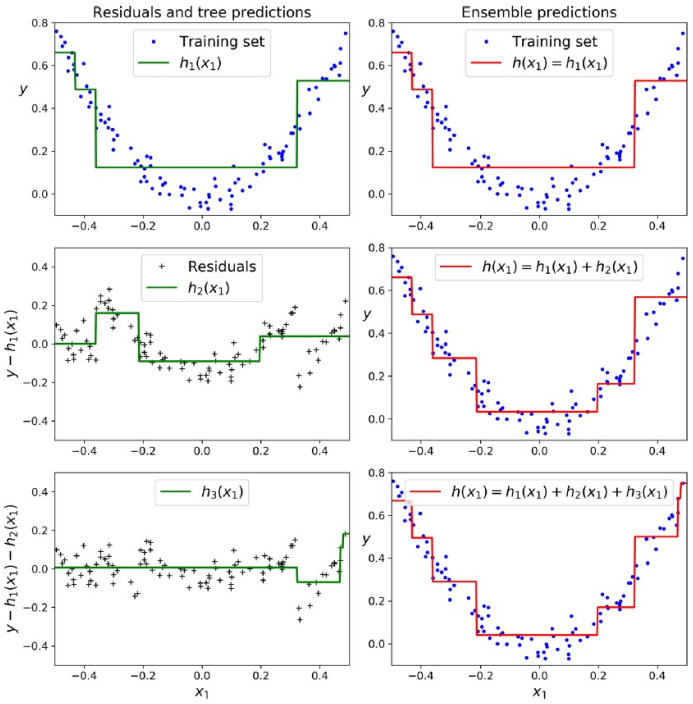


Figure 5.5: Illustrating the working principle of gradient boosting from the work of Geron [167]. The first predictor (top left) is trained normally, then each consecutive predictor (middle left and lower left) is trained on the previous predictor's residuals; the right column shows the resulting ensemble's predictions.

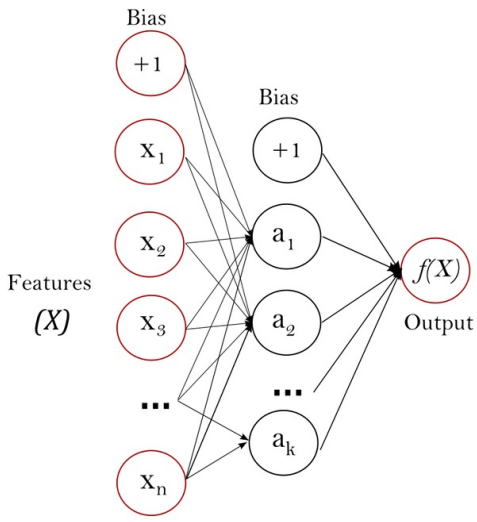


Figure 5.6: Illustrating the working principle of multi-layer perceptron with one hidden layer.

and each tries to correct the previous iteration. In gradient boosting, the next learner tries to minimize the residual error of the previous predictor (see Figure 5.5).

- **Multilayer Perceptron (MLP) Regressor:** Multilayer perceptron is an artificial neural network supervised learning algorithm that can learn non-linear functions for both classification and regression. Between the input and output layers, which contain the independent and dependent variables, there may be one or more nonlinear layers called hidden layers. The leftmost layer, called the input layer, consists of a set of neurons $\{x_i|x_1, x_2, ..., x_m\}$ representing the input features. Each neuron in the hidden layer transforms the values from the previous layer with a weighted linear summation $w_1x_1 + w_2x_2 + ... + w_mx_m$, followed by a nonlinear activation function $g(\cdot) : R \rightarrow R$ - like the hyperbolic tangent function. The output layer receives the values from the last hidden layer and transforms them into output values. Multilayer perceptron regressors are capable of learning non-linear relationships. Drawbacks are that MLPs have a non-convex loss function and multiple local minima, they are sensitive to feature scaling, and the large number of hyperparameters to scale. Therefore, both model training and hyperparameter tuning are meticulous tasks.

Based on the evaluation metrics of the preliminary models, random forest regression and gradient boosting regression performed best and were selected for optimization.

Optimizing the hyperparameters of the models: Hyperparameters are the parameters of the model that are not directly fitted in the learning process. However, it is possible to search for optimized hyperparameters by cross-validation on the training data. Once the candidate model is selected, the parameter space, search method, cross-validation scheme, and score function must be chosen. We used an exhaustive grid search with 10-fold cross-validation, scored by the value of the coefficient of determination. That is, each combination of the proposed hyperparameters was applied and the mean and standard deviation of the coefficient of determination were calculated based on 10-fold cross validation. Finally, the

best combination was selected. These parameters were selected in several iterative steps, always in the neighborhood of the best parameters of the previous step, to find a model where all the selected hyperparameters are intermediate values. In the case of random forest regression, the number of estimators (individual trees), the maximum depth of each tree, the minimum sample on a branch before splitting, and the minimum number of samples per leaf were optimized in the ranges, 150-200, 12-18, 3-7, and 2-5, respectively. For gradient boosting regression, the number of estimators, the maximum depth of each tree, the learning rate, and the ratio of subsamples were optimized in the ranges 1-3, 100-300, 0.05-0.15, and 0.5-1.0, respectively.

Testing the optimized models and analyzing the predictions: After hyperparameter optimization, the models were evaluated on the test dataset. The evaluation is based on the same metrics (RMSE, AARD and R^2) as the model development. The predictions of the test set and the full dataset were compared using the relative deviation distribution to check for potential differences between the two sets. The relative deviation distribution and residue analysis were used to determine if certain functional groups were more prevalent in the outliers. The latter includes analysis of the application domain. Finally, the accurate incorporation of the temperature dependence of density in DES with the same compositions but at different temperatures was also checked by visual analysis.

5.2.3 Results and discussion

5.2.3.1 Dataset

The first thing is the preliminary analysis of the composed dataset. The data set has been assembled from the experimental measurements available in the literature. First, the histogram of the density values is drawn to see the distribution of the density and to check for possible outliers. (See Figure 5.7.) This plot also shows the estimated normal distribution based on the mean and standard deviation of the data set. The density of the data set does not follow a perfect normal distribution, there are three main peaks around the mean. However, the machine learning models applied are not linear regression models, and the normal distribution of the target variable is not required. In addition, the data set has a few high-density outliers. The small number of them does not allow a good generalization of the model, and thus the trained model is not likely to predict well in this range. Therefore, it was decided to remove these outliers from the dataset. This was done by calculating the z-score of the density data and removing the absolute values greater than 3.0. This removes about 0.3 % (5 data points) of the data, reducing the maximum value from 1.63 g/cm^2 to 1.35 g/cm^3 (for comparison, the third quartile is 1.21 g/cm^3) (see Table 5.1). The density distribution of the cleaned dataset is shown in Figure 5.7.

As mentioned earlier, instead of extensive feature engineering, functional group contributions were chosen to generate the input variables. The considered features are summarized in Table 5.2, together with their count in the database. A commonly used grouping of molecular structures is used, with a distinction made between functional groups present in chain or aromatic ring structures, and whether they are present in a HBA or HBD compound.

In addition to outliers, missing values in functional groups are also taken into account.

Table 5.1: Statistics of the original and trimmed dataset. Trimming the 5 outliers did not significantly change the distribution.

	Original dataset		Trimmed dataset	
	Density [g/cm^3]	Temperature [°C]	Density [g/cm^3]	Temperature [°C]
N	1426		1421	
Mean	1.1484	48.85	1.1472	48.94
Std.	0.0896	20.01	0.0875	19.99
Min	0.8990	10.00	0.8990	10.00
25 %	1.0914	30.00	1.0910	30.00
50 %	1.1517	45.00	1.1512	45.00
75 %	1.2184	65.00	1.2173	65.00
Max	1.6300	100.00	1.3500	100.00

Since the data were collected from the literature, it is not guaranteed that all possible functional groups are adequately represented. Table 5.2 counts the occurrences of different functional groups in the data set. While all functional groups in non-aromatic structures are represented (with hydroxyl and methyl functions being the most common and ketones the least common), many functionalities in aromatic groups are missing. This means that the model cannot be trained to predict the properties of these groups and they should be excluded from the model. This reduces the number of structural features to 30.

The last consideration is temperature. It is important to include the correct temperature dependence of the density in the model, as this is required for later use in the more complex prediction of solubilities. Again, the method presented here is based on the experimentally available systems. This covers the range of 20-100 °C. However, this is not uniformly available for every DES, as melting and decomposition temperatures vary from compound to compound. Since machine learning models, especially the tree-based models used here, perform poorly in extrapolation, good results are expected only in the range where experimental data are available.

With this preliminary analysis of the data set, the following expectations for the models

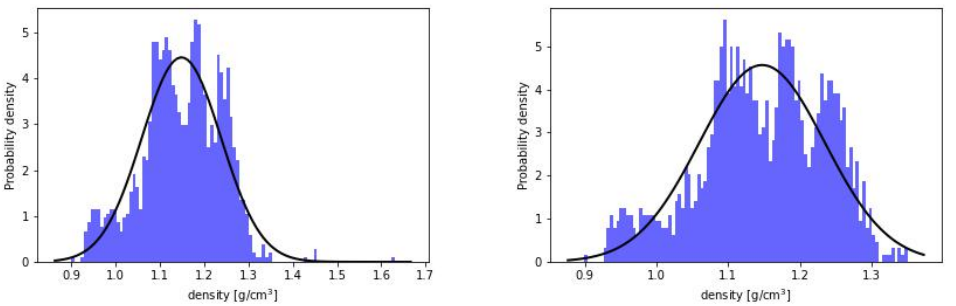


Figure 5.7: a) Distribution of the original density dataset. b) Distribution of the density dataset after removing the $z \geq 3.0$ outliers.

can be made: The model should perform well in the density range of 0.89-1.35 g/cm^3 and temperature range of 20-100 °C on DES containing the available functional groups listed in Table 5.2.

Table 5.2: Included functional groups and their count in the dataset. Red highlighting indicates missing functional groups. While non-aromatic groups are well represented, most aromatic structures are poorly described, except for simple rings.

Non-aromatic groups			
HBA		HBD	
Group	Count	Group	Count
CH3-	1373	CH3-	299
-OH	899	-OH	2573
-COOH	69	-COOH	397
-NH2	165	-NH2	111
HOC-	8	HOC-	158
-COO- (ester group)	740	-COO-	15
-COHN- (amide group)	16	-COHN-	12
-O- (ether group)	444	-O-	333
Aromatic groups			
HBA		HBD	
Group	Count	Group	Count
CH3-	2	CH3-	29
-OH	5	-OH	129
-COOH	**0**	**-COOH**	**0**
-NH2	**0**	**-NH2**	**0**
HOC-	**0**	**HOC-**	**0**
-COO- (ester group)	**0**	**-COO-**	**0**
-COHN- (amide group)	**0**	**-COHN-**	**0**
-O- (ether group)	**0**	-O-	105

5.2.3.2 Preliminary models and model optimization

To develop an optimal density prediction model, preliminary models are tested. Four regression models are selected: decision tree, random forest, gradient boosting, and multilayer perceptron. Models of increasing complexity are chosen because it is not known in advance how complex the underlying problem is. In this way, an optimal model can be found that describes the relationship between functional groups and density, but does not overfit the data or consume too much computing power. The preliminary models are taken from the Scikit-learn library with the default hyperparameters (except for the multi-layer perceptron). For the multi-layer perceptron: 1000 maximum iterations, 3 hidden layers with 30 nodes per layer and the limited-memory Broyden-Fletcher-Goldfarb-Shanno optimizer algorithm are used, based on previous work in our group. The evaluation metrics of the preliminary models are summarized in Table 5.3 and the fit of the predictions on the training dataset with 10-fold cross validation is shown in Figure 5.8.

In addition to the mean of the 10-fold cross validation of the statistical metrics, the standard deviation of the measurements is also taken into account. A higher standard deviation indicates a higher difference of prediction in different folds, which is a sign of overfitting of the model.

The decision tree has the second lowest average error and the third highest average coefficient of determination. On the other hand, it has the highest deviation of both from all models, indicating a large difference between the performance in the different folds and thus overfitting. Figure 5.8a also shows a large number of outliers, even though the average error is small. These results indicate that the decision tree may not have the flexibility to describe the diversity of possible DES structures, or at least to describe certain ranges.

The random forest has the lowest error and the highest coefficient of determination among all preliminary models. At the same time, the deviation of the prediction error is half that of the single decision tree. This is reasonable because using simple decision trees in an ensemble helps with overfitting. The same effect can be seen in Figure 5.8b: there are still outliers, but the errors are smaller.

On average, gradient boosting has the highest error and the lowest coefficient of determination. On the other hand, the deviation between the different folds is the lowest. In addition, Figure 5.8c shows that there are fewer outliers in the predictions and the error is slightly higher but more evenly distributed. This indicates that the model is good at describing the underlying relationship, but that the model needs to be optimized.

The multilayer perceptron has the second highest error and the second lowest coefficient of determination. The cross-validation error is also intermediate compared to other models. This model also has relatively few outliers and the error pattern is similar to gradient boosting (see 5.8d and 5.8c).

Based on these results, random forest and gradient boosting regression are selected for hyperparameter optimization. As mentioned earlier, decision tree seems to overfit, and instead of regularizing a single tree, it is simpler to use an ensemble model. Accordingly, random forest has the best metrics. Although gradient boosting has the highest average error, it has the second lowest deviation during cross-validation, and the simplicity of the

model makes it a promising candidate. The multi-layer perceptron model had relatively good results and is a versatile model, but due to the high computational cost and large number of hyperparameters, it is discarded from further exploration.

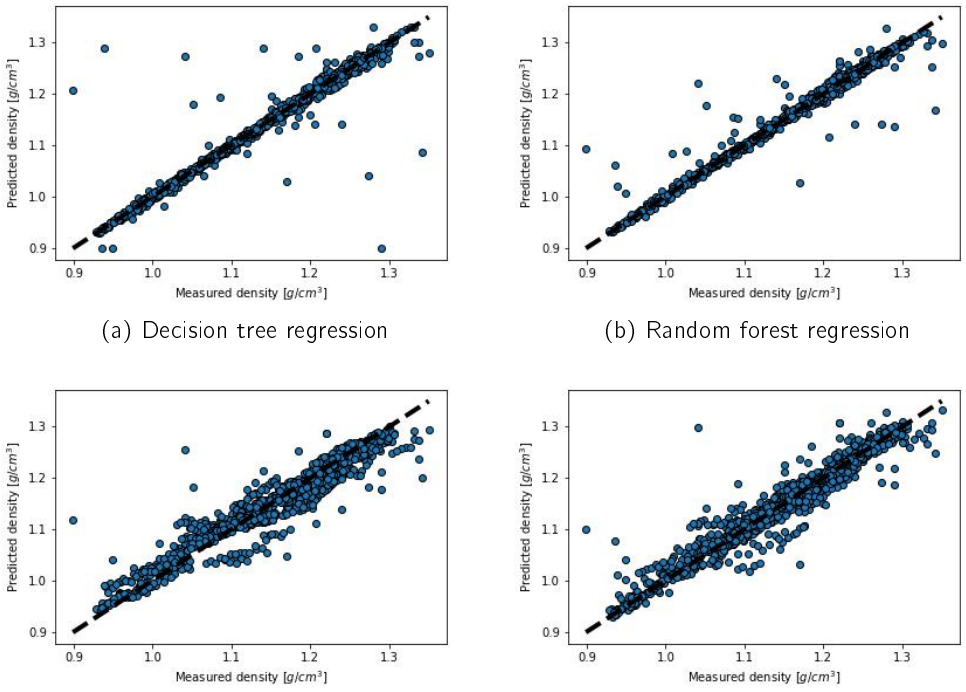
Table 5.3: Evaluation metrics of the different preliminary (non-optimized) machine learning models: Means and standard deviations of 10-fold cross-validation. Random forest regression has the best mean accuracy. Gradient boosting regression has the smallest standard deviation, indicating no overfitting.

Model	RMSE		AARD [%]		R^2	
	Mean	Std.	Mean	Std.	Mean	Std.
Decision tree regression	0.01923	0.01384	0.59	0.22	0.927	0.090
Random forest regression	0.01412	0.00693	0.49	0.14	0.967	0.028
Gradient boosting regression	0.02546	0.00354	1.54	0.15	0.913	0.023
Multi-layer perceptron regression	0.02274	0.00475	1.25	0.21	0.929	0.028

Grid search is used to tune the hyperparameters of the two selected models. The selected models have relatively low computational requirements and a small number of hyperparameters, so grid search is easier to use for optimization than random search of possible hyperparameter combinations.

In the random forest model, the number of individual trees, the depth of individual trees, the minimum number of samples in a branch to split, and the minimum number of samples in a leaf are tuned. Since the last three have no constraints in the default model (no maximum depth, one sample in the leaf), the hyperparameter tuning here is the regularization of the model for better generalization without too much loss of accuracy. The hyperparameters are regularized in a two-step grid search in the range of n_estimators=150-250, max_depth=12-18, min_samples_split=3-7, min_samples_leaf=2-5. The two-step search is performed in such a way that the selected values are the internal point of the already narrowed grid. The final selected hyperparameters are: max_depth=16, min_samples_leaf=2, min_samples_split=5, n_estimators=200. Based on Table 5.4, the regularization slightly decreased the accuracy of the model, while the deviation in the cross-validation did not change. Figure 5.9a shows no significant changes in the outliers. This indicates that the hyperparameter tuning did not significantly improve the performance of the model. However, with the regularization of the models, it will be less likely to overfit the test data.

In the gradient boosting model, the maximum depth of each tree, the number of trees, the learning rate, and the subsampling ratio are tuned. By decreasing the maximum depth of each tree, decreasing the subsample ratio, and increasing the number of estimators, the goal is to better generalize the model. By increasing the learning rate, the goal is to speed up model training without losing accuracy. Again, the grid search is performed in two iterative steps to find internal values in the searched region. Such values are found for the ratio of subsamples and learning rate. However, the maximum depth of the trees could not be decreased without significant loss of accuracy, and increasing the number of estimators after a certain point did not significantly increase the accuracy either. The final hyperparameters are learning_rate=0.25, n_estimators=300, max_depth=3 and subsample_ratio=0.8. In the table 5.4, the metrics slightly increased with the tuning of the hyperparameters, while the deviation within the cross-validation decreased. Figure 5.9b shows that in general the resid-



(a) Decision tree regression

(b) Random forest regression

(c) Gradient boosting regression

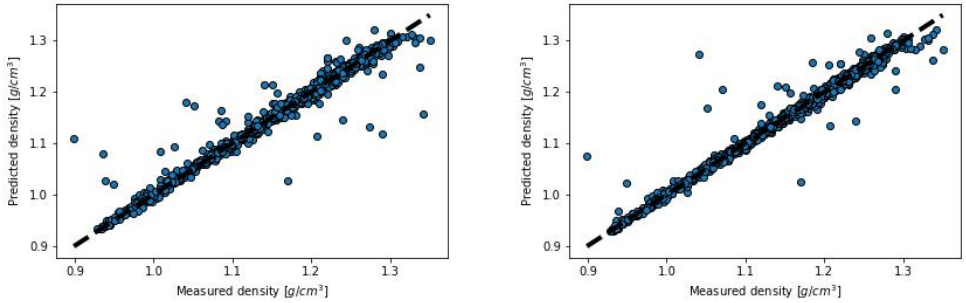
(d) Multi-layer perceptron regression

Figure 5.8: Predictions vs. experimental density values of the training set in the different preliminary (non-optimized) machine learning models. While decision tree and random forest have lower average prediction error with more outliers, gradient boosting and multilayer perceptron have higher average error but fewer outliers.

uals decreased slightly, but some outliers are still present. Overall, these results indicate that the optimization improved the performance of the gradient boosting regression, which now surpasses the results of the random forest on the training data.

5.2.3.3 Final models and prediction analysis

Table 5.5 shows the performance of the optimized models on the test and full data sets. The random forest regression model has a relative error of about 0.5 % in predicting the test data points, with a coefficient of determination of 0.989. For the full data set, the relative error is 0.364 %, while the coefficient of determination is 0.988. The gradient boosting regression has a relative error of 0.386 % on the test with a coefficient of determination of 0.991. For the entire data set, the relative error is 0.279 % and the coefficient of determination is 0.997. While the gradient boosting model performs slightly better, both models have good accuracy with errors close to or below 0.5 %. Figure 5.10 shows that only a few outliers remain in the final models, and even those have relatively small residual errors. It can also be seen that the outliers are the same data points in both models.



(a) Random forest regression (b) Gradient boosting regression

Figure 5.9: Predictions vs. experimental density values of the training set in the two hyperparameter optimized machine learning models. Both models have a low mean error and few outliers.

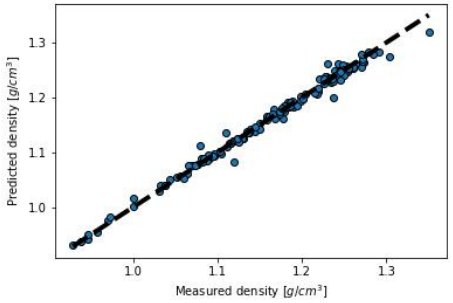
Table 5.4: Evaluation metrics of the two hyperparameter optimized machine learning models. The random forest has the same metrics after hyperparameter optimization, indicating regularization without losing much predictive power. In the case of gradient boosting, the accuracy increased while the cross-validation error decreased.

Model	RMSE		AARD [%]		R^2	
	Mean	Std.	Mean	Std.	Mean	Std.
Preliminary models						
Random forest regression	0.00025	0.00021	0.49	0.14	0.967	0.028
Gradient boosting regression	0.00066	0.00018	1.54	0.15	0.913	0.023
Optimized models						
Random forest regression	0.00028	0.00022	0.60	0.14	0.963	0.029
Gradient boosting regression	0.00020	0.00016	0.49	0.09	0.974	0.021

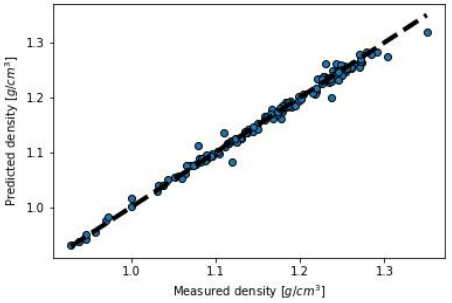
Table 5.5: Evaluation metrics of the two hyperparameter-optimized machine learning models on the test set and on the full dataset. In addition to the overall good accuracy, the small difference between the results of the full dataset and the test data shows minimal overfitting.

Model	RMSE	AARD [%]	R^2
Random forest regression - test set	9.25e-03	0.505	0.989
Random forest regression - all data	9.84e-03	0.364	0.988
Gradient boosting regression - test set	8.44e-03	0.386	0.991
Gradient boosting regression - all data	8.43e-03	0.279	0.997

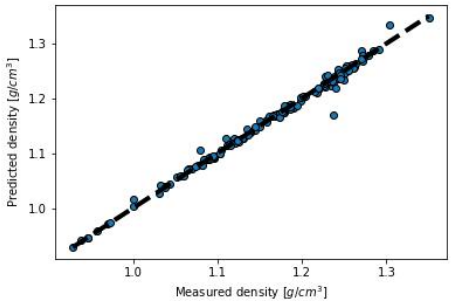
Figure 5.11 shows the learning curves of the two optimized models. A learning curve is a plot of the model performance (described by a loss function, in this case the RMSE) on the training and validation sets (or in this case the test set) as a function of the training set size. It is a tool to find out how much the model benefits from adding more training data and whether the model suffers from variance of bias error. As more data is added to the training set, the error of the training prediction will initially increase as it becomes harder to generalize the predictions. At the same time, the error of the test set decreases because



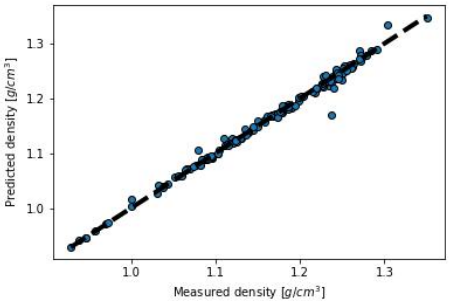
(a) Random forest regression on test set



(b) Random forest regression on all data



(c) Gradient boosting regression on test set



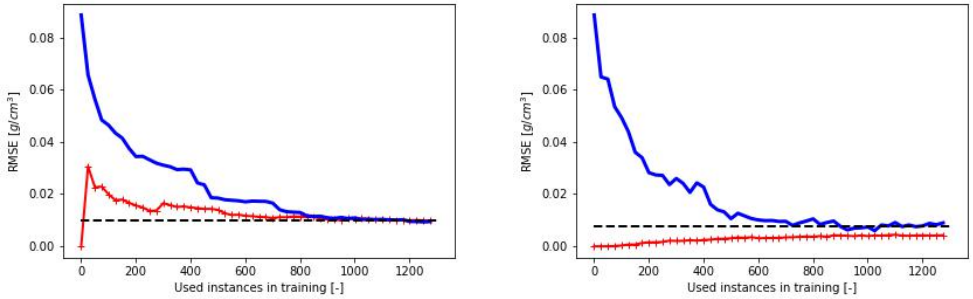
(d) Gradient boosting regression on all data

Figure 5.10: Predictions vs. experimental density values of the test set and the full dataset in the two hyperparameter optimized machine learning models. Each model has a small error and few outliers. The outliers are present in the test data set.

the model generalizes better with more data. The two errors should merge as more data is added (as we see in Figure 5.11), and both reach a plateau as the model reaches its limit. If the error of the training set continuously decreases, it means that there is not enough data to build a well generalized model. A large gap between the training and test curves indicates overfitting of the model, as the model performs much better on the training data. If the gap is small but the loss function is high, the model is underfitting and a more complex model should be used.

In Figure 5.11, both models reach a plateau at a low loss function, and the train and test curves also converge. This means that the models do not have significant under- or overfitting. In the case of the random forest model, the plateau is reached around an RMSE value of 0.01 g/cm^3 with between 800 and 900 included data points in the training set. In the case of gradient boosting, the gap between the training and test RMSE is slightly larger, indicating a slight overfitting, but the loss function at the plateau is slightly lower, around 0.0075 g/cm^3 RMSE. The model reaches a plateau at a higher number of data points, between 900 and 1000.

Overall, both models train sufficiently on the available database and are complex enough to describe the underlying relationships between input and output parameters, while the regularization introduced during hyperparameter optimization prevents overfitting.



(a) Random forest regression

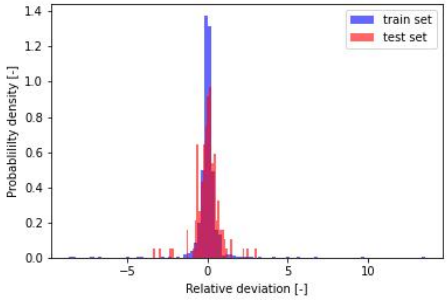
(b) Gradient boosting regression

Figure 5.11: Learning curves of the two hyper-parameter optimized machine learning models. The merging error of the training dataset (red curve) and the test dataset (blue curve) indicates the sufficient size of the data.

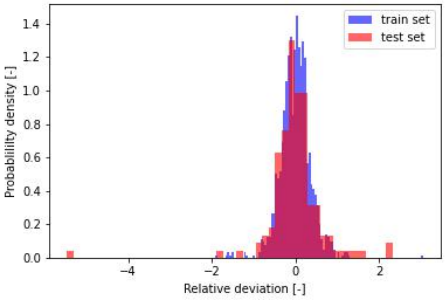
Figure 5.12 shows the distribution of the relative error of the predictions in the train and test set and the residual error of the predictions as a function of density. In both models, the deviation follows a normal distribution around zero values with a small number of outliers. There are no clusters at specific non-zero values that would indicate an artifact in the model. In the random forest model, the largest relative error is 13.5 % and it belongs to a data point in the training set. However, there are only 10 data points in the data set with a relative error equal to or greater than 5.0 %. In the gradient boosting model, the error is smaller; the highest relative error is -5.5 % and only 4 data points have equal or higher error than 2.0 %. The outliers of the random forest and gradient boosting models are shown in the Table 5.6 and 5.7.

Figures 5.12b and 5.12b show the residual error of the predictions as a function of density. In both models, the residual error seems to be mostly independent of the value of the predicted property. However, in the random forest model, some points show a negative correlation with the density. Based on the structure and temperature of the outliers (cf. Figures 5.12c and 5.6) there is no indication that certain functional groups or compounds would cause the error, as the outliers show a great variety in structure. However, the outliers are all low temperature data points. This indicates that the random forest model has lower accuracy for low temperature samples. This is reasonable since there are fewer DES stable in the 20-25 °C range and therefore the model is likely to be less accurate here. On the other hand, the same trend is not present for the gradient boosting method, which generalizes better even with smaller amounts of data.

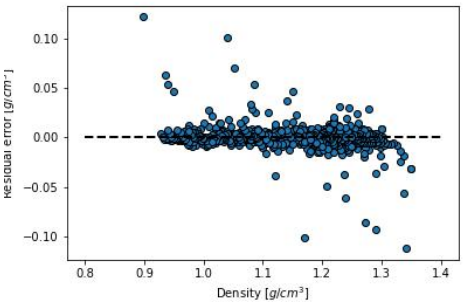
Finally, the temperature dependence of the density is evaluated. 4 DES are selected: choline chloride-urea (molar ratio 1:2), choline chloride-ethylene glycol (molar ratio 1:2), tetrahexylammonium bromide-glycerol (molar ratio 1:2), and betaine lactic acid (molar ratio 1:2) in the temperature range of 20 to 90 °C. The experimental and predicted densities as a function of temperature are plotted in Figure 5.13. In this system, only the random forest prediction of choline chloride-urea had a largely outlying value at 30 °C (see Figure 5.13a). The rest of the predictions follow the experimental trend. Although gradient boosting generally has smaller errors, in the case of choline chloride urea and the model has artifacts where the



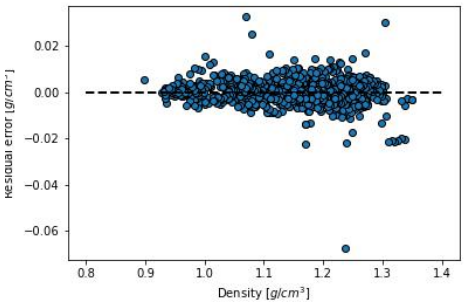
(a) Random forest regression - relative deviation



(b) Gradient boosting regression - relative deviation



(c) Random forest regression - density vs. residue



(d) Gradient boosting regression - density vs. residue

Figure 5.12: Relative deviation in the test set and all data and the residual vs. density plot of all data of the two hyperparameter optimized machine learning models. There is no grouping in the outliers, but in the random forest model the outliers show a negative correlation with density.

density does not decrease with temperature (see Figures 5.13b and 5.13f), but forms a plateau. This could be a problem when the model is combined with theory-based models such as PC-SAFT, because conflicting physicochemical rules cause errors in the modeling. This issue needs to be further investigated in the future. In addition, the errors tend to increase at extreme temperatures. This is consistent with the findings on outlier data points (cf. Tables 5.6, 5.7 and Figure 5.13). Since tree-based models tend to extrapolate poorly, it is expected that predictions at the boundary of the feature space will be less accurate.

5.2.4 Conclusions

The goal of this work was to develop a density prediction model for deep eutectic solvents by combining machine learning and group contribution methods.

A dataset of 1426 instances from the literature was combined and the number of different functional groups of the compounds was calculated. The input variables of the model are

Table 5.6: Samples with a relative error greater than 5 % in the random forest model on the full data set. There is no specific compound among the outliers, but all points belong to low temperatures.

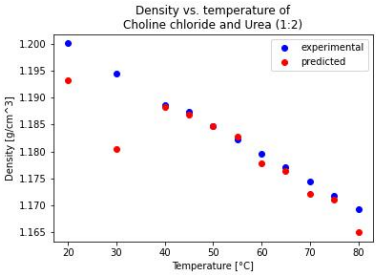
ID	HBA	HBD	Temperature [°C]	Relative error [%]
178	menthol	decanoic acid	25.0	13,5
17	ethylammonium chloride	acetamide	25.0	9,7
183	menthol	thymol	25.0	6,8
11	choline chloride	butanediol	20.0	6,7
181	lidocaine	menthol	25.0	5,8
548	Choline chloride	Acetamide	30.0	5,0
18	ethylammonium chloride	2,2,2-trifluoroacetamide	25.0	-6,8
26	2-(Diethylamino)ethanol hydrochloride	2,2,2-trifluoroacetamide	40.0	-7,2
12	choline chloride	2,2,2-trifluoroacetamide	25.0	-8,3
176	tetra n-butyl ammonium bromide	glycerol	25.0	-8,6

Table 5.7: Samples with relative error above 2 % in the gradient boosting model with the full dataset. There is no specific compound among the outliers, but all points belong to low temperatures.

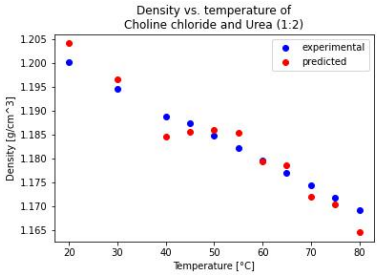
ID	HBA	HBD	Temperature [°C]	Relative error [%]
174	tetra n-butyl ammonium bromide	ethyleneglycol	25.0	3,0
44	choline chloride	Fructose	25.0	2,3
29	choline chloride	phenol	40.0	2,3
881	Choline chloride	Oxalic acid	30.0	-5,5

the number of different functional groups in the HBA and HBD compounds, the molar ratio of the compounds and the temperature. After removing outliers, the dataset of 1421 data points has information about 38 HBA and 71 HBD compounds, 132 combinations of them in 229 different molar ratios in the temperature range of 20-100 °C and about the density range of 0.899-1.35 g/cm^3. Due to the small data set, the randomly selected 90 % of the data set was used for training and only 10 % for testing the built models.

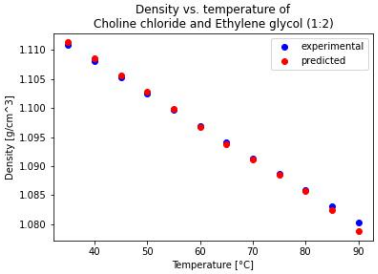
For the machine learning task, simple decision tree, multi-layer perceptron, random forest, and gradient boosting regression models were considered. After initial testing, the latter two were selected for optimization as the final models. The random forest regression model yielded a mean absolute relative deviation of 0.505 % and a coefficient of determination of 0.989 on the test set. The gradient boosting regression model yielded 0.386 % mean absolute relative deviation and 0.991 coefficient of determination on the test set. Based on the learning curves, the size of the data set is sufficient to train these models. Based on the residual errors of the predictions, there are no systematic errors in the model, but the random forest model has a higher error at low temperatures (20-30 °C). The built models



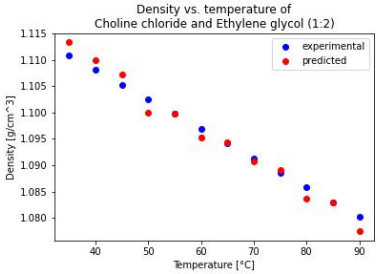
(a) Random forest regression - temperature dependence of density



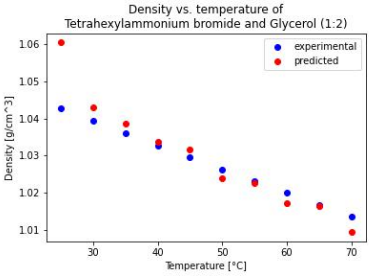
(b) Gradient boosting regression - temperature dependence of density



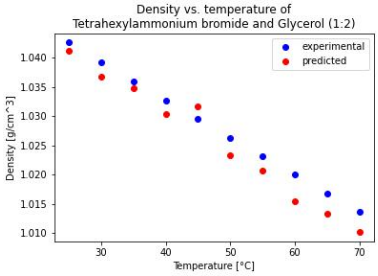
(c) Random forest regression - temperature dependence of density



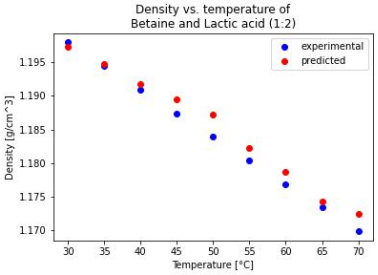
(d) Gradient boosting regression - temperature dependence of density



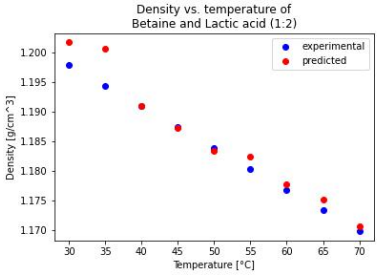
(e) Random forest regression - temperature dependence of density



(f) Gradient boosting regression - temperature dependence of density



(g) Random forest regression - temperature dependence of density



(h) Gradient boosting regression - temperature dependence of density

Figure 5.13: Temperature dependence of the density of selected DES systems in the two hyper-parameterized machine learning models. The models incorporate the temperature dependence well, with some deviations at the extremes.

also describe the temperature dependence of the density well, but the gradient boosting model has plateaus in the temperature-density relation at certain points. Correction of this error will be necessary in the future to combine the proposed models with theory-based models.

5.3 Viscosity prediction

5.3.1 Introduction

Viscosity is one of the most important properties of DES to model. Its accurate prediction is necessary because it determines the utility of solvent candidates via the mass transfer and energy dependence of the material flow. It's important from both the chemical and engineering feasibility aspects of the process. As with other properties of DES, the strong hydrogen bonding interactions influence viscosity. DES typically have very high viscosity, which is one of the limiting factors in their application. Accurate determination of liquid viscosity is already difficult due to the liquid phase momentum transfer mechanism in the system, but exceptionally strong secondary interactions make modeling even more challenging.

Consequently, theoretical methods, although providing a sound theoretical background to the underlying mechanism, often operate with very large errors [168, 169]. On the other hand, simple correlations and semi-empirical methods often give better results [170, 171, 172].

