Investigating the Interaction of Chitin in Organic Electrolyte Solutions Using Molecular Dynamics



Thesis presented for the degree of

Master of Science

by *Lenard Leslie Carroll*

Supervisor: Dr. Gerhard Venter.

Co-Supervisor: Assoc. Prof. Anwar Jardine

Department of Chemistry

University of Cape Town

July 2019

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Investigating the Interaction of Chitin in Organic Electrolyte Solutions Using Molecular Dynamics

Abstract by Lenard Leslie Carroll

The dissolution, hydrolysis and fermentation of biopolymers afford biofuels, an alternative source of energy. Unfortunately, biopolymers have hydrogen bonding networks that are difficult to disrupt and dispersion forces to overcome, all of which make it insoluble in most common organic solvents and water under moderate conditions.

Much work has been devoted to improving the dissolution of biopolymers, via alternative solvents or by developing new ground-breaking processes. One alternative solvent that has become quite popular in biomass dissolution studies are ionic liquids (ILs). Ionic liquids are attractive solvents due to its broad range of uses and advantageous properties. ILs have been promising in its use in separation, extraction, catalysis, lubricants, fuel cells, batteries and liquid crystal research. ILs also have low vapour pressure, which implies low toxicity with respect to their clean-up. While many ILs are produced under environmentally unfriendly conditions, more studies are being done on finding ways to synthesise these species using the 12 design principles of Green Chemistry.

While a plethora of studies has been done on the experimental dissolution of cellulose in ionic liquids, similar studies have been minimal for chitin. As such, a computational investigation on the dissolution of chitin in ionic liquids and organic electrolyte solutions (OESs) is presented here. OESs consists of an ionic liquid and an additional aprotic organic molecular solvent, known henceforth as a co-solvent. These mixtures are considered as some ILs have high viscosities, which decreases its ability to effectively dissolve biopolymers, but by adding co-solvents to the IL, the mixture's viscosity decreases, potentially improving on the solubility of the biopolymer.

Computationally, the dissolution of chitin was modelled through molecular dynamics simulations as implemented in the AMBER MD code, by studying the separation of two 4-methyl-β-D-N-acetylglucosamine-(1$\rightarrow$4')-1'-methyl-β-D-N'-acetylglucosamine ($(GlcNAc)_2Me_2$) molecules (chosen as the model for chitin) in various solvent systems using potential of mean force calculations. The ionic liquids of choice were 1-butyl-3-methylimidazolium acetate ($[C_4C_1im][CH_3COO]$) and 1-butyl-3-methylimidazolium methyl sulfate ($[C_4C_1im][CH_3SO_4]$), two ILs that have experimental physical properties available, a requirement for MD simulation validation. The co-solvents chosen were dimethyl carbonate, propylene carbonate and γ-valerolactone, three structurally similar bio-based solvents.

The solvation of a $(GlcNAc)_2Me_2$ monomer was also studied in this project via radial distribution functions, interaction energies and hydrogen bond analyses, as to support the results produced from the separation study. Additionally, the experimental swelling of chitin was investigated as to compare it to the interaction energy results, acting as further validation of the computational results.

The computational results suggest that the $(GlcNAc)_2Me_2$ monomer will interact more favourably with pure $[C_4C_1im][CH_3COO]$, followed by the 8:2 $[C_4C_1im][CH_3COO]$:co-solvent OESs, the 2:8 $[C_4C_1im][CH_3COO]$:co-solvent OESs and then the pure co-solvents. The solvation study agrees with this trend. The PMF results also show that a $(GlcNAc)_2Me_2$ dimer will separate spontaneously in all the solvent systems, with the least amount of thermodynamic work required (to separate) in pure $[C_4C_1im][CH_3COO]$, and the most in pure dimethyl carbonate.

Acknowledgements

This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number 106067), and for that I would like to sincerely thank the National Research Foundation (NRF) and the University of Cape Town (UCT) for making it possible for me to do my master's degree in chemistry. Your help has allowed me to work under the supervision of Doctor Gerhard Venter and Associate Professor Anwar Jardine. The advice, mentorship and help that I have received from my supervisors are invaluable. I believe the skills and knowledge that I have learned under your supervision has only helped me exponentially grow as a scientist, and I appreciate that with all my heart. Working in the Scientific Computing Research Unit (SCRU) lab has also been a tremendous experience, as such, I would like to thank the SCRU lab and the Centre for High Performance Computing (CHPC) for the computing time and -resources. Most importantly, I would like to thank my partner, Mahlatse Ndou. You have stood by my side through my lowest of lows and you were there helping me celebrate my highest of highs. With you in my life I don't have to ask more from it, I have everything.

Thank You!

Table of Contents

Abstract **I**
Acknowledgements **II**
Table of Contents **III**
List of Tables **V**
List of Figures **VI**
List of Abbreviations **X**
List of Common Symbols and Constants **XII**
Glossary **XIV**

1 Introduction **1**
1.1 The Need for Biofuels 1
1.2 The Problem with the Production of Biofuels 1
1.3 The Crystal Structures and -Forms of Cellulose and Chitin 2
1.3.1 Processes Used For Biomass Dissolution 4
1.4 Ionic Liquids: Its Properties, History, Application, Advantages and Disadvantages 5
1.4.1 Biomass Dissolution using Ionic Liquids 8
1.4.2 Biomass Dissolution Mechanism in Ionic Liquids: the Role of the Ions 9
1.5 A Better Biomass Dissolution with the Addition of Co-Solvents 11
1.6 Research Focus 11
1.7 Aims and Objectives 12
1.7.1 Computational Aims 12
1.7.2 Experimental Aims 12
1.7.3 Computational Objectives 12
1.7.4 Experimental Objectives 13
1.8 Choice of Ionic Liquids, Co-solvents and Organic Electrolyte Solution Ratios 13

2 Experimentation **16**
2.1 Synthesis of $[C_4C_1im][CH_3COO]$ 16
2.1.1 NMR Data Analysis of $[C_4C_1im][CH_3COO]$ 16
2.2 Dissolution and Swelling 18
2.2.1 Calculating the Swelling Capacity of Chitin 19
2.2.1.1 Swelling Capacity Results 20

3 Quantum Mechanics **22**
3.1 Operators 22
3.2 Approximations to Schrödinger's Equation 23
3.3 Spin Functions, Spatial Functions and Slater Determinants 24
3.4 Hartree-Fock Theory 25
3.5 Møller-Plesset Perturbation Theory 27
3.6 Basis Sets 27
3.7 Geometry Optimisation With Quantum Mechanics 29

4 Molecular Mechanics and -Dynamics **31**
4.1 Force Fields 31
4.2 Electrostatic Interactions 32
4.3 Van der Waals Interactions 33
4.4 Force Fields in AMBER 33
4.5 Some Background on Molecular Dynamics 34
4.5.1 Verlet Algorithm 35
4.5.2 Leapfrog Algorithm 36
4.5.3 Particle Mesh Ewald Method: Ewald Summation 36

5 Force Field Parameterisation **39**
5.1 Introduction to Theory 39
5.1.1 Parameterisation of a System 39
5.1.2 Minimisation, Heating, Equilibration and Production 40
5.1.2.1 Periodic Boundary Conditions and Imaging 40
5.2 Computational Details 41
5.3 Building and Parameterising the Pure Solvent Systems 42

6 Solvation of $(GlcNAc)_2Me_2$ **53**
6.1 Introduction to Theory 53
6.1.1 Radial Distribution Functions 53
6.1.2 Interaction Energy 53
6.1.3 Hydrogen Bond Analysis 53
6.2 Computational Details 54
6.3 Solvation Results 55
6.3.1 Interaction Energy Results 56
6.3.2 RDF Results 63
6.3.3 Hydrogen Bond Analysis Results 72

7 Potential of Mean Force Calculations Using Umbrella Sampling **77**
7.1 Introduction to Theory 77
7.1.1 Potential of Mean Force 77
7.1.2 Umbrella Sampling 77
7.1.3 Using the Weighted Histogram Analysis Method to Obtain Unbiased PMFs 78
7.2 Computational Details 79
7.3 Potential of Mean Force Calculation Analysis 81
7.3.1 Pure Solvent PMF Results 86
7.3.2 OES PMF Results 88

8 Conclusion **93**

9 Future Work **95**

10 References **97**

Apppendix 1 **104**

Appendix 2 **117**

List of Tables

Table 2.1: The swelling capacity of chitin in various solvent systems as heated to 80 °C
Table 2.2: The molar swelling capacity of chitin in various solvent systems as heated to 80 °C
Table 5.1: Comparison of the computational and experimental density of propylene carbonate, dimethyl carbonate, γ-valerolactone, 1-butyl-3-methylimidazolium acetate and 1-butyl-3-methylimidazolium methyl sulfate
Table 5.2: Comparison of the computational and experimental heat of vaporisation of propylene carbonate, dimethyl carbonate, γ-valerolactone and 1-butyl-3-methylimidazolium acetate
Table 5.3: Comparison of the computational and experimental self-diffusion coefficients of the GAFF2 parameterised 1-butyl-3-methylimidazolium methyl sulfate
Table 5.4: Comparison of the computational and experimental density of the scaled 1-butyl-3-methylimidazolium acetate and 1-butyl-3-methylimidazolium methyl sulfate ionic liquids
Table 5.5: Comparison of the computational and experimental heat of vaporisation of the scaled 1-butyl-3-methylimidazolium acetate ionic liquids
Table 5.6: Comparison of the computational and experimental self-diffusion coefficients of the scaled 1-butyl-3-methylimidazolium methyl sulfate IL
Table 5.7: The total interaction energy between a $(GlcNAc)_2Me_2$ molecule and various solvent molecules
Table 5.8: The total interaction energy between acetate and 1-butyl-3-methylimidazolium from the solvation of a $(GlcNAc)_2Me_2$ molecule by various solvent molecules
Table 5.9: Frequency of occurrence (%) of the intraring O3 (Reducing)-H$\cdots$O5 (Non-Reducing) intramolecular hydrogen bond from $(GlcNAc)_2Me_2$, the number of average intramolecular hydrogen bonds formed in a $(GlcNAc)_2Me_2$ molecule and the number of average intermolecular hydrogen bonds formed between a $(GlcNAc)_2Me_2$ molecule and various solvent molecules
Table 5.10: The relative binding free energy of two $(GlcNAc)_2Me_2$ molecules in various solvent systems
Table 6.1: The total solvation energy between a $(GlcNAc)_2Me_2$ molecule and various solvent components
Table 6.2: The average partial charge for the oxygen atoms in the pure solvents
Table 6.3: The total normalised (to 1020 ion pairs) interaction energy between acetate and 1-butyl-3-methylimidazolium from the solvation of a $(GlcNAc)_2Me_2$ molecule by various solvent systems
Table 6.4: Number of average intermolecular hydrogen bonds formed between the carbonyl oxygen atoms from the solvent components and hydrogen bond donating substituents of $(GlcNAc)_2Me_2$
Table 6.5: Frequency of occurrence (%) of the intraring O3 (Reducing)-H$\cdots$O5 (Non-Reducing) intramolecular hydrogen bond from $(GlcNAc)_2Me_2$, the number of average intramolecular hydrogen bonds formed in a $(GlcNAc)_2Me_2$ molecule, the number of average intermolecular hydrogen bonds formed between a $(GlcNAc)_2Me_2$ molecule and various solvent components, and the frequency of occurrence of all cation (using the -N-C-N- carbon atom) to $(GlcNAc)_2Me_2$ intermolecular hydrogen bonds
Table 7.1: The well-depth and energy barriers of $(GlcNAc)_2Me_2$ separation in pure solvent
Table 12.1: Comparison of the bond lengths of 1-butyl-3-methylimidazolium from this work to the experimentally known bond lengths of the cation from 1-butyl-3-methylimidazolium chloride

List of Figures

Figure G.1: Flow chart depicting the Confab algorithm.
Figure G.2: A general molecular orbital diagram highlighting the HOMO and LUMO.
Figure G.3: Primitive cubic unit cell with the lattice points indicated at the corners of the box.
Figure 1.1: (From left-to-right) a schematic illustration of the significant hydrogen bonds between stacked cellulose strands (left) and between adjacent cellulose strands (right).
Figure 1.2: The I_β crystalline form of cellulose.
Figure 1.3: The crystalline form of α-chitin.
Figure 1.4: A schematic illustration of the hydrogen bonding differences between (a) α-chitin and (b) β-chitin.
Figure 1.5: The structures of chitin (left) and chitosan (right).
Scheme 1: The viscose process of cellulose degradation.
Scheme 2: The Lyocell process of cellulose degradation.
Figure 1.6: A chart of applications and uses of ionic liquids.
Figure 1.7: The structures of some of the common cations (left) and anions (right) used to make up an ionic liquid.
Figure 1.8: The proposed dissolution mechanism of cellulose in ionic liquid.
Figure 1.9: An illustration highlighting the repulsive nature of anions and cations during cellulose dissolution.
Figure 1.10: The structure of 4-methyl-β-D-N-acetylglucosamine-(1$\rightarrow$4')-1'-methyl-β-D-N'-acetylglucosamine ($(GlcNAc)_2Me_2$).
Figure 1.11: The structure of 1-butyl-3-methylimidazolium acetate.
Figure 1.12: The structure of 1-butyl-3-methylimidazolium methyl sulfate.
Figure 1.13: The structures of dimethyl carbonate (left), propylene carbonate (middle) and γ-valerolactone (right).
Figure 2.1: Labelled two-dimensional structure of 1-butyl-3-methylimidazolium.
Figure 2.2: ^{1}H-NMR spectrum of 1-butyl-3-methylimidazolium acetate.
Figure 2.3: ^{13}C-NMR spectrum of 1-butyl-3-methylimidazolium acetate
Figure 2.4: An illustration of the swelling capacity determination procedure.
Figure 3.1: The radial dependence of Gaussian- (left) and Slater type functions for the 1s orbital.
Figure 3.2: A direct comparison of the Gaussian type orbital (GTO) function to the Slater type orbital (STO) function for the 1s orbital.
Figure 4.1: The radial dependence of Gaussian- (left) and Slater type functions for the 1s orbital.
Figure 4.2: A direct comparison of the Gaussian type orbital (GTO) function to the Slater type orbital (STO) function for the 1s orbital.
Figure 5.1: The density profiles of the a) $[C_4C_1im][CH_3COO]$ 6-31G* parameter set and b) $[C_4C_1im][CH_3SO_4]$ 6-31+G* parameter set.
Figure 5.2: The geometry optimised structures of propylene carbonate (left), dimethyl carbonate (middle) and γ-valerolactone (right).
Figure 5.3: The geometry optimised structure of 1-butyl-3-methylimidazolium as computed at the MP2/6-31G* level of theory.
Figure 5.4: A top down view of the geometry optimised structures of the acetate anion at the MP2/6-31+G* (left) and MP2/6-31G* (right) levels of theory.
Figure 5.5: The 8 unique conformers of 1-butyl-3-methylimidazolium produced from the geometry optimisation of the 26 conformers obtained from the `Confab/Openbabel` conformational search.
Figure 5.6: The 8 unique conformers of 1-butyl-3-methylimidazolium produced from the geometry optimisation of the 26 conformers obtained from the `Confab/Openbabel` conformational search.