Global correlation models are currently only available for ionic liquids [173, 174]. In viscosity modeling of DES there are two main approaches: the application of IL models by parameterizing to DES and forming a global model in the form of e.g. reduced Arrhenius type viscosity model or the development of complex theoretical models with the combination of theoretical viscosity models with equation of state models.

Although simple models with general parameters would be needed for the large number of possible DES, high accuracy global viscosity modeling is still a challenge. Modeling becomes even more challenging when ternary and quaternary mixtures are considered.

The simple correlation approaches used in DES modeling come from work with ionic liquids. Based on the correlation of Vogel [175] and the correlation of Daubert and Danner [176], Haghbakhsh and Raessi proposed a simple empirical correlation method for ILs and their aqueous mixtures [177]. The model for pure ILs, using 4 parameters and based on 245 viscosity measurements of 8 ILs, achieved an absolute average relative deviation (AARD) of 2.97 %. For aqueous mixtures, using 6 parameters on 512 data points, it reached 6.72 % AARD. Later, Bakhtyari et al. used the same approach for DES by combining the Lydersen-Joback-Reid group contribution method [153] [154] with the Lee-Kessler mixing rules [155] to calculate and use critical pressure and temperature as input parameters. The correlation part used 2 parameters as fitting constants. This model was applied to 1308 data points from 156 DES and achieved 10.4 % AARD. The same group reported the application of the Vogel-Fulcher-Tamman model [158] for viscosity calculation using 3 DES-specific fitting constants. The latter method achieved 1.7 % AARD. In their recent work, Haghbakhsh et al. investigated the application of different empirical viscosity models (Arrhenius-like, Grunberg-Nissan [178], Jouyban-Acree [179], McAllister [180] and Preferential Solvation [181]) to choline chloride-ethylene glycol systems [160]. Many models were able to achieve good accuracy, even for the mixtures of DES and alcohols, but all models required experimental data to determine the model parameters. The problem with these models is that optimizing the many parameters requires experimental data on the specific systems for high accuracy, and using only global parameters results in poor accuracy. In addition, with component-specific parameterization, they cannot be used to predict novel systems.

The other approach is to combine theoretical viscosity models with Equation of State (EoS) models. Here, EoS models are used to provide parameters to the viscosity model in order to improve the accuracy of the latter. This approach was first applied to pure alcohols by Parvaneh et al. [182]. They combined three theoretical viscosity models, the free volume theory [183], Eyring's theory [184] and the friction theory [185] with four EoS, the Peng-Robinson [186], Redlich-Kwong-Soave [187], Perturbed-Chain statistical associating fluid theory [188], and cubic plus association [189]. The parameters of the viscosity models were partially calculated by EoS and the rest were fitted to viscosity measurements. The applied data set contained information on 9 alcohols in the temperature and pressure range of 203-373 K and 0.1-118 MPa with 1090 viscosity data points. Both the free volume theory and the friction theory achieved good accuracy with the CPA and PC-SAFT EoS (AARD 1.70-2.75 %). However, the large number of parameters required many experimental measurements. This method was implemented by Haghbakhsh et al. in two consecutive studies [93, 173].

In their first work, they combined the friction theory with CPA and PC-SAFT EoS to calculate the viscosity of 27 DES [173]. In this approach they used the Lydersen-Joback-Reid group contribution method [153, 154] with the Lee-Kesler mixing rules [155] to calculate the critical properties and accentric factors. They used the two EoS to determine the attractive and repulsive pressure parameters in the viscosity model. Experimental density measurements were used to adjust the 5-5 parameters of the two EoS. Finally, the 5 adjustable parameters of the friction theory were fitted to the experimental viscosity measurements. In their work, they predicted the viscosity of 27 DES based on 590 density and 253 viscosity measurements of these systems, achieving 4.4 % AARD with both EoS.

In their second paper, they used the same approach and data set, but used the free volume theory as the viscosity model [93]. In this approach, the CPA and PC-SAFT are auxiliary equations to calculate the density needed in the residual term in the viscosity equation. With this approach, both combinations achieved an AARD of 2.7 %. These examples show that the use of theoretical models provides accurate and theoretically sound predictions, but it requires a large number of experimental measurements and its application to novel systems is very limited.

To date, no global viscosity model has been proposed that can make accurate predictions without prior experimental data on the system. Given the large number of possible DES and the need for computational screening of their properties for effective design, their application is severely hampered. In this chapter, I explore a similar approach that I have used effectively for density modeling. To do so, I use the experimental dataset assembled by Bakhtyari et al. for their viscosity model [190]. I set up a qualitative structure-property relationship model by describing the DES structure by the numbers of different functional groups in their compounds instead of complicated feature generation steps. By using the general structure as an input variable, novel systems become describable, which is a great advantage over previous modeling efforts.

5.3.2 Machine learning workflow

The same workflow and models were used for viscosity prediction as for density prediction. There are only two differences compared to the density section. First, the viscosity data set

was taken from the work of Bakhtyari et al. [190]. Second, the data set was not truncated based on the standard deviation (z-score), but at a fixed value of 200 mPas.

5.3.3 Results and discussion

5.3.3.1 Dataset

The initial data set consists of 1306 data points from 22 previously published studies. The dataset covers a viscosity range of 5.2-17645 mPaS and a temperature range of 4-95 °C (see Table 5.8). However, the distribution of the data shows a strong tail at higher values (see figure 5.14). While 75 % of the data points are below 208 mPaS, the last 25 % cover the range 208-17645 mPas. This means that the high viscosity range is heavily undersampled and no accurate predictions can be expected here. (This has also been proven by preliminary prediction experiments, which are not reported here.) Since the dataset does not follow a normal distribution at all, trimming the dataset based on the standard deviation is not an option (see Table 5.8: the average viscosity is 366 mPas with a standard deviation of 1263 mPas). To overcome this sampling problem, I dropped high viscosity data points by excluding instances above 200 mPas. This reduced the number of data points from 1306 to 971. The average density decreased from 366 mPas to 32 mPas, while the standard deviation decreased from 1263 mPas to 45 mPas, see table 5.8. This is still not a normal distribution, but the models used do not require normality. On the other hand, this trimming results in a much better described range of viscosity (see figure 5.14).

As with density modeling, the temperature dependence of viscosity is an important consideration to include in the model. Similarly, the available data depend on the melting and decomposition temperatures of DES, so the information available is different for each system. Fortunately, trimming does not affect the described temperature range or temperature distribution. Both the original and the trimmed dataset describe instances in the temperature range 4-95 °C (see Table 5.8). This indicates that the high viscosity instances are related to the included compounds rather than lower temperatures. As a result, information about the structural diversity is lost with the trimming rather than the temperature described.

Table 5.8: Statistics of the original and trimmed viscosity data sets. There are significant changes in both the size and distribution of the data set.

	Original dataset		Trimmed dataset	
	Viscosity [mPas]	Temperature [°C]	Viscosity [mPas]	Temperature [°C]
N	1306		971	
Mean	365.92	43.28	31.93	44.49
Std.	1262.84	18.09	44.83	18.82
Min	5.20	4.00	5.20	4.00
25 %	33.89	30.00	28.07	30.00
50 %	70.00	40.00	47.60	45.00
75 %	208.00	55.00	84.85	60.00
Max	17645.50	95.00	197.60	95.00

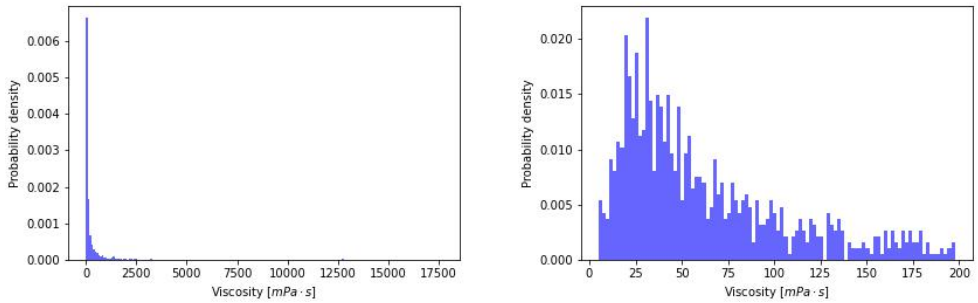


Figure 5.14: a) Distribution of the original viscosity data set. b) Distribution of the viscosity data set after trimming at 200 mPas. The original data set does not have a normal distribution, so in order to have a usable data set, the high values are trimmed at a fixed point.

The same feature engineering is used as for density modeling. The functional groups present in the DES compounds are taken into account to generate the input variables of the model. A commonly used grouping of molecular structures is used, differentiating between functional groups present in chain or aromatic ring structures, and whether it is in a HBA or HBD compound. The functional groups considered and their abundance in the database are summarized in Table 5.9.

Besides the available viscosity values, the available functional groups are the main limitation of the model (the model cannot predict DES with functional groups for which it has no information, for which it is not trained). Since the data are taken from the literature, not every functional group is adequately represented. Table 5.9 shows that many considered functional groups are not present in the database at all, and the limitations are even greater for groups in aromatic rings. While most functional groups in non-aromatic structures are represented, with the exception of aldehyde and amide groups in HBAs and aldehyde and ester groups in HBDs, most functionalities in aromatic structures are missing. Carboxyl, amine, ester, and amide groups present on aromatic groups are completely missing from the database. In addition, data on aromatic ether and hydroxyl groups are also very limited. These groups are excluded from model development and the model will not have predictive capabilities for novel systems containing these groups. This reduced the number of structural features to 22.

After this pre-processing of the data set, the model will provide information in the following range: Prediction can be made in the viscosity range of 5.2-197.6 mPas and temperature range of 4-95 °C for DES containing functional groups listed in Table 5.9.

Table 5.9: Included functional groups and their count in the dataset. Red highlighting indicates missing functional groups. Many functional groups are missing from the viscosity database, limiting the application domain of predictive models.

Non-aromatic groups			
HBA		HBD	
Group	Count	Group	Count
CH3-	286	CH3-	182
-OH	548	-OH	264
-COOH	4	-COOH	130
-NH2	19	-NH2	66
HOC-	**0**	**HOC-**	**0**
-COO- (ester group)	56	**-COO- (ester group)**	**0**
-COHN- (amide group)	**0**	-COHN- (amide group)	6
-O- (ether group)	56	-O- (ether group)	123
Aromatic groups			
HBA		HBD	
Group	Count	Group	Count
CH3-	2	CH3-	9
-OH	5	-OH	59
-COOH	**0**	**-COOH**	**0**
-NH2	**0**	**-NH2**	**0**
HOC-	**0**	**HOC-**	**0**
-COO- (ester group)	**0**	**-COO- (ester group)**	**0**
-COHN- (amide group)	**0**	**-COHN- (amide group)**	**0**
-O- (ether group)	3	-O- (ether group)	17

5.3.3.2 Preliminary models and model optimization

The model training is done in two steps: first, a number of preliminary models are tested, then the best candidates are optimized by hyperparameter tuning. For preliminary training, the same four regression models are selected as for density modeling: decision tree, random forest, gradient boosting, and multilayer perceptron. Models of increasing complexity are chosen because it is not known in advance how complex the underlying problem is. This allows us to find an optimal model that describes the relationship between functional groups and density, but does not overfit the data or consume too much computing power. The preliminary models are taken from the Scikit-learn library with the default hyperparameters (except for the multi-layer perceptron). For the multi-layer perceptron, 1000 maximum iterations, 3 hidden layers with 30 nodes per layer, and the limited-memory Broyden-Fletcher-Goldfarb-Shanno optimizer algorithm are used, based on preliminary experiments in our group. The evaluation metrics of the preliminary models are summarized in Table 5.10 and the fit of the predictions from the training data set with 10-fold cross validation is shown in Figure 5.15.

In addition to the mean of the 10-fold cross validation of the statistical metrics, the standard deviation of the measurements is also taken into account. A higher standard deviation

indicates a higher difference of prediction in different folds, which is a sign of overfitting of the model.

The decision tree had the highest average error and the lowest average coefficient of determination. In addition, the standard deviation of both metrics was the highest among the tested models, indicating a large difference between cross-validation and overfitting.

Figure 5.15a also shows that the error increases with viscosity. This could be due to less data at higher viscosity values. These results suggest that the decision tree may not be flexible enough to describe the variety of possible DES structures, especially at higher viscosity values.

The random forest has the second lowest error and the second highest coefficient of determination among all preliminary models. At the same time, the deviations of the prediction error and the coefficient of determination are the lowest of all tested models. This indicates that the ensemble method helps with overfitting (as expected). The improvement is also visible in Figure 5.15b: the error decreases at higher viscosity values. This is reasonable, as using simple decision trees in an ensemble helps with overfitting.

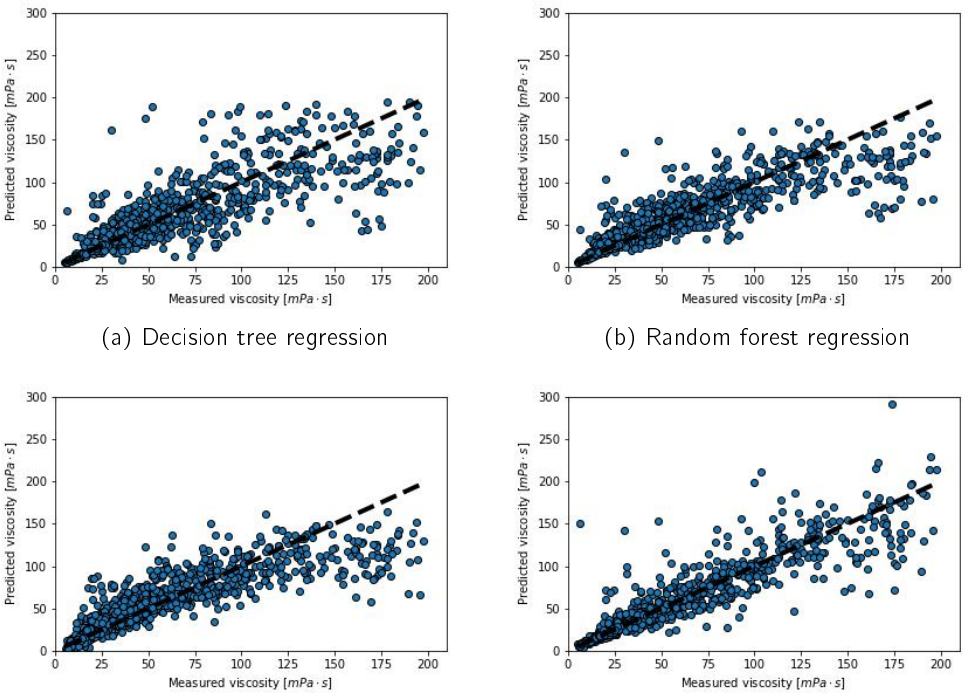
On average, gradient boosting has the second highest error and the second lowest coefficient of determination. The deviation of both metrics is slightly higher than for random forest and multi-layer perceptron. Figure 5.15c shows a systematic underprediction of viscosity values at higher viscosity values.

The multi-layer perceptron has the lowest error and the highest coefficient of determination. In addition, the cross-validation error of both is the smallest of all tested models. This model has more outliers than previous models, but the error pattern shows no systematic error (see Figure 5.15d).

Based on these results, random forest regression and multilayer perceptron regression are selected for hyperparameter optimization. As mentioned earlier, decision tree seems to overfit and instead of regularizing a single tree, it is simpler to use an ensemble model. Accordingly, random forest has the second best metrics. Interestingly, gradient boosting underperforms compared to random forest. Although both models use an ensemble of trees, the gradient boosting model works through iterations, which can be a drawback on smaller datasets. The multilayer perceptron model had the best metrics. The fact that it was the best for viscosity and not for density may be due to the more complex nature of the underlying problem.

Table 5.10: Evaluation metrics of the various preliminary (non-optimized) machine learning models. Large errors are observed for each model. Random forest and multilayer perceptron have the best accuracy and consistency between cross-validations.

Model	RMSE		AARD [%]		R^2	
	Mean	Std.	Mean	Std.	Mean	Std.
Decision tree regression	726.64	165.25	28.6	6.4	0.639	0.067
Random forest regression	540.70	125.48	25.9	4.9	0.732	0.049
Gradient boosting regression	617.40	135.97	34.8	4.1	0.693	0.055
Multi-layer perceptron regression	412.80	125.37	20.6	0.1	0.796	0.056



(a) Decision tree regression

(b) Random forest regression

(c) Gradient boosting regression

(d) Multi-layer perceptron regression

Figure 5.15: Predictions vs. experimental viscosity values of the training set in the different preliminary (non-optimized) machine learning models. Random forest regression and multilayer perceptron have lower error, but the error appears to increase with viscosity in each model.

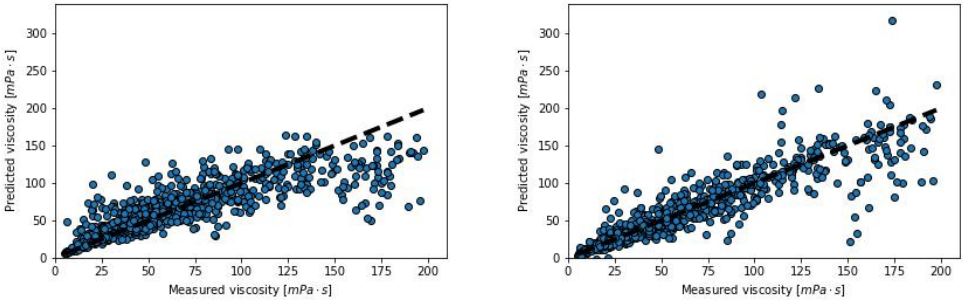
Grid search is used to tune the hyperparameters of the two selected models. While the random forest has relatively low computational requirements, the multi-layer perceptron has many adjustable parameters and is computationally demanding. Therefore, I used a grid

search in both cases. For the random forest, I tried many combinations for optimization. For the multi-layer perceptron, I only tried a handful of combinations close to the expected optimum based on my preliminary models.

In the random forest model, the number of individual trees, the depth of individual trees, the minimum number of samples in a branch to split, and the minimum number of samples in a leaf are tuned. Since the last three have no constraints in the default model (no maximum depth, one sample in the leaf), the hyperparameter tuning here is the regularization of the model for better generalization without losing too much accuracy. The hyperparameters are regularized in a two-step grid search in the range of n_estimators=150-250, max_depth=12-17, min_samples_split=3-7, min_samples_leaf=2-5. The search is done in such a way that the selected values are the internal point of the already narrowed grid. The final selected hyperparameters are: max_depth=15, min_samples_leaf=2, min_samples_split=3, n_estimators=250. Based on the Table 5.11, the regularization slightly decreased the accuracy of the model, while the deviation in the cross-validation slightly decreased. Figure 5.16a shows no significant changes in the outliers. This indicates that the hyperparameter tuning did not significantly improve the performance of the model. However, with the regularization of the models, it will be less likely to overfit the test data.

In the multi-layer perceptron, the activation function, the solver, the number of hidden layers, and the number of neurons per layer are tuned. Due to the large number of parameters, the goal is not only regularization of the model, but also accuracy. In the first step, the activation function and the solver are selected in a model with 3 hidden layers and 30 nodes per layer. In this step, logistic, hyperbolic tangent, and rectified linear activation functions are tested. For solvers, the limited-memory Broyden-Fletcher-Goldfarb-Shanno, Adam, and stochastic gradient descent algorithms are tested. After the initial screening, rectified linear activation functions and limited-memory Broyden-Fletcher-Goldfarb-Shanno are selected for further modeling. The number of hidden layers in the range of 1-8 and the number of nodes per layer in the range of 1-60 are tested in multiple iterative steps until a stable value is reached with the internal values of the test set. The nodes in the hidden layers are kept the same in each layer, as previous studies have not found a significant difference in accuracy by varying the layout of the neurons in this way. The final layout is a 3 hidden layer neural network with 20 neurons per layer. The results of the final model on the training set with 10-fold cross validation are shown in Table 5.11 and Figure 5.16b. The accuracy of the model increased slightly with the hyperparameter tuning. At the same time, the cross-validation error also increased, indicating an overfitting of the model.

In general, both models still perform poorly after optimization compared to the density model and previous efforts in the field. Table 5.11 shows that the optimized random forest regressor has an AARD of 28.7 % and a coefficient of determination of only 0.715 based on the average of 10-fold cross validation. Similarly, the multilayer perceptron model has an AARD of 18.4 % and a coefficient of determination of 0.807. These results indicate that the recommended models will perform poorly. Since the optimization does not help the accuracy, this is probably due to the insufficient size and quality of the applied dataset.



(a) Random forest regression

(b) Multi-layer perceptron regression

Figure 5.16: Predictions vs. experimental viscosity values of the training set in the two hyperparameter optimized machine learning models. The error decreased, but still higher deviation at high viscosities

Table 5.11: Evaluation metrics of the two hyperparameter optimized machine learning models. Hyperparameter tuning did not improve the accuracy of the models.

Model	RMSE		AARD [%]		R^2	
	Mean	Std.	Mean	Std.	Mean	Std.
Preliminary models						
Random forest regression	540.70	125.48	25.9	4.9	0.732	0.049
Multi-layer perceptron regression	412.80	125.37	20.6	0.1	0.796	0.056
Optimized models						
Random forest regression	572.05	124.87	28.7	4.9	0.715	0.055
Multi-layer perceptron regression	398.11	218.91	18.4	3.4	0.807	0.093

5.3.3.3 Final models and prediction analysis

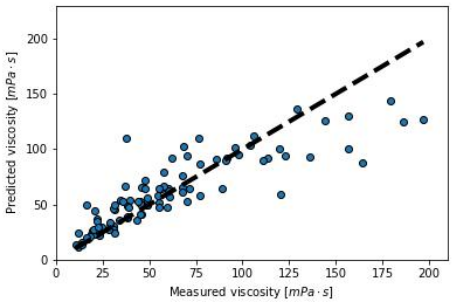
Table 5.12 shows the performance of the optimized models on the test and full data sets. The random forest regression model has a relative error of about 27 % in predicting the test data points with a coefficient of determination of 0.760. For the full data set, the relative error is 17.81 %, while the coefficient of determination is 0.886. These metrics clearly indicate that the model is not suitable for quantitative predictions. In addition, the difference between the results on the full dataset and the test dataset indicates the overfitting of the model. The difference in accuracy between the train and test sets after regularization indicates that the size of the dataset is also insufficient to generalize the model. In addition, Figure 5.17a and Figure 5.17a show that the outliers also primarily belong to the test set.

The multilayer perceptron had higher accuracy (see Table 5.12). The model on the test set has 14.94 % AARD and 0.886 coefficient of determination. On all data, the model has 12.23 % AARD and 0.939 coefficient of determination. Although there is some overfitting,

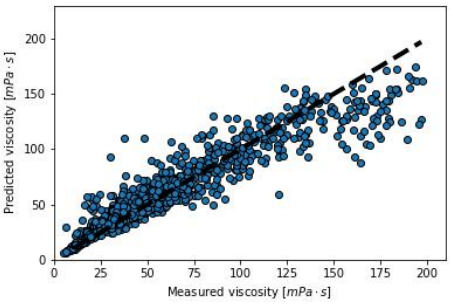
the multilayer perceptron performs much better than the random forest. Figure 5.17d shows that the outliers are also more evenly distributed between the training and test sets, which also indicates the smaller overfitting. Nevertheless, this means that the reason for the low accuracy of the second model is also the insufficient size of the data. Again, this model can still be considered as qualitatively accurate.

Table 5.12: Evaluation metrics of the two hyperparameter optimized machine learning models on the test set and on the full dataset. The model performs significantly worse on the test dataset, indicating overfitting.

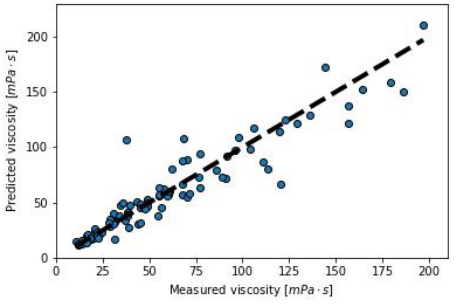
Model	RMSE	AARD [%]	R^2
Random forest regression - test set	21.74	26.8	0.760
Random forest regression - all data	15.70	17.8	0.886
Multi-layer perceptron - test set	14.49	14.9	0.886
Multi-layer perceptron - all data	11.04	12.2	0.939



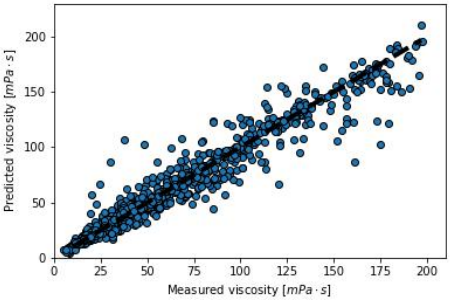
(a) Random forest regression on test set



(b) Random forest regression on all data



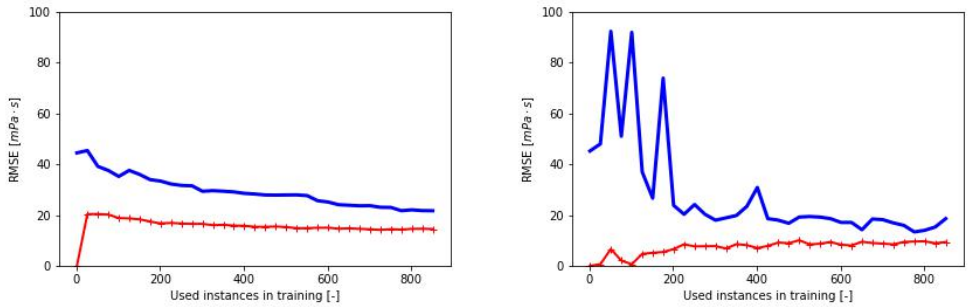
(c) Multi-layer perceptron regression on test set



(d) Multi-layer perceptron regression on all data

Figure 5.17: Predicted vs. experimental viscosity values of the test set and the full data set in the two hyperparameter optimized machine learning models. The final models still have large errors, especially at higher viscosity values.

Figure 5.18 shows the learning curves of the two optimized models. For both models, the loss function (RMSE) of the test set decreases as the size of the training set increases and approaches the loss function of the training set, but the error does not reach a plateau before all the data is used up. The difference in the loss function is greater for the random forest model than for the multilayer perceptron, suggesting greater overfitting in the former, which is consistent with previous findings. The fact that the models did not reach a stable error using all the training data means that the model is not properly trained due to the small data set. This means that the model could include more complexity that better describes the underlying relationship, but there is not enough data to represent it. This can only be corrected by including more data points. In the case of the random forest, the RMSE decreases steadily until all the training data is used up and reaches about 22 mPas RMSE. In the case of the multi-layer perceptron, the decrease of the loss function is also steady until the complete test set is reached, and the final RMSE is around 15 mPas (ignoring the fluctuation of the loss function of the test set). For the random forest, the final difference between the training and test RMSE is about 8 mPas, indicating overfitting. For the multi-layer perceptron, this difference is about 3 mPas, indicating a small overfitting. Overall, both models train poorly on the available database. The models are believed to be sufficiently complex to describe the underlying relationships between input and output parameters, but with the small dataset, proper regularization was not possible, resulting in overfitting. More data would be needed to further develop the models.



(a) Random forest regression (b) Multi-layer perceptron regression

Figure 5.18: Learning curves of the two hyperparameter optimized machine learning models. The merging error of the training dataset (red curve) and the test dataset (blue curve) indicates that the size of the dataset is not sufficient to train the models.

Due to the high error of the models and the clearly insufficient amount of data, no further analysis is performed. Since the models are not properly trained and only qualitative predictions are possible at this point, it does not make sense to analyze either the model errors or the temperature dependence of the viscosity predictions (i.e., the density error is larger than the temperature effect).

5.3.4 Conclusions

In this section, I investigated the predictive modeling of DES viscosities by combining the group contribution method with machine learning. A dataset of 1308 viscosity measurements was taken from the literature and the number of different functional groups in the compounds was calculated and used as input parameters. Additional input parameters were the molar ratio of the HBA and HBD compounds and the temperature. The final data set contained 971 points in the viscosity range of 5.2-197.6 mPas and the temperature range of 4-95 °C. Due to the small dataset, the randomly selected 90 % of the dataset was used for training and only 10 % for testing the built models. For the machine learning task, simple decision tree, multi-layer perceptron, random forest, and gradient boosting regression models were considered. After initial testing, the multi-layer perceptron was selected as the best performing model for optimization as the final predictor. The final multilayer perceptron model yielded an absolute average relative deviation of 12.23 % and a coefficient of determination of 0.939. The learning curve analysis indicates that the large error is caused by the insufficient size of the data set.

5.4 Solubility prediction

The solubility of the actors of the reaction determines the applicability of certain solvents, therefore a robust and accurate solubility model is necessary to design applications in DES. In addition, it is clear from the experimental work in Chapter 3 that the solvation of the reactants has the greatest effect on reaction performance in enzymatic reactions. The problem is that due to the strong hydrogen bonding in DES and the presence of salts in the solvent itself, currently available models (e.g. the Hansen solubility parameters) do not give correct predictions of solubility in DES.

Research on modeling solubility in DES has been rather limited. Since DES are considered to be a good candidate for CO_2 absorption, most studies have focused on this topic [90, 63, 31, 99, 91]. Dissolution of active pharmaceutical ingredients or other complex molecules is less discussed in the literature [35, 191, 28].

The main challenge in using a data-driven model for solubility, as with density and viscosity, is the lack of available experimental data. The lack of data is twofold. First, much less effort is put into collecting high quality and extensive data on different NADES and substrates. (Screening substrate solubilities in high-viscosity NADES is meticulous work due to slow equilibration). Second, a substrate-NADES system is at least ternary, which means that the dimensions of the feature space are higher. (The amount of data required to achieve the same resolution grows exponentially with the dimension of the feature space). Therefore, purely data-driven solubility prediction methods won't be feasible in the near future.

This led to the idea of combining a data-driven technique with a theoretical model to take advantage of both. Since PC-SAFT has been used in the past to effectively predict the solubility of ternary and quaternary DES systems [88, 89, 92], and the PC-SAFT parameters of the DES can be optimized based on the density values of the DES [93, 192]. In addition, an accurate and widely applicable DES density model is already available, and the two can be combined to determine the solubility in a DES system. More specifically, the density model could be used to determine the parameters of the DES in the PC-SAFT model, which is then used to determine the solid-liquid equilibria of the substrate-NADES system.

The detailed basics of PC-SAFT EoS are described by Gross and Sadowski [88, 188]. To summarize the basics of PC-SAFT: The PC-SAFT model is an EoS that models the thermodynamic behavior of complex fluids. It is an extension of the SAFT EoS, which was developed to model the thermodynamics of simple fluids such as gases and liquids [193].

The PC-SAFT model was developed to address some of the limitations of the SAFT equation of state. One of the major limitations of SAFT is its inability to accurately model the behavior of fluids with non-spherical molecules, such as polymers. This is because SAFT is based on the idea of treating a fluid as a collection of "hard spheres" that interact through a pair potential. This approach is not well suited for modeling the interactions of non-spherical molecules.

The PC-SAFT model overcomes this limitation by using the perturbation theory to describe the interactions between molecules. This allows the model to more accurately capture the behavior of non-spherical molecules, and to better predict the thermodynamic properties of complex fluids over a wide range of temperatures and pressures. This makes it a useful tool

for a variety of applications, including the design of chemical processes and the prediction of the performance of materials in different environments.

The calculation of solubility by PC-SAFT is nicely summarized in the work of Orellana et al. [194]. The equation of state is organized into terms that account for different types of intermolecular interactions:

- the hard chain reference
- dispersion
- association

The EoS is expressed in terms of the Helmholtz residual energy (Equation 5.4) because other thermodynamic properties can be derived from this.

$$a^{res} = a^{\ hc} + a^{\ disp} + a^{\ assoc} \tag{5.4}$$

where a^{hc} represents the Helmholtz energy related to the hard chain contribution, a^{disp} is related to the dispersive attraction of the nonspherical molecule contribution, and a^{assoc} is related to the association contributions. The a^{hc} hard chain contribution has two parameters: the segment diameter (σ_i) and the number of segments per chain (m_i). The a^{disp} dispersive contribution has one parameter, the dispersion energy parameter (u_i/k). Two additional parameters are needed for the associative contribution: the associating energy (ϵ^{A_i,B_i}) and the effective associating volume (κ^{A_i,B_i}). These parameters can be determined by fitting density and vapor pressure data.

Cross-interactions between two molecules i and j are calculated using the conventional Berthelot-Lorenz combination rules (Equations 5.5 and 5.6):

$$\sigma_{ij} = 1/2(\sigma_i + \sigma_j) \tag{5.5}$$

$$u_{ij} = \sqrt{u_i u_j}(1 - \kappa_{ij}) \tag{5.6}$$

where k_{ij} is an adjustable binary interaction parameter. In associating compounds, the cross-associating interactions are expressed by Equations 5.7 and 5.8, proposed by Wolbach and Sandler [195]:

$$\epsilon^{A_i,B_j} = \frac{\epsilon^{A_i,B_i} + \epsilon^{A_j,B_j}}{2} \tag{5.7}$$

$$\kappa^{A_i,B_j} = \sqrt{\kappa^{A_i,B_i}\kappa^{A_j,B_j}}\left(\frac{\sqrt{\sigma_{ii}\sigma_{jj}}}{\frac{1}{2}(\sigma_{ii}+\sigma_{jj})}\right)^3 \tag{5.8}$$

The solubility of the substrate in the solvent can be described by the solid-liquid equilibrium relationship (Equation 5.9), assuming a pure solid phase and neglecting the influence of the different heat capacities of the pure solid and the liquid:

$$x_i^L = \frac{\varphi_{0i}^L}{\varphi_{0i}^L} exp\Big\{-\frac{\delta h_{0i}^{SL}}{RT}\Big((1 - \frac{T}{T_{0i}^{SL}}\Big)\Big\} \tag{5.9}$$

where φ_{0i}^{L} is the fugacity coefficient of the pure substrate and φ_{0i}^{L} is the fugacity coefficient of the substrate in the mixture, δh_{0i}^{SL} is the enthalpy of the pure substrate T_{0i}^{SL} is the melting point of the substrate and T is the temperature of the system. These melting properties of the substrate have to be determined experimentally or estimated by other modeling methods. For example, Orellana et al. used both experimental results [196] and group contribution based estimates [197] to determine gallic acid properties [194]. The PC-SAFT method is used to calculate the required fugacity coefficients by Equation 5.10:

$$RT\ln\varphi_i = \left\{\frac{\delta n a^{res}}{\delta n_i}\right\}_{T,V,n_{k\neq i}} - RTlnZ \tag{5.10}$$

where n_i is the moles of component i, n is the total moles in the mixture, R is the universal gas constant and Z is the compressibility factor. The compressibility factor is also calculated by PC-SAFT according to Equation 5.11:

$$Z = 1 + \rho\left(\frac{\delta(a^{res}/RT)}{\delta\rho}\right)_{T,n_i} \tag{5.11}$$

where ρ is the density at the given pressure and temperature.

With the DES density model, the parameters for the solvent itself can be calculated based on the predictions. Solute parameters must also be determined. Fortunately, these are more readily available than for DES, since the experimental data for common compounds are more extensive.

The applicability of the solubility model overlaps with the applicability of the density model, but the available information on the solute itself must also be taken into account. Based on previous publications, the relative deviation of the PC-SAFT model from experimental values can reach up to 10 % [93, 192, 194]. Combined with the uncertainty of the density model, this can lead to error propagation. However, since the error in density prediction is an order of magnitude smaller than the error in PC-SAFT, this propagation should not be a primary concern.

5.5 Linking results to hypotheses

In this chapter, I explored the application of machine learning combined with group contribution methods to model the density and viscosity models for predicting DES behavior. In addition, I provided recommendations on how to combine the density model with PC-SAFT to predict the solubility of substrates in DES solvents. The two recommended density models, a random forest and a gradient boosting regressor, achieved good accuracy on a density data set of 1426 points. They can predict density values over a wide range of structures and temperatures with a relative error of 0.5 %. Learning curves and error analysis show that there is no systematic error in the system or reduced accuracy for certain functional groups, and the data was sufficient for optimal training of the proposed models. The viscosity model was also feasible with the available data. A random forest and a multi-layer perceptron regressor were built on a data set of 1306 points. However, the accuracy of the model was much lower, with a relative error of 12 %. The learning curve of the final model indicates that the available data is not sufficient to build more accurate models. This is consistent with the high number of missing functional groups in the original viscosity data set and the wide distribution of viscosity values. The proposed density model can also describe the parameters of the PC-SAFT method. The possible bottleneck of the process is the accuracy of the PC-SAFT model rather than the density model.

In this chapter, I explored how to predict NADES behavior in a structured way and how to minimize the amount of empirical data required to do so. My hypothesis was that data-driven, machine learning-based methods can be used to predict the density and viscosity of bulk media, and that there is sufficient data available in the literature to develop accurate models. My results indicate that the assumptions were correct in the case of density modeling. I was able to develop a practical density model for various DES. In the case of the viscosity, the proposed model has high error, however, the analysis of the error and learning curves shows that the problem is the insufficient amount of data rather the approach itself. In its current form, the model can serve as a qualitative predictor for DES in the lower viscosity range and with moderate structural limitations. The method of solubility modeling by combining the density machine learning model with PC-SAFT has also been proposed, but no practical application has been demonstrated.

As data-driven methods are proposed, the data that drive the models is the biggest limitation. The density model is well trained in the available feature space (the variety of functional groups for which we have information), but it is not applicable to novel functional groups (see table 5.2). The same is true for viscosity models, but the accuracy is low even for known groups. In machine learning, "bigger is always better": the only way to increase accuracy is to increase the viscosity data set. As DES is an intensively researched field, this problem will be solved in the future. The proposed combined solubility model should also be put into practice. The methodology proposed in this chapter could be applied to any physicochemical property of (NA)DES, provided that sufficient data are available. The combination of theoretical and data-driven methods expands the possibilities even further, as the example of solubility modeling shows.

Chapter 6

Case study: deacetylation of mannosylerythritol lipids in NADES

"All sorts of things can happen when you're open to new ideas and playing around with things."

Stephanie Kwolek

6.1 Preface

As mentioned in the Introduction, this research started with the idea of enzymatic deacetylation of mannosylerythritol lipids in NADES. A water-free medium was needed because water induces side reactions in the system. MELs are biosurfactants; therefore, a benign medium would also be beneficial. However, the first attempts with commonly used NADES were not successful for the reaction, so a selection method was needed. Now, although the predictive models still lack accuracy, the framework can be used to select from the known NADES. Solubility is the most important selection criterion. Low viscosity is advantageous but not a priority. Based on the simulation results of Chapter 4, I expect a stable enzyme and normal activity in most of the available systems. With these conditions, I test a few selected systems for practical application.

In this work, I test my first and second hypotheses once again. I show that the effect of NADES can be described by the specific interactions between NADES-substrate, NADES-NADES and NADES-enzyme and that the effects of NADES on enzymatic reactions can be related to substrate solubility/solvation, media viscosity and changes in enzyme structure. Previously, I concluded that solubility plays the primary role, and while low viscosity is optimal (from a process point of view), no mass transfer limitation is observed in viscous systems. In addition, based on the results of Chapters 3 and 4, side reactions are more likely than changes in activity due to enzyme-NADES interactions.

In this chapter I discuss the results of enzymatic deacetylation of mannosylerythrithol lipids in different NADES systems. I compare the performance of hydrophilic and hydrophobic, acid-containing and not containing NADES and two selected reference solvents to find an optimal

solvent for the reaction. Performance is measured by the output of the most valuable, fully deacetylated MEL compound. In addition to optimal yield, cost and environmental aspects of the solvent are discussed.

The content of this chapter has also been submitted to Applied Microbiology and Biotechnology with the title "Deacetylation of mannosylerythritol lipids in hydrophobic natural deep eutectic solvents". Co-authors of the paper are Jonas Cassimon, Iris Cornet, Erik C. Neyts and Pieter Billen. Jonas Cassimon contributed to the development of the methodology, the formal analysis of the results and the writing of the original draft. Iris Cornet, Erik C. Neyts, and Pieter Billen contributed to the supervision of the project and to the review and editing in the writing process. In addition, Pieter Billen obtained financial support for the project.

6.2 Abstract

Mannosylerythritol lipids (MELs) are a promising group of biosurfactants due to their high fermentation yield, self assembling properties and biological activities. During the fermentation, a mixture of MELs with different levels of acylation is formed. From this mixture, the fully deacetylated form (MEL-D) is the most valuable. To decrease the environmental impact of the deacetylation and to exclude toxic contaminants, an enzymatic process with natural deep eutectic solvents (NADES) was developed in this study.

We tested the deacetylation of a purified MELs mixture with immobilized *Candida antarctica* lipase B enzyme and 2-ethylhexanol as co-substrate in 140 hours long reactions. As solvent, we compared four NADES, both hydrophilic (choline chloride-ethylene glycol) and hydrophobic (menthol-octanoic acid, thymol–menthol and thymol–coumarin), and as reference, pure 2-ethylhexanol (the co-substrate) and toluene (a commonly used volatile organic compound). We monitored the change in concentrations of the different MEL compounds by normal phase high performance liquid chromatography coupled with evaporative light scattering detector.

We reached the highest yield of MEL-D in the pure 2-ethylhexanol as solvent. However, the toluene and hydrophobic NADES systems had similar yields after 140 hours. In the hydrophilic NADES no conversion was observed and in menthol–octanoic acid we experienced a side reaction which led to the formation of triacylated MELs. The pathway of the reaction was the same in every system: MEL-A converted into MEL-C and then into MEL-D, while MEL-B directly converted into MEL-D. The main difference between the systems was the speed of turning MEL-A into MEL-C, which also determined the remaining amount of A and C compounds at the end of the reaction in the different solvents.

Our results indicate that deacetylation of MELs mixtures in NADES as solvent is possible with comparable yields to common organic solvents and even to pure co-substrate. Hydrophobic NADES without carboxylic acid compounds facilitate the reaction the most. This method has a lower cost than using pure co-substrate as solvent and is environmentally less harmful than toluene, yet it has comparable yield of the target product.

6.3 Introduction

6.3.1 Mannosylerythritol lipids

Mannosylerythritol lipids (MELs) are a promising class of biosurfactants [198]. They represent a green and circular alternative as they are produced from renewable resources by microorganisms [199]. MELs are composed of a 4-O-β-D-mannopyranosyl-erythritol hydrophilic head and one or more fatty acid chains as hydrophobic tails. They are produced by fungi from the genus *Ustilago* or yeasts form the genus *Pseudozyma* in the fermentation of vegetable oils with high yields of over 100 g/L [200]. The MEL class consists of many similar molecules. The class is divided on their degree of acetylation, on the amount of fatty acid chains and sometimes on their chirality [201]. The term MEL is often used for the

commonly produced diacetylated MEL. These MELs are named MEL-A, MEL-B, MEL-C or MEL-D according to their degree of acetylation at C6' and C4' (See Figure 6.1).

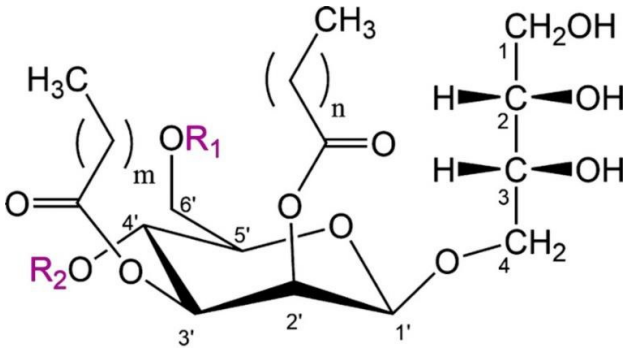


Figure 6.1: General structure of mannosylerythritol lipids (MELs). MEL-A: R1 = R2 = Ac; MEL-B: R1 = Ac, R2 = H; MEL-C: R1 = H, R2 = Ac; MEL-D: R1 = R2 = H.

MELs are a subject of interest due to their excellent interfacial properties, highly biodegradable structure and non-toxicity [202]. On top of this, MELs exhibit biomedical effects, such as skin care properties, repair of damaged hair and anti-inflammatory properties [203, 204, 200]. Because of these properties, MELs have potential applications in cosmetics, food, pharmaceutical and environmental protection. For most applications, the fully deacetylated MEL-D is the most suitable, because of its higher solubility in water, its excellent interfacial activity even at lower concentrations and, most importantly, because it cannot be deacetylated further [205]. The other MELs have the potential to lose an acetyl group and release acetic acid in the product. This is to not only to be avoided because of the change in interfacial properties and polarity, but also due to the strong smell of acetic acid.

The amount of MEL-D produced by the yeasts or fungi is very low compared to the amount of other MELs. For this reason, a method was developed by Fukuoka et al. in 2011 to produce MEL-D from the other MELs via a deacetylation reaction [205]. This method utilizes immobilized *Candida antarctica* lipase B (iCALB) enzyme as a biocatalyst and ethanol as solvent for the alcoholysis reaction. Fukuoka et al. managed to convert more than 99% of MEL-B into MEL-D after seven days and saw a partial conversion of MEL-A into MEL-C. Goossens and Wijnants optimized this enzymatic reaction in 2019 by testing different organic solvents (C2-C8 alcohols) and their water content [206]. They discovered that water has a negative effect on the reaction and that the chirality of the erythritol has a large impact on the reaction. They demonstrated that the alcohol substrate has an effect on the reaction rate of specific deacetylations. The conversion of MEL-A into MEL-C is the fastest in 2-ethylhexanol, whereas MEL-C converts the fastest into MEL-D in isomyl alcohol. The overall yield of MEL-D starting from a mixture of the four types of MELs was highest using 2-ethylhexanol.

6.3.2 Natural deep eutectic solvents

NADES emerged in the recent years as a green alternative of common organic solvents due to their benign properties [14]. NADES are the mixture of two or more naturally derived

compounds (like primary metabolites or resources from waste streams, such as quaternary ammonium salts, amines, sugars, organic and amino acids) [13], which form strong secondary interactions in the form of an intermolecular hydrogen bonding network [11]. This results in the stabilization of the mixture in the liquid state (with the minimum freezing temperature at the eutectic composition), facilitating its use as solvent. NADES have a low toxicity, they are non-volatile, non-flammable, biodegradable and often biocompatible [16, 17]. One drawback of NADES is their often high viscosity (due to the strong secondary interactions in the system), which may induce mass transfer limitation in applications [61]. Additionally, NADES are considered as designer solvents, which means that their properties can be tailored to the specific application by changing the constituents and their ratio in the mixture [1, 19].

Due to their beneficial properties and versatility, the application of NADES is considered in many fields. The possible fields of application were reviewed by Zhang et al [5]. In practice, NADES are currently used in electrochemistry [23] and separation processes [24]. However, their use in biochemistry is also considered. The first enzymatic reaction was described by Gorke et al. in 2008 [109], but since then many experimental and simulation studies were conducted on the subject [107, 108, 105, 106]. Recently Xu et al. wrote a review about the experimental research of the field [41]. Two main concerns about conducting enzymatic reactions in NADES are the enzyme-solvent interactions (changes in enzyme structure or complete denaturation) [112] and the viscosity of the solvent [54]. As many applied NADES have constituents which can form strong interactions with the enzyme residues (e.g., urea or carboxylic acids), their use can change the behavior of the catalyst. However, experimental studies observed similar or, in some cases, even higher activities of the enzyme than in common organic solvents [113, 109, 70]. Additionally, simulations also confirmed the stability of the enzyme structure in NADES [111, 110, 68, 112]. The other possible issue is the viscosity, as many NADES have significantly higher viscosities than common solvents [61, 54]. Nevertheless, experimental researches did not report on mass transfer limitations in NADES [113, 109, 70].

In the production of biosurfactants it would be reasonable to also make the production process more environmental friendly and the product free from organic solvent traces. NADES would be a possible pathway for that, as this group of green solvents was already applied in other enzymatic reactions. However, to date no research was reported on the deacetylation of MELs in NADES. We investigate if NADES can be an alternative of common organic solvents in the process of deacetylation. In addition, we compare how the reaction rates change with the different solvents. As benchmark, we compared the results of 2-ethylhexanol (co-substrate) and toluene (commonly used organic solvent in lipase catalysis) as solvents to NADES. As NADES, we considered choline chloride-ethylene glycol as an often studied hydrophilic NADES and three hydrophobic NADES (thymol-octanoic acid, thymol-menthol and menthol-coumarin). We did this differentiation as we did not know how well the biosurfactants will dissolve in the systems of different polarity.

6.4 Methods and Materials

6.4.1 Materials

Choline chloride (99 %, Thermo scientific, China), ethylene glycol (Laboratory reagent grade, $\geq$ 99%, Fisher chemical, USA), DL-menthol (99 %, Janssens Chimica, Belgium), octanoic acid (99 %, Acros organics, Germany), thymol (Laboratory reagent grade, Fisher chemical, India) and coumarin ($\geq$ 99%, Thermo scientific, France) were used for the preparation of the NADES systems. We used 2-ethylhexanol (99%, Acros organics, Germany) and Immozyme CALB-T2-150XL immobilized lipase enzyme (Chiralvision) for the catalytic reaction. Toluene (Laboratory reagent grade, $\geq$ 99%, Fisher chemical, UK) was used as a reference solvent for the deacetylation reactions. For the preparatory flash chromatography and analytical HPLC we used methanol (HPLC grade $\geq$ 99.8%, Fisher chemical, UK), isopropanol (HPLC grade, $\geq$ 99.8%, Chem-Lab NV, Belgium), acetone (HPLC grade, $\geq$ 99.8%, Fisher chemical, UK), dichloromethane (HPLC grade, $\geq$ 99.8%, Fisher chemical, Germany) and formic acid ($\geq$ 98%, Acros organics, Germany). For the water content measurement we used Hydranal-composite 5 (Honeywell Fluka)

6.4.2 Production of MELs

To obtain a concentrated MEL mixture, we used the product of Goossens et al. [201]. The detailed fermentation and isolation procedure is discussed in their work [201]. In our research we obtained the concentrated MEL-enriched phase together with some yeast cells and water residue. This mixture was first dissolved in ethyl acetate and dried by anhydrous Na_2SO_4. The solution was filtered through a Whatman paper filter. The ethyl acetate solvent was removed by rotary evaporation at 60 °C and 300 mbar. This resulted a brown viscous liquid, still containing vegetable oil and free fatty acids from the fermentation. To remove these, the mixture was dissolved in an n-hexane:methanol:water 1:6:3 (v:v:v) mixture. The aqueous bottom phase was collected and washed twice with n-hexane. The water and methanol were again evaporated using a rotary evaporator, resulting once more in a brown viscous liquid. This concentrated crude MEL mixture still contained a small amounts of residual free fatty acids, residual oil and triacylated MELs. To purify the samples from these residues, the mixture was separated by flash chromatography. For that we used a flash chromatography system (BUCHI Pure C-815 Flash with an ELSD detector) with a 25 g silica column (Chromabond® Flash RS 25 SiOH, 40 – 63 μm). As eluents, dichloromehtane, isopropanol and methanol were used. The separation method is in the supplementary information (See Figure A.23 in the Appendix). To acquire pure MEL samples, the flash chromatography appropriate samples of the above mentioned method were combined. For the pure samples of MEL-A and MEL-B, an additional purification step was made on the combined MEL-A-MEL-B samples, described in the supplementary of information (See Figure A.24 in the Appendix).

6.4.3 HPLC quantification

To quantitatively determine the concentration of each type of MEL, an HPLC method is developed. 20 μL of sample diluted with dichloromethane to 1000 ppm (all MELs combined) was injected in a Nova Pak® silica column (Waters, 60 Å, 4 μm, 3.9 mm x 150 mm) protected with a μPorasil^{TM} guard column (Waters, 10 μm, 3.9 mm x 20 mm). As eluents, dichloromehtane and isopropanol were used, both spiked with 0.16% of formic acid. The detailed HPLC method is described in the supplementary information (See Figure A.25 in the Appendix). An Agilent 1260 infinity II HPLC system was coupled with an Agilent 1260 infinity II ELDS with the nebulizer. The ELSD temperature was set to 30 °C, the evaporator chamber temperature to 30 °C and the carrier gas flowrate to 1.3 standard liters per minute (SLM). An example of the obtained chromatogram is shown in Fig. 6.2.

To determine the concentration and retention of each type of MEL in a mixture, the HPLC system with ELSD was calibrated using a standard of each type of MEL, described in the earlier section. The concentration of the MELs was calibrated in the range of 50-400 ppm.

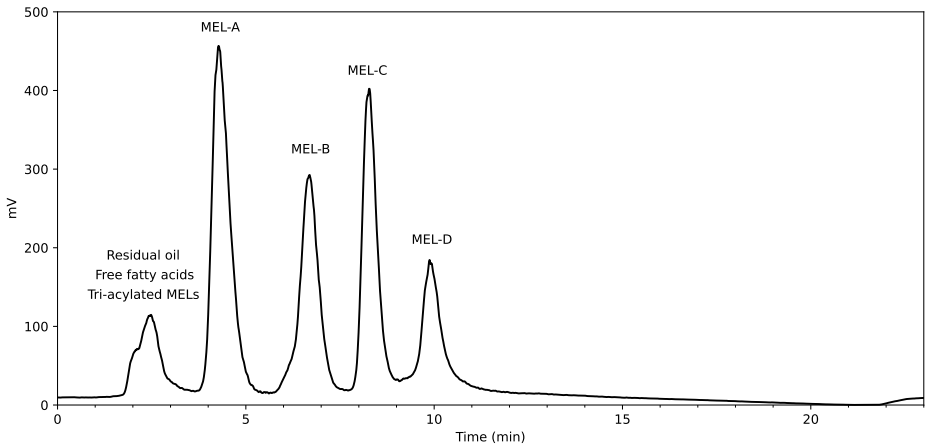


Figure 6.2: Example of a HPLC chromatogram from a MEL separation.

6.4.4 Preparation of NADES

The components of the NADES are weighed to get the required molar ratio and put together in an Erlenmeyer flask. The Erlenmeyer flask is sealed with a glass stopper to limit the amount of water from the air that dissolves in the NADES. The sealed flask was heated to 80 °C in a glycerol bath and held at this temperature while stirring until a clear liquid was obtained. After cooling to room temperature, the water content of the NADES was measured with a Karl-Fischer titration (Mettler Toledo V30 Volumetric KF Titrator) to make sure it is below 1%.

6.4.5 Deacetylation reaction

The goal of this reaction is to deacetylate MEL-A, MEL-B and MEL-C to produce MEL-D. The simplified reaction scheme is shown in Figure 6.3 with iCALB as enzyme and 2-ethylhexanol as alcohol substrate. The following reaction setup and conditions were used. The reactions were done in glass vials that were placed in a glycerol bath and held at a constant temperature of 60 °C throughout the reaction. The reaction mixture contained 5 g of solvent (reference solvent or NADES), 300 mg of 2-ethylhexanol and 300 mg of crude MEL mixture. The vials were put in the glycerol bath and stirred for five minutes to let the reaction mixture reach the reaction temperature. The reaction was started by adding 30 mg of enzyme to the mixture The immobilized enzyme was crushed with a mortar and pestle to increase the accessible surface of the beads before the reaction. The crushed enzyme was sieved, and the particles collected between 75 μm and 355 μm were used for the reaction. The reaction was sampled by taking 100 μL from the reaction mixture by an auto-pipette and the samples were diluted with 4 ml of dichloromethane. To stop the reaction, the dichloromethane was preliminary cooled to -17 °C and the diluted samples were filtered with a 45 μm PTFE syringe filter to remove the enzyme from the mixture. The samples were analyzed by HPLC and the concentration of each type of MEL was calculated.

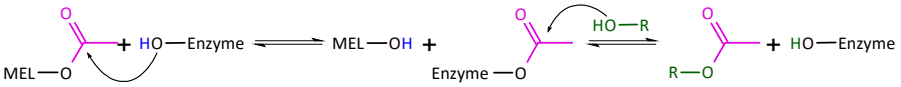


Figure 6.3: Reaction mechanism of MELs deacetylation.

6.5 Results

The performance of the different solvents were compared by the conversion of different MEL compounds and the final yield of the main product, MEL-D. First, we plotted the relative amount of the different compounds in the mixture as the function of time. From that we calculated the initial reaction rate, assuming first order reaction, and the yield of the different compounds after 40 and 140 hours. According to Goossens [207], MEL-A reacts into MEL-C, while MEL-B and MEL-C react into MEL-D and there is no direct conversion of MEL-A into MEL-B. Our results indicate a similar pathway in every tested solvent system. This is supported by the fact that MEL-B got depleted relatively early in every solvent, while MEL-C showed different changes in the relative concentration, depending on the conversion rate of MEL-A in the different solvent systems. Therefore, we also assume that there was no direct conversion of MEL-A into MEL-B. This is possibly related to the steric hindrance of the C4 carbon in the ring, shielding the carboxyl group until the same group is available at the more accessible C6 group (See Figure 6.4).

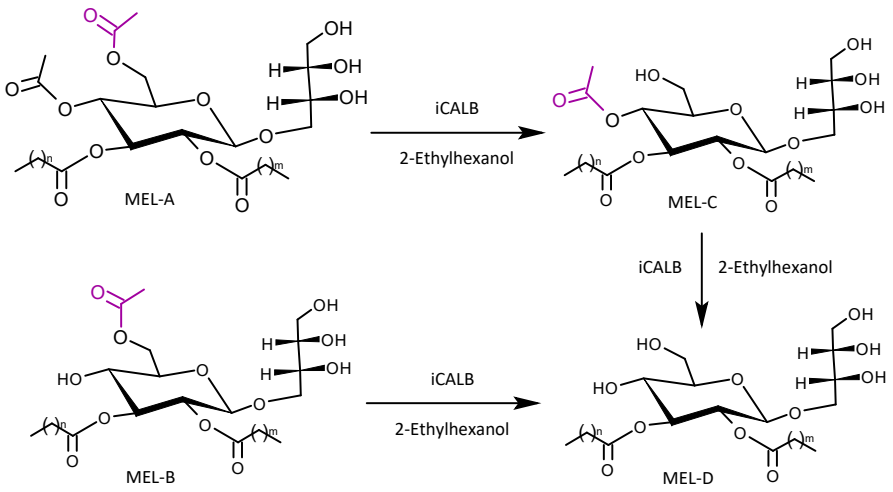


Figure 6.4: Reaction pathways of the enzymatic deacetylation of a MEL mixture with 2-ethylhexanol.

6.5.1 Deacetylation in reference solvents

First we tested the conversion of MELs in 2-ethylhexanol. As we use the cosubstrate in large excess here, we expect the fastest conversion and the highest yield from all the tested systems. This assumption is correct, the depletion of MEL-A and MEL-B are the fastest, while the final yield of MEL-D was the highest (52.7 % relative concentration in all MELs) among all the tested systems (See Figure 6.5). The reaction of MEL-B in 2-ethylhexanol finishes relatively early, after 40 hours no more MEL-B is observable in the mixture. The conversion of MEL-A is also fast, after 80 hours only 10 % remains in the system. However, the conversion of MEL-C into MEL-D is slower, which leads to the increase of MEL-C relative concentration in the system (from the initial 10% up to 42% by the end of the reaction). Together with the calculated reaction rates, this means that the conversion of the C compounds goes much slower than the conversion of the B compound. Only about 5% of the MEL-C is converted further to MEL-D. The reason behind this is the better accessibility of the acetyl group on the C6 atom, which facilitates the much faster conversion of MEL-B into D and MEL-A into C (in line with the findings of Fukuoka et al. [205]).

Toluene was chosen as a second solvent because it is a commonly used volatile organic compound in lipase reactions. It usually has good yields and reaction rates as it has low viscosity and immobilized enzymes preserve their activity well in toluene. According to our expectations, we get good conversion rates and similar final yield of MEL-D to 2-ethylhexanol (See Figure 6.6). The final yield of MEL-D is slightly below the value in 2-ethylhexanol (cf. 51.7% vs. 52.7%). Additionally, the conversion of MEL-A to MEL-C is slower and the relative concentration of MEL-C remains constant during the reaction. The reaction rates are all lower than in the pure co-substrate but that was expected, due to the lower concentration of the 2-etyhlhexanol. Although the reaction is overall slower in the beginning, in this system the MEL-C seems to be more reactive. This is possibly due to the changed interaction between the solvent and the substrate.

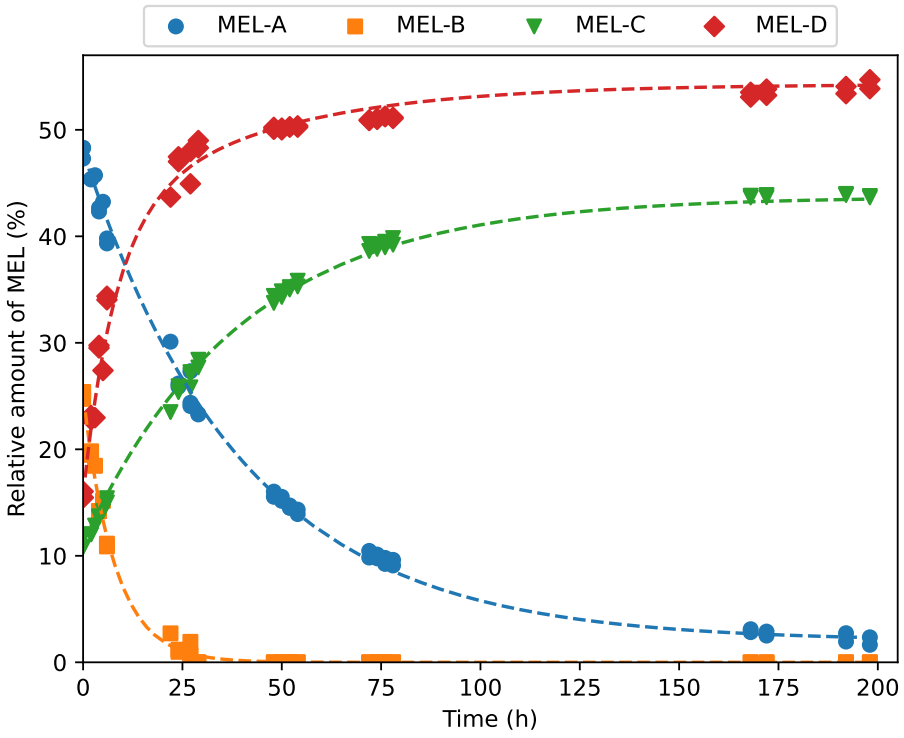


Figure 6.5: Conversion of each type of MEL in 2-ethylhexanol.

6.5.2 Deacetylation in NADES

The first NADES we tested was the choline chloride-ethylene glycol in eutectic composition (1:2 molar ratio). This system is already used successfully in lipase reactions, providing long thermal stability to the enzyme [68, 113]. However, due to the high viscosity of the system (49 $mPa \cdot s$ at 25 °C), we expect slower reaction due to the possible mass transfer limitation. Against our expectations, there is no visible conversion in the system. Therefore, after 40 hours, the sampling was stopped. As the reaction actors and the solvent form a heterogeneous system, we assume that the reaction does not occur because the actors are present in different phases. This leads us to focus on more apolar NADES which might be a better solvent for the substrates. The possible denaturation of the enzyme was also considered, but discarded as other studies show the activity of the same enzyme in this NADES [113].

To handle the issue of solubility, we considered hydrophobic NADES next. The first system was the eutectic ratio (3:2 molar ratio) of menthol and octanoic acid. This system has a lower viscosity than the choline chloride-ethylene glycol system (15 mPas at 25 °C), which also makes it a better fit for the reaction. However, the ratio of the different MEL compounds did not change over time. Moreover, the absolute amount of the four MEL compounds decreased during the reaction, while the peak related to the triacetylated MELs increased (See Figure A.26). This happens, because the presence of carboxylic acid induces

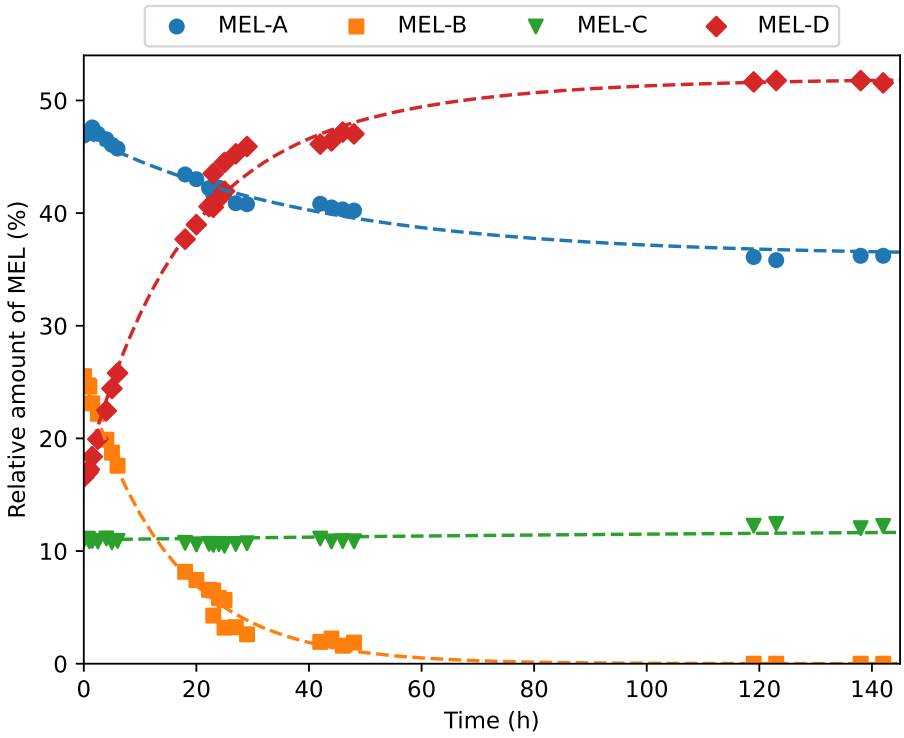


Figure 6.6: Conversion of each type of MEL in toluene.

the inverse reaction, i.e., esterification of the mannosyl hydroxyl group, which results in triacetylated MELs (See Figure A.27). This phenomenon is already described by Recke et al. [208].

Since the carboxylic acid compound induced side reactions, we tested a system without acid constituents. The first system was the 1:1 molar ratio mixture of thymol-menthol. This system has a relatively high viscosity of 53 mPas at 25 °C, which is higher than the choline chloride-ethylene glycol system (49 mPas) at the same temperature. Nevertheless, in this system we observed comparable yields and reaction rates than in toluene (See Figure 6.7). While the final yield stays slightly below the level in toluene (49.6 % vs. 51.7 %), we observe a faster conversion of MEL-A into MEL-C. However, the conversion rate of MEL-C into MEL-D is lower than in toluene or in 2-ethylhexanol, what increases the relative concentration of the MEL-C during the reaction. The conversion of MEL-B into MEL-D is also slightly slower and the system reaches a full depletion only after 100 hours. This is possibly due to the higher viscosity of the system.

The final system we tested was the 1:1 molar ratio of thymol-coumarin. This system has a lower viscosity (29 $mPa \cdot s$ at 25 °C). With this we expect higher initial reaction rate. The results (See Figure 6.8) partially confirm our expectations, as the depletion of MEL-B is faster. As in MEL-B the acetyl group is in a more accessible position on the C6 atom, the lower viscosity has a greater effect on the conversion rate. Faster reaction rates are measured also in the conversion of MEL-A into MEL-C and MEL-C into MEL-D. However,

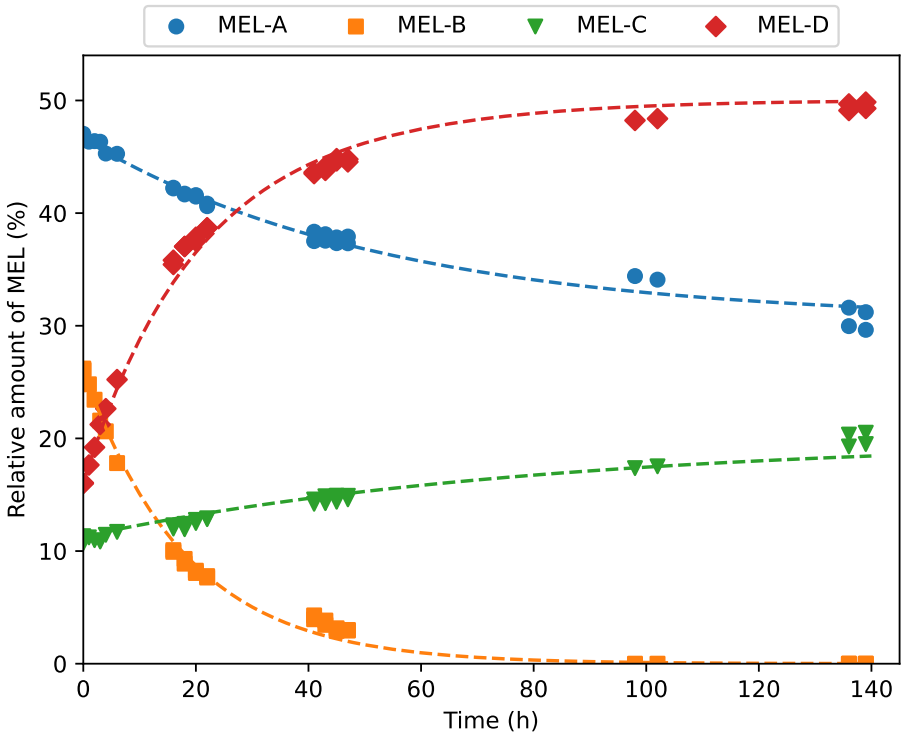


Figure 6.7: Conversion of each type of MEL in thymol-menthol NADES.

the final yields do not change significantly. MEL-D reaches a 49.9 % relative concentration, while the amount of MEL-C slightly increases to 19.5 %. Overall, from all the tested eutectic systems this last NADES has the highest conversion yield of MEL-D and also the highest reaction rates . Moreover, it also has higher initial reaction rates than toluene. Despite the higher rates, the thymol-coumarin system does not reach a higher yield of MEL-D than toluene.

6.6 Discussion

6.6.1 Feasibility of the deacetylation in NADES

In this study we compare the deacetylation of MELs mixture with iCALB and 2-ethylhexanol in various NADES and reference solvents. Among the tested solvent systems there are hydrophilic and hydrophobic NADES, containing a quaternary ammonium salt (choline chloride), diol (ethylene glycol), carboxylic acid (octanoic acid), terpenes (menthol and thymol) and coumarin. From these systems, the hydrophilic NADES formed by choline chloride and ethylene glycol and the hydrophobic menthol-octanoic acid proved to be inefficient. The former one due to the lack of solubility of the substrates and the latter due to reverse reaction induced by the presence of the carboxylic acid. However, the thymol-menthol and

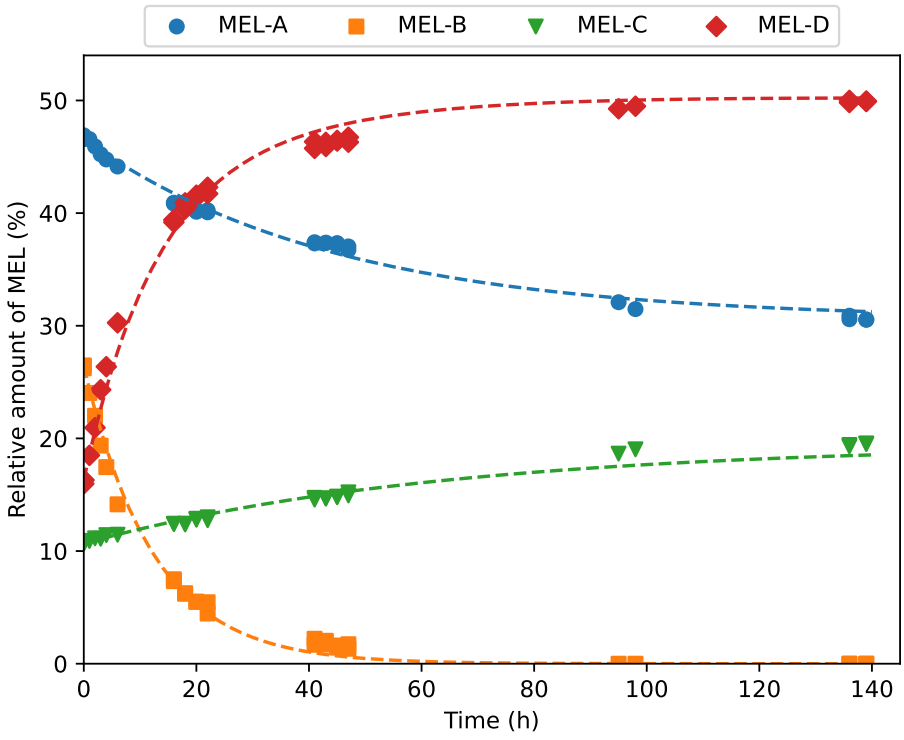


Figure 6.8: Conversion of each type of MEL in thymol–coumarin NADES.

thymol-coumarin systems are appropriate media for the effective and selective production of the deacetylated MEL compound. The best results are achieved with the thymol-coumarin system, which has a comparable yield to the toluene reference and even higher initial reaction rates than in toluene. These results prove that NADES can be appropriate reaction media for processing MEL biosurfactants.

6.6.2 Difference in initial reaction rate and final yields

In Table 6.1 we compare the final yields and initial reaction rates of the solvents, which facilitated the target reaction. Our hypothesis was that the high viscosity of the NADES will play a major role in the reaction rates. However, the measured reaction rates do not show a correlation with the viscosity, the two and three fold changes in viscosity (cf. 49, 53 and 29 $mPa \cdot s$ at 25 °C viscosity of choline chloride-urea, thymol-menhtol and thymol-coumarin systems, respectively) only result in marginal changes in the initial reaction rates (See Table 6.1). The solubility seems to have a much larger effect. In the case of the hydrophilic NADES, the lack of solubility of the substrates completely blocks the reaction. Similarly, the slightly polar NADES systems have a positive effect on the conversion of MEL A into MEL-C, but makes the reaction of MEL-C into MEL-D slower. This is possibly due to the better accessibility of the acetyl group on the C4 atom of the mannosyl group in the more apolar solvent, and better accessibility of the C6 acetyl group in a polar one

(cf. Figure 6.4 and Table 6.1). The selection of the optimal NADES system will require additional research, as here we only examined a few commonly applied NADES.

The plateau at the end of the reaction will also need further investigation. The reaction stops after 140 hours, however, 50 % of the substrate is still available. This could indicate the denaturation of the enzyme. Although earlier studies reported the general long-term stability of enzymes in NADES systems even at elevated temperatures [113, 68], the eutectic solvents applied in this study were not investigated for the thermal stability of the enzyme.

Table 6.1: Reaction rate and summary of the deacetylation reaction in different solvents.

Concentration 0 hour	MEL-A [%]	MEL-B [%]	MEL-C [%]	MEL-D [%]
2-ethylhexanol	47.8	25.1	11.4	15.8
Toluene	46.9	25.5	11.0	16.6
Thymol-Menthol	46.8	26.1	11.0	16.0
Thymol-Coumarin	46.7	26.3	10.8	16.1
Concentration 140 hours	MEL-A [%]	MEL-B [%]	MEL-C [%]	MEL-D [%]
2-ethylhexanol	5.0	0.0	42.3	52.7
Toluene	36.2	0.0	12.1	51.7
Thymol-Menthol	30.4	0.0	20.1	49.6
Thymol-Coumarin	30.6	0.0	19.5	49.9
Reaction rate (%/h)	A→C	B→D	C→D	Overall D
2-ethylhexanol	-1.148	-3.191	-0.343	3.533
Toluene	-0.259	-1.661	-0.252	1.914
Thymol-Menthol	-0.301	-1.431	-0.194	1.625
Thymol-Coumarin	-0.358	-2.140	-0.219	2.359

6.6.3 Economy and utilization

Compared to the reaction in pure co-substrate, the yield in NADES systems decreases only slightly. Two reasons to use the co-substrate only in stoichiometric amounts are the toxicity and economic feasibility. Although 2-ethylhexanol has low toxicity, it oxidizes into 2-hexanoic acid, which has teratogenic effect [209]. Additionally, the price of 2-ethylhexanol is 2495 USD/ton [210], while toluene is 1016 USD/ton [211]. In comparison, the material cost of the thymol-menthol system is 1457 USD/ton and in case of thymol-coumarin, it is 1473 USD/ton [212, 213, 214]. From that aspect the use of NADES could be the middle ground between economic considerations and environmental safety.