Figure 5.7: The three torsions selected for fitting in 1-butyl-3-methylimidazolium.
Figure 5.8: The torsional profile fitting of torsions na-c3-c3-c3 (left), c3-c3-c3-c3 (middle) and cc-na-c3-c3 (right). The red profile is for the MM torsional profile and the blue is for QM.
Figure 5.9: The C1-O1-S1-O2 torsion of methyl sulfate (left), the torsional profile fitting of the methyl sulfate anion as computed using the 6-31+G* basis set (middle) and the torsional profile fitting of methyl sulfate as computed using the 6-31G* basis set (right). The red profile is for the MM torsional profile and the blue is for QM.
Figure 5.10: A simulation box consisting of 256 propylene carbonate molecules.
Figure 5.11: The original MSD versus τ plot to calculate the self-diffusion coefficient of 1-butyl-3-methylimidazolium (left), and the truncated MSD versus time change plot for 1-butyl-3-methylimidazolium (right).
Figure 6.1: The list of atoms from the solvent components and chitin monomer used in the radial distribution function calculations and hydrogen bond analyses.
Figure 6.2: The structure of a minimised $(GlcNAc)_2Me_2$ molecule.
Figure 6.3: An initial simulation box built in `Packmol` containing one centred $(GlcNAc)_2Me_2$ (blue) molecule surrounded by 816 propylene carbonate (grey) molecules and 204 1-butyl-3-methylimidazolium acetate (red) ion pairs.
Figure 6.4: Time series of non-bonded energies between the $(GlcNAc)_2Me_2$ monomer and a) pure propylene carbonate, b) pure acetate, c) pure 1-butyl-3-methylimidazolium, d) acetate from 8:2 IL:γV, e) $[C_4C_1im]^+$ from 8:2 IL:γV, f) γ-valerolactone from 8:2 IL:γV, g) dimethyl carbonate from 2:8 IL:DC, h) acetate from 2:8 IL:DC and i) $[C_4C_1im]^+$ from 2:8 IL:DC.
Figure 6.5: Total solvation energy versus ideal total solvation energy plot for the 8:2 and 2:8 IL:co-solvent OESs.
Figure 6.6: Total solvation energy versus swelling capacity (per mass) plot.
Figure 6.7: The electrostatic- and van der Waals energy between propylene carbonate and the $(GlcNAc)_2Me_2$ molecule in the 8:2 1-butyl-3-methylimdiazolium acetate:propylene carbonate OES.
Figure 6.8: The electrostatic- and van der Waals energy between propylene carbonate and the $(GlcNAc)_2Me_2$ molecule in the 9:1 1-butyl-3-methylimdiazolium acetate:propylene carbonate OES.
Figure 6.9: The equilibrated simulation boxes used for solvating a $(GlcNAc)_2Me_2$ monomer with 8:2 IL:PC. Here the IL ion pairs are coloured pink, while the propylene carbonate molecules are represented in purple and the carbohydrate as green.
Figure 6.10: A time series of the average non-bonded energies for the interaction between a $(GlcNAc)_2Me_2$ monomer with propylene carbonate molecules from the four 8:2 IL:PC simulation boxes.
Figure 6.11: A time series of the average non-bonded interaction energy for the interaction between a $(GlcNAc)_2Me_2$ monomer with all the ion pairs/molecules from 8:2 IL:PC.
Figure 6.12: Radial distribution functions and integrated RDFs between C4 carbon atom from 1-butyl-3-methylimidazolium and the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer. The shaped-lines (or those which are not purely a solid representation) show the integration of RDFs.
Figure 6.13: A snapshot of the 1-butyl-3-methylimdazolium C4 carbon atoms (green spheres) solvating the $(GlcNAc)_2Me_2$ monomer within 6 Å (away from it) for the $[C_4C_1im][CH_3COO]$ IL.
Figure 6.14: Radial distribution functions and integrated RDFs between the carbonyl (carbon) atom from acetate or sulfur atom from methyl sulfate and the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer. The shaped-lines (or those which are not purely a solid representation) show the integration of RDFs.
Figure 6.15: A snapshot of the acetate carbonyl (carbon) atoms (purple spheres) solvating the $(GlcNAc)_2Me_2$ monomer within 6 Å (away from it) for the $[C_4C_1im][CH_3COO]$ IL.

Figure 6.16: Radial distribution functions and integrated RDFs between the carbonyl (carbon) atoms from the co-solvents and the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer. The shaped-lines (or those which are not purely a solid representation) show the integration of RDFs.
Figure 6.17: A snapshot of the propylene carbonate carbonyl (carbon) atoms (brown spheres) solvating the $(GlcNAc)_2Me_2$ monomer within 6 Å (away from it) for the pure propylene carbonate simulation box.
Figure 6.18: Ratio of propylene carbonate carbonyl (carbon) atoms and $[C_4C_1im][CH_3COO]$ carbonyl- and -N-C-N- carbon atoms coordinating around the glycosidic oxygen atom from $(GlcNAc)_2Me_2$.
Figure 6.19: Charge of the solvent species coordinating around the $(GlcNAc)_2Me_2$ monomer at various separations. In the top figure $[C_4C_1im][CH_3COO]$ is used, while for the bottom figure $[C_4C_1im][CH_3SO_4]$ was chosen.
Figure 6.20: Radial distribution functions involving the carbonyl oxygen atoms (or those of o-type) with 1.a) the N2 non-reducing nitrogen atom, 1.b) the N2 reducing nitrogen atom, 2.a) the O3 non-reducing oxygen atom, 2.b) O3 reducing oxygen atom, 3.a) O6 non-reducing oxygen atom and 3.b) O6 reducing oxygen atom from the $(GlcNAc)_2Me_2$ monomer.
Figure 6.21: A snapshot of the methyl sulfate anions solvating around the $(GlcNAc)_2Me_2$ 6 Å away from it.
Figure 6.22: A snapshot of the methyl sulfate anions solvating around the $(GlcNAc)_2Me_2$ 6 Å away from it.
Figure 6.23: Snapshots of all the (left) acetate ions, (middle) 1-butyl-3-methylimidazolium cations and (right) γ-valerolactone molecules from 8:2 IL:γV solvating the $(GlcNAc)_2Me_2$ monomer 6 Å away from it.
Figure 7.1: An example of poor- and good histogram overlap.
Figure 7.2: The crystal structure of chitin.
Figure 7.3: The gas-phase separation of two $(GlcNAc)_2Me_2$ molecules at 3.50 Å and 29.75 Å, respectively.
Figure 7.4: The histograms created by `WHAM` for the gas-phase separation of two $(GlcNAc)_2Me_2$ molecules.
Figure 7.5: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in the gas-phase.
Figure 7.6: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in TIP3P water.
Figure 7.7: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in 1-butyl-3-methylimidazolium acetate at 298.15 K, using production runs of 50 ns (red) and 100 ns (blue).
Figure 7.8: Natural logarithm of the RDF between two $(GlcNAc)_2Me_2$ molecules in 1-butyl-3-methylimidazolium acetate at 353.15 K, after 150 ns of production simulation time.
Figure 7.9: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in 1-butyl-3-methylimidazolium acetate at 353.15 K for a production run of 50 ns.
Figure 7.10: A snapshot of the immediate solvation of the $(GlcNAc)_2Me_2$ dimer in $[C_4C_1im][CH_3SO_4]$ separated at 5.00 Å.
Figure 7.11: A snapshot of the immediate solvation of the $(GlcNAc)_2Me_2$ dimer in $[C_4C_1im][CH_3SO_4]$ separated at 6.50 Å.
Figure 7.12: The histograms created by WHAM for the separation of two $(GlcNAc)_2Me_2$ molecules in pure γ-valerolactone (at 353.15 K.)
Figure 7.13: The PMF curves of separating a $(GlcNAc)_2Me_2$ dimer in pure $[C_4C_1im][CH_3COO$, $[C_4C_1im][CH_3SO_4]$, γ-valerolactone, propylene carbonate and dimethyl carbonate.