Due to the high viscosity of NADES and their strong interaction with the substrates, separation of the end product is an important aspect. This separation step is not in the scope of this paper. However, the literature discusses many feasible methods for the product recovery. For example, the review of [4] lists the recovery of various products, including biochemical products. Applicable strategies for the recovery of biosurfactants are solid phase [215] and liquid-liquid extraction [216], supercritical CO_2 extraction or anti solvents [217, 218]. We note that these methods are tested on the lab scale, and therefore the best method for economically relevant scales are still to be determined. Considering the low water solubil-

ity of the final thymol-menthol and thymol-coumarin systems and the amphiphilic nature of the MEL compounds, after the reaction a simple water extraction could be the most straightforward method. Again, this is yet to be tested.

6.7 Linking results to hypotheses

In this chapter, I studied the enzymatic deacetylation of mannosylerythritol lipids in different NADES. Following the selection framework, I selected four NADES and evaluated their performance based on the yield of the fully deacetylated MEL-D compound. Some NADES had comparable yields to the reference organic solvent, but at lower cost and with less environmental impact. The NADES had minor differences in performance, but viscosity showed no correlation with yield over time. However, side reactions were observed in the acidic NADES system.

I investigated the hypothesis that the effects of NADES on enzymatic reactions can be linked to substrate solubility/solvation, viscosity of the media and changes in the enzyme structure. Based on previous results, solubility and potential side reactions appeared to be the most important factors. The results in Section 6.5.2 support these considerations. Comparing the results of hydrophobic and hydrophilic NADES, solubility indeed had the largest effect. Viscosity did not show any correlation to the reaction rate, in the two effective NADES system, thymol-coumarin (η=29 mPas) and thymol-menthol (η=53 mPas), the initial reaction rate was twice as high in the more viscous system. In the thymol-octanoic acid system a side reaction was observed between the MELs and the acid compound. These results support that solubility and side reactions are the most relevant properties to consider in solvent selection. These findings are in line with the vinyl laurate transesterification results in Chapter 3. Additionally, the amount of MEL-A and MEL-C compounds at the end of the reaction was significantly different in NADES and in reference solvents. This indicates that the accessibility of the two acetyl groups changes with the solvent.

The changing conversion rates of the MEL compounds in different NADES suggest that the interaction between the media and the substrate also alters the accessibility of the different sites of the substrate. (See the difference in MEL-A and MEL-C compounds at the end of the reaction in the NADES and reference solvents.) This suggests that a single value, such as solubility, is not sufficient to describe the changes in reaction rate associated with enzyme-NADES interactions. Nevertheless, the final conversion of MEL-D was not significantly different in the cases investigated, but in other reactions the yield of the product may also change. Related to this, the exact solubility values of the different MEL compounds were not determined in this research. However, these solubility values would provide a better understanding of the solvation effect of NADES. Further research on the specific interactions will be necessary in the future.

Chapter 7

Conclusions and Outlook

"There's only one thing that can save a man from madness and that's uncertainty."

Dmitry Glukhovsky

7.1 Conclusion

7.1.1 Summary of the research

To realize the potential of NADES as green, economical and designer solvents in biochemistry, it is necessary to predict their effect on biocatalysis with as little experimental effort as possible. This is necessary to select the right media for specific applications. The aim of this research was to develop a framework for predicting the behavior of NADES via their descriptive properties and to provide a proof of concept for the prediction of these properties.

Due to the high number of potential NADES and enzymes, this research is limited to the study of *Candida antarctica* lipase B enzyme and few selected NADES for the experimental part. This limitation is rationalized by two considerations. First, the available literature showed very similar results for different enzymes in NADES and the general behavior of NADES is defined at the molecular level (the hydrogen bonding), so the specific systems studied are less important. Second, the scope of this research is to provide a proof of concept, for which the study of a selected system is sufficient. Obviously, considering the large number of possible NADES systems, the potential wider application in the future would require additional investigation. Due to the large number of potential NADES systems, it was necessary to include previous experimental systems. The density database used here is from the work of Haghbakhsh et al. [160]. The viscosity database is from the work of Bakhtyari et al. [190]. The optimized force fields for the molecular dynamics simulation are taken from the work of Doherty and Acevedo [76]. These works are proven but time-consuming steps of the necessary data acquisition.

This research aimed to develop a structured methodology for natural deep eutectic solvent selection and formulation for enzymatic reactions. To outline the development of a

holistic model, I examined the interactions of NADES with other actors, the relationship of NADES behavior to reaction outcome, and many different modeling techniques and approaches to describe these phenomena. In short, to describe such a methodology, I had to answer how NADES primarily affect the enzymatic reactions; how to describe these effects quantitatively; what methods are necessary to predict these quantitative values; and how much experimental data are needed for these predictions. In the previous chapters, I explored these questions, often in parallel, and drew conclusions about several hypotheses in the same chapter. Therefore, before giving my recommendation for a holistic approach, I will review the original research questions and my answers to them at the end of my investigation.

Q1 How do NADES influence enzymatic reactions?

H1 The effect of NADES on biocatalysis can be described by the specific interactions between NADES-substrate, NADES-NADES and NADES-enzyme.

It is clear from the literature that NADES compounds and other actors in the mixture interact primarily through hydrogen bonding and that an intermolecular hydrogen bonding network exists in the media. As a simple model of the effect of NADES on biocatalysis, the interactions between NADES compounds, between NADES and substrates of the reaction, and between NADES and the enzyme are considered. The latter can be further distinguished based on the nature of the interaction. Stabilization/denaturation of the enzyme is possible as well as changes in the active sites leading to altered enzyme activity. Finally, side reactions with the medium itself can occur.

Together with **H2** (where we linked these interactions to quantitative properties), this framework seemed appropriate to describe the effect of NADES on the enzymatic reactions (see Chapter 3). However, the results showed that the mass transfer limitation induced by the strong NADES-NADES interactions (i.e., viscosity) played a smaller role than expected based on the high viscosity of NADES (see Chapters 3 and 6). Similarly, NADES-enzyme interactions have a small effect on the outcome of the reaction. In the experiments, it was difficult to distinguish the effect of solvation from that of enzyme denaturation (see Chapter 3). In molecular dynamics simulations, while the previously described surface stabilization was observed, there was no clear evidence of changes in the active sites of the enzyme (see Chapter 4). Interestingly, changing the molar ratio of the NADES did not significantly affect the structure of the NADES, nor did it induce stronger interactions with the enzyme (see Chapter 4). Overall, all simulations and experimental results were interpreted in terms of the proposed interactions.

Based on simulations and experimental results, the specific interactions between NADES-substrate, NADES-NADES and NADES-enzyme are sufficient to discuss the effects of NADES on the enzymatic reactions. At the same time, only small effects of NADES-NADES and NADES-enzyme interactions were observed in the systems tested. Apparently, the NADES-substrate interactions primarily determined the outcome of the enzymatic reactions tested.

Q2 How can these effects be expressed by discrete and quantitative properties?

H2 The effects of NADES on enzymatic reactions can be linked to substrate solubility/solvation, viscosity of the media and changes in the enzyme structure.

The direct interactions between the various actors in the mixture can be described by quantitatively measurable and predictable properties that are independent of each other. The interactions between NADES compounds are characterized by the viscosity of the solvents. Viscosity was expected to play a significant role in the behavior of NADES, since mass transfer limitation is considered a major problem of these highly viscous media. However, the results in Chapters 3 and 6 indicate that the viscosity of the solvent has little effect on the reaction. The solubility or solvation energies of the substrates characterize the interaction between the media and the reaction actors. Of course, good solubility of the substrate in the solvent is a fundamental requirement anyway. Finally, changes in active sites and enzyme activity can be measured experimentally and by computer simulation. These effects are considered to be independently determinable, but the determination can be difficult due to overlapping effects in the experiments or due to complex simulation methods.

Separating the various effects on the experimental results of biocatalysis (e.g., initial rate activity or yield) can be difficult (see Chapter 3), again because of the overlapping effects. The viscosity and solubility of NADES and NADES-substrate systems are measurable and quantitatively predictable (see Chapters 3 and 5). However, accurate experimental determination of solubility is difficult because the viscous NADES system requires long equilibration times. With MD simulation it is possible to separate the effect of the NADES itself on the enzyme (see Chapter 4).

Two elements of the hypothesis require further investigation. First, the effect of NADES on the enzyme is difficult to summarize in a single predictable and quantitative metric. The structural changes of the enzyme can be simulated and described in detail, but the parameters used (e.g. RMSD, distance of active site residues) are difficult to interpret and compare between different enzymes. Second, describing the qualitative relationship between the system properties (viscosity, solubility) and the experimental metrics of the biocatalytic reaction (e.g., initial rate activity of the reaction) is difficult and has not been demonstrated in this research. At this point, only a qualitative prediction of the reaction performance based on the NADES behavior is possible.

Linking the effects of NADES on enzymatic reactions to solvation, viscosity, and changes in enzyme structure provides a holistic framework. All three can be measured experimentally and predicted by models. The NADES effect on the enzyme can be simulated with high accuracy, but at this point the results cannot be expressed in a single, quantitative and easily predictable metric. However, experimentally relating their contribution to the outcome of the reaction is difficult because they mask each other's effect on the enzymatic reaction. In addition, solubility was found to be the most important aspect, with viscosity and enzyme-solvent interactions having little effect.

Q3 How can the effect of NADES on biocatalysis be predicted in a structured way?

H3 The relevant properties (solubility, viscosity, structural changes of the enzyme) can be predicted by a multiscale model combining molecular, macroscale and data-driven modeling approaches.

The various elements mentioned above can be predicted separately in a multi-scale and multi-step modeling approach. While the interactions between the enzyme and the NADES require a molecular-scale model, the bulk properties of the solvent can be

predicted by macroscale or data-driven approaches. The latter also requires a comprehensive and structured database of available experimental data (from the literature). In addition, these data are necessary for the validation of the predictions.

Many models have been proposed to predict the properties of NADES and their interactions with other compounds (see Chapter 2). Here, a data-driven method, a combination of a machine learning with a group contribution method, is selected for predicting the density of NADES. The developed model is capable of predicting densities with an average relative error of less than 0.5 %. The predicted densities are used in a PC-SAFT (macroscale model) as input for model parameter fitting, and then the PC-SAFT model itself can be used for solubility prediction (see chapters 2 and 5). Unfortunately, the solubility model was not feasible within the time frame of this project, but the methodology is already available for selected substrate-NADES systems (see Chapter 5). A viscosity model of NADES was also built using a similar data-driven approach. However, the predictions in its current state are less accurate (average relative error of 15 %) and for this property the lack of experimental data is still an issue, as shown by the learning curve of the final model (see Chapter 5). The stabilizing and activating effect of NADES on the enzyme itself is also simulated by molecular dynamics for the very early stage of the system (first 200 ns) (see Chapter 4). However, this method is too computationally expensive for screening large numbers of systems. In addition, optimized force fields for NADES are hardly available. It is reasonable to check the enzyme stability in NADES after viscosity and solubility screening only for the selected candidate systems.

In this research, a multi-scale model is proposed to describe the effects of NADES on enzymatic reactions. Machine learning models are combined with PC-SAFT and molecular dynamics simulations. The machine learning models use existing databases of experimental density and viscosity values to predict the densities and viscosities of novel systems. The predicted densities are used to fit the parameters of the PC-SAFT model, which after parameterization can calculate the phase equilibrium, i.e. the solubility of different substrates in the NADES. The molecular dynamics simulation uses force fields optimized for NADES to determine the structural changes of the enzyme induced by the NADES. At this point, however, the viscosity model provides quantitative accuracy at best, and the PC-SAFT model has yet to be integrated into the framework. In addition, the rationality of molecular dynamics simulation over experimental screening is not established at this stage, given the computational resources required.

Q4 How to minimize the amount of empirical data required for the predictions?

H4 In the data-driven submodels, analysis of the database, model accuracy, and prediction error can be used to determine the currently available application domain and the amount of additional data needed.

The designer nature and vast number of possible NADES systems make building a comprehensive model very challenging. Determining the limitations and potential of the models is equally challenging. The amount of data required depends on the expected domain and the accuracy required. Since model development is limited by the available data, it is more useful at this point to reverse the question and examine what the currently available data is sufficient for. For data-driven models, the analysis of the structural diversity in the available database describes the domain in which the

model can be expected to be accurate. In the real application domain, the required accuracy can be determined by error and outlier analysis of the predictions. In the case of machine learning models, large data sets are required to train the model. In Chapter 5, the accuracy of the final models and the calculated learning curves show that the current size of the dataset is sufficient to make qualitatively accurate predictions only in the case of density. Nevertheless, the proposed viscosity model has the potential to become an accurate qualitative prediction tool, but more data and possibly more fine-tuning of the model will be necessary (see Chapter 5).

Unfortunately, the data-driven density model was not combined with the PC-SAFT solubility prediction, but the good accuracy of the density model (Chapter 5) and previous work of PC-SAFT (Chapter 2) suggest qualitative accuracy. In previous work, PC-SAFT had an order of magnitude higher error than our proposed density model, so our model should not cause any limitations. In addition, PC-SAFT also requires the parameterization of the solute. This step is not covered in this work, but data are more readily available for commonly used chemicals than for NADES (see Chapter 2).

The parameterization of the molecular and macroscopic models also requires experimental data (e.g., force fields for molecular dynamics simulations), but the analysis of the data requirements of these models is beyond the scope of this research. In molecular dynamics simulations, it is important to have an accurately parameterized force field, since standard force fields do not include the strong secondary interactions that are characteristic of NADES (see Chapter 4). Even then, experimental validation is required to see if the simulation gives good agreement with the experiments (or to see if the modeled system is stable in reality) (see Chapter 4). Since these simulations are already computationally intensive and therefore time-consuming, obtaining experimental data for certain systems may be an alternative.

The proposed molecular dynamics simulation and machine learning models are capable of making accurate predictions. In the case of molecular dynamics, a parameterized force field is required, which requires experimental data on the system of interest. In the case of machine learning models, there is currently only enough data for accurate density predictions. In the case of viscosity, only qualitative predictions are possible. The prior analysis of the data set and the posterior analysis (learning curve and prediction error) of the data driven models describe the (accurately predictable) domain of the models. The analysis shows that the limiting factor of the viscosity model is the insufficient data. Nevertheless, the feasibility of the models is demonstrated and the accuracy will increase with future data.

7.1.2 Proposed structured methodology

The recommended methodology is shown in Figure 7.1. The framework is divided into three distinct steps: screening properties, checking enzyme stability and validation.

- **Screening**

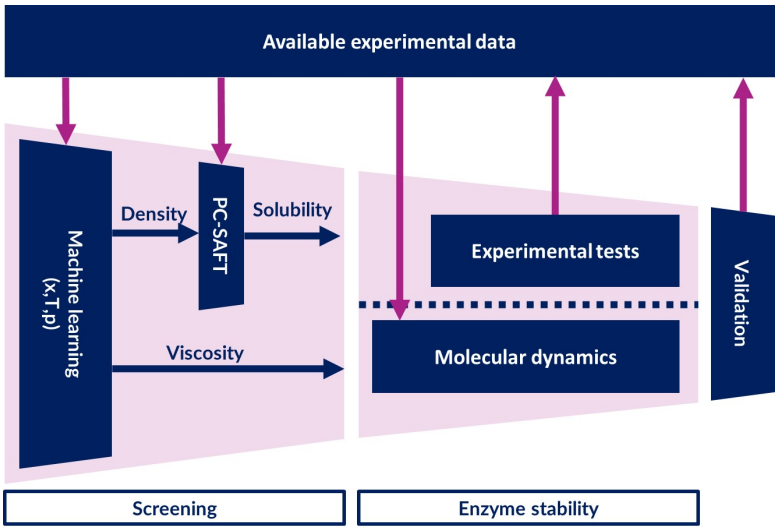


Figure 7.1: Proposed multi-scale screening model for selecting NADES for enzymatic reactions.

In the screening step, the solubility of the substrates in NADES and the viscosity of NADES can be determined. Since the methods proposed here use relatively small computational resources, these models can be applied to a large number of instances (i.e. screening). These models are based on data-driven and macroscale approaches. The screening steps can be applied without prior experimental measurements, since the structures of the applied NADES are described by the model (see Chapter 5) and physicochemical data of the solutes are available for the parameterization of the PC-SAFT component (or the parameters of the solutes are already available from other sources). Currently, the models do not have a graphical user interface, but with basic knowledge of Python and Jupyter Notebook, the user can directly use the code.

- **Solubility prediction**

  For the solubility prediction, the user needs the chemical structure and molar composition of the NADES and either the PC-SAFT parameters of the solute or the density and enthalpy of vaporization of the substrate to fit the PC-SAFT parameters. The user must examine the structure of the NADES to see if the proposed density model covers each functional group present. If it does, the user can use the algorithm in the proposed model to decompose the NADES compounds into functional number counts, which are the input variables to the model. The temperature input variable is set to predict densities in the range of 20 to 100 °C. The trained machine learning algorithm can then be applied to the instance to predict the density of the systems. The predicted densities at different temperatures are used to fit the parameters of the PC-SAFT method for the given NADES. Then, with the parameters of the NADES and the substrate, the solid-liquid equilibria can be calculated and from that the solubility at a given temperature. Currently, pressure dependence is not included because most NADES data are available at atmospheric pressure.

- **Viscosity prediction**

  The viscosity prediction works similarly. For viscosity prediction, the user needs the chemical structure and molar composition of the NADES. The user must check the structure of the NADES to see if the proposed viscosity model covers each functional group present. If this is the case, the user can use the algorithm in the proposed model to decompose the NADES compounds into functional number counts, which are the input variables of the model. The user also needs to set a temperature, but at this point the model cannot consistently predict the temperature dependence of viscosity. The trained machine learning algorithm can then be applied to the instances to predict the viscosity of the systems.

- **Enzyme stability**

  The enzyme stability determination is based on molecular dynamics simulation. In theory, enzyme stability determination does not require any experimental work. The user needs the structural data of the applied NADES, optimized force field parameters of the NADES and the structure of the enzyme. If optimized force field parameters are not available, the user can use the general force field parameters directly or parameterize the force field by approximate quantum mechanical calculations or empirical modification of potentials to fit physicochemical properties (e.g. density, viscosity, heat of vaporization) if the latter are available. Force field optimization is beyond the scope of this proposed model. However, if an optimized and validated force field is not available, the user will need experimental data from the NADES at least for the validation of a generic force field. Assuming that all elements are available, the user must follow the methodology described in Chapter 4. For the calculations, the user will also need access to sufficient computing power. The exact requirements will depend on the size and length of the simulation, the hardware available, and the time available. (Although a direct comparison is not meaningful, for a sense of proportion, the final production simulations in Chapter 4 were run on an Intel Xeon E5-2680 v4 2.40GHz with 28 threads at a speed of 19,190 ns/day).

  Alternatively, the user can determine enzyme stability experimentally using the methodology presented in Chapter 3 or other enzyme assays available in the literature. Experimental measurements allow the study of enzyme behavior on the time scale of seconds, minutes, and hours, which is often more relevant to practical applications. Experimental screening is also easier when the exact mechanism of the enzyme-NADES interaction is not of interest. To get a complete picture of enzyme behavior in NADES and their interactions, simulation and experimental work should be done in parallel.

- **Validation**

  As the NADES candidates are selected based on the computational results (both screening and enzyme stability), the physicochemical properties and behavior of the enzyme need to be experimentally validated. In addition to validation, the collected experimental data can be used to expand the database and thus train the data-driven models. The PC-SAFT method requires parameterization data for the NADES and substrate compounds, and molecular dynamics force fields must also be adapted to NADES based on experimental results. Given the small size of the current data sets, the modeling process is iterative. As new experimental data are acquired, the accuracy of various model elements can be increased in the future. The connection of the models with the different databases is present in each step, which is also shown in figure 7.1.

7.2 Further considerations

In the methodology described above, all considerations of the initial hypotheses have been successfully implemented. The effect of NADES on biocatalysis is approached from the aspect of the hydrogen-bond network; the effect of the network is described by specific properties of the system; these properties are predicted with multi-scale modeling and minimal experimental work. This has certain limitations and consequent shortcomings that should be addressed in future research. In addition, there are some open questions that could be explored in future studies. The following issues require further work in this area:

- **Complexity of the model:**

  For more accurate predictions, the complexity of the solubility part of the model should be increased. At the same time, for faster and more applicable results, the enzyme-NADES interaction part of the model should be simplified.

  In the second hypothesis, I assumed that the interactions between the NADES compounds and other actors in the system could be described in a simplified way by considering only separate, quantitatively measurable physicochemical properties. This simplification kept the model manageable in terms of computational requirements and necessary data input (which is rather limited at this point). Basically, it is a trade-off between accuracy and speed/feasibility of the model. On the one hand, the accuracy of the model is reduced and important interaction schemes between the actors may be overlooked. For the latter, the changes induced by the NADES-substrate interactions are the best example. The different initial reaction rates of vinyl laurate transesterification in Chapter 3 and the different final ratio of MEL-C and MEL-A compounds in Chapter 6 are related to the changed interaction strength and sites between the NADES and the substrate, but a single metric like solubility does not describe these important differences. Unfortunately, there is currently not enough data on NADES solvation to apply a more complex approach. One possibility would be to determine the solvation energies of the substrates by molecular dynamics simulation and to analyze the results directly or to incorporate the results into a predictive model. Alternatively, if enough data is available, machine learning models could be built and analyzed to determine the major interaction sites of the system.

  A similar complexity problem is the structural change of the enzyme during the reaction. Molecular dynamics simulations provide accurate information about the enzyme structure, but at a high computational cost, and their results cannot be summarized in a single metric. How to incorporate the results of these simulations into decision making remains an open question. Therefore, from a practical point of view, experimental work provides more concrete information at this stage. Currently, only a handful of papers discuss the effect of NADES on enzyme structures, and only for a few enzymes and a small set of NADES. In the future, a more general model of NADES-enzyme interaction will be needed for predictive screening, but this will require more data and better understanding. Molecular dynamics simulation will continue to be a useful tool to achieve the latter. So far, the results point in the direction that NADES cannot penetrate the enzyme structure [68, 110, 111, 112]. If this could be verified as a general behavior, the model could be greatly simplified by abandoning detailed structural modeling of the protein and focusing on surface interactions.

- **Accuracy and domain of machine learning models:**

  For an accurate viscosity model, more experimental data are needed. The application of additional feature generation approaches should be considered.

  The accuracy and the application domain of machine learning models are primarily determined by the available data and the design choices regarding the input variables (consequently, the feature generation). In this work, both were determined at the beginning of the project. We used the data available in the literature and we used group contribution based feature generation. This setup worked well for density prediction (our model uses a simpler algorithm but has similar accuracy to the best one in the literature [159]). In the case of viscosity, theoretical models based on experimental data (e.g., the work of Bakhtyari et al. [190]) still greatly exceed our model in accuracy. (By experimentally based, I mean that the model requires experimental results on the specific NADES it is trying to predict, and is therefore not applicable for predicting novel systems). The learning curve analysis of our viscosity model showed that the main problem is the lack of data. As more data becomes available with the advancement of the field, the accuracy of the model will increase. In addition, the outlier analysis shows that there are no specific functional groups with outstanding errors in either the density or the viscosity model, although certain functional groups, especially functional groups in aromatic rings, are often underrepresented or completely missing from the database. Again, additional data on these functional groups will solve this problem.

  The use of the group contribution approach in feature engineering has been tentatively established for this study. On the one hand, it provides an explicable and flexible model. With the functional groups, any novel system can be predicted and their contribution to hydrogen bonding can also be described. On the other hand, many groups are poorly described and consequently the accuracy of the associated NADES will be low. Excluding such groups also reduces the versatility of the model. From this point of view, the revision of the feature engineering and the use of more fundamental descriptors should be considered in the future. More generic feature engineering has already been presented (e.g. the work of Halder et al. [162]).

- **Combining machine learning with PC-SAFT:**

  An accurate density model is available for PC-SAFT, the model must be integrated, and substrate data is also required.

  The PC-SAFT method can predict solubility with high accuracy when parameterized with experimental data, but when combined with machine learning results, error propagation must be considered. Our current model has a coefficient of determination of 0.997 and an average relative error of 0.28 %. In the work of Orellana et al.[194], the error of the PC-SAFT model for solubility in aqueous DES was 5.65 %, which is an order of magnitude larger than our density model. From this point of view, error propagation should not be a problem. Nevertheless, the density and solubility models need to be integrated in the future.

  In addition, the PC-SAFT model needs to be applied to the substrate. For this purpose, PC-SAFT parameters must also be obtained for these compounds. For the relevant compounds (e.g. pharmaceutical compounds, food additives), physicochemical data are more readily available or PC-SAFT parameters are already determined. Thus, compilation of the necessary substrate data set will be less difficult than for NADES.

- **Molecular dynamics:**

  The role of molecular dynamics simulation should shift from direct structural prediction to a design tool for modeling enzyme structural changes and solvation.

  The role of molecular dynamics simulation in the framework needs to be changed. As it was already mentioned in the 'Complexity of the model' section, the currently available studies do not suggest significant changes around the active site of the enzyme [68, 110, 111, 112]. Molecular dynamics simulation is also computationally too demanding and requires too much system-specific preparation to be used for screening, at least at the current price of computational power. The role of molecular dynamics simulation should be reconsidered in the following two cases. First, when unexpected results arise from experiments, it would be reasonable to perform a simulation to reveal possible changes in the enzyme structure. Second, molecular dynamics simulation still plays an important role in understanding the behavior of NADES at the molecular level. Based on this understanding, simplified models can be built. In addition to the enzyme structure, molecular dynamics can also be used to simulate solvation processes. Therefore, in the future, simulations should be used to calculate the solvation free energies of the systems and to develop a (possibly data-driven) model based on the information obtained. Such applications in ionic liquids have been reviewed by Varela et al. [219].

- **Handling structural data:**

  A more structured, curated database would be needed. Structural data of NADES compounds should be included in a way that incorporates the stereochemistry of the molecules and explicitly counts ionic species.

  Since predictive models are largely based on experimental data, managing the available data is an important step in making accurate predictions. How the data is measured, collected and processed has the greatest impact on the accuracy of the model. From this perspective, a structured and curated database is necessary. In our models, we used a relatively small database of about 2000 data points collected from the literature. However, the current dataset does not take into account the means of measurement and the integrity of the data. In addition, the dataset was compiled in 2021, the most recent reports are not included, and the dataset is not kept up to date.

  The other consideration is the encoding of the structure of the compounds. In the current models, the structures are expressed by SMILES codes. However, this does not take into account the stereochemistry of the chemicals. Both from the literature and from our results in Chapters 3 and 6 regarding solvation effects, it is clear that the sites of interactions play a role in the outcome of the reaction. Consequently, the inclusion of stereochemistry is necessary for more accurate results. In addition, current feature generation does not account for ionic compounds. Many group contribution methods have problems incorporating ionic compounds, but salts play an important role in NADES, so the functional groups should be expanded in the future. Moreover, other research groups are already working on such differentiation (e.g. the machine learning work of [220] or the extension of Hansen solubility parameters by [221]).

7.3 Outlook

In addition to unresolved issues, there are considerations that arose during the research and therefore are not included in the originally proposed methodology. Based on the findings discussed above, the following topics could be considered to further advance the prediction of NADES behavior:

- **Accessible and maintained DES database:**

  An online accessible and maintained database would not only accelerate the development of data-driven models, but would also connect experimentalists and theoreticians from different fields.

  One of the biggest challenges of this project was accessing and validating information about DES. The information currently available on DES is very fragmented. Data can be collected from different papers reported in the literature, but this doesn't guarantee coverage of the whole field, and it can also lead to conflicting data, partly due to measurement error and partly due to different measurement methods (e.g. some researchers report water content along with density or viscosity values, others do not). This is true not only for experimental physicochemical data, but also for DES-optimized force fields. There are review articles on certain aspects of the field, but they suffer from the same problems described above. At the same time, DES show a large variety in structure. Considering the applicable compounds (e.g., variations of quaternary ammonium salts, polyols, organic and amino acids), widely varying molar ratios, ternary and quaternary mixtures with different water contents, the number of possible compositions is likely in the range of 10^6 [1]. Currently, data sets are only available in separate studies, such as the work of Zhang [5], Smith [1], or Raessi [160].

  Therefore, the creation of an online database of DES should be considered, similar to initiatives in other fields, such as PoLyInfo for polymeric materials or Protein Data Bank for large biological molecules. Such a database would help both experimentalists to share data and achieve higher utilization, and modelers to access traceable, unified, and comparable data to improve their models. As discussed in earlier chapters, PC-SAFT and MD simulations also rely on experimental data for parameterization and validation, so such a database would not only benefit research on data-driven methods. A database would also make results and measurement techniques comparable for properties such as density, viscosity, solubility (CO_2, metal oxides, drugs), surface tension, ionic, freezing point, ionic conductivity as a function of water content, temperature, and pressure. As the field of DES is developing rapidly, it would also be necessary to keep the database up to date. As a personal opinion, the most suitable research group to maintain such a database would be the Des.solve research group [222]. They have already published databases in their articles, have an extensive international network of DES research groups, and are themselves doing fundamental research on the properties and behavior of NADES.

- **Expanding molecular dynamics simulation**

  As computational resources increase, molecular dynamics simulations may become available for screening processes. Developments in force field optimization

may also lead to more widespread use of this type of simulation. Molecular dynamics could be used to study the process of solvation and the reaction itself.

Molecular dynamics simulations require computational resources that are not readily available for screening. Enzymatic simulations, especially in NADES systems that require long equilibration times, require lengthy simulations of large systems. The necessary computing power is available at Tier 1 and Tier 2 facilities for a single NADES system. Testing hundreds or thousands of systems would be difficult to prioritize. However, with the decrease in the price of computing power (i.e., Moore's Law), the required computing power will be available in lower tier equipment in the near future.

The other limiting factor for the widespread application of molecular dynamics simulation for NADES is the lack of accurate force field for various NADES compounds. Conventional force fields require further optimization given the highly non-ideal behavior of NADES, the asymmetry of their components, and the complexity of their molecular electrostatic interactions. The optimization of the force fields uses approximate quantum mechanical calculations, often in the gas phase, or the empirical modification of potentials to match experimental properties. The optimization protocols are not always reproducible, and semi-automation often results in inconsistent and unreliable parameters. Nevertheless, recent work has reported successful optimization of OPLS force fields using both empirical and algorithmic approaches [76, 142, 223, 224]. Doherty et al. and Spittle et al. demonstrated the optimization of the force field by scaling the atomic charges to fit the bulk phase physical properties of NADES [76, 142]. Zhong et al. optimized an OPLS-based force field to simulate the NADES system using a genetic algorithm [223]. Velez and Acevedo discuss the combination of artificial neural networks with ab initio molecular dynamics to generate high precision force fields by training the machine learning algorithm on the simulation trajectories [224].

With an adequate force field, it would also be possible to calculate the solvation energies of the substrate. Methods such as thermodynamic integration or Bennett Acceptance Ratio are suitable for calculating the solvation free energy of a given compound and media, but these methods have not yet been applied to DES. Published results focus on the specific interaction sites rather than solvation free energy calculations [225]. The simulation of interaction sites and interaction energies would help to develop better predictive models, either by applying the results at the theoretical level or by using the simulation data as training for the models. The reaction process itself could also be studied by molecular dynamics simulation via the application of reactive fore fields, however such research has not yet been published in relation to NADES or enzymes [226].

Appendix A

Supplementary information

A.1 Biocatalytic transesterification of vinyl laurate in NADES

The supporting information contains 3 figures.

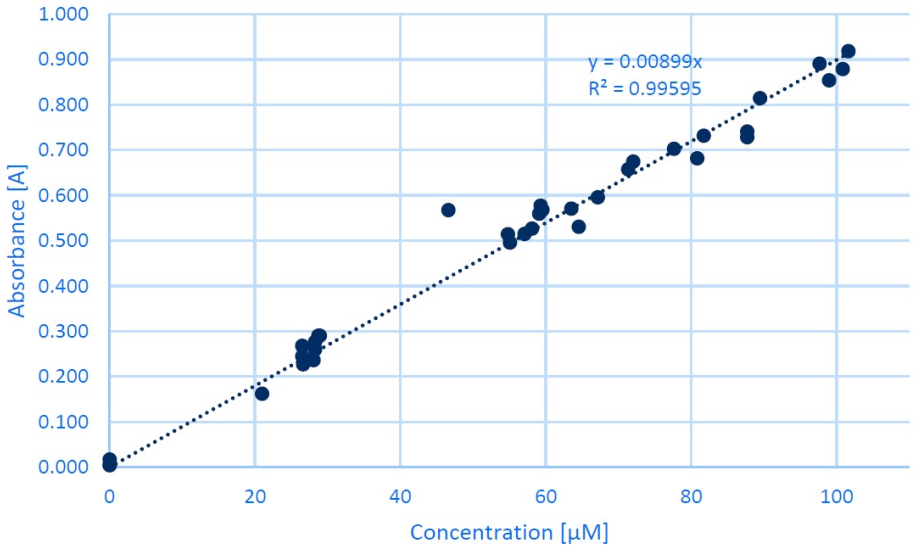


Figure A.1: Calibration curve of the absorbance of vinyl laurate in relation to the concentration with the fitted linear equation and the coefficient of determination.

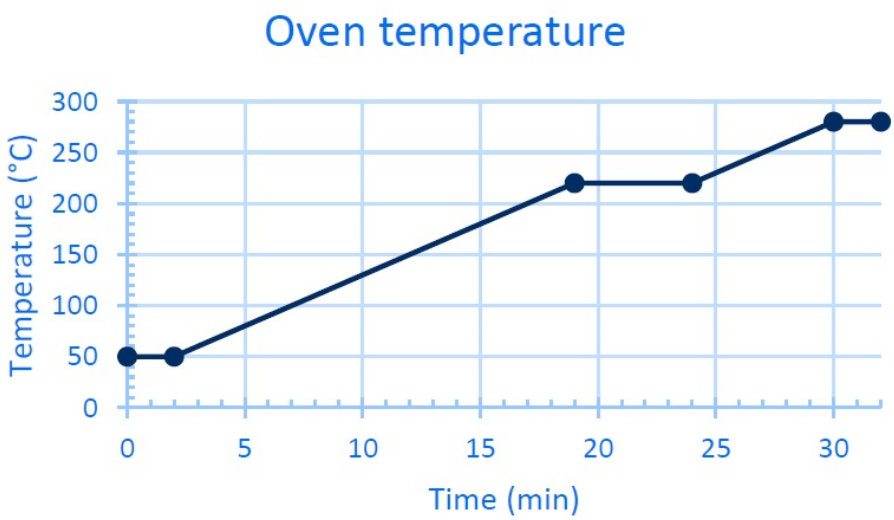


Figure A.2: Oven temperature of the gas chromatograph for the used method for the GC-MS.

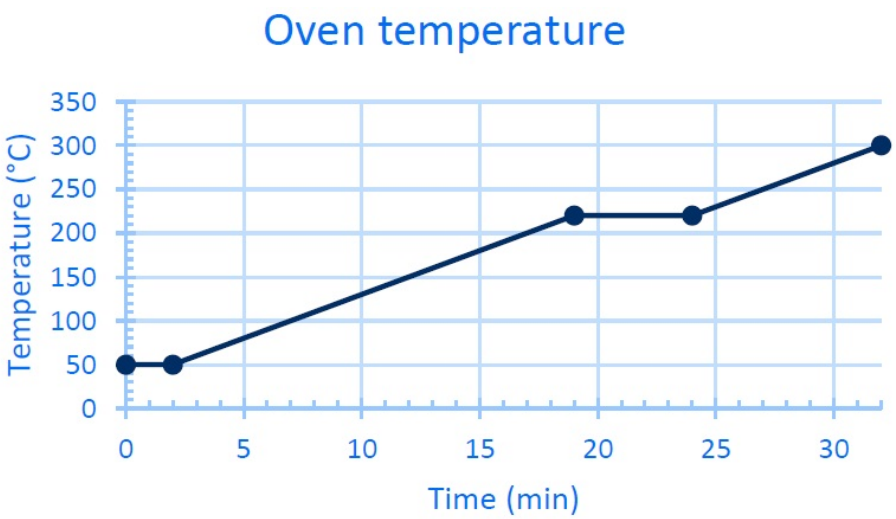


Figure A.3: Oven temperature of the gas chromatograph for the used method for the GC-FID.

A.2 Effect of NADES with non-eutectic composition on enzyme stability

The supporting information contains 19 figures. These give additional information about center-of-mass (COM) and atom-atom radial distribution functions (RDF) of the NADES systems, graphical representation of the enzyme structure in NADES, the root-mean-square deviation (RMSD) of atomic positions and root-mean-square fluctuation of atomic positions (RMSF) of enzyme residues, the changes in the secondary structure of the enzyme, the changes in distance between α5 and α10 helices and COM RDF of the HBD compounds around a selected group of surface residues of the enzyme and the whole enzyme.

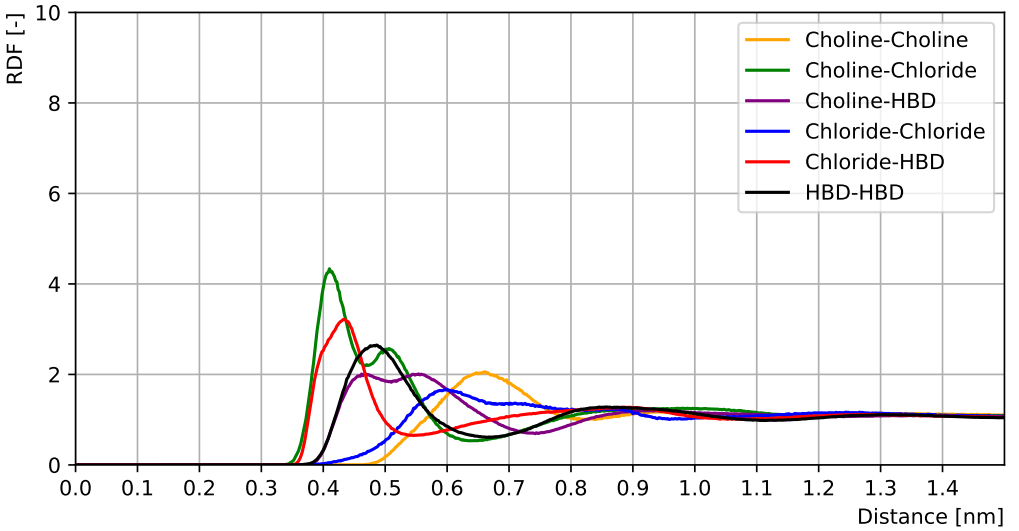
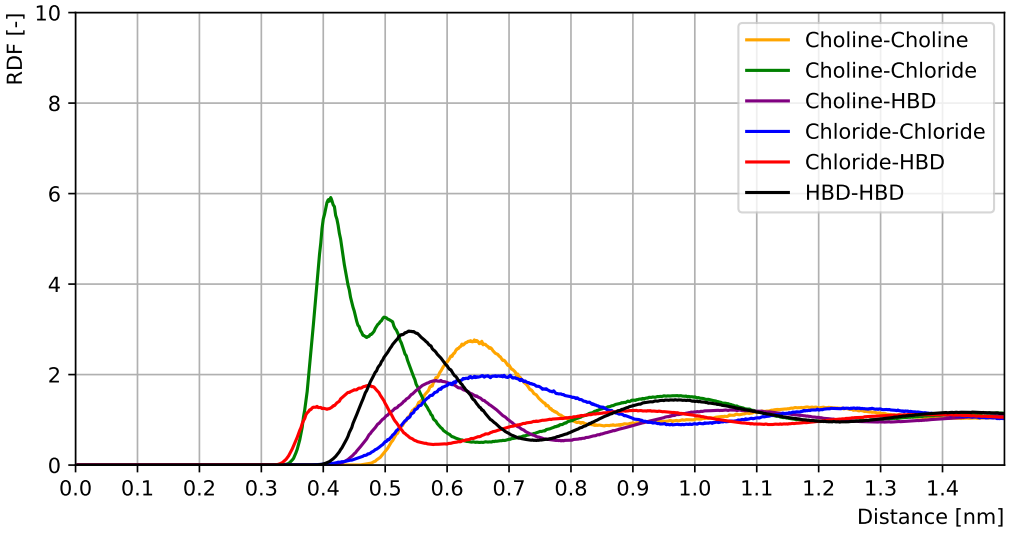
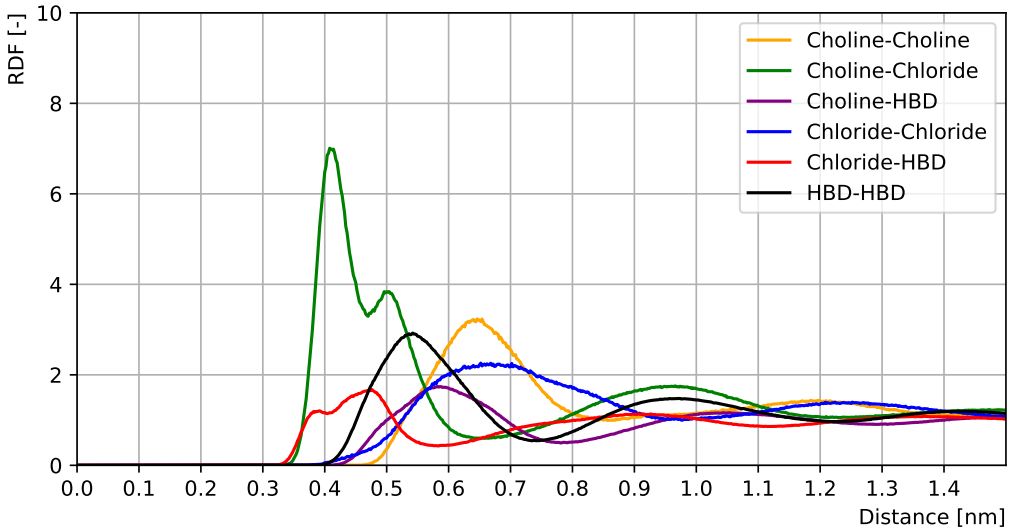


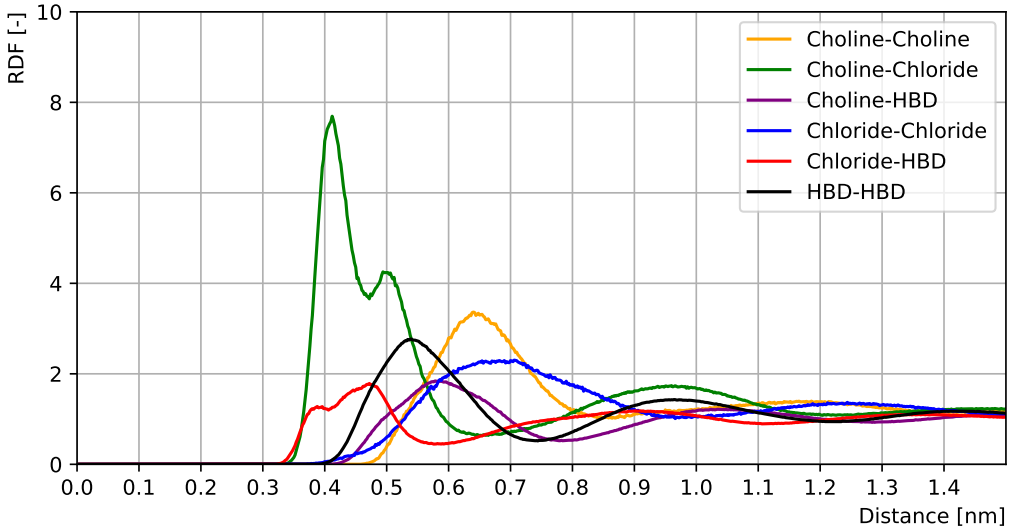
Figure A.4: COM RDF of CCUR12 system



(a)

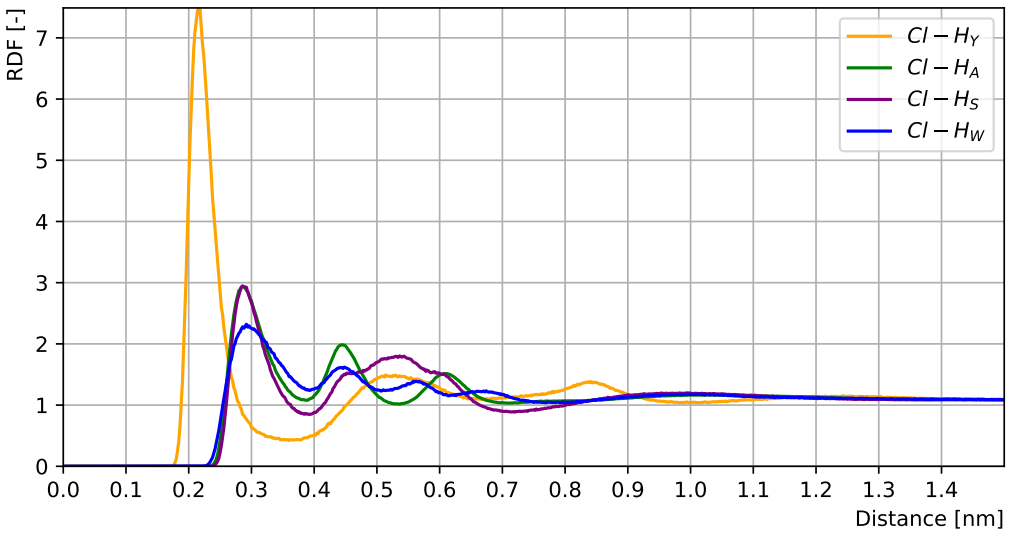


(b)

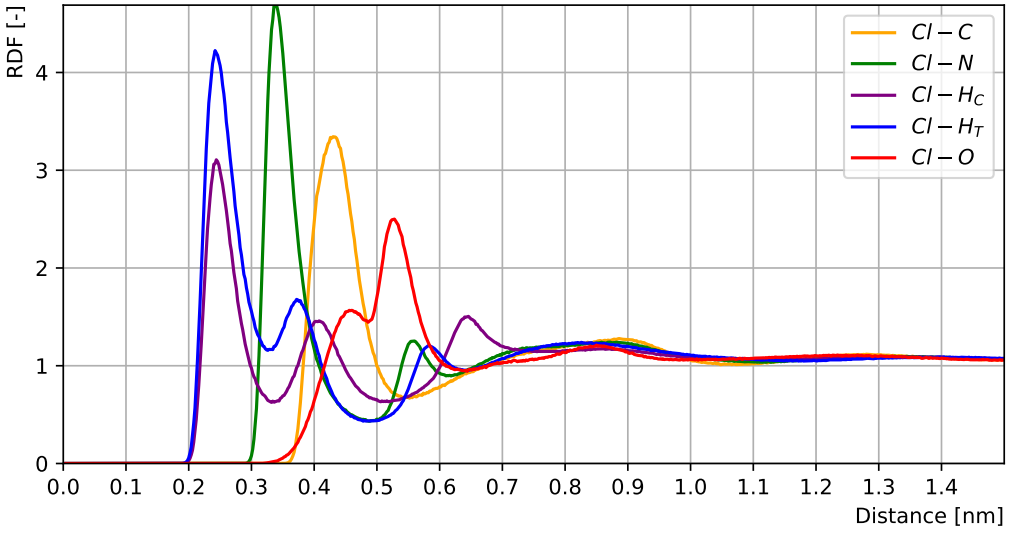


(c)

Figure A.5: COM RDF of (a) CCGLY12 , (b) CCGLY13 and (c) CCGLY14 NADES systems

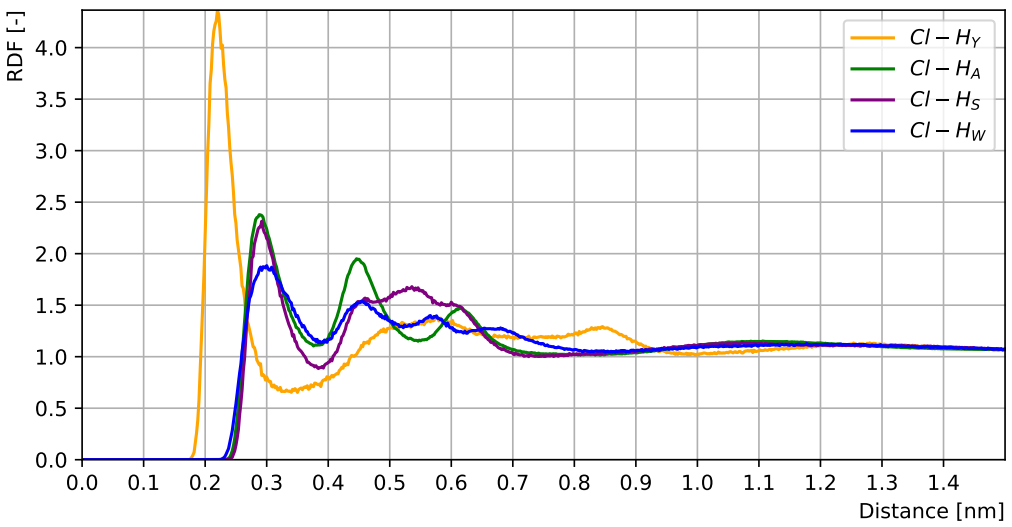


(a)

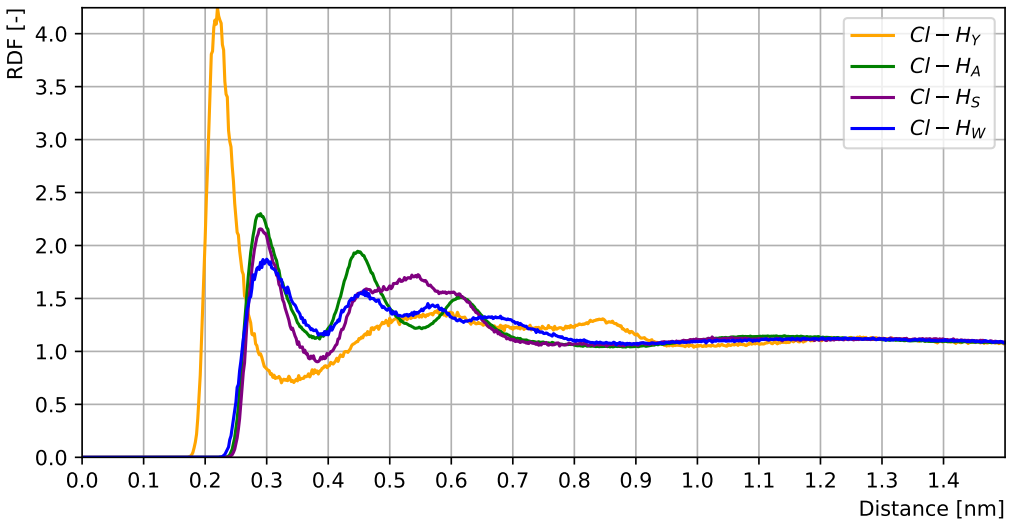


(b)

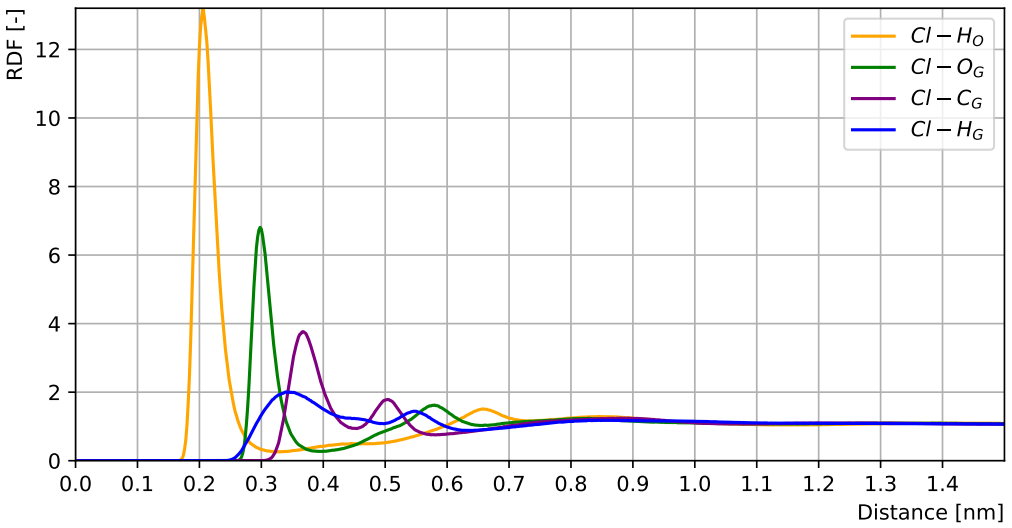
Figure A.6: Atom-atom RDFs of CCUR12 system: (a) HBA and (b) HBD


RDF [-]
Distance [nm]
$Cl - H_Y$
$Cl - H_A$
$Cl - H_S$
$Cl - H_W$

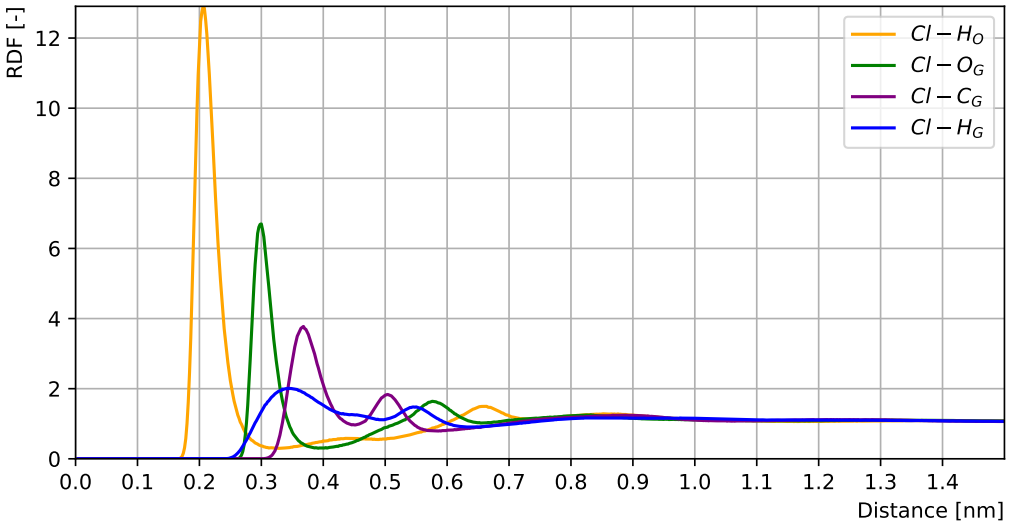
(a)


RDF [-]
Distance [nm]
$Cl - H_Y$
$Cl - H_A$
$Cl - H_S$
$Cl - H_W$

(b)

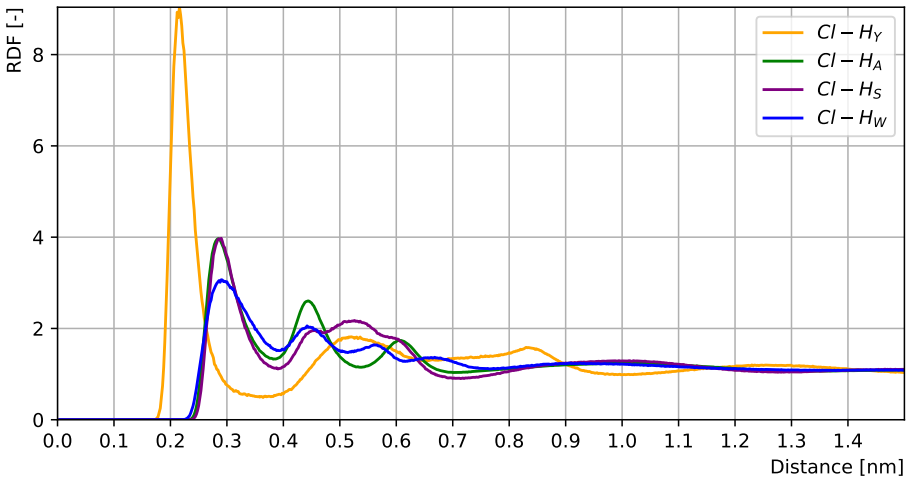


(c)

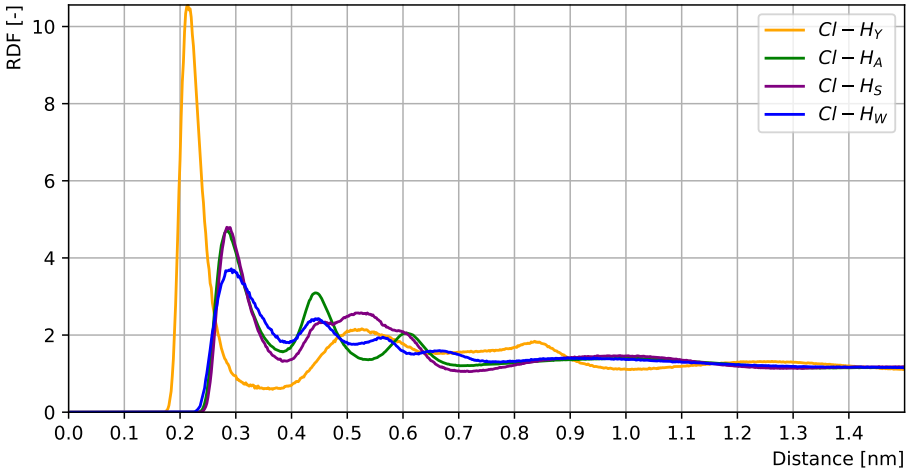


(d)

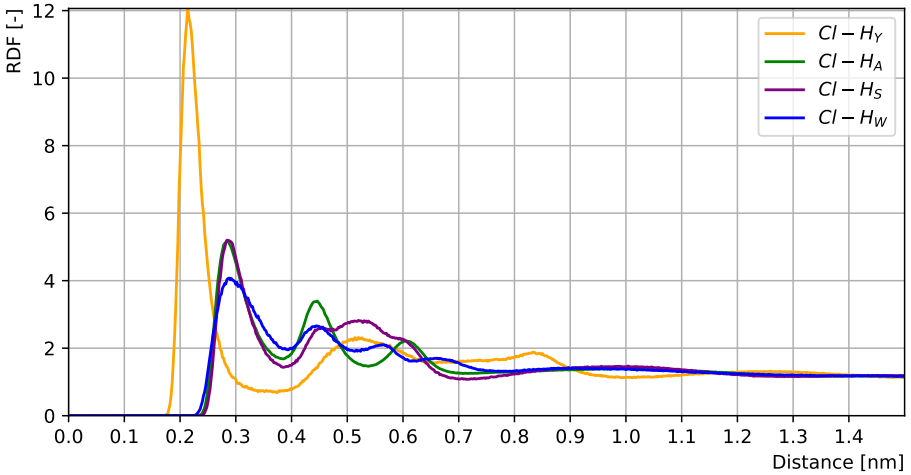
Figure A.7: Atom-atom RDFs of (a) CCEG13 HBA, (b) CCEG14 HBA, (c) CCEG13 HBD and (d) CCEG14 HBD


RDF [-]
Distance [nm]
$Cl - H_Y$
$Cl - H_A$
$Cl - H_S$
$Cl - H_W$

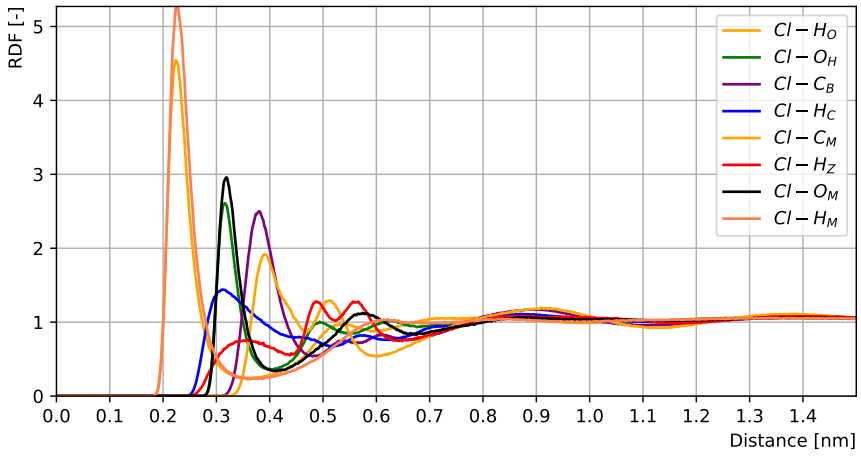
(a)


RDF [-]
Distance [nm]
$Cl - H_Y$
$Cl - H_A$
$Cl - H_S$
$Cl - H_W$

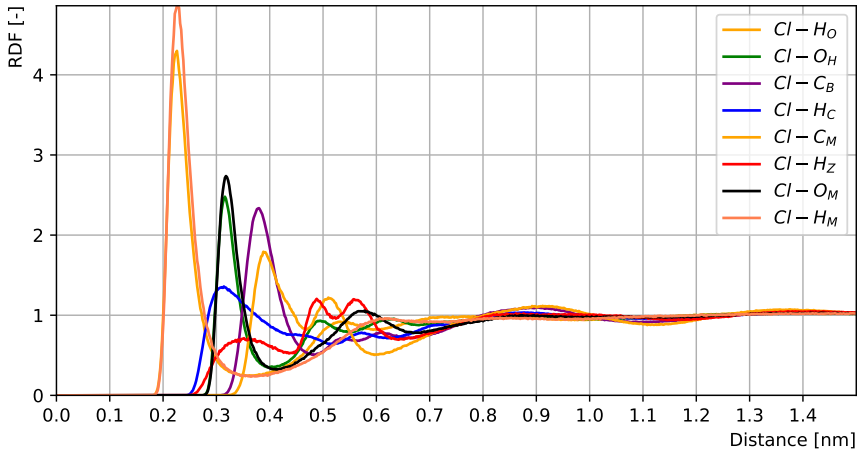
(b)


RDF [-]
Distance [nm]
$Cl - H_Y$
$Cl - H_A$
$Cl - H_S$
$Cl - H_W$

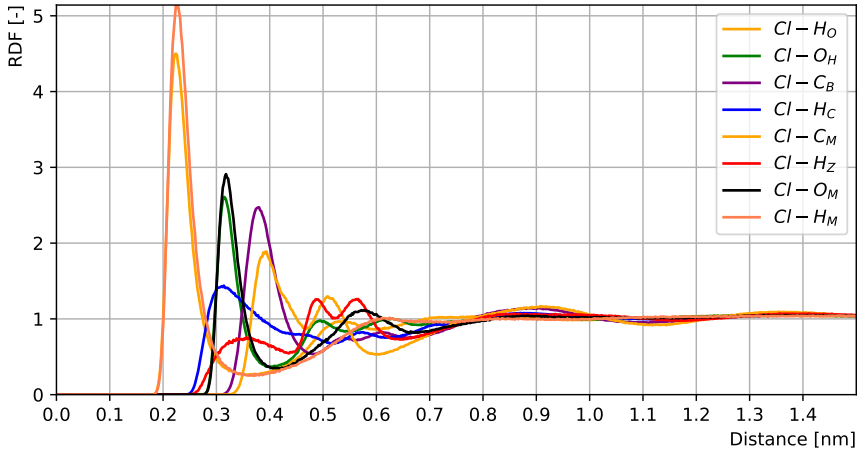
(c)



(d)

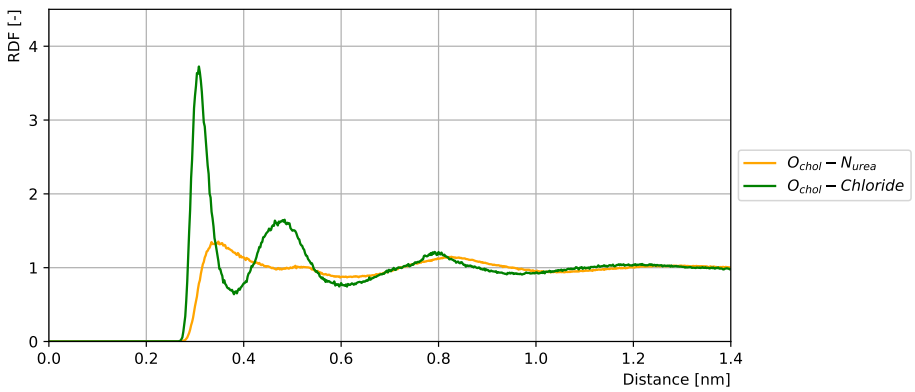


(e)

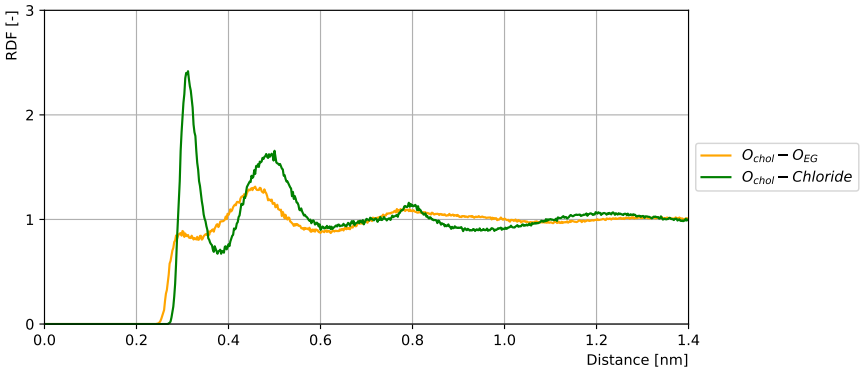


(f)

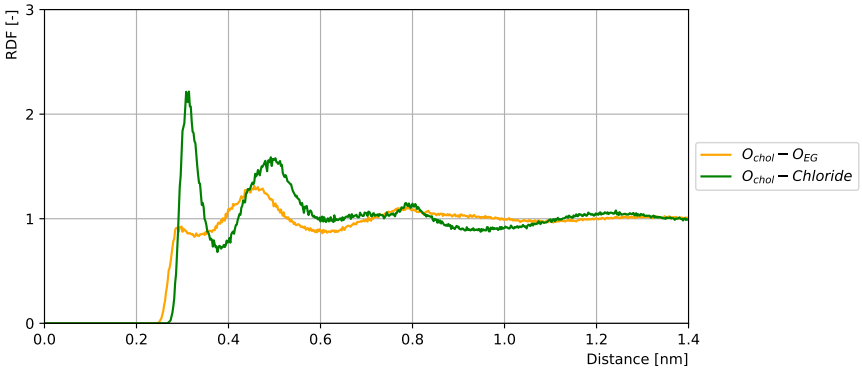
Figure A.8: Atom-atom RDFs of (a) CCGLY12 HBA, (b) CCGLY13 HBA, (c) CCGLY14 HBA , (d) CCGLY12 HBD, (e) CCGLY13 HBA and (f) CCGLY14 HBD


RDF [-]
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Distance [nm]
$O_{chol} - N_{urea}$
$O_{chol} - Chloride$

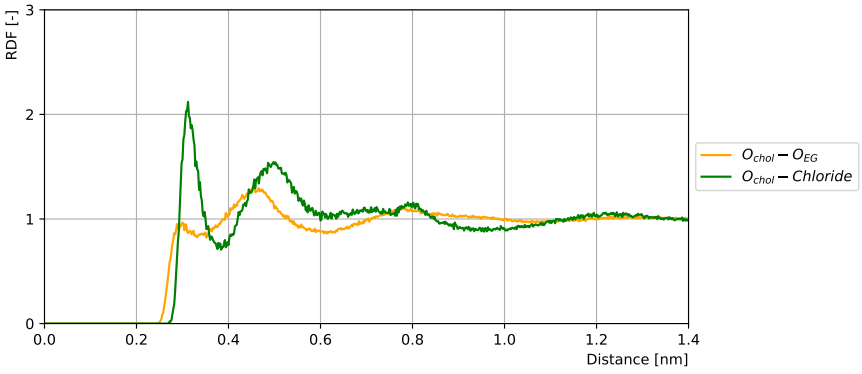
(a)


RDF [-]
Distance [nm]
$O_{chol} - O_{EG}$
$O_{chol} - Chloride$

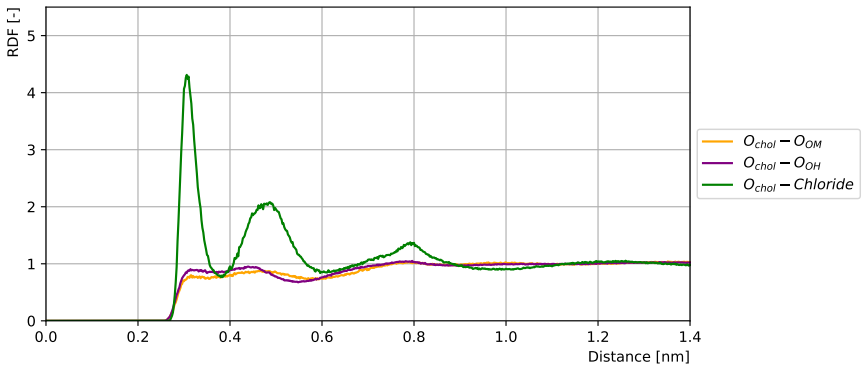
(b)


RDF [-]
Distance [nm]
$O_{chol} - O_{EG}$
$O_{chol} - Chloride$

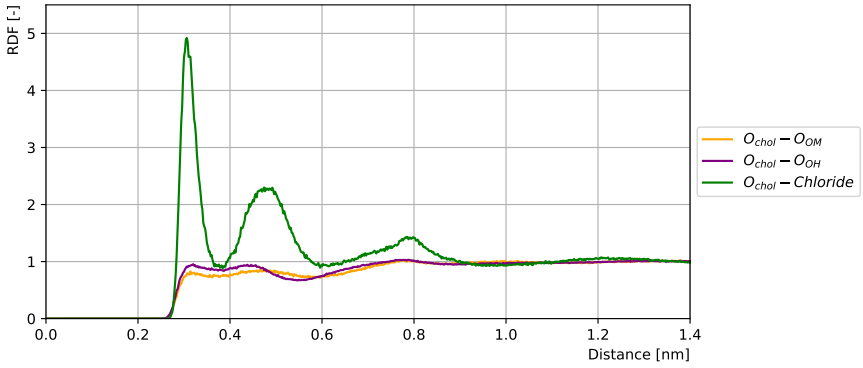
(c)


RDF [-]
Distance [nm]
$O_{chol} - O_{EG}$
$O_{chol} - Chloride$

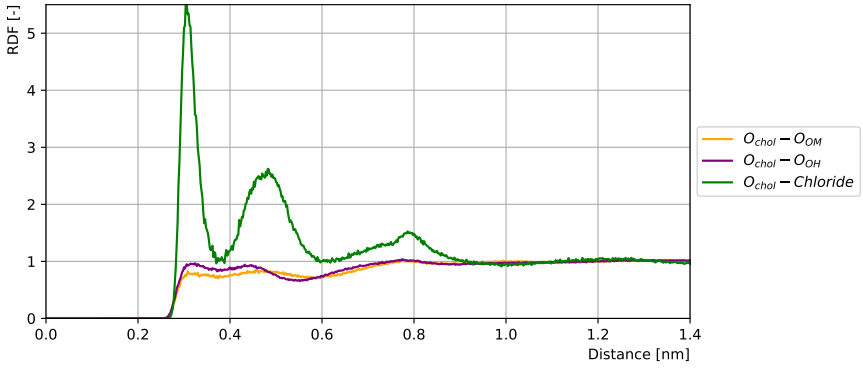
(d)



(e)



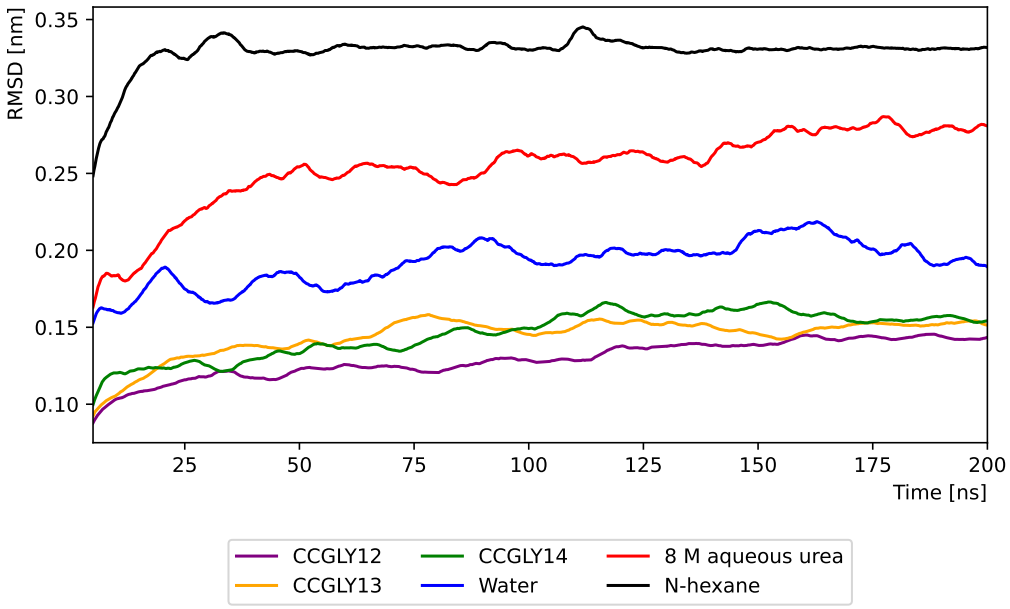
(f)



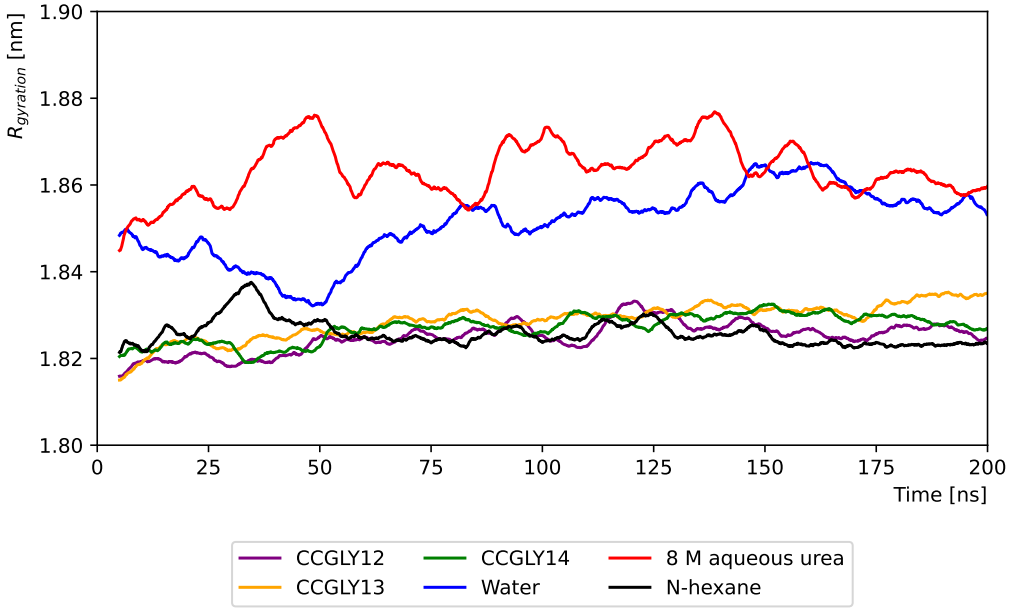
(g)

Figure A.9: Atom-atom RDFs between the oxygen of the hydroxyl group in choline and the nitrogen atoms of the amine groups in urea in (a) CCUR12, the oxygen atoms of hydroxyl groups in ethylene glycol in (b) CCEG12, (c) CCEG13 and (d) CCEG12 and the oxygen atoms of hydroxyl groups in glycerol in (e) CCGLY12, (f) CCGLY13 and (g) CCGLY14 systems. The RDF of the chloride anion was also plotted to compare the compounds direct interactions and the interactions via the anion.