Figure 7.14: An illustration of simulation boxes created for the 8:2 and 2:8 IL:co-solvent OESs, in which the (blue) $(GlcNAc)_2Me_2$ dimer (separated at 3.50 Å) is directly surrounded by co-solvent (green) molecules and then ionic liquid (red) ion pairs.
Figure 7.15: An illustration of simulation boxes created for the 8:2 and 2:8 IL:co-solvent OESs, in which the (blue) $(GlcNAc)_2Me_2$ dimer (separated at 3.50 Å) is directly surrounded by ion pairs (red) and then co-solvent (green) molecules.
Figure 7.16: The potential of mean force curves for the separation of two $(GlcNAc)_2Me_2$ monomers in the 8:2 and 2:8 IL:co-solvent OESs. The simulation boxes used here was for the co-solvent starting close to the dimer.
Figure 7.17: The potential of mean force curves for the separation of two $(GlcNAc)_2Me_2$ monomers in the 8:2 and 2:8 IL:co-solvent OESs. The simulation boxes used here was for the ion pairs starting close to the dimer.
Figure 7.18: a) Interaction energy between each $(GlcNAc)_2Me_2$ monomer at various radial separations. b) Solvation energy between the solvent and the dimer (single entity) at various monomer-monomer radial separations. c) Number of average intermolecular hydrogen bonds between the two separate monomers at various radial separations. d) Number of average intramolecular hydrogen bonds in the dimer (single entity) at various monomer-monomer radial separations. The mixtures chosen here are those fitting the description of Figure 7.14.
Figure 12.1: Partial charges of propylene carbonate as calculated by the R.E.D. Server at the HF/6-31G* level of theory.
Figure 12.2: Partial charges of dimethyl carbonate as calculated by the R.E.D. Server at the HF/6-31G* level of theory.
Figure 12.3: Partial charges of γ-valerolactone as calculated by the R.E.D. Server at the HF/6-31G* level of theory.
Figure 12.4: Partial charges of 1-butyl-3-methylimidazolium as calculated by the R.E.D. Server at the HF/6-31G* level of theory.
Figure 12.5: The partial charges of acetate as calculated by the R.E.D. Server at the HF/6-31+G* (left) and HF/6-31G* (right) levels of theory.
Figure 12.6: The partial charges of methyl sulfate as calculated by the R.E.D. Server at the HF/6-31G* (left) and HF/6-31+G* (right) levels of theory.
Figure 12.7: The PMF curve of separating two $(GlcNAc)_2Me_2$ monomers in pure $[C_4C_1im][CH_3COO]$ at 353.15 K in which the production phase was run for 150 ns.

List of Abbreviations

AD - Anaerobic Digestion
AFM - Atomic Force Microscopy
AIMD - *Ab Initio* Molecular Dynamics
AMBER - Assisted Model Building with Energy Refinement
$[AMIM]^+$ - 1-Allyl-3-methylimidazolium
$[C_4C_1py]Cl$ - 1-Butyl-3-methylpyridinium Chloride
B3LYP - Becke 3-Parameter (Exchange), Lee, Yang and Parr
BSSE - Basis Set Superposition Error
$[C_2C_1im]^+$/$[EMIM]^+$ - 1-Ethyl-3-methylimidazolium
$[C_4C_1im]^+$/$[BMIM]^+$- 1-Butyl-3-methylimidazolium
$[C_4C_1pip]^+$ - N-Butyl-N-methylpiperidium
$[C_4C_1pyr]^+$ - N-Butyl-N-methylpyrrolodinium
$[C_2OHMIM]$ - 1-(2-hydroxyethyl)-3-methylimidazolium
$[CH_3COO]^-$/Ac - Acetate
$[CH_3SO_4]^-$ - Methyl Sulfate
CHPC - Centre for High Performance Computing
CP-MAS - Cross-Polarization Magic Angle Spinning
CPU - Central Processing Unit
DC - Dimethyl Carbonate
DFT - Density Functional Theory
DMAc - Dimethylacetamide
DMSO - Dimethyl Sulfoxide
DSC - Differential Scanning Calorimetry
$[Et_4N][GeCl_3]$ - Tetraethylammonium Trichlorogermanate
$[Et_4N][SnCl_3]$ - Tetraethylammonium Trichlorostannanide
$[EtNH_3][NO_3]$ - Ethylammonium Nitrate
FT-IR - Fourier-Transform Infrared Spectroscopy
GA - Genetic Algorithm
GAFF - General Amber Force Field
Glc - (1,4)-Dimethoxy-β-D-glucose
$(GlcNAc)_2Me_2$ - 4-Methyl-β-D-N-acetylglucosamine-(1$\rightarrow$4')-1'-methyl-β-D-N'-acetylglucosamine
GPU - Graphical Processing Unit
GTO - Gaussian Type Orbital
γV - γ-Valerolactone
HB - Hydrogen Bond
HF - Hartree-Fock
HOMO - Highest Occupied Molecular Orbital
IE - Interaction Energy
IL - Ionic Liquid
LiCl - Lithium Cloride
LUMO - Lowest Unoccupied Molecular orbital
MAX - Maximum
MD - Molecular Dynamics
$[Me_2PO_4]^-$ - Dimethyl Phosphate
MM - Molecular Mechanics
MMFF - Merck Molecular Force Field

$[MMIM]^+$ - 1,3-Dimethylimidazolium Dimethyl
MO - Molecular Orbital
MP - Møller-Plesset
MSD - Mean Square Displacement
NMMO - N-Methylmorpholine-N-oxide
NaOH - Sodium Hydroxide
NMR - Nuclear Magnetic Resonance
OES - Organic Electrolyte Solution
PBC - Periodic Boundary Condition
PC - Propylene Carbonate
PES - Potential Energy Surface
PMEMD - Particle Mesh Ewald Molecular Dynamics
PDF - Probability Distribution Function
PMF - Potential of Mean Force
QM - Quantum Mechanics
QUILL - Queen's University Ionic Liquids Laboratory
RDF - Radial Distribution Function
RED - Relative Energy Difference
R.E.D. - RESP ESP charge Derive
RESP - Restrained ElectroStatic Potential
RPM - Rotations Per Minute
RT - Room Temperature
SCF - Self-Consistent Field
SCRU - Scientific Computing Research Unit
SDF - Spatial Distribution Function
SEM - Scanning Electron Microscopy
SMD - Steered Molecular Dynamics
STO - Slater Type Orbitals
TGA - Thermogravimetric Analysis
TI - Thermodynamic Integration
TIP3P - Transferable Intermolecular Potential with 3 Points
WHAM - Weighted Histogram Analysis Method
XRD - X-Ray powder Diffraction