Figure A.10: Graphical representation of the CALB enzyme structure in CCEG12 after 200 ns proudction run. The color coding shows the different structural parts of the enzyme: α-helix (purple), β-sheet (yellow) and random coil (cyan).

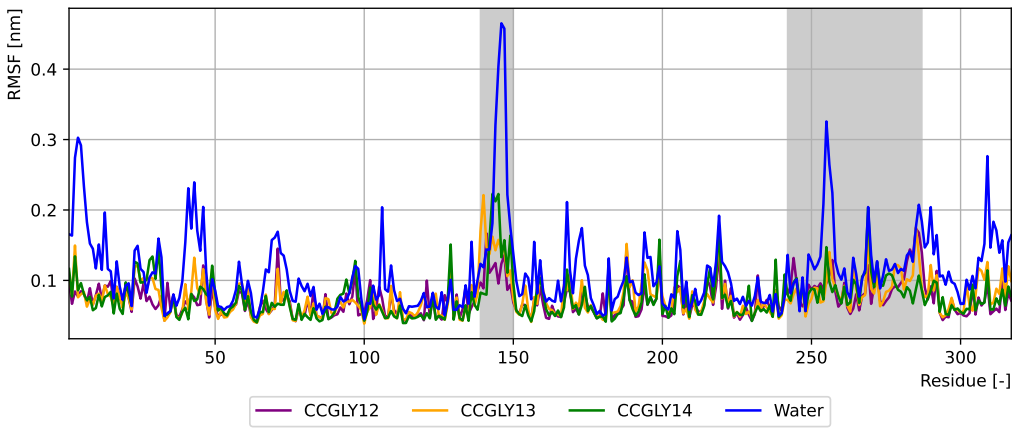


(a)



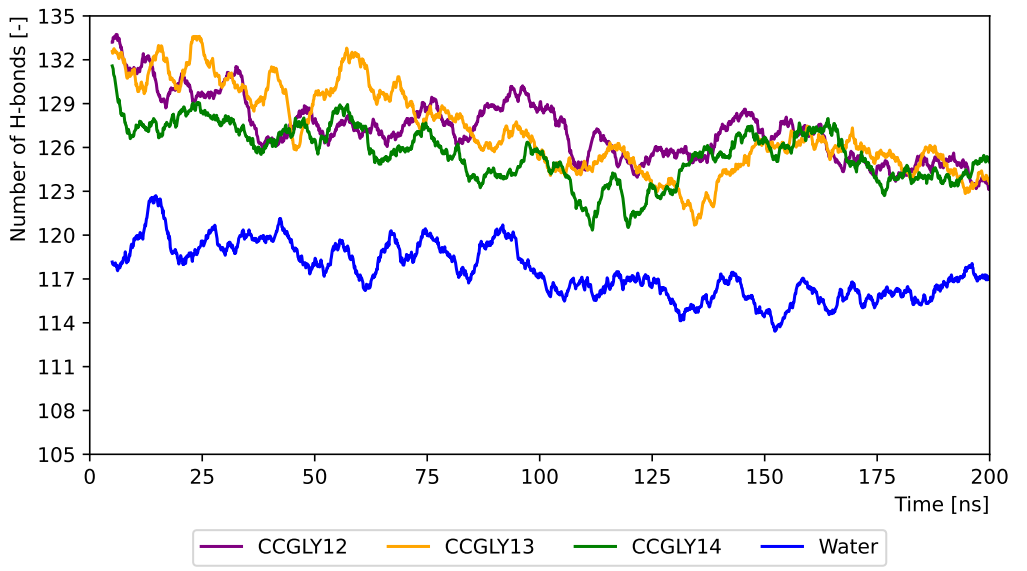
(b)

Figure A.11: (a) RMSD and (b) R_g values of CCGLY systems



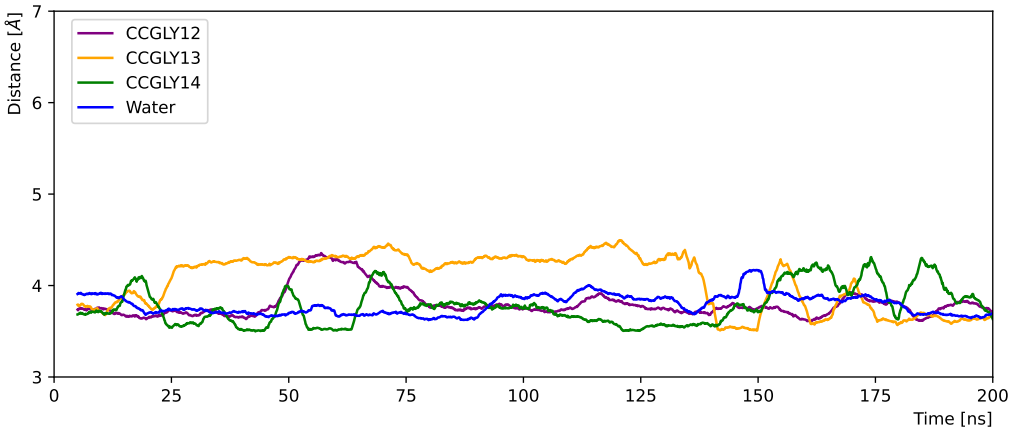
(a)

Figure A.12: RMSF values of CCGLY systems

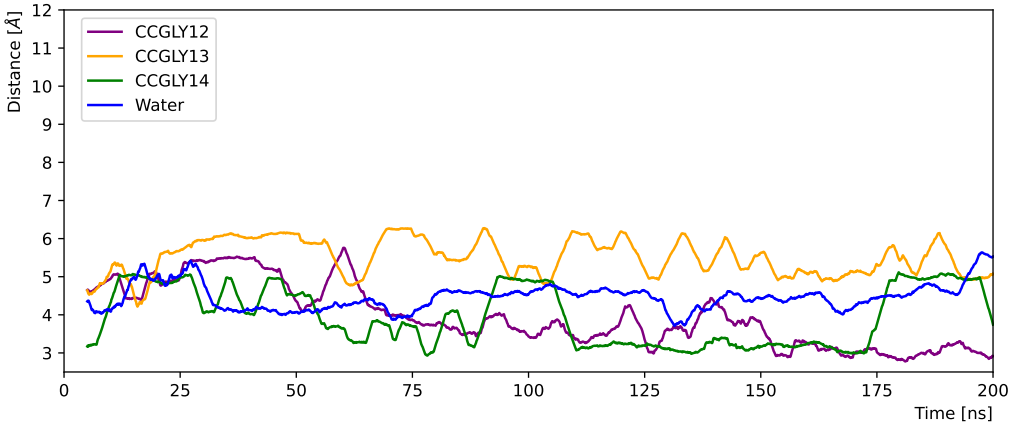


(a)

Figure A.13: Number of intra-main chain hydrogen bonds in the enzyme in CCGLY systems

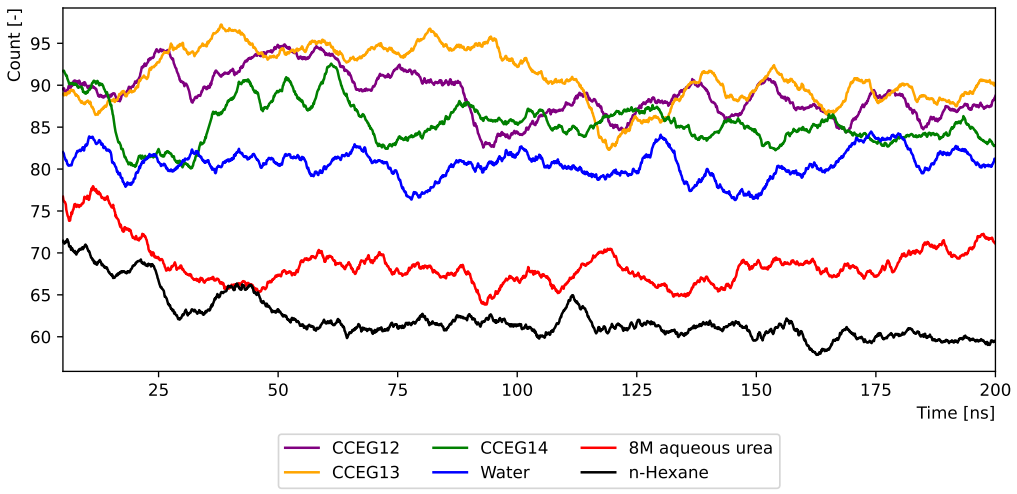


(a)

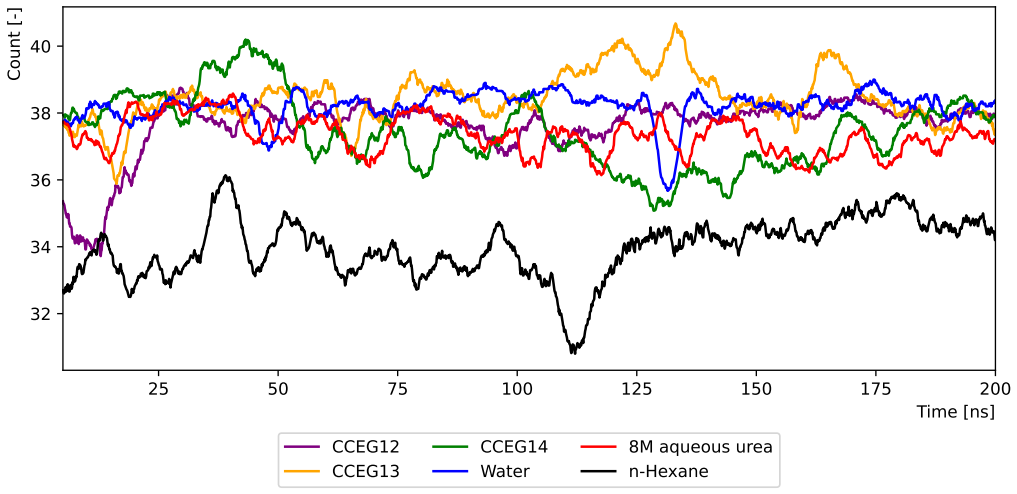


(b)

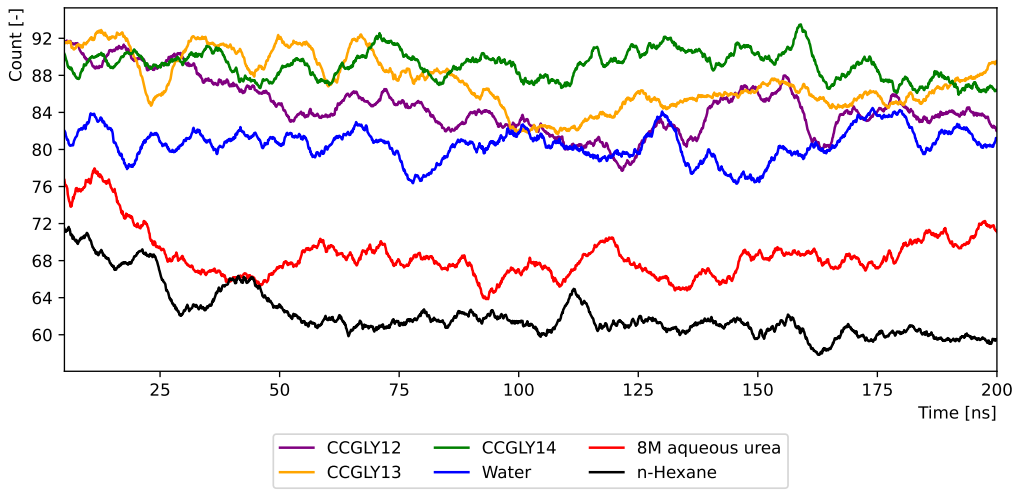
Figure A.14: Catalytic triad distances in CCGLY systems: (a) ASP187-HIS224 and (b) HIS224-SER105


Count [-]
95
90
85
80
75
70
65
60
25
50
75
100
125
150
175
200
Time [ns]
CCEG12
CCEG13
CCEG14
Water
8M aqueous urea
n-Hexane

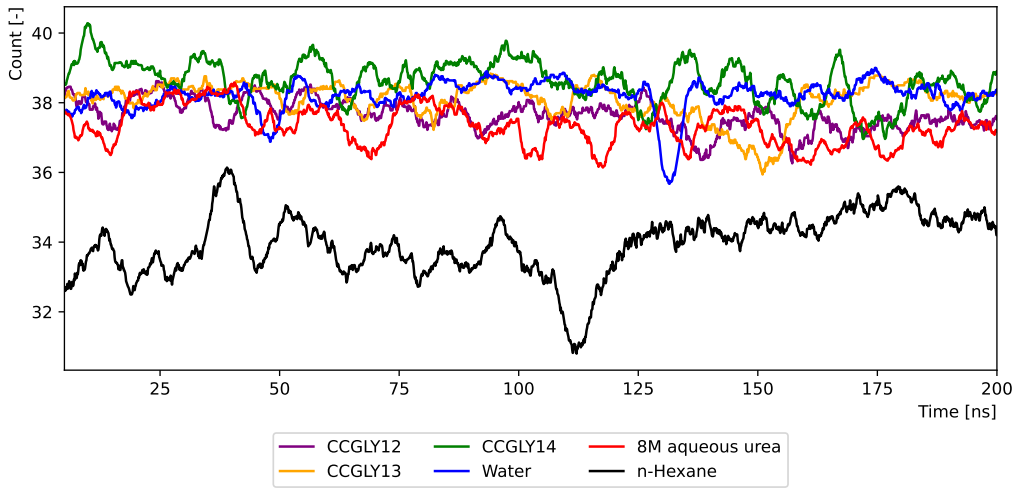
(a)


Count [-]
40
38
36
34
32
25
50
75
100
125
150
175
200
Time [ns]
CCEG12
CCEG13
CCEG14
Water
8M aqueous urea
n-Hexane

(b)

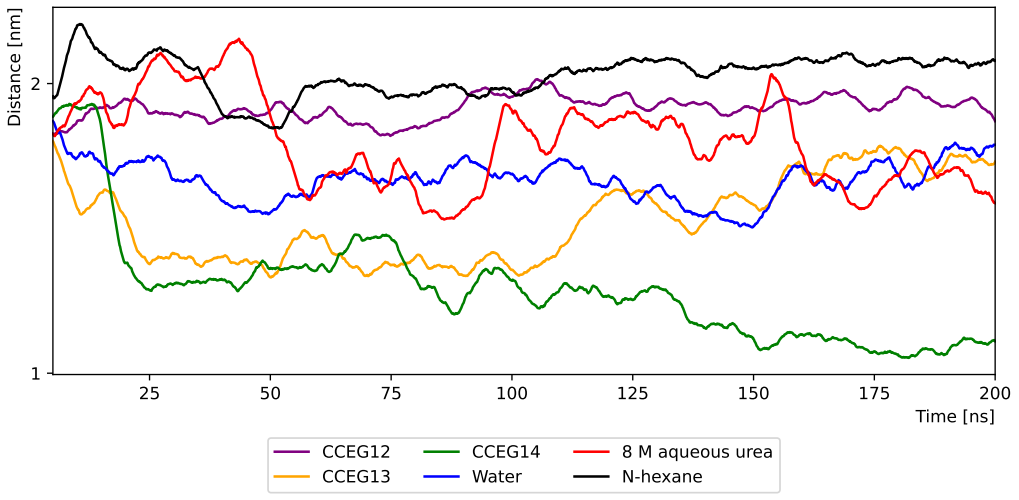


(c)

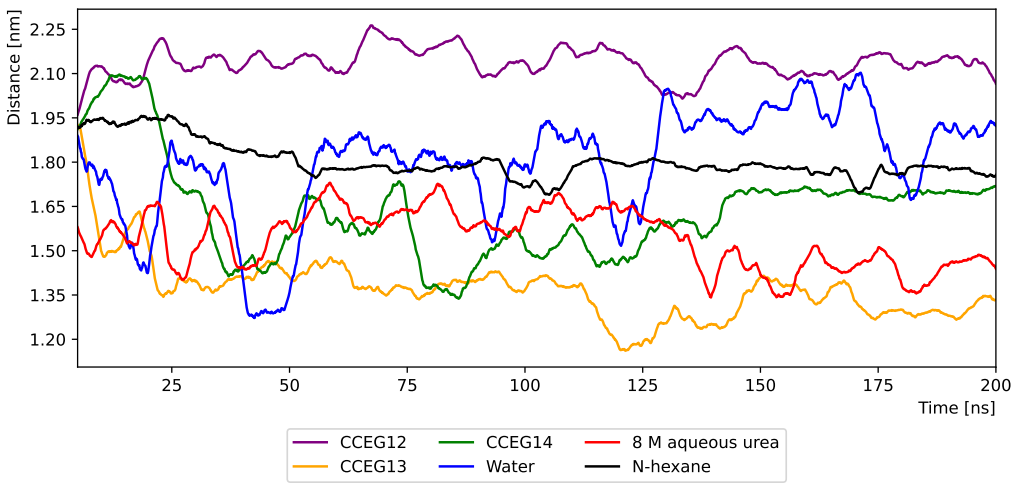


(d)

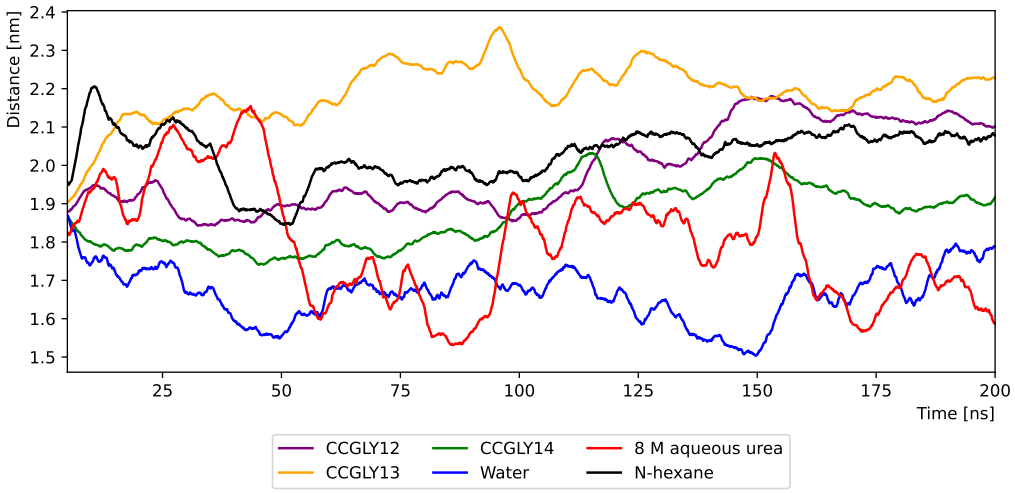
Figure A.15: Changes of secondary structure of the enzyme in CCEG and CCGLY systems: (a) number of residues in α-helix in CCEG, (b) number of residues in β-sheet in CCEG, (c) number of residues in α-helix in CCGLY and (d) number of residues in β-sheet in CCGLY


Distance [nm]
2
1
25
50
75
100
125
150
175
200
Time [ns]
CCEG12
CCEG13
CCEG14
Water
8 M aqueous urea
N-hexane

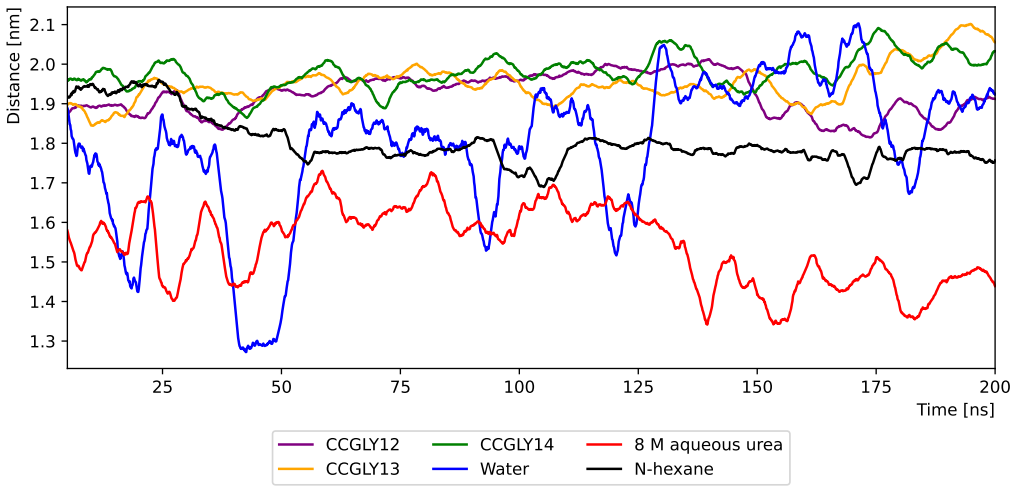
(a)


Distance [nm]
2.25
2.10
1.95
1.80
1.65
1.50
1.35
1.20
25
50
75
100
125
150
175
200
Time [ns]
CCEG12
CCEG13
CCEG14
Water
8 M aqueous urea
N-hexane

(b)

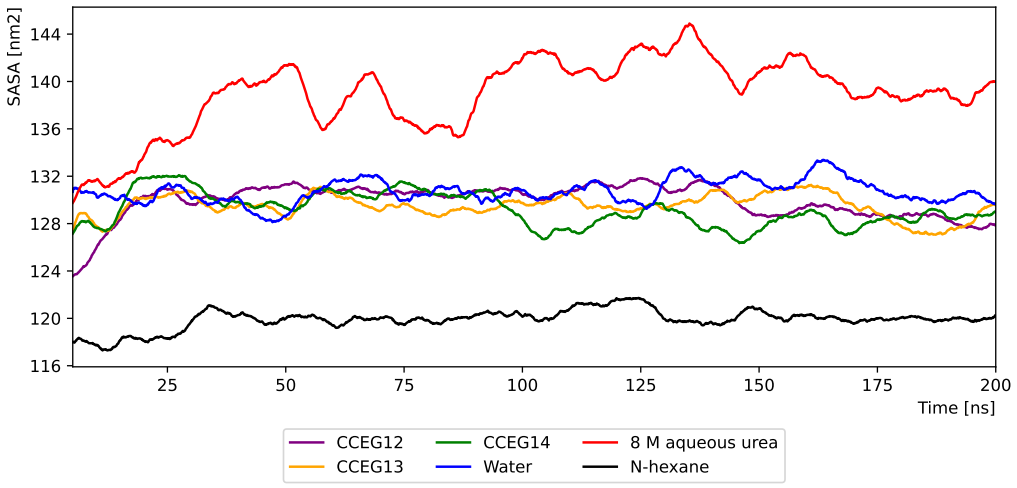


(c)

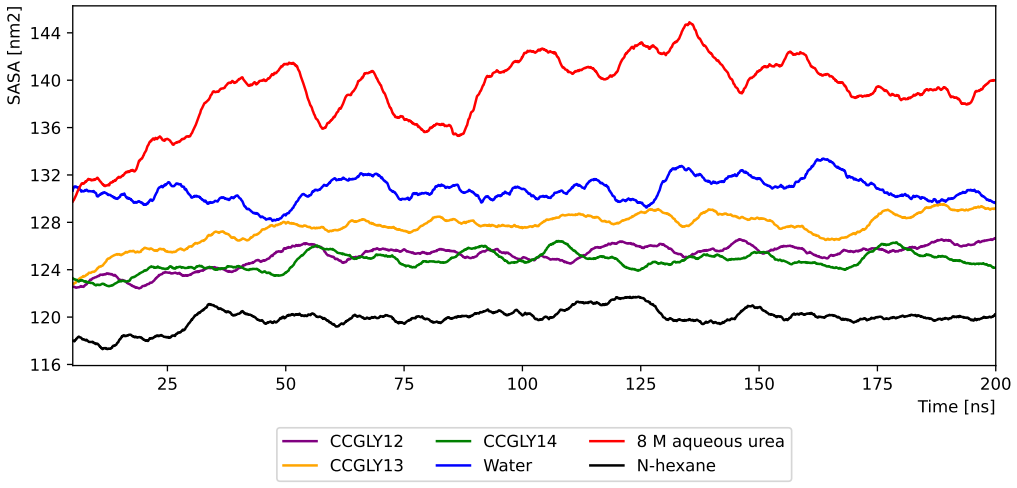


(d)

Figure A.16: Changes of helical distances between α5 and α10 helices of the enzyme in CCEG and CCGLY systems: (a) distance between residues GLY142 and LEU278 in CCEG, (b) distance between residues ALA146 and ALA287 in CCEG, (c) distance between residues GLY142 and LEU278 in CCGLY and (d) distance between residues ALA146 and ALA287 in CCEG

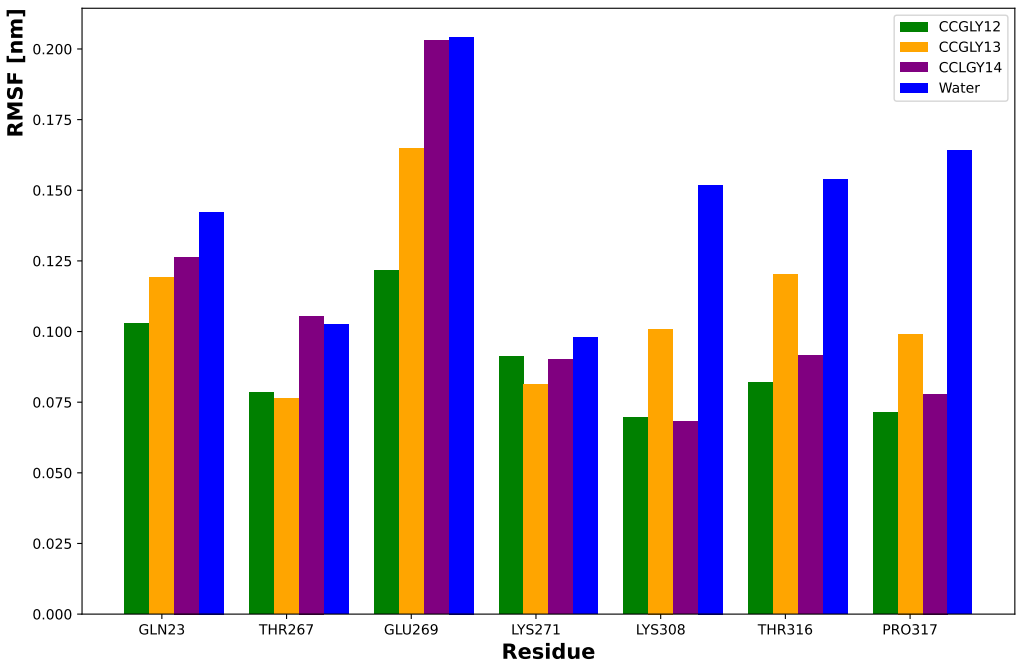


(a)



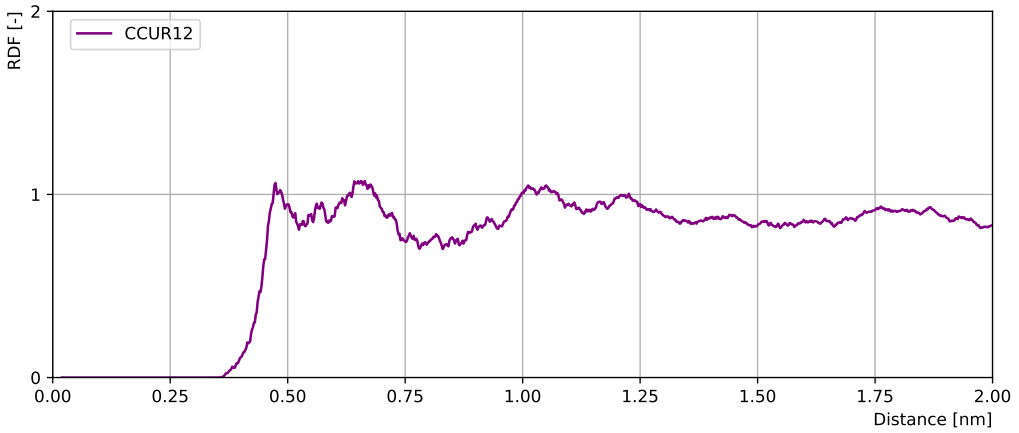
(b)

Figure A.17: SASA in (a) CCEG and (b) CCGLY systems

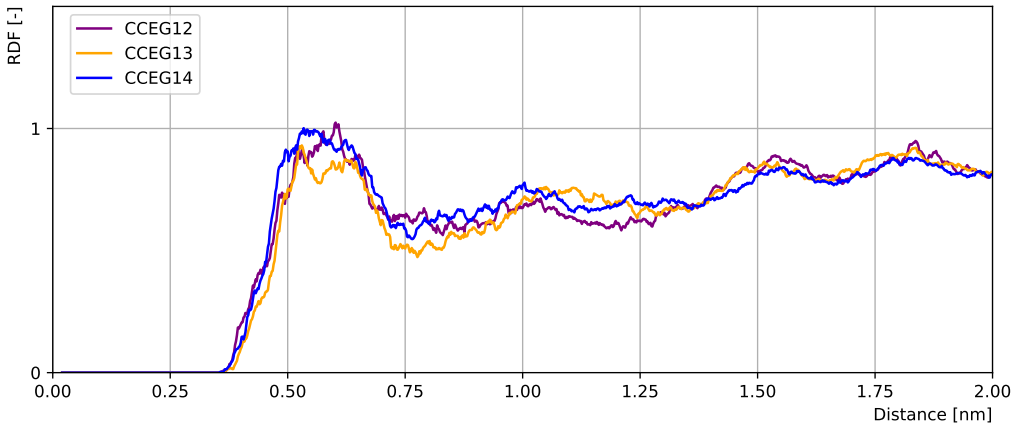


(a)

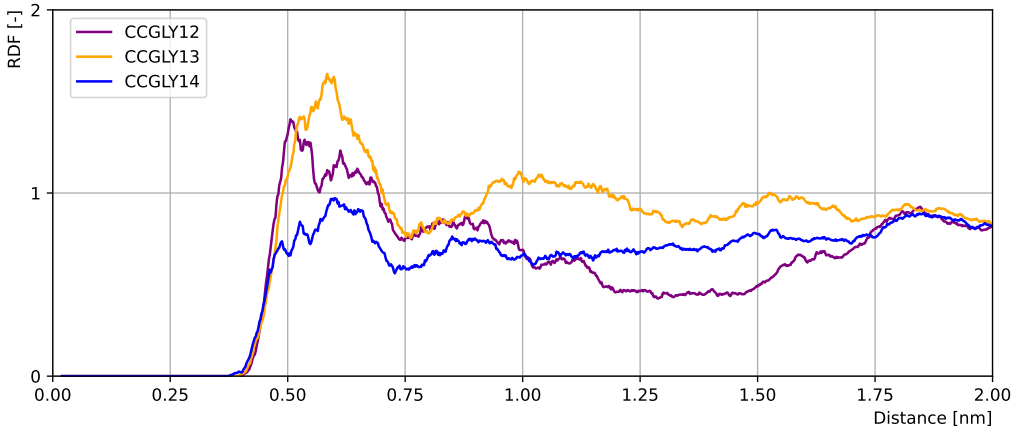
Figure A.18: RMSF of surface residues in CCGLY systems



(a)

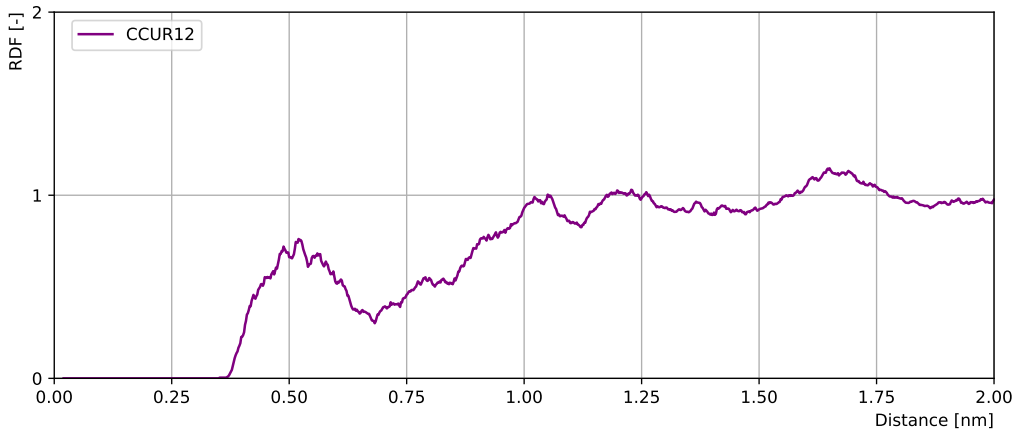


(b)

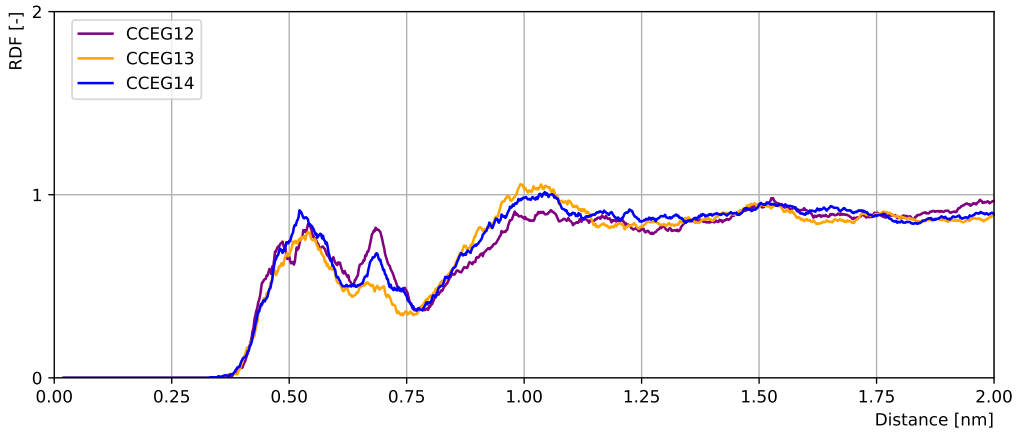


(c)

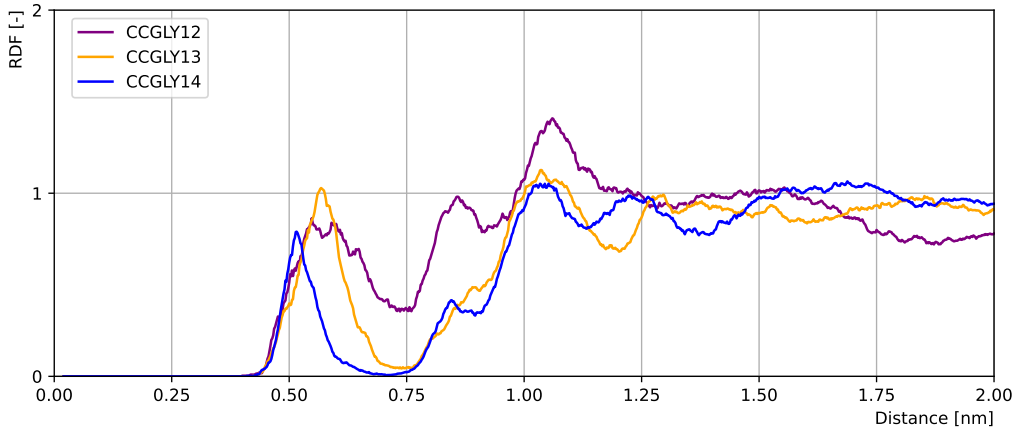
Figure A.19: Center-of-mass RDF of surface residue GLN23 and HBD molecules in (a) CCUR, (b) CCEG and (c) CCGLY systems



(a)

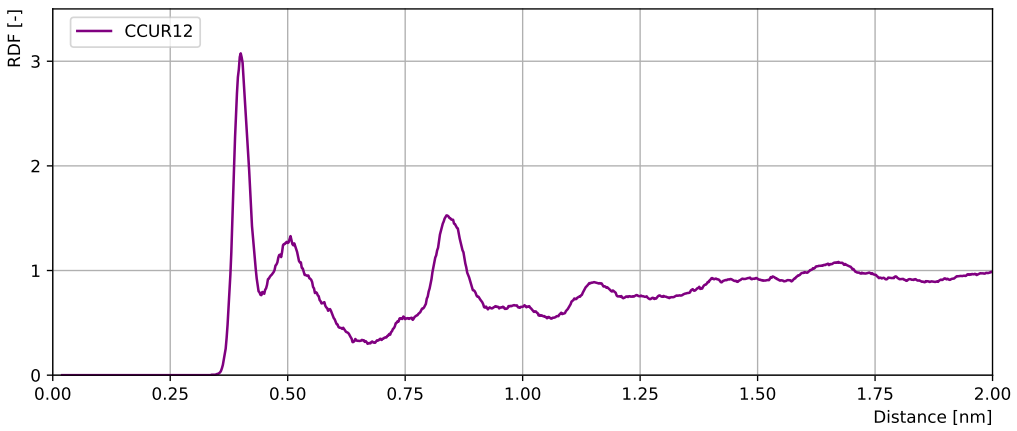


(b)

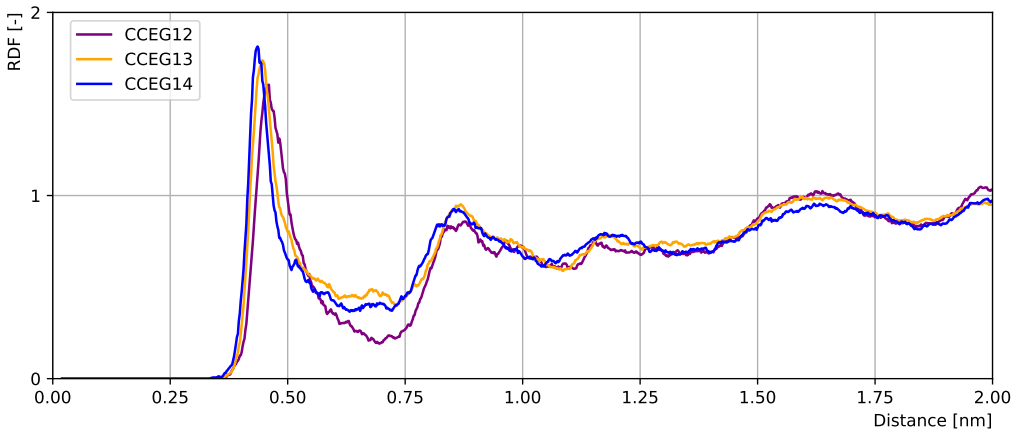


(c)

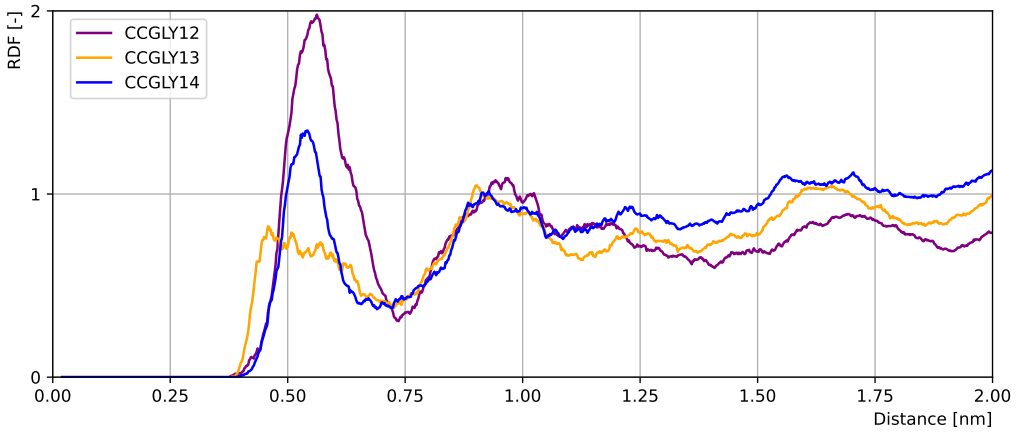
Figure A.20: Center-of-mass RDF of surface residue LYS271 and HBD molecules in (a) CCUR, (b) CCEG and (c) CCGLY systems



(a)

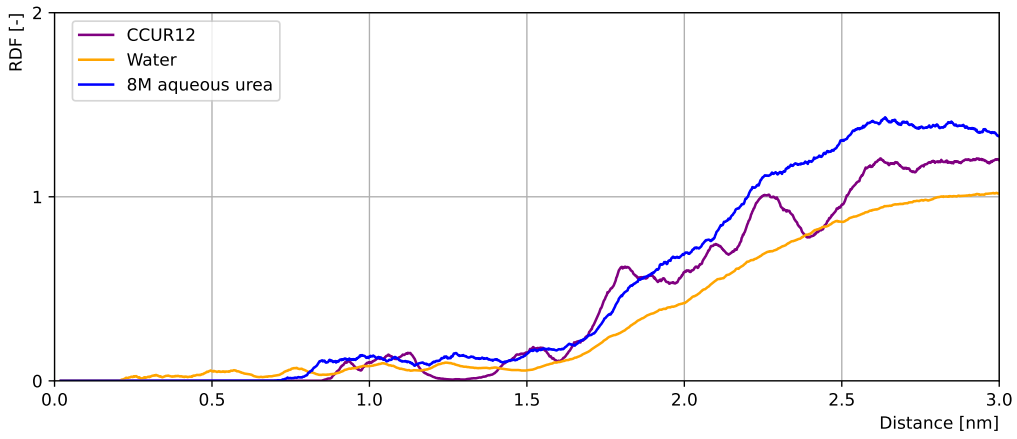


(b)

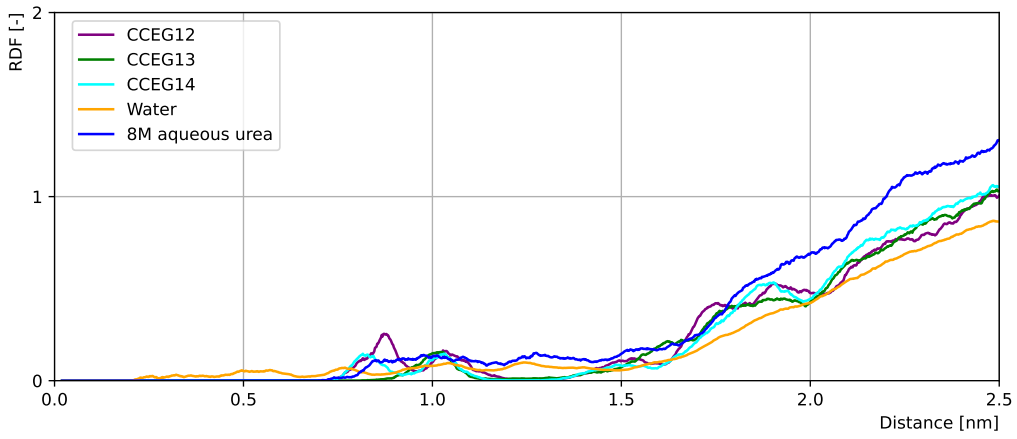


(c)

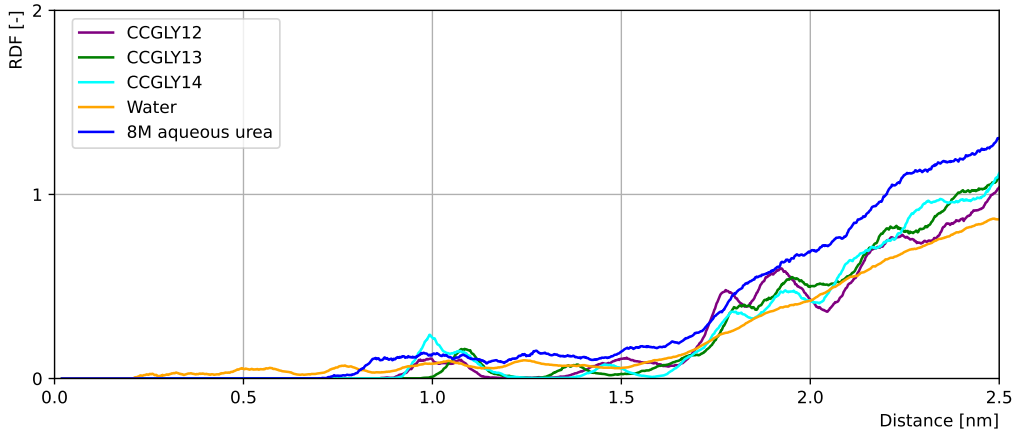
Figure A.21: Center-of-mass RDF of surface residue LYS308 and HBD molecules in (a) CCUR, (b) CCEG and (c) CCGLY systems



(a)



(b)



(c)

Figure A.22: Center-of-mass RDF of the enzyme and HBD molecules in (a) CCUR, (b) CCEG and (c) CCGLY and reference systems

A.3 Case study: deacetylation of mannosylerythritol lipids in NADES

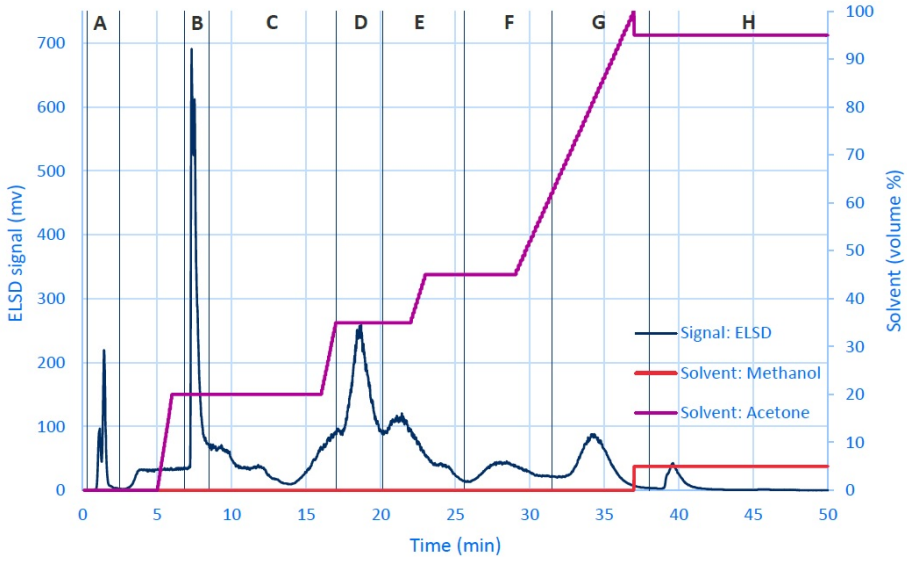


Figure A.23: Obtained chromatogram from the flash separation of a crude MEL mixture with the modified method. Fractions: A: Residual oil; B: Free fatty acids; C: Tri-acylated MELs; D: MEL-A; E: MEL-B; F: MEL-C; G: MEL-D; H: Residue.

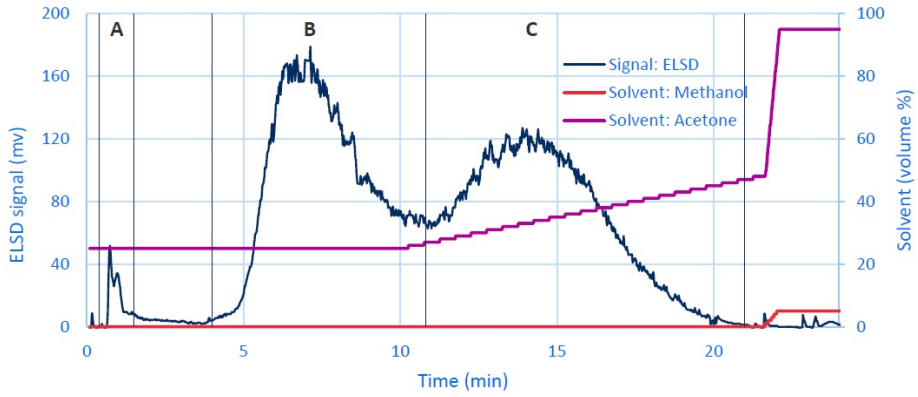


Figure A.24: Flash chromatogram obtained from the separation of MEL A and MEL B. Fractions: A: tri acylated MELs; B: MEL A; C: MEL B.

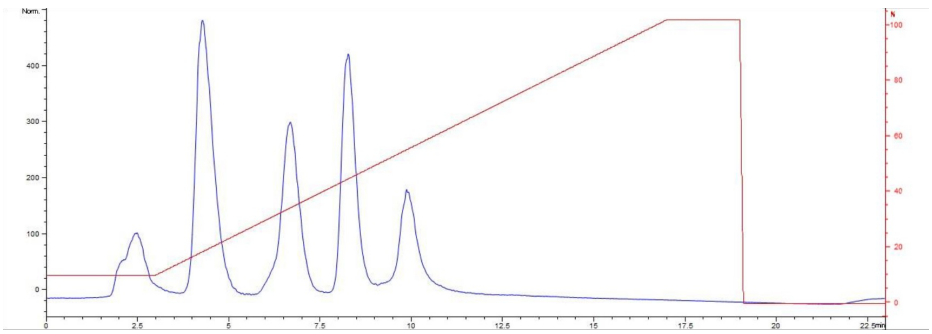


Figure A.25: Obtained chromatogram by injecting a mixture of MEL-A, MEL-B, MEL-C and MEL-D using the modified method with spiked eluent. Blue = ELSD signal; Red = gradient: % isopropanol. Peaks: 2.5 min: plastic residue; 4.5min: MEL-A; 6.2min: MEL-B; 8.5min: MEL-C; 10min: MEL-D.

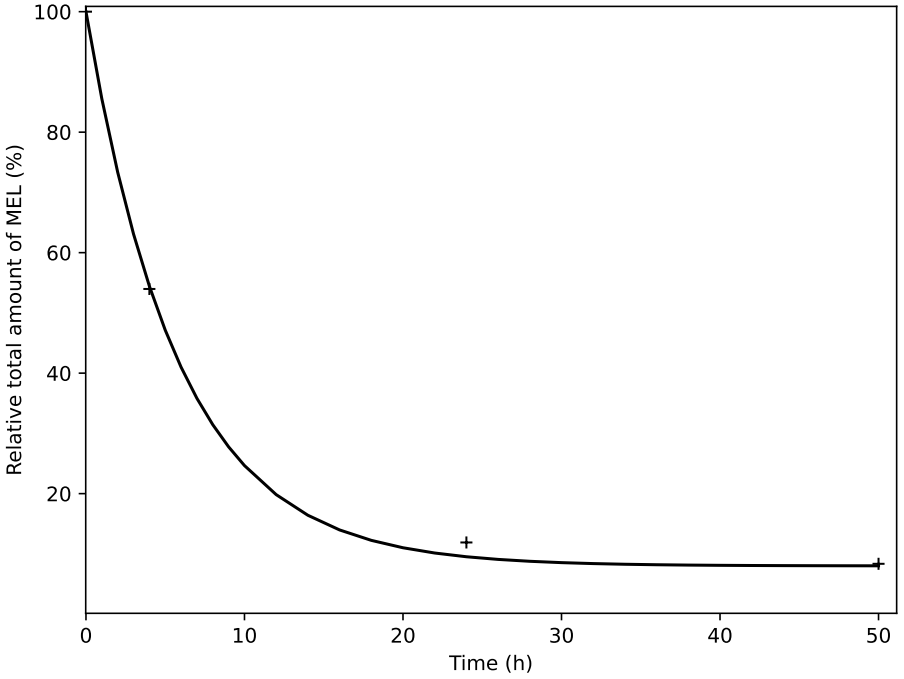


Figure A.26: Decrease of the original MEL content in the menthol – octanoic acid NADES.

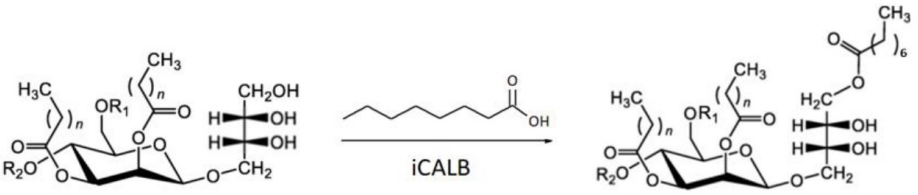


Figure A.27: Reaction of octanoic acid with MEL compounds and formation of triacylated compounds.

Appendix B

Personal information

"Well, I am a dilettante. It's only in England that dilettantism is considered a bad thing. In other countries it's called interdisciplinary research."

Brian Eno

B.1 Education

Budapest University of Technology and Economics, Bachelor's Degree, Chemical Engineering, 2011-2015, Grade: Good, Bachelor thesis: Preparation and characterization of thin chitosan-gelatin coatings

Budapest University of Technology and Economics, Master's Degree, Chemical Engineering, 2015-2017, Grade: Degree with honour, Master thesis: Investigation of the interactions between active ingredients and polymer stabilizers in nanodispersions

University of Antwerp, PhD, Applied Engineering, 2019-2022, Thesis: A structured methodology for NADES selection and formulation for enzymatic reactions

B.2 Experience

InnoStudio Inc., Budapest, Hungary, Intern-Research fellow, Formulation of polymer stabilized nanodispersions by flow chemistry, 2014-2017

Polinvent Ltd., Gyál, Hungary, Research and Development Chemical engineer, Development of UV-cured methacrylate resin systems, 2017-2018

University of Antwerp, Antwerp, Belgium, Research fellow, Development of structured methdology for NADES selection and formulation for enzymatic reactions, 2018-2022

B.3 Conference contributions

1. events in Flanders Focusing on Engineering and Chemical Technology, second cycle (eFFECT2), 2019, national conference, poster presentation: The green solvents of tomorrow: Natural deep eutectic solvents in biocatalytic reactions

2. 31st European Biomass Conference & Exhibit (EUBCE), 2019, Lisbon, Portugal, international conference, oral presentation: Enzymatic conversion of mannosylerythritol lipids in natural deep eutectic solvents

3. 1st International Meeting on Deep Eutectic Systems (IMDES), 2019, Lisbon, Portugal, international conference, poster presentation: Enzymatic Conversion Of Mannosylerythritol Lipids In Natural Deep Eutectic Solvents

4. 1st International GREENERING Conference, 2021, Costa de Caparica, Portugal, online international conference, poster presentation: Towards the mechanistic understanding of natural deep eutectic solvents effect on enzyme catalyzed reactions

5. 2nd International Meeting on Deep Eutectic Systems (IMDES), 2021, Lisbon, Portugal, online international conference, oral presentation: Towards the mechanistic understanding of natural deep eutectic solvents effect on enzyme catalyzed reactions

6. 32nd European Symposium on Computer Aided Process Engineering (ESCAPE-32), 2022, Toulouse, France, international conference, oral presentation: Predicting the density of natural deep eutectic solvents by the combination of a group-contribution method and artificial neural networks

7. 9th International Conference on Engineering for Waste and Biomass Valorisation (WasteEng), 2022, Copenhagen, Denmark, international conference, oral presentation: Understanding Biocatalysis in Natural Deep Eutectic Solvents: Transesterifications with Candida Antarctica Lipase B in Various Solvents

B.4 Publications

1. Kovács, A., Neyts, E.C., Cornet, I., Wijnants, M., Billen, P., 2020. Modeling the Physicochemical Properties of Natural Deep Eutectic Solvents. ChemSusChem 13, 3789–3804. https://doi.org/10.1002/cssc.202000286

2. Kovács, A., Yusupov, M., Cornet, I., Billen, P., Neyts, E.C., 2022. Effect of natural deep eutectic solvents of non-eutectic compositions on enzyme stability. Journal of Molecular Liquids 366, 120180. https://doi.org/10.1016/j.molliq.2022.120180

3. Sztancs, G., Kovács, A., Toth, A.J., Mizsey, P., Billen, P., Fozer, D., 2021. Catalytic hydrothermal carbonization of microalgae biomass for low-carbon emission power generation: the environmental impacts of hydrochar co-firing. Fuel 300, 120927. https://doi.org/10.1016/j.fuel.2021.120927

Bibliography

[1] Emma L. Smith, Andrew P. Abbott, and Karl S. Ryder. Deep eutectic solvents (dess) and their applications. *Chemical Reviews*, 114(21):11060–11082, oct 2014. 6, 16, 44, 63, 94, 139, 161

[2] Tracy El Achkar, Hélène Greige-Gerges, and Sophie Fourmentin. Basics and properties of deep eutectic solvents: a review. *Environmental Chemistry Letters*, 19(4):3397–3408, apr 2021. 6

[3] Laura J. B. M. Kollau, Mark Vis, Adriaan van den Bruinhorst, A. Catarina C. Esteves, and Remco Tuinier. Quantification of the liquid window of deep eutectic solvents. *Chemical Communications*, 54(95):13351–13354, 2018. 6

[4] Benworth B. Hansen, Stephanie Spittle, Brian Chen, Derrick Poe, Yong Zhang, Jeffrey M. Klein, Alexandre Horton, Laxmi Adhikari, Tamar Zelovich, Brian W. Doherty, Burcu Gurkan, Edward J. Maginn, Arthur Ragauskas, Mark Dadmun, Thomas A. Zawodzinski, Gary A. Baker, Mark E. Tuckerman, Robert F. Savinell, and Joshua R. Sangoro. Deep eutectic solvents: A review of fundamentals and applications. *Chemical Reviews*, 121(3):1232–1285, dec 2020. 6, 7, 148

[5] Qinghua Zhang, Karine De Oliveira Vigier, Sebastien Royer, and Francois Jerome. Deep eutectic solvents: syntheses, properties and applications. *Chemical Society Reviews*, 41(21):7108, 2012. 6, 16, 17, 19, 20, 22, 23, 25, 44, 63, 64, 94, 139, 161

[6] Andrew P. Abbott, Glen Capper, David L. Davies, Raymond K. Rasheed, and Vasuki Tambyrajah. Novel solvent properties of choline chloride/urea mixtures. *Chemical Communications*, (1):70–71, nov 2002. 6, 19, 20, 44

[7] David R. Lide;. *Handbook of Data on Common Organic Compounds*. CRC Press, 1994. 6

[8] Richard J. Lewis. *Hawley's Condensed Chemical Dictionary*. Wiley, 1997. 6

[9] Siwen Zhu, Hongping Li, Wenshuai Zhu, Wei Jiang, Chao Wang, Peiwen Wu, Qi Zhang, and Huaming Li. Vibrational analysis and formation mechanism of typical deep eutectic solvents: An experimental and theoretical study. *Journal of Molecular Graphics and Modelling*, 68:158–175, jul 2016. 6, 18, 28, 44

[10] Chao Zhang, Yongzhong Jia, Yan Jing, Huaiyou Wang, and Kai Hong. Main chemical species and molecular structure of deep eutectic solvent studied by experiments with dft calculation: a case of choline chloride and magnesium chloride hexahydrate. *Journal of Molecular Modeling*, 20(8), jul 2014. 6, 28, 44

[11] Attila Kovacs, Pieter Billen, Iris Cornet, Marc Wijnants, and Erik C. Neyts. Modeling the physicochemical properties of natural deep eutectic solvents - a review. *ChemSusChem*, may 2020. 6, 15, 44, 64, 95, 139

[12] Yang Liu, J. Brent Friesen, James B. McAlpine, David C. Lankin, Shao-Nong Chen, and Guido F. Pauli. Natural deep eutectic solvents: Properties, applications, and perspectives. *Journal of Natural Products*, 81(3):679–690, mar 2018. 6

[13] Young Hae Choi, Jaap van Spronsen, Yuntao Dai, Marianne Verberne, Frank Hollmann, Isabel W.C.E. Arends, Geert-Jan Witkamp, and Robert Verpoorte. Are natural deep eutectic solvents the missing link in understanding cellular metabolism and physiology? *Plant Physiology*, 156(4):1701–1705, jun 2011. 6, 17, 44, 63, 94, 139

[14] Alexandre Paiva, Rita Craveiro, Ivo Aroso, Marta Martins, Rui L. Reis, and Ana Rita C. Duarte. Natural deep eutectic solvents - solvents for the 21st century. *ACS Sustainable Chemistry & Engineering*, 2(5):1063–1071, mar 2014. 6, 16, 17, 44, 64, 94, 138

[15] Filippo Maria Perna, Paola Vitale, and Vito Capriati. Deep eutectic solvents and their applications as green solvents. *Current Opinion in Green and Sustainable Chemistry*, 21:27–33, feb 2020. 6

[16] Maan Hayyan, Yves Paul Mbous, Chung Yeng Looi, Won Fen Wong, Adeeb Hayyan, Zulhaziman Salleh, and Ozair Mohd-Ali. Natural deep eutectic solvents: cytotoxic profile. *SpringerPlus*, 5(1), jun 2016. 6, 44, 139

[17] Zhen Yang. Toxicity and biodegradability of deep eutectic solvents and natural deep eutectic solvents, nov 2019. 6, 44, 139

[18] María Francisco, Adriaan van den Bruinhorst, and Maaike C. Kroon. Low-transition-temperature mixtures (LTTMs): A new generation of designer solvents. *Angewandte Chemie International Edition*, 52(11):3074–3085, feb 2013. 6, 16, 17

[19] Manuela Panic, Marina Cvjetko Bubalo, and Ivana Radojcic Redovnikovic. Designing a biocatalytic process involving deep eutectic solvents. *Journal of Chemical Technology & Biotechnology*, 96(1):14–30, sep 2020. 6, 44, 139

[20] Coby J. Clarke, Wei-Chien Tu, Oliver Levers, Andreas Bröhl, and Jason P. Hallett. Green and sustainable solvents in chemical processes. *Chemical Reviews*, 118(2):747–800, jan 2018. 6, 44

[21] Philip G. Jessop. Green/alternative solvents. In *Encyclopedia of Sustainable Technologies*, pages 611–619. Elsevier, 2017. 6

[22] Junxiong Wu, Qinghua Liang, Xiaoliang Yu, Qiu-Feng Lü, Lianbo Ma, Xianying Qin, Guohua Chen, and Baohua Li. Deep eutectic solvents for boosting electrochemical energy storage and conversion: A review and perspective. *Advanced Functional Materials*, 31(22):2011102, mar 2021. 6

[23] Andrew P. Abbott, John Collins, Ian Dalrymple, Robert C. Harris, Reena Mistry, Fulian Qiu, James Scheirer, and William R. Wise. Processing of electric arc furnace dust using deep eutectic solvents. *Australian Journal of Chemistry*, 62(4):341, 2009. 6, 44, 139

[24] Yuntao Dai, Jaap van Spronsen, Geert-Jan Witkamp, Robert Verpoorte, and Young Hae Choi. Ionic liquids and deep eutectic solvents in natural products research: Mixtures of solids as extraction solvents. *Journal of Natural Products*, 76(11):2162–2173, nov 2013. 6, 22, 24, 44, 139

[25] Alessandra Procentese, Francesca Raganati, Giuseppe Olivieri, Maria Elena Russo, Lars Rehmann, and Antonio Marzocchella. Deep eutectic solvents pretreatment of agro-industrial food waste. *Biotechnology for Biofuels*, 11(1), feb 2018. 6

[26] Baranse Ozturk, Charles Parkinson, and Maria Gonzalez-Miquel. Extraction of polyphenolic antioxidants from orange peel waste using deep eutectic solvents. *Separation and Purification Technology*, 206:1–13, nov 2018. 6

[27] Yu-Loong Loow, Eng Kein New, Ge Hoa Yang, Lin Yang Ang, Luther Yang Wei Foo, and Ta Yeong Wu. Potential use of deep eutectic solvents to facilitate lignocellulosic biomass utilization and conversion. *Cellulose*, 24(9):3591–3618, jun 2017. 6

[28] Joan G. Lynam, Narendra Kumar, and Mark J. Wong. Deep eutectic solvents' ability to solubilize lignin, cellulose, and hemicellulose; thermal stability; and density. *Bioresource Technology*, 238:684–689, aug 2017. 6, 130

[29] Yu-Loong Loow, Ta Yeong Wu, Ge Hoa Yang, Lin Yang Ang, Eng Kein New, Lee Fong Siow, Jamaliah Md. Jahim, Abdul Wahab Mohammad, and Wen Hui Teoh. Deep eutectic solvent and inorganic salt pretreatment of lignocellulosic biomass for improving xylose recovery. *Bioresource Technology*, 249:818–825, feb 2018. 6

[30] Hua Zhao and Gary A. Baker. Ionic liquids and deep eutectic solvents for biodiesel synthesis: a review. *Journal of Chemical Technology & Biotechnology*, 88(1):3–12, nov 2012. 6

[31] Mohammad Karimi, Akbar Jodaei, Asiyeh Khajvandi, Amirhosein Sadeghinik, and Rahim Jahandideh. In-situ capture and conversion of atmospheric CO2 into nano-CaCO3 using a novel pathway based on deep eutectic choline chloride-calcium chloride. *Journal of Environmental Management*, 206:516–522, jan 2018. 6, 130

[32] Chunyan Ma, Shokat Sarmad, Jyri-Pekka Mikkola, and Xiaoyan Ji. Development of low-cost deep eutectic solvents for CO 2 capture. *Energy Procedia*, 142:3320–3325, dec 2017. 6

[33] Andrew P. Abbott, David Boothby, Glen Capper, David L. Davies, and Raymond K. Rasheed. Deep eutectic solvents formed between choline chloride and carboxylic acids: versatile alternatives to ionic liquids. *Journal of the American Chemical Society*, 126(29):9142–9147, jul 2004. 6, 19, 23, 25, 44

[34] Anika Söldner, Julia Zach, and Burkhard König. Deep eutectic solvents as extraction media for metal salts and oxides exemplarily shown for phosphates from incinerated sewage sludge ash. *Green Chemistry*, 21(2):321–328, 2019. 6

[35] Andrew P. Abbott, Glen Capper, David L. Davies, Katy J. McKenzie, and Stephen U. Obi. Solubility of metal oxides in deep eutectic solvents based on choline chloride. *Journal of Chemical & Engineering Data*, 51(4):1280–1282, jun 2006. 6, 130

[36] Zheng-Meng Jiang, Lan-Jin Wang, Zhao Gao, Bo Zhuang, Qiang Yin, and E-Hu Liu. Green and efficient extraction of different types of bioactive alkaloids using deep eutectic solvents. *Microchemical Journal*, 145:345–353, mar 2019. 6

[37] Mohamad Hamdi Zainal-Abidin, Maan Hayyan, Gek Cheng Ngoh, and Won Fen Wong. From nanoengineering to nanomedicine: A facile route to enhance biocompatibility of graphene as a potential nano-carrier for targeted drug delivery using natural deep eutectic solvents. *Chemical Engineering Science*, 195:95–106, feb 2019. 6

[38] Francisca Mano, Marta Martins, Isabel Sá-Nogueira, Susana Barreiros, João Paulo Borges, Rui L. Reis, Ana Rita C. Duarte, and Alexandre Paiva. Production of electrospun fast-dissolving drug delivery systems with therapeutic eutectic systems encapsulated in gelatin. *AAPS PharmSciTech*, 18(7):2579–2585, feb 2017. 6

[39] Michael Zakrewsky, Amrita Banerjee, Sanjana Apte, Theresa L. Kern, Mattie R. Jones, Rico E. Del Sesto, Andrew T. Koppisch, David T. Fox, and Samir Mitragotri. Choline and geranate deep eutectic solvent as a broad-spectrum antiseptic agent for preventive and therapeutic applications. *Advanced Healthcare Materials*, 5(11):1282–1289, mar 2016. 6

[40] Magdalena Pätzold, Sascha Siebenhaller, Selin Kara, Andreas Liese, Christoph Syldatk, and Dirk Holtmann. Deep eutectic solvents as efficient solvents in biocatalysis. *Trends in Biotechnology*, 37(9):943–959, sep 2019. 6

[41] Pei Xu, Gao-Wei Zheng, Min-Hua Zong, Ning Li, and Wen-Yong Lou. Recent progress on deep eutectic solvents in biocatalysis. *Bioresources and Bioprocessing*, 4(1), jul 2017. 6, 17, 25, 44, 45, 58, 64, 89, 139

[42] Zaira Maugeri and Pablo Domínguez de María. Whole-cell biocatalysis in deep-eutectic-solvents/aqueous mixtures. *ChemCatChem*, 6(6):1535–1537, mar 2014. 6

[43] Adepu K. Kumar, Bhumika S. Parikh, and Mohanty Pravakar. Natural deep eutectic solvent mediated pretreatment of rice straw: bioanalytical characterization of lignin extract and enzymatic hydrolysis of pretreated biomass residue. *Environmental Science and Pollution Research*, 23(10):9265–9275, jun 2015. 6

[44] Andrey Shishov, Aleksei Pochivalov, Lawrence Nugbienyo, Vasil Andruch, and Andrey Bulatov. Deep eutectic solvents are not only effective extractants. *TrAC Trends in Analytical Chemistry*, 129:115956, aug 2020. 7

[45] Andrew P. Abbott, Glen Capper, David L. Davies, Helen L. Munro, Raymond K. Rasheed, and Vasuki Tambyrajah. Preparation of novel, moisture-stable, lewis-acidic ionic liquids containing quaternary ammonium salts with functional side chains. *Chemical Communications*, (19):2010–2011, 2001. 16, 63

[46] Andrew P. Abbott, Glen Capper, and Stephen Gray. Design of improved deep eutectic solvents using hole theory. *ChemPhysChem*, 7(4):803–806, apr 2006. 19

[47] Andrew P. Abbott, Robert C. Harris, Karl S. Ryder, Carmine D'Agostino, Lynn F. Gladden, and Mick D. Mantle. Glycerol eutectics as sustainable solvent systems. *Green Chem.*, 13(1):82–90, 2011. 19, 22, 23, 24

[48] Mukhtar A. Kareem, Farouq S. Mjalli, Mohd Ali Hashim, and Inas M. AlNashef. Phosphonium-based ionic liquids analogues and their physical properties. *Journal of Chemical & Engineering Data*, 55(11):4632–4637, nov 2010. 19, 20, 24

[49] Zaira Maugeri and Pablo Dominguez de Maria. Novel choline-chloride-based deep-eutectic-solvents with renewable hydrogen bond donors: levulinic acid and sugar-based polyols. *RSC Adv.*, 2(2):421–425, 2012. 19, 20, 24

[50] Ingo Krossing, John M. Slattery, Corinne Daguenet, Paul J. Dyson, Alla Oleinikova, and Hermann Weingärtner. Why are ionic liquids liquid? a simple explanation based on lattice and solvation energies. *Journal of the American Chemical Society*, 128(41):13427–13434, oct 2006. 20, 22

[51] Carmine D'Agostino, Robert C. Harris, Andrew P. Abbott, Lynn F. Gladden, and Mick D. Mantle. Molecular motion and ion diffusion in choline chloride based deep eutectic solvents studied by 1h pulsed field gradient nmr spectroscopy. *Physical Chemistry Chemical Physics*, 13(48):21383, 2011. 22

[52] Andrew P. Abbott, John C. Barron, Gero Frisch, Stephen Gurman, Karl S. Ryder, and A. Fernando Silva. Double layer effects on metal nucleation in deep eutectic solvents. *Physical Chemistry Chemical Physics*, 13(21):10224, 2011. 22, 23

[53] K. Shahbaz, F.S. Mjalli, M.A. Hashim, and I.M. AlNashef. Prediction of the surface tension of deep eutectic solvents. *Fluid Phase Equilibria*, 319:48–54, apr 2012. 22, 35, 36, 95, 97

[54] Yujiao Xie, Haifeng Dong, Suojiang Zhang, Xiaohua Lu, and Xiaoyan Ji. Effect of water on the density, viscosity, and co_2 solubility in choline chloride/urea. *Journal of Chemical & Engineering Data*, 59(11):3344–3352, nov 2014. 22, 139

[55] Baokun Tang and Kyung Ho Row. Recent developments in deep eutectic solvents in chemical sciences. *Monatshefte für Chemie - Chemical Monthly*, 144(10):1427–1454, aug 2013. 22

[56] Andrew P. Abbott, Robert C. Harris, and Karl S. Ryder. Application of hole theory to define ionic liquids by their transport properties. *The Journal of Physical Chemistry B*, 111(18):4910–4913, mar 2007. 23, 70

[57] Amulya K. N. Reddy John O'M. Bockris. *Volume 1: Modern Electrochemistry*. Springer US, June 1998. 23, 24

[58] Wenjing Li, Zhaofu Zhang, Buxing Han, Suqin Hu, Jinliang Song, Ye Xie, and Xiaosi Zhou. Switching the basicity of ionic liquids by CO2. *Green Chemistry*, 10(11):1142, 2008. 24

[59] Dhawal Shah and Farouq S. Mjalli. Effect of water on the thermo-physical properties of reline: An experimental and molecular simulation based approach. *Physical Chemistry Chemical Physics*, 16(43):23900–23907, aug 2014. 24

[60] Xiangqian Meng, Karine Ballerat-Busserolles, Pascale Husson, and Jean-Michel Andanson. Impact of water on the melting temperature of urea + choline chloride deep eutectic solvent. *New Journal of Chemistry*, 40(5):4492–4499, 2016. 24

[61] Anita Yadav and Siddharth Pandey. Densities and viscosities of choline chloride + urea deep eutectic solvent and its aqueous mixtures in the temperature range 293.15 k to 363.15 k. *Journal of Chemical & Engineering Data*, 59(7):2221–2229, jun 2014. 24, 49, 65, 69, 70, 139

[62] Jessica M. Rimsza and L. Rene Corrales. Adsorption complexes of copper and copper oxide in the deep eutectic solvent 2:1 urea - choline chloride. *Computational and Theoretical Chemistry*, 987:57–61, may 2012. 25, 28, 58

[63] Xiaoyong Li, Minqiang Hou, Buxing Han, Xiaoling Wang, and Lizhuang Zou. Solubility of co2 in a choline chloride + urea eutectic mixture. *Journal of Chemical & Engineering Data*, 53(2):548–550, feb 2008. 25, 130

[64] Wen Cheng Su, David Shan Hill Wong, and Meng Hui Li. Effect of water on solubility of carbon dioxide in (aminomethanamide + 2-hydroxy-n,n,n-trimethylethanaminium chloride). *Journal of Chemical & Engineering Data*, 54(6):1951–1955, jun 2009. 25, 65

[65] Henry G. Morrison, Changquan C. Sun, and Sesha Neervannan. Characterization of thermal behavior of deep eutectic solvents and their potential as drug solubilization vehicles. *International Journal of Pharmaceutics*, 378(1-2):136–139, aug 2009. 25

[66] Marta Faggian, Stefania Sut, Beatrice Perissutti, Valeria Baldan, Iztok Grabnar, and Stefano Dall'Acqua. Natural deep eutectic solvents (nades) as a tool for bioavailability improvement: Pharmacokinetics of rutin dissolved in proline/glycine after oral administration in rats: Possible application in nutraceuticals. *Molecules*, 21(11):1531, nov 2016. 25

[67] Tomasz Jelinski and Piotr Cysewski. Application of a computational model of natural deep eutectic solvents utilizing the cosmo-rs approach for screening of solvents with high solubility of rutin. *Journal of Molecular Modeling*, 24(7), jun 2018. 25, 32

[68] Hassan Monhemi, Mohammad Reza Housaindokht, Ali Akbar Moosavi-Movahedi, and Mohammad Reza Bozorgmehr. How a protein can remain stable in a solvent with high content of urea: insights from molecular dynamics simulation of candida antarctica lipase b in urea : choline chloride deep eutectic solvent. *Physical Chemistry Chemical Physics*, 16(28):14882, 2014. 25, 29, 45, 58, 64, 65, 73, 77, 79, 81, 82, 86, 87, 89, 139, 144, 148, 158, 160