List of Common Symbols and Constants

∞ - Infinity symbol
* - The addition of polarisation functions to a basis set or the conjugate of a complex number
$^\circ$ - Degrees
°C - Degree Celsius
a_j - Acceleration of a system j
Å - Ångstrom (1 Å = 1×10^{-10} m)
cal - Calories
cm - Centimetre ($\times10^{-2}$ m)
D_j - Self-diffusion coefficient of a system j
∂ - Partial derivative
e - Euler's number = 2.7182...
e^- - An electron
E_j - Total energy of a system j
ϵ_0 - Vacuum permittivity = 8.854×10^{-12} F.m^{-1}
ε - Well-depth in the van der Waals energy
$\sum_j$ - The sum (or addition) with respect to all j systems
f - Femto ($\times10^{-15}$)
F_j - Force of a system j
g - Grams
ΔG - Gibbs Free Energy
$\hbar$ - Reduced Planck's constant = $\frac{6.626\times10^{-34}\text{m}^2\text{kg/s}}{2\pi}$
$\Delta_{\text{vap}}H$ - Heat of Vaporisation
i - Imaginary number = $\sqrt{-1}$
J - Joules
k - Kilo ($\times1000$)
K - Kelvin (Temperature in °C + 273.15)
k_B - Boltzmann's constant = 1.381×10^{-23} m^2.kg.s^{-2}.K^{-1}
ln - The natural logarithmic function
m_j - Mass of a system j
M - Mega ($\times10^6$)
M_j - Molar mass of a system j
mol - Moles
n - Nano ($\times10^{-9}$)
η_j - Viscosity of a system j
$N!$ - $N\times N-1\times N-2\times N-3$ ... $\times3\times2\times1$
N_A - Avogadro's number = 6.022×10^{22} molecules.mol^{-1}
p - Pico ($\times10^{-12}$)
ρ_j - Density of a system j
q_j - Partial charge of a system j
r_{jk} - Total distance between systems j and k
R - Universal gas constant = 8.314 J/mol.K
s - Seconds
$\int$ - Integral
t - Time
π - Pi = 3.1415...
µ - Micro ($\times10^{-6}$)

v_j - Velocity of a system j
V_j - Volume of a system j
ω_j - A target torsional angle from system j
Ψ_j - Wavefunction of a system j

Glossary

- Ballistic motion describes the collision (nature) of atoms, specifically at the start of a production calculation. It is thus (in a way) the movement of atoms/components during the continuous bombardment of surrounding components.
- Bio-based materials are wholly- or partly derived from materials of biological origin. For example, γ-valerolactone is derived from cellulose, which is isolated from plants, wood, cotton, etc.
- A biodegradable substance, is one that can be broken down (decomposed) by bacteria or other living organisms, thus avoiding pollution.
- Boundary effects are the effects caused/experienced on a simulation box and its containing atoms when it is not enclosed by a boundary (like certain solvents are enclosed in a glass bottle).
- The Confab algorithm is a process used to determine the lowest energy conformers of an ionic/molecular component. The process is summarised in Figure G.1 (O'Boyle, N.; Vandermeersch, T.; Flynn, C.; Maguire, A.; Hutchison, G. *J. Cheminform.* **2011**, *3*, 1-9) below:

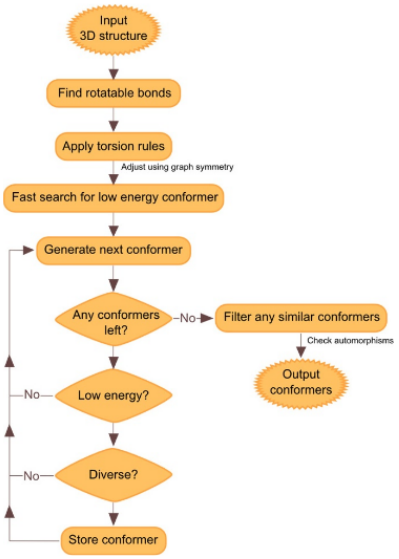


Figure G.1: Flow chart depicting the Confab algorithm.

- Dipole-dipole interactions are the attractive interaction that exists between two dipolar molecules through space, in which the one molecule is characterised as partially negative and the other as partially positive.
- A dipole-induced dipole interaction is the interaction between one dipolar molecule and another molecule which has a dipole induced into it (through an ion/dipolar molecule).
- The Dirac delta function $\delta(t)$ is defined as $\lim_{\varepsilon\to 0}\frac{H(t)-H(t-\varepsilon)}{\varepsilon}$, where $H(t)$ is known as the Heaviside function with $H(t) = \begin{cases} 0 & t<0 \\ \frac{1}{2} & t=0 \\ 1 & t>0 \end{cases}$.

- Electron affinity can be described as the change in energy of a neutral atom when an electron is added to it.
- Electronic structure describes how electrons are arranged and occupies the available energy levels of an atom.
- An ensemble is set up by constructing a large number of imaginary members with three properties constant throughout. There is, for example, the NVT ensemble in which the number of members, volume and temperature is kept constant throughout the simulation.
- Ergodicity being achieved means that over a sufficiently long time, a system in some region of phase space with microstates of the same energy will become the same as the average over the statistical ensemble.
- Fossil fuels are formed from buried dead organisms being decomposed and compressed over centuries and contain energy that originates in ancient photosynthesis.
- In MO (molecular orbital) theory, the HOMO (highest occupied molecular orbital) is identified as a fully-occupied orbital in the molecular orbital diagram that is higher in energy compared to all the other occupied orbitals. The LUMO (lowest unoccupied molecular orbital) is identified as a fully-unoccupied orbital in the MO diagram that is lowest in energy compared to all the other unoccupied orbitals. See Figure G.2 for an illustration of this.

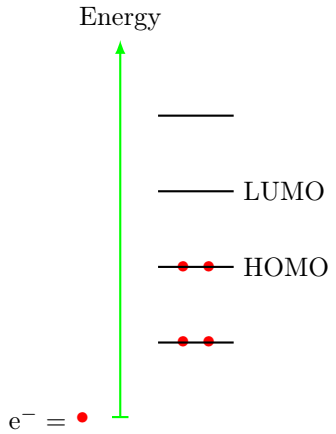


Figure G.2: A general molecular orbital diagram highlighting the HOMO and LUMO.

- Hydrophobicity is defined as the measure of the inability of a substance to dissolve in water or lack of intent to interact with water.
- Hygroscopicity is the capacity of a system to react with the moisture content of air, by either releasing or absorbing water vapour.
- Simplistically, Langevin dynamics is used to describe the motion of a particle dissolved in a liquid. The force experienced by a particle using Langevin dynamics is represented by $F_i(t) = m_i \frac{dv_i}{dt} = -\gamma v_i + R_i(t)$, where m_i is the mass of particle i, v_i is the velocity of particle i, γ is the frictional coefficient on the particle and $R_i(t)$ is the random force on the particle i at time t.
- A lattice is defined as an ordered-array of points which describes the arrangement of particles to form a crystal. Regarding a unit cell, which is the smallest part of a crystal which when regularly repeated through translation in three dimensions creates the whole crystal, a lattice is identified as the points/location of particles on the unit cell. For example, consider the

primitive cubic unit cell in Figure G.3.

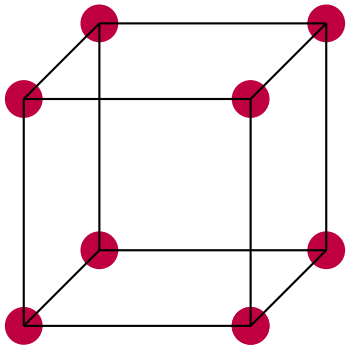


Figure G.3: Primitive cubic unit cell with the lattice points indicated at the corners of the box.

- A multipole is defined as any of several forms of a monopole with no pole strength or net charge.
- Partition functions express the number of thermally accessible states that a system provides to carriers (such as electrons). Alternatively, it can be defined as the measure of the volume occupied by a system in phase space.
- Physiochemical properties is a collection of both physical and chemical properties.
- A reaction coordinate is the shortest path along the lowest potential energy surface moving from the reactant well to the product well.
- The self-consistent field (SCF) method consists of a serious of steps. The process starts off by guessing wave functions for the occupied MOs, which is necessary to construct one-electron operators. Differentials are solved using these guessed states, which subsequently outputs a new set of wave functions, presumably different to the guessed ones. From the new wave functions, new one-electron operators are created which resultantly produces electron charge (probability) densities. This process continues until the difference between the newly determined set and immediately preceding set are below some threshold criterion. At this point, the final set of wave functions are known as the converged SCF orbitals. Using a tight SCF convergence improves on the accuracy, convergence and increases the number of SCF cycles used, but extends on computational cost as well.
- The SHAKE algorithm which is based on the Verlet algorithm (see Chapter 4), is used for coordinate and thus bond constraints.
- Thermally labile substances are substances that are (more) susceptible to change, destruction or decomposition in response to heat.
- Generally the trapezoid rule is defined as approximating the region underneath a function as trapezoids to obtain the area. Assuming that a function is represented as $P(r)$, the integration of it is represented as $A = \int_{\mathrm{L}}^{\mathrm{U}} P(r)$, in which A is the area underneath the curve, L Å is the lower radial distance boundary and U Å is the upper radial distance boundary. In the trapezoidal method, the area gets approximated as $A = \int_{\mathrm{L}}^{\mathrm{U}} P(r) \approx \sum_i (\frac{\Delta r}{2})[P(r_i) - P(r_{i-1})]$, for which $P(r_i)$ is the value of the function at radial distance r_i and Δr is related to the number of subintervals chosen.
- Vapour pressure is the pressure exerted by the gaseous or vapour form of a substance over its solid and/or liquid phase.

[Page intentionally left blank.]

1 Introduction

1.1 The Need for Biofuels

Currently, there exists several sources of energy, all available for the use in energising our vehicles and electrical devices. Some of these sources of energy are wind, solar and geothermal; however one of the primary sources used is fossil fuel energy. Not only is the burning of fossil fuels environmentally unfavourable, but its resources are depleting rapidly.[1]

Due to the rapid depletion of fossil fuel resources, its use has become more expensive, with price increases occurring regularly. Yet, even with the knowledge of price increases and resource depletion, fossil fuels are still highly depended on as alternatives for it are expensive to initially set up or use on a regular basis.[2] This disadvantage makes the search and substitution of fossil fuel energy significantly more difficult.

Fortunately, a few decades ago, it was discovered that by dissolving, hydrolysing and fermenting biomass, one can produce biofuels, such as ethanol, that can act as "greener" alternatives to fossil fuels.[3] The label of "green" is used here as this process can potentially obey the 12 design steps of Green Chemistry.[4] The 12 design steps of Green Chemistry is explained in more depth later on.

Biofuels can be produced from sustainable and virtually inexhaustible biomasses, such as cellulose or chitin. The former can be isolated from wood, cotton or sugarcane and the latter from the exoskeletons of crustaceans and insects or cell walls of fungi.[5] Unfortunately, the growing of plant-based biomasses and the processing of these into biofuels consume a lot of energy, time and finances, which makes biofuels less favourable than fossil fuels as a source of energy.[6] Moreover, a key factor in the effective- and environmentally favourable production of biofuels has not presently been achieved.

1.2 The Problem with the Production of Biofuels

Biomass consists of many polymeric strands tightly packed or locked together due to strong intermolecular- and intramolecular interactions as well as interlayer dispersion forces, such as stacking interactions.[7] In the biofuel production process, biomasses are dissolved and subsequently hydrolysed, outputting simpler sugars which when fermented produces biofuels. While it was once widely assumed that hydrogen bond interactions mainly accounted for the rigid and stable structure of biomasses as well as the tight packing of these strands, recent quantum mechanical calculations have shown that stacking interactions can also greatly contribute to these observations.[8] Hydrogen bonding originates from electrostatics and charge transfer, while stacking interactions originate from van der Waals interactions, π-π interactions and hydrophobic forces. The latter component has not been explored to the same extent that the former has in terms of carbohydrate-carbohydrate interactions.[9–10] These forces that exist within and between biomass strands cannot be overcome using common organic solvents or water under moderate conditions,[11] as such, this makes the dissolution process difficult to implement. As can be seen in Figure 1.1, intramolecular- and intermolecular hydrogen bonding interactions are plentiful in cellulose.

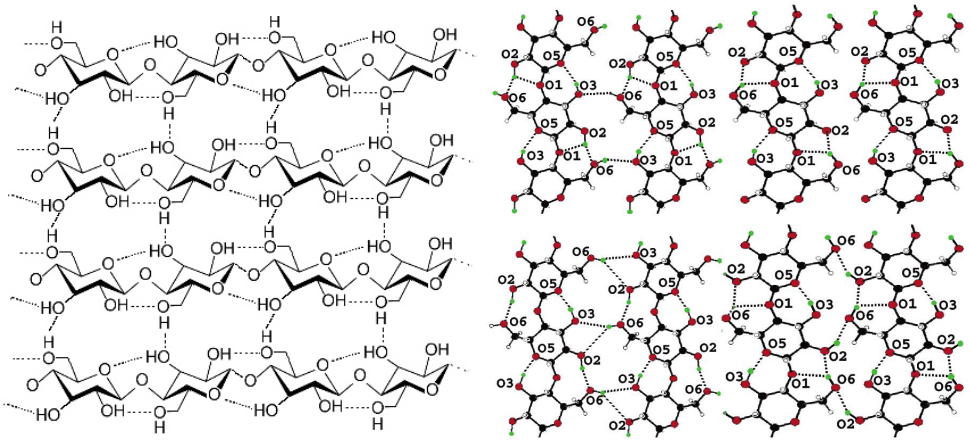


Figure 1.1: A schematic illustration of the significant hydrogen bonds between stacked cellulose strands (left) and between adjacent cellulose strands (right).[12]

The hydrogen bonding network shown for adjacent cellulose strands in Figure 1.1 was described by Nishiyama *et al.*[12] using synchrotron X-ray and neutron fibre diffraction, illustrating the great number of stabilising hydrogen bonds available in cellulose, all of which contribute together to make the biopolymer a highly stable material. Some of these hydrogen bonds include the O6-H$\cdots$O2 intermolecular- and intramolecular hydrogen bonds, the O3-H$\cdots$O5 intramolecular hydrogen bond, the O6-H$\cdots$O3 intramolecular- and intermolecular hydrogen bonds, and so forth.