[69] Erwann Durand, Jerome Lecomte, Bruno Barea, Eric Dubreucq, Robert Lortie, and Pierre Villeneuve. Evaluation of deep eutectic solvent–water binary mixtures for lipase-catalyzed lipophilization of phenolic acids. *Green Chemistry*, 15(8):2275, 2013. 25

[70] Ben-Pei Wu, Qing Wen, Hong Xu, and Zhen Yang. Insights into the impact of deep eutectic solvents on horseradish peroxidase: Activity, stability and structure. *Journal of Molecular Catalysis B: Enzymatic*, 101:101–107, mar 2014. 26, 45, 64, 89, 139

[71] Gregorio Garcia, Mert Atilhan, and Santiago Aparicio. An approach for the rationalization of melting temperature for deep eutectic solvents from dft. *Chemical Physics Letters*, 634:151–155, aug 2015. 28

[72] Durgesh V. Wagle, Carol A. Deakyne, and Gary A. Baker. Quantum chemical insight into the interactions and thermodynamics present in choline chloride based deep eutectic solvents. *The Journal of Physical Chemistry B*, 120(27):6739–6746, jun 2016. 28

[73] Ryan Stefanovic, Michael Ludwig, Grant B. Webber, Rob Atkin, and Alister J. Page. Nanostructure, hydrogen bonding and rheology in choline chloride deep eutectic solvents as a function of the hydrogen bond donor. *Physical Chemistry Chemical Physics*, 19(4):3297–3306, 2017. 28

[74] Ruh Ullah, Mert Atilhan, Baraa Anaya, Majeda Khraisheh, Gregorio García, Ahmed ElKhattat, Mohammad Tariq, and Santiago Aparicio. A detailed study of cholinium chloride and levulinic acid deep eutectic solvent system for CO2 capture via experimental and molecular simulation approaches. *Physical Chemistry Chemical Physics*, 17(32):20941–20960, 2015. 28, 29, 64, 65

[75] Gregorio Garcia, Mert Atilhan, and Santiago Aparicio. The impact of charges in force field parameterization for molecular dynamics simulations of deep eutectic solvents. *Journal of Molecular Liquids*, 211:506–514, nov 2015. 29, 65

[76] Brian Doherty and Orlando Acevedo. Opls force field for choline chloride-based deep eutectic solvents. *The Journal of Physical Chemistry B*, 122(43):9982–9993, aug 2018. 29, 64, 65, 66, 67, 69, 70, 151, 162

[77] Hui Sun, Yan Li, Xue Wu, and Guohui Li. Theoretical study on the structures and properties of mixtures of urea and choline chloride. *Journal of Molecular Modeling*, 19(6):2433–2441, feb 2013. 29, 64

[78] Sasha L. Perkins, Paul Painter, and Coray M. Colina. Experimental and computational studies of choline chloride-based deep eutectic solvents. *Journal of Chemical & Engineering Data*, 59(11):3652–3662, oct 2014. 29

[79] Anuradha Das, Suman Das, and Ranjit Biswas. Density relaxation and particle motion characteristics in a non-ionic deep eutectic solvent (acetamide + urea): Time-resolved fluorescence measurements and all-atom molecular dynamics simulations. *The Journal of Chemical Physics*, 142(3):034505, jan 2015. 29, 64

[80] Andreas Klamt. The cosmo and cosmo-rs solvation models. *WIREs Computational Molecular Science*, 1(5):699–709, apr 2011. 30

[81] Cosmo-rs theory, the basics. 31

[82] Sarwono Mulyono, Hanee F. Hizaddin, Inas M. Alnashef, Mohd A. Hashim, Anis H. Fakeeha, and Mohamed K. Hadj-Kali. Separation of BTEX aromatics from n-octane using a (tetrabutylammonium bromide + sulfolane) deep eutectic solvent – experiments and COSMO-RS prediction. *RSC Advances*, 4(34):17597, 2014. 31

[83] Hanee F. Hizaddin, Anantharaj Ramalingam, Mohd Ali Hashim, and Mohamed K.O. Hadj-Kali. Evaluating the performance of deep eutectic solvents for use in extractive denitrification of liquid fuels by the conductor-like screening model for real solvents. *Journal of Chemical & Engineering Data*, 59(11):3470–3487, sep 2014. 31

[84] Andreia S. L. Gouveia, Filipe S. Oliveira, Kiki A. Kurnia, and Isabel M. Marrucho. Deep eutectic solvents as azeotrope breakers: Liquid–liquid extraction and COSMO-RS prediction. *ACS Sustainable Chemistry & Engineering*, 4(10):5640–5650, sep 2016. 31

[85] Tayeb Aissaoui, Yacine Benguerba, and Inas M. AlNashef. Theoretical investigation on the microstructure of triethylene glycol based deep eutectic solvents: COSMO-RS and TURBOMOLE prediction. *Journal of Molecular Structure*, 1141:451–456, aug 2017. 31

[86] Franziska Bezold, Maria E. Weinberger, and Mirjana Minceva. Assessing solute partitioning in deep eutectic solvent-based biphasic systems using the predictive thermodynamic model cosmo-rs. *Fluid Phase Equilibria*, 437:23–33, apr 2017. 31

[87] Liliana P. Silva, Luis Fernandez, João H. F. Conceição, Mónia A. R. Martins, Adriel Sosa, Juan Ortega, Simão P. Pinho, and João A. P. Coutinho. Design and characterization of sugar-based deep eutectic solvents using conductor-like screening model for real solvents. *ACS Sustainable Chemistry & Engineering*, 6(8):10724–10734, jun 2018. 32

[88] Joachim Gross and Gabriele Sadowski. Perturbed-chain saft: An equation of state based on a perturbation theory for chain molecules. *Industrial & Engineering Chemistry Research*, 40(4):1244–1260, feb 2001. 32, 130

[89] Sergey P. Verevkin, Aleksandra Yu. Sazonova, Alla K. Frolkova, Dzmitry H. Zaitsau, Igor V. Prikhodko, and Christoph Held. Separation performance of BioRenewable deep eutectic solvents. *Industrial & Engineering Chemistry Research*, 54(13):3498–3504, mar 2015. 33, 130

[90] Lawien F. Zubeir, Christoph Held, Gabriele Sadowski, and Maaike C. Kroon. Pc-saft modeling of CO2 solubilities in deep eutectic solvents. *The Journal of Physical Chemistry B*, 120(9):2300–2310, feb 2016. 33, 130

[91] Carin H.J.T. Dietz, Dannie J.G.P. van Osch, Maaike C. Kroon, Gabriele Sadowski, Martin van Sint Annaland, Fausto Gallucci, Lawien F. Zubeir, and Christoph Held. PC-SAFT modeling of CO 2 solubilities in hydrophobic deep eutectic solvents. *Fluid Phase Equilibria*, 448:94–98, sep 2017. 33, 130

[92] Paula V.A. Pontes, Emanuel A. Crespo, Monia A.R. Martins, Liliana P. Silva, Catarina M.S.S. Neves, Guilherme J. Maximo, Miriam Dupas Hubinger, Eduardo A.C. Batista, Simao P. Pinho, Joao A.P. Coutinho, Gabriele Sadowski, and Christoph Held. Measurement and pc-saft modeling of solid-liquid equilibrium of deep eutectic solvents of quaternary ammonium chlorides and carboxylic acids. *Fluid Phase Equilibria*, 448:69–80, sep 2017. 34, 130

[93] Reza Haghbakhsh, Khalil Parvaneh, Sona Raeissi, and Alireza Shariati. A general viscosity model for deep eutectic solvents: The free volume theory coupled with association equations of state. *Fluid Phase Equilibria*, 470:193–202, aug 2018. 34, 119, 130, 132

[94] S. Abbot, C. M. Hansen, and H. Yamamoto. *Hansen Solubility Parameters in Practice*. Hansen-Solubility.com, 2015. 35

[95] Zheng Li and Ping I. Lee. Investigation on drug solubility enhancement using deep eutectic solvents and their derivatives. *International Journal of Pharmaceutics*, 505(1-2):283–288, may 2016. 35

[96] K. Shahbaz, F.S. Mjalli, M.A. Hashim, and I.M. AlNashef. Prediction of deep eutectic solvents densities at different temperatures. *Thermochimica Acta*, 515(1-2):67–72, mar 2011. 35, 95

[97] K. Shahbaz, S. Baroutian, F.S. Mjalli, M.A. Hashim, and I.M. AlNashef. Densities of ammonium and phosphonium based deep eutectic solvents: Prediction using artificial intelligence and group contribution techniques. *Thermochimica Acta*, 527:59–66, jan 2012. 35, 93, 95

[98] K. Shahbaz, F.S. Ghareh Bagh, F.S. Mjalli, I.M. AlNashef, and M.A. Hashim. Prediction of refractive index and density of deep eutectic solvents using atomic contributions. *Fluid Phase Equilibria*, 354:304–311, sep 2013. 35, 95

[99] Nouman Rafique Mirza, Nathan J. Nicholas, Yue Wu, Sandra Kentish, and Geoffrey W. Stevens. Estimation of normal boiling temperatures, critical properties, and acentric factors of deep eutectic solvents. *Journal of Chemical & Engineering Data*, 60(6):1844–1854, jun 2015. 36, 95, 130

[100] F.S. Ghareh Bagh, K. Shahbaz, F.S. Mjalli, I.M. AlNashef, and M.A. Hashim. Electrical conductivity of ammonium and phosphonium based deep eutectic solvents: Measurements and artificial intelligence-based prediction. *Fluid Phase Equilibria*, 356:30–37, oct 2013. 37

[101] Idowu Adeyemi, Mohammad R.M. Abu-Zahra, and Inas M. AlNashef. Physicochemical properties of alkanolamine-choline chloride deep eutectic solvents: Measurements, group contribution and artificial intelligence prediction techniques. *Journal of Molecular Liquids*, 256:581–590, apr 2018. 37, 97

[102] Monia A. R. Martins, Simao P. Pinho, and Joao A. P. Coutinho. Insights into the nature of eutectic and deep eutectic mixtures. *Journal of Solution Chemistry*, 48(7):962–982, aug 2018. 44

[103] Yang-Lei Chen, Xun Zhang, Ting-Ting You, and Feng Xu. Deep eutectic solvents (DESs) for cellulose dissolution: a mini-review. *Cellulose*, 26(1):205–213, nov 2018. 44

[104] Hua Zhao, Gary A. Baker, and Shaletha Holmes. New eutectic ionic liquids for lipase activation and enzymatic preparation of biodiesel. *Organic & Biomolecular Chemistry*, 9(6):1908, 2011. 44, 45

[105] Kai-Hua Zhao, Yu-Zheng Cai, Xiao-Sheng Lin, Jun Xiong, Peter Halling, and Zhen Yang. Enzymatic synthesis of glucose-based fatty acid esters in bisolvent systems containing ionic liquids or deep eutectic solvents. *Molecules*, 21(10):1294, sep 2016. 44, 45, 139

[106] Aleksandra Misan, Jelena Nadpal, Alena Stupar, Milica Pojic, Anamarija Mandic, Robert Verpoorte, and Young Hae Choi. The perspectives of natural deep eutectic solvents in agri-food sector. *Critical Reviews in Food Science and Nutrition*, 60(15):2564–2592, aug 2019. 44, 139

[107] Sascha Siebenhaller, Claudia Muhle-Goll, Burkhard Luy, Frank Kirschhöfer, Gerald Brenner-Weiss, Ekkehard Hiller, Michael Günther, Steffen Rupp, Susanne Zibek, and Christoph Syldatk. Sustainable enzymatic synthesis of glycolipids in a deep eutectic solvent system. *Journal of Molecular Catalysis B: Enzymatic*, 133:S281–S287, nov 2016. 44, 45, 139

[108] Beatrice Kleiner and Ulrich Schörken. Native lipase dissolved in hydrophilic green solvents: A versatile 2-phase reaction system for high yield ester synthesis. *European Journal of Lipid Science and Technology*, 117(2):167–177, jan 2015. 44, 45, 139

[109] Johnathan T. Gorke, Friedrich Srienc, and Romas J. Kazlauskas. Hydrolase-catalyzed biotransformations in deep eutectic solvents. *Chemical Communications*, (10):1235, 2008. 44, 45, 51, 55, 58, 64, 75, 89, 139

[110] Binbin Nian, Chen Cao, and Yuanfa Liu. How candida antarctica lipase b can be activated in natural deep eutectic solvents: experimental and molecular dynamics studies. *Journal of Chemical Technology & Biotechnology*, 95(1):86–93, oct 2019. 44, 45, 58, 65, 73, 77, 79, 82, 89, 139, 158, 160

[111] Mohamed Shehata, Aise Unlu, Ugur Sezerman, and Emel Timucin. Lipase and water in a deep eutectic solvent: Molecular dynamics and experimental studies of the effects of water-in-deep eutectic solvents on lipase stability. *The Journal of Physical Chemistry B*, 124(40):8801–8810, sep 2020. 44, 45, 55, 58, 65, 73, 75, 77, 82, 86, 139, 158, 160

[112] Attila Kovács, Maksudbek Yusupov, Iris Cornet, Pieter Billen, and Erik C. Neyts. Effect of natural deep eutectic solvents of non-eutectic compositions on enzyme stability. *Journal of Molecular Liquids*, 366:120180, nov 2022. 44, 57, 62, 139, 158, 160

[113] E. Durand, J. Lecomte, B. Barea, G. Piombo, E. Dubreucq, and P. Villeneuve. Evaluation of deep eutectic solvents as new media for candida antarctica b lipase catalyzed reactions. *Process Biochemistry*, 47(12):2081–2089, dec 2012. 45, 47, 48, 51, 55, 58, 64, 89, 139, 144, 148

[114] Ibrahim Juneidi, Maan Hayyan, Mohd Ali Hashim, and Adeeb Hayyan. Pure and aqueous deep eutectic solvents for a lipase-catalysed hydrolysis reaction. *Biochemical Engineering Journal*, 117:129–138, jan 2017. 45

[115] L. Goujard, P. Villeneuve, B. Barea, J. Lecomte, M. Pina, S. Claude, J. Le Petit, and E. Ferre. A spectrophotometric transesterification-based assay for lipases in organic solvent. *Analytical Biochemistry*, 385(1):161–167, feb 2009. 45, 48

[116] Qian Wang, Xiaoqian Yao, Yanrong Geng, Qing Zhou, Xingmei Lu, and Suojiang Zhang. Deep eutectic solvents as highly active catalysts for the fast and mild glycolysis of poly(ethylene terephthalate)(PET). *Green Chemistry*, 17(4):2473–2479, 2015. 58

[117] Devleena Shivakumar, Joshua Williams, Yujie Wu, Wolfgang Damm, John Shelley, and Woody Sherman. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *Journal of Chemical Theory and Computation*, 6(5):1509–1519, apr 2010. 58

[118] Usman L. Abbas, Qi Qiao, Manh Tien Nguyen, Jian Shi, and Qing Shao. Molecular dynamics simulations of heterogeneous hydrogen bond environment in hydrophobic deep eutectic solvents. *AIChE Journal*, 68(1), aug 2021. 58

[119] Ana S.D. Ferreira, Rita Craveiro, Ana Rita Duarte, Susana Barreiros, Eurico J. Cabrita, and Alexandre Paiva. Effect of water on the structure and dynamics of choline chloride/glycerol eutectic systems. *Journal of Molecular Liquids*, 342:117463, nov 2021. 65, 71

[120] Oliver S. Hammond, Daniel T. Bowron, and Karen J. Edler. The effect of water upon deep eutectic solvent nanostructure: An unusual transition from ionic mixture to aqueous solution. *Angewandte Chemie*, 129(33):9914–9917, may 2017. 65

[121] Supreet Kaur, Aditya Gupta, and Hemant K. Kashyap. How hydration affects the microscopic structural morphology in a deep eutectic solvent. *The Journal of Physical Chemistry B*, 124(11):2230–2237, feb 2020. 65, 71

[122] Liel Sapir and Daniel Harries. Restructuring a deep eutectic solvent by water: The nanostructure of hydrated choline chloride/urea. *Journal of Chemical Theory and Computation*, 16(5):3335–3342, mar 2020. 65

[123] Erik Lindahl, Marc James Abraham, Berk Hess, and David Van Der Spoel. *GROMACS 2021.3 Manual*, 2021. 66, 69

[124] Sander Pronk, Szilard Pall, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, feb 2013. 66

[125] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilard Pall, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, sep 2015. 66

[126] L. Martinez, R. Andrade, E. G. Birgin, and J. M. Martinez. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, oct 2009. 66

[127] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, feb 1996. 66

[128] J. Uppenberg and T.A. Jones. THE SEQUENCE, CRYSTAL STRUCTURE DETERMINATION AND REFINEMENT OF TWO CRYSTAL FORMS OF LIPASE b FROM CANDIDA ANTARCTICA, may 1994. 66, 86

[129] Jonas Uppenberg, Mogens Trier Hansen, Shamkant Patkar, and T.Alwyn Jones. The sequence, crystal structure determination and refinement of two crystal forms of lipase b from candida antarctica. *Structure*, 2(4):293–308, apr 1994. 66

[130] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, jan 1996. 66

[131] Erin M. Duffy, Daniel L. Severance, and William L. Jorgensen. Urea: Potential functions, logp, and free energy of hydration. *Israel Journal of Chemistry*, 33(3):323–330, 1993. 66

[132] Leela S. Dodda, Israel Cabeza de Vaca, Julian Tirado-Rives, and William L. Jorgensen. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research*, 45(W1):W331–W336, apr 2017. 66

[133] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, jan 2007. 67

[134] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, dec 1981. 67

[135] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An N·log(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, jun 1993. 67

[136] Paul Bauer, Berk Hess, and Erik Lindahl. Gromacs 2022 manual. 2022. 68, 69

[137] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From algorithms to applications*. Academic Press, 2002. 69

[138] Frank Eisenhaber, Philip Lijnzaad, Patrick Argos, Chris Sander, and Michael Scharf. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284, mar 1995. 69

[139] Wouter G. Touw, Coos Baakman, Jon Black, Tim A. H. te Beek, E. Krieger, Robbie P. Joosten, and Gert Vriend. A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43(D1):D364–D368, oct 2014. 69

[140] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, dec 1983. 69

[141] Rhoda B. Leron, Allan N. Soriano, and Meng-Hui Li. Densities and refractive indices of the deep eutectic solvents choline chloride + ethylene glycol or glycerol and their aqueous mixtures at the temperature ranging from 298.15 to 333.15 k. *Journal of the Taiwan Institute of Chemical Engineers*, 43(4):551–557, jul 2012. 69, 70

[142] Stephanie Spittle, Derrick Poe, Brian Doherty, Charles Kolodziej, Luke Heroux, Md Ashraful Haque, Henry Squire, Tyler Cosby, Yong Zhang, Carla Fraenza, Sahana Bhattacharyya, Madhusudan Tyagi, Jing Peng, Ramez A. Elgammal, Thomas Zawodzinski, Mark Tuckerman, Steve Greenbaum, Burcu Gurkan, Clemens Burda, Mark Dadmun, Edward J. Maginn, and Joshua Sangoro. Evolution of microscopic heterogeneity and dynamics in choline chloride-based deep eutectic solvents. *Nature Communications*, 13(1), jan 2022. 70, 162

[143] Yong Zhang, Henry Squire, Burcu Gurkan, and Edward J. Maginn. Refined classical force field for choline chloride and ethylene glycol mixtures over wide composition range. *Journal of Chemical & Engineering Data*, jan 2022. 70

[144] Lindong Weng and Mehmet Toner. Janus-faced role of water in defining nanostructure of choline chloride/glycerol deep eutectic solvent. *Physical Chemistry Chemical Physics*, 20(35):22455–22462, 2018. 75

[145] Hassan Monhemi. Protein simulation in supercritical CO2: The challenge of force field. *Journal of Molecular Liquids*, 343:117662, dec 2021. 77

[146] Amedeo Caflisch and Martin Karplus. Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure*, 7(5):477–S2, may 1999. 77

[147] Peter Carlqvist, Robert Eklund, Karl Hult, and Tore Brinck. Rational design of a lipase to accommodate catalysis of baeyer-villiger oxidation with hydrogen peroxide. *Journal of Molecular Modeling*, 9(3):164–171, jun 2003. 82

[148] Alberto Gutierrez, Lorena Zamora, Cristina Benito, Mert Atilhan, and Santiago Aparicio. Insights on novel type v deep eutectic solvents based on levulinic acid. 156(9):094504. 94

[149] Andrew P. Abbott. Deep eutectic solvents and their application in electrochemistry. 36:100649. 94

[150] Xiaoxia Li and Kyung Ho Row. Development of deep eutectic solvents applied in extraction and separation. 39(18):3505–3520. 94

[151] Kapil Mamtani, Kaveh Shahbaz, and Mohammed M. Farid. Deep eutectic solvents – versatile chemicals in biodiesel production. 295:120604. 94

[152] Emmanuel A. Oke and Sushma P. Ijardar. Advances in the application of deep eutectic solvents based aqueous biphasic systems: An up-to-date review. 176:108211. 94

[153] J. O. Valderrama and P. A. Robles. Critical properties, normal boiling temperatures, and acentric factors of fifty ionic liquids. *Industrial Engineering Chemistry Research*, 46(4):1338–1344, jan 2007. 95, 118, 119

[154] Jose O. Valderrama, Wilson W. Sanga, and Juan A. Lazzus. Critical properties, normal boiling temperature, and acentric factor of another 200 ionic liquids. *Industrial Engineering Chemistry Research*, 47(4):1318–1330, jan 2008. 95, 118, 119

[155] L. Oellrich U. Plöcker J.M. Prausnitz R. Langhorst S. Zeck H. Knapp, R. Döring. Vapor-liquid equilibria for mixtures of low boiling substances. *DECHEMA Chemistry Data Series*, VI, 1982. 95, 118, 119

[156] Harold G. Rackett. Equation of state for saturated liquids. *Journal of Chemical Engineering Data*, 15(4):514–517, oct 1970. 95

[157] Aliakbar Roosta and Roghayeh Bardool. A simple correlation for estimating the viscosity of pure ionic liquids and their binary mixtures. *Industrial Engineering Chemistry Research*, 56(15):4600–4610, apr 2017. 95, 96

[158] Reza Haghbakhsh, Roghayeh Bardool, Ali Bakhtyari, Ana Rita C. Duarte, and Sona Raeissi. Simple and global correlation for the densities of deep eutectic solvents. *Journal of Molecular Liquids*, 296:111830, dec 2019. 95, 96, 118

[159] Reza Haghbakhsh, Mehrdad Taherzadeh, Ana Rita C. Duarte, and Sona Raeissi. A general model for the surface tensions of deep eutectic solvents. *Journal of Molecular Liquids*, 307:112972, jun 2020. 95, 96, 159

[160] Reza Haghbakhsh, Sona Raeissi, and Ana Rita C. Duarte. Group contribution and atomic contribution models for the prediction of various physical properties of deep eutectic solvents. *Scientific Reports*, 11(1), mar 2021. 95, 96, 99, 118, 151, 161

[161] Xiao-Jing Hou, Liu-Ying Yu, Yan-Xu Wang, Ke-Jun Wu, and Chao-Hong He. Comprehensive prediction of densities for deep eutectic solvents: A new bonding-group interaction contribution scheme. *Industrial & Engineering Chemistry Research*, 60(35):13127–13139, aug 2021. 95, 96

[162] Amit Kumar Halder, Reza Haghbakhsh, Iuliia V. Voroshylova, Ana Rita C. Duarte, and M. Natalia D. S. Cordeiro. Density of deep eutectic solvents: The path forward cheminformatics-driven reliable predictions for mixtures. *Molecules*, 26(19):5779, sep 2021. 95, 97, 159

[163] Aboozar Khajeh and Mehdi Shakourian-Fard. Chemical structure-based models for prediction of density of ammonium and phosphonium-based deep eutectic solvents. *Journal of Molecular Liquids*, 343:117595, dec 2021. 95, 97

[164] Farouq S. Mjalli, Kaveh Shahbaz, and Inas M. AlNashef. Modified rackett equation for modelling the molar volume of deep eutectic solvents. *Thermochimica Acta*, 614:185–190, aug 2015. 96

[165] Leonidas Constantinou and Rafiqul Gani. New group contribution method for estimating properties of pure compounds. *AIChE Journal*, 40(10):1697–1710, oct 1994. 96

[166] Mohammad Atif Faiz Afzal, Aditya Sonpal, Mojtaba Haghighatlari, Andrew J. Schultz, and Johannes Hachmann. A deep neural network model for packing density predictions and its application in the study of 1.5 million organic molecules. *Chemical Science*, 10(36):8374–8383, 2019. 96

[167] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media. 102, 103

[168] John G. Kirkwood, Frank P. Buff, and Melvin S. Green. The statistical mechanical theory of transport processes. III. the coefficients of shear and bulk viscosity of liquids. *The Journal of Chemical Physics*, 17(10):988–994, oct 1949. 118

[169] Peter T. Cummings and Denis J. Evans. Nonequilibrium molecular dynamics approaches to transport properties and non-newtonian fluid rheology. *Industrial & Engineering Chemistry Research*, 31(5):1237–1252, may 1992. 118

[170] Dongshun Deng, Xiaobang Liu, and Bao Gao. Physicochemical properties and investigation of azole-based deep eutectic solvents as efficient and reversible so2 absorbents. 56(46):13850–13856. 118

[171] Catarina Florindo, M. Margarida Oliveira, Luis C. Branco, and Isabel M. Marrucho. Carbohydrates-based deep eutectic solvents: Thermophysical properties and rice straw dissolution. 247:441–447. 118

[172] Shokat Sarmad, Yujiao Xie, Jyri-Pekka Mikkola, and Xiaoyan Ji. Screening of deep eutectic solvents dess as green co2 sorbents: from solubility to viscosity. 41(1):290–301. 118

[173] Reza Haghbakhsh, Khalil Parvaneh, and Alireza Shariati. Viscosities of pure ionic liquids using combinations of free volume theory or friction theory with the cubic, the cubic plus association, and the perturbed-chain statistical associating fluid theory equations of state at high pressures. *Industrial & Engineering Chemistry Research*, 56(8):2247–2258, feb 2017. 118, 119

[174] N. V. K. Dutt, Y. V. L. Ravikumar, and K. Yamuna Rani. REPRESENTATION OF IONIC LIQUID VISCOSITY-TEMPERATURE DATA BY GENERALIZED CORRELATIONS AND AN ARTIFICIAL NEURAL NETWORK (ANN) MODEL. 200(12):1600–1622. 118

[175] D. H. Vogel. The law of relation between the viscosity of liquids and the temperature. *Physikalische Zeitschrift*, 22:645, 1921. 118

[176] R. P. Danner T. E. Daubert. *Physical and thermodynamic properties of pure chemicals : data compilation*. Taylor & Francis, 1989. 118

[177] Reza Haghbakhsh and Sona Raeissi. Two simple correlations to predict viscosities of pure and aqueous solutions of ionic liquids. *Journal of Molecular Liquids*, 211:948–956, nov 2015. 118

[178] L. GRUNBERG and ALFRED H. NISSAN. Mixture law for viscosity. 164(4175):799–800. 118

[179] Abolghasem Jouyban, Jafar Soleymani, Farshad Jafari, Mehry Khoubnasabjafari, and William E. Acree. Mathematical representation of viscosity of ionic liquid + molecular solvent mixtures at various temperatures using the jouyban–acree model. 58(6):1523–1528. 118

[180] R. A. McAllister. The viscosity of liquid mixtures. 6(3):427–431. 118

[181] Alif Duereh, Yoshiyuki Sato, Richard Lee Smith, and Hiroshi Inomata. Application of the preferential solvation viscosity model to binary liquid mixtures: Aqueous, non-aqueous, ionic liquid, and deep eutectic solvent systems. 58(32):14991–15002. 118

[182] Khalil Parvaneh, Reza Haghbakhsh, and Mohammad Reza Rahimpour. High pressure viscosity modeling of pure alcohols based on classical and advanced equations of state. *Journal of the Taiwan Institute of Chemical Engineers*, 58:57–70, jan 2016. 119

[183] Masatoshi Yoshimura, Christian Boned, Guillaume Galliéro, Jean-Patrick Bazile, Antoine Baylaucq, and Hideharu Ushiki. Influence of the chain length on the dynamic viscosity at high pressure of some 2-alkylamines: Measurements and comparative study of some models. 369(2-3):126–137. 119

[184] Lei Qun-Fang, Hou Yu-Chun, and Lin Rui-Sen. Correlation of viscosities of pure liquids in a wide temperature range. 140(1-2):221–231. 119

[185] Sergio E Quiñones-Cisneros, Claus K Zéberg-Mikkelsen, and Erling H Stenby. One parameter friction theory models for viscosity. 178(1-2):1–16. 119

[186] Ding-Yu Peng and Donald B. Robinson. A new two-constant equation of state. 15(1):59–64. 119

[187] Giorgio Soave. Equilibrium constants from a modified redlich-kwong equation of state. 27(6):1197–1203. 119

[188] Joachim Gross and Gabriele Sadowski. Application of the perturbed-chain SAFT equation of state to associating systems. 41(22):5510–5515. 119, 130

[189] Simin Keshtkari, Reza Haghbakhsh, Sona Raeissi, Louw Florusse, and Cor J. Peters. Vapor-liquid equilibria of isopropyl alcohol + propylene at high pressures: Experimental measurement and modeling with the CPA EoS. 84:182–189. 119

[190] Ali Bakhtyari, Reza Haghbakhsh, Ana Rita C. Duarte, and Sona Raeissi. A simple model for the viscosities of deep eutectic solvents. *Fluid Phase Equilibria*, 521:112662, oct 2020. 119, 120, 151, 159

[191] Sara Hajebrahimi and Aliakbar Roosta. Solubility of acetaminophen in aqueous solutions of three natural deep eutectic solvents (NADESs) and individual components of the NADESs. *Journal of Molecular Liquids*, 316:113867, oct 2020. 130

[192] Reza Haghbakhsh, Sona Raeissi, Khalil Parvaneh, and Alireza Shariati. The friction theory for modeling the viscosities of deep eutectic solvents using the CPA and PC-SAFT equations of state. *Journal of Molecular Liquids*, 249:554–561, jan 2018. 130, 132

[193] W.G. Chapman, K.E. Gubbins, G. Jackson, and M. Radosz. Saft: Equation-of-state solution model for associating fluids. 52:31–38. 130

[194] Bruno Sepulveda-Orellana, Nicolas F. Gajardo-Parra, Hoang T. Do, Jose R. Perez-Correa, Christoph Held, Gabriele Sadowski, and Roberto I. Canales. Measurement and PC-SAFT modeling of the solubility of gallic acid in aqueous mixtures of deep eutectic solvents. 66(2):958–967. 131, 132, 159

[195] Jeffrey P. Wolbach and Stanley I. Sandler. Using molecular orbital calculations to describe the phase behavior of cross-associating mixtures. 37(8):2917–2928. 131

[196] Fatima L. Mota, Antonio J. Queimada, Simao P. Pinho, and Eugenia A. Macedo. Aqueous solubility of some natural phenolic compounds. 47(15):5182–5189. 132

[197] Jorge Marrero and Rafiqul Gani. Group-contribution based estimation of pure component properties. 183-184:183–208. 132

[198] Joseph Irudayaraj Arutchelvi, Sumit Bhaduri, Parasu Veera Uppara, and Mukesh Doble. Mannosylerythritol lipids: a review. *Journal of Industrial Microbiology & Biotechnology*, 35(12):1559–1570, aug 2008. 137

[199] Ibrahim M. Banat, Andrea Franzetti, Isabella Gandolfi, Giuseppina Bestetti, Maria G. Martinotti, Letizia Fracchia, Thomas J. Smyth, and Roger Marchant. Microbial biosurfactants production, applications and future potential. *Applied Microbiology and Biotechnology*, 87(2):427–444, apr 2010. 137

[200] Tomotake Morita, Tokuma Fukuoka, Tomohiro Imura, and Dai Kitamoto. Production of mannosylerythritol lipids and their application in cosmetics. *Applied microbiology and biotechnology*, 97(11):4691–4700, jun 2013. 137, 138

[201] Eliane Goossens, Marc Wijnants, Dirk Packet, and Filip Lemière. Enhanced separation and analysis procedure reveals production of tri-acylated mannosylerythritol lipids by pseudozyma aphidis. *Journal of Industrial Microbiology and Biotechnology*, 43(11):1537–1550, nov 2016. 137, 140

[202] Ana Letícia Silva Coelho, Paulo Emílio Feuser, Bruno Augusto Mattar Carciofi, Cristiano José de Andrade, and Débora de Oliveira. Mannosylerythritol lipids: antimicrobial and biomedical properties. *Applied Microbiology and Biotechnology*, 104(6):2297–2318, jan 2020. 138

[203] Tomotake Morita, Masaru Kitagawa, Shuhei Yamamoto, Atsushi Sogabe, Atsushi Sogabe, Tokuma Fukuoka, and Dai Kitamoto. Glycolipid biosurfactants, mannosylerythritol lipids, repair the damaged hair. *Journal of oleo science*, 59(5):267–272, may 2010. 138

[204] Yosuke Morita, Satoshi Tadokoro, Masao Sasai, Dai Kitamoto, and Naohide Hirashima. Biosurfactant mannosyl-erythritol lipid inhibits secretion of inflammatory mediators from rbl-2h3 cells. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1810(12):1302–1308, dec 2011. 138

[205] Tokuma Fukuoka, Takashi Yanagihara, Tomohiro Imura, Tomotake Morita, Hideki Sakai, Masahiko Abe, and Dai Kitamoto. Enzymatic synthesis of a novel glycolipid biosurfactant, mannosylerythritol lipid-d and its aqueous phase behavior. *Carbohydrate research*, 346(2):266–271, feb 2011. 138, 143

[206] Elian Goossens and Marc Wijnants. Method for the regioselective deacetylation of mannosylerythritol lipids, 2016. 138

[207] Eliane Goossens. *Lipase-catalyzed modifications of mannosylerythritol lipids and their characterization*. PhD thesis, University of Antwerp, 2019. 142

[208] Verena K. Recke, Catharina Beyrle, Melanie Gerlitzki, Rudolf Hausmann, Christoph Syldatk, Victor Wray, Harukuni Tokuda, Nobutaka Suzuki, and Siegmund Lang. Lipase-catalyzed acylation of microbial mannosylerythritol lipids (biosurfactants) and their characterization. *Carbohydrate Research*, 373:82–88, may 2013. 145

[209] E. J. Ritter, W. J. Scott, J. L. Randall, and J. M. Ritter. Teratogenicity of di(2-ethylhexyl) phthalate, 2-ethylhexanol, 2-ethylhexanoic acid, and valproic acid, and potentiation by caffeine. 35(1):41–46. 148

[210] www.chemanalyst.com. 2-ethylhexanol price trend and forecast (accessed: February 1, 2023). 148

[211] www.chemanalyst.com. Toluene price trend and forecast (accessed: February 1, 2023). 148

[212] www.indiamart.com. Thymol, for industrial, a.b. enterprises (accessed: February 1, 2023). 148

[213] madeinchina.com. Dl-menthol cas 89-78-1 with best price (accessed: February 1, 2023)). 148

[214] madeinchina.com. Factory supply food grade coumarin cas 91-64-5 with low price (accessed: February 1, 2023). 148

[215] Kyung Min Jeong, Min Sang Lee, Min Woo Nam, Jing Zhao, Yan Jin, Dong-Kyu Lee, Sung Won Kwon, Ji Hoon Jeong, and Jeongmi Lee. Tailoring and recycling of deep eutectic solvents as sustainable and efficient extraction media. 1424:10–17. 148

[216] Asli Isci and Martin Kaltschmitt. Recovery and recycling of deep eutectic solvents in biomass conversions: a review. 12(S1):197–226. 148

[217] Min Woo Nam, Jing Zhao, Min Sang Lee, Ji Hoon Jeong, and Jeongmi Lee. Enhanced extraction of bioactive natural products using tailor-made deep eutectic solvents: application to flavonoid extraction from flos sophorae. 17(3):1718–1727. 148

[218] Mohammad Chand Ali, Jia Chen, Haijuan Zhang, Zhan Li, Liang Zhao, and Hongdeng Qiu. Effective extraction of flavonoids from lycium barbarum l. fruits by deep eutectic solvents-based ultrasound-assisted extraction. 203:16–22. 148

[219] L.M. Varela, T. Mendez-Morales, J. Carrete, V. Gomez-Gonzalez, B. Docampo-Alvarez, L.J. Gallego, O. Cabeza, and O. Russina. Solvation of molecular cosolvents and inorganic salts in ionic liquids: A review of molecular dynamics simulations. 210:178–188. 160

[220] Ahmadreza Roosta, Reza Haghbakhsh, Ana Rita C. Duarte, and Sona Raeissi. Machine learning coupled with group contribution for predicting the density of deep eutectic solvents. *Fluid Phase Equilibria*, 565:113672, feb 2023. 160

[221] Yumi Kobayashi, Sonoko Tokishita, and Hideki Yamamoto. Determination of hansen solubility parameters of ionic liquids by using walden plots. *Industrial & Engineering Chemistry Research*, 59(32):14217–14223, jul 2020. 160

[222] Des.solve research group. 161

[223] Xiang Zhong, Caroline Velez, and Orlando Acevedo. Partial charges optimized by genetic algorithms for deep eutectic solvent simulations. 17(5):3078–3087. 162

[224] Caroline Velez and Orlando Acevedo. Simulation of deep eutectic solvents: Progress to promises. 12(4). 162

[225] Dhawal Shah, Ulan Mansurov, and Farouq S. Mjalli. Intermolecular interactions and solvation effects of dimethylsulfoxide on type III deep eutectic solvents. 21(31):17200–17208. 162

[226] Thomas P Senftle, Sungwook Hong, Md Mahbubul Islam, Sudhir B Kylasa, Yuanxia Zheng, Yun Kyung Shin, Chad Junkermeier, Roman Engel-Herbert, Michael J Janik, Hasan Metin Aktulga, Toon Verstraelen, Ananth Grama, and Adri C T van Duin. The reaxff reactive force-field: development, applications and future directions. 2(1). 162