1.3 The Crystal Structures and -Forms of Cellulose and Chitin

Cellulose has two different naturally-occurring crystalline forms, known as I_α and I_β. These mainly differ according to the stacking of the strands. However, pretreating cellulose under various conditions can result in more crystalline forms. For example, cellulose III_I is produced by pretreating cellulose I_β with liquid ammonia.[13] A representation of the crystalline form of cellulose I_β can be seen in Figure 1.2.

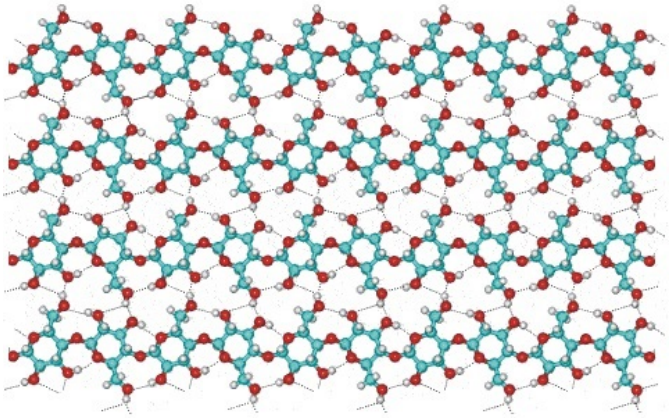


Figure 1.2: The I_β crystalline form of cellulose.[14]

Similarly, naturally occurring α and β crystalline forms are also found for chitin. In the former, the chitin chains are arranged in parallel fashion ($\uparrow\uparrow$) and is mostly found in arthropods and crustaceans, while the latter is arranged in an antiparallel fashion ($\uparrow\downarrow$) and is mainly obtained from marine diatoms.[15−16] This is different from the naturally occurring crystalline forms of cellulose, as both cellulose I crystalline forms are parallel in chain orientation.[17] In a 2017 study by Kaya *et al.*, a third naturally occurring crystalline form of chitin (isolated from the cocoon of the moth *Orgyia dubia*) was authenticated using ^{13}C CP-MAS NMR, XRD, FT-IR, Raman spectroscopy, TGA, DSC, SEM, AFM, chitinase digestive testing, elemental analysis and quantum mechanical calculations. This crystalline form is known as γ-chitin and has a $\uparrow\downarrow\uparrow$ orientation of chitin chains.[15−16,18] The presence of amino acetyl groups and the subsequent hydrogen bonds formed because of these groups has made chitin a difficult to dissolve biopolymer in most common solvents.[15−16] See Figure 1.3 for an illustration of the crystalline form of α-chitin, and Figure 1.4 for an illustration of the hydrogen bonding network differences between α-chitin and β-chitin.

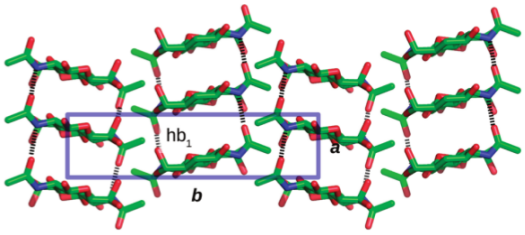


Figure 1.3: The crystalline form of α-chitin.[19]

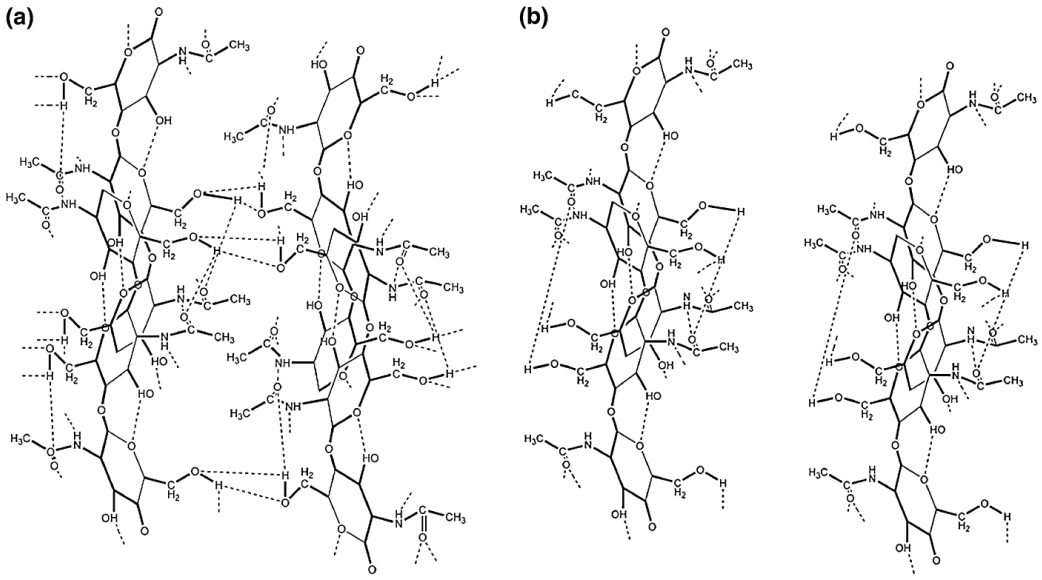


Figure 1.4: A schematic illustration of the hydrogen bonding differences between (a) α-chitin and (b) β-chitin.[20]

Since α-chitin has more hydrogen bonds available to stabilise its structure compared to β-chitin, it should require more thermodynamic work from the solvents to disrupt the hydrogen bonds of α-chitin than β-chitin, implying that the former is more difficult to dissolve. This is confirmed experimentally.[15]

As a result of chitin's poor solubility in most solvent systems, it is often modified to produce derivatives with a higher solubility.[21] One such derivation procedure is deacetylation, in which sodium hydroxide (NaOH) is added to convert the amino acetyl groups on the polymer into amine groups. Resultantly, this affords the polymer known as chitosan, which is more soluble than chitin in most solvents. This is suggested to be as a result of the greater basicity of its amine groups relative to the N-acetyl groups in chitin, which leads to a more favourable amine protonation to afford $[NH_3R]^+$, and facilitates in a greater interaction with the solvent.[22−23] See Figure 1.5 for an illustration of the structures of chitin and chitosan, respectively.

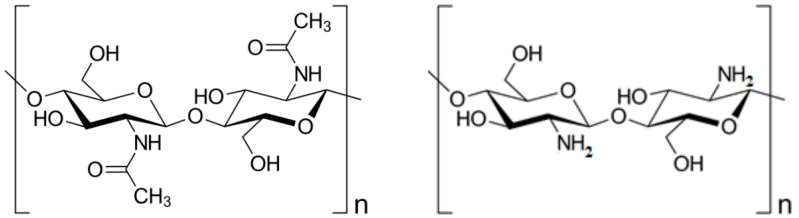


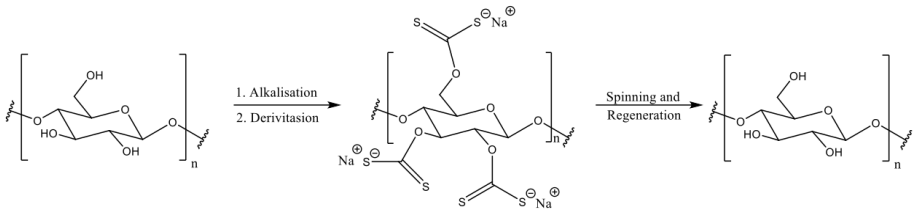
Figure 1.5: The structures of chitin (left) and chitosan (right).

Aside from the stacking and chain orientation of biomass being visually different in its various forms, these differences also result in vastly varying interaction, degrees of degradation and dissolution. Recent experiments have shown that it is easier to degrade cellulose III_I than cellulose I_β, based on the difference in hydrogen bonding network rearrangement, and a change in stacking interactions.[21,24]

In addition to the predominant O-H$\cdots$O type hydrogen bonds and stacking forces, X-ray and neutron diffraction studies have also indicated that C-H$\cdots$O type hydrogen bonding greatly contributes to the stabilising intrasheet- and intersheet interactions in cellulose.[12,13,25] Parthasarathi *et al.*'s computational study highlights that as the number of cellulose strands used for stacking (intrasheet and intersheet) increases, the interactions between the cellulose monomers will reach a point where the stacking forces start contributing more towards the stabilisation energy than the hydrogen bonding does.[13] This is especially significant for the I_β crystal form. From their study, it also appears that when smaller crystal models are used, cellulose III_I is more stable than cellulose I_β, but as the model size increases, so does the stability of cellulose I_β, to the point where it becomes more stable than the cellulose III_I model. This can explain why the latter degrades easier.[13]

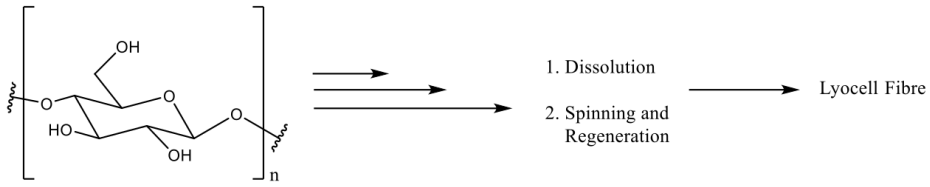
1.3.1 Processes Used for Biomass Dissolution

Despite these strong interactions, biomass can still be dissolved; however, it generally requires the use of specialist solvent systems and harsh conditions to do so. The viscose and Lyocell processes are mainly used for cellulose dissolution; and while most of the following discussion will be focused on cellulose (which has been studied more often), much of it can be transferred to chitin as well. For the viscose process (Scheme 1), a prior chemical modification has to be carried out on the biopolymer to produce cellulose xanthate; however, this can lead to severe environmental pollution.[26−27]



Scheme 1: The viscose process of cellulose degradation.[27]

The Lyocell process (Scheme 2) is the modernised and acceptable alternative to the viscose process. Here NMMO (N-methylmorpholine-N-oxide) is used as the solvent for cellulose processing. NMMO is low in toxicity, is biodegradable and has a ±98% recovery rate. In this process, unfortunately, high temperatures are required, there is a high thermal instability and oxidative side reactions may occur, all of which are unwanted.[27]



Scheme 2: The Lyocell process of cellulose degradation.[27]

Two solvent-free processes (outside of using water) that are gaining recognition are hydrothermal liquefaction and anaerobic digestion (AD). Hydrothermal liquefaction is the processing of biomass in a hot, pressurised water environment long enough for the biomass to thermochemically convert into crude liquid fuels. Typical hydrothermal processing conditions are 523-647 K and 4-22 MPa. While this technique seems highly promising, not enough information is available to accurately assess the advantages and disadvantages of this process.[28] Anaerobic digestion is a natural process used for breaking down organic materials (such as biopolymers) using microorganisms in a closed system such as high rate bioreactors. Unfortunately, performing AD in a one stage system rapidly decreases the pH in the reactor and greatly produces volatile fatty acids, which disturbs the activity of methanogenic bacteria.[29–30] It is as a result of the disadvantages mentioned in the above that ionic liquids have been considered for biomass dissolution.

1.4 Ionic Liquids: Its Properties, History, Application, Advantages and Disadvantages

Room temperature ionic liquids (RTILs) are molten salts consisting of a cation and anion that are poorly coordinated and are liquid below 100 °C.[31] Ionic liquids are classified as designer compounds as a change in the size of the ionic components, symmetry, presence of hydrogen bonding between the cation and anion and consequently how the IL's anion and cation interact, can lead to the alteration of its physiochemical properties.[32–38]

A short overview of the history of ionic liquids is presented below; for a more in-depth look at its history, see the review paper by Welton.[39] One of the earliest reports of ionic liquids dates back to the year 1914, when Walden was searching for liquid molten salts and resultantly discovered that $[EtNH_3][NO_3]$ has a low melting point of 12 °C.[40] For the following 7 decades, ionic liquid research and discovery was moving at a slow rate, but notable publications made during this time was by

Parshall who used low-melting systems such as $[Et_4N][GeCl_3]$ and $[Et_4N][SnCl_3]$ as solvents for platinum catalysed hydrogenation reactions;[41] and by Ford, who explored the toxicity and antimicrobial activity of triethylhexylammonium triethylhexylboride.[42]

The interest in ionic liquids greatly started to increase in the 1980s. Evans *et al.* studied the activity of alkaline phosphatase in water-$[EtNH_3][NO_3]$ mixtures and found that at lower concentrations, the IL has a beneficial effect on the enzyme's activity, but at a higher concentration it is inactive.[43] Evans' study was picked up by Poole who decided to delve deeper into the use of ionic liquids as a stationary phase in gas-liquid chromatography.[44] In the same decade, the Wilkes' group also looked at using 1,3-dialkylimidazolium containing ionic liquids and suggested $[C_2C_1im]^+$ to be a preferred cation for good transport properties.[45] The imidazolium-based cations are still favourably used in ionic liquid research today.[46] Wilkes further explored the applications of ionic liquids, this time as both a solvent and a catalyst in Friedel-Craft reactions,[47] with Jaeger and Tucker demonstrating a similar advantageous use of ionic liquids in Diels-Alder reactions in their 1989 paper.[48] The work done by Jaeger and Tucker came a decade before the Welton and Seddon groups highlighted the potential of ILs for these reactions. "Soon after, the first university-industry collaborative research centre dedicated to IL research was created, the Queen's University Ionic Liquids Laboratory (QUILL)."[49]. In the last few decades, more applications of ionic liquids have been discovered, advantages have been streaming in, but disadvantages have also been found.

Ionic liquids are of specific interest in this project due to their broad and extensive range of use, as well as its capabilities as a solvent. It has been used as a reaction medium, a catalyst and in separation and extraction.[50–52] Additionally, ionic liquids can be used at low to high-temperature ranges without resulting in the degradation of the solvent. Since ILs consists purely of ionic components, these species are conductive and might have high electrical stability, which is sought after for conductors and batteries as electrolytes.[53] Additionally, many ILs have a negligibly low vapour pressure due to the strong electrostatic interaction that exists between the ionic components. Most ionic liquids are also low in volatility, which implies a poor distillation, and is thus favourable for the extraction of poorly volatile or thermally labile compounds.[54] A more extensive list of applications and uses can be seen in Figure 1.6, which shows why ILs are an attractive solvent.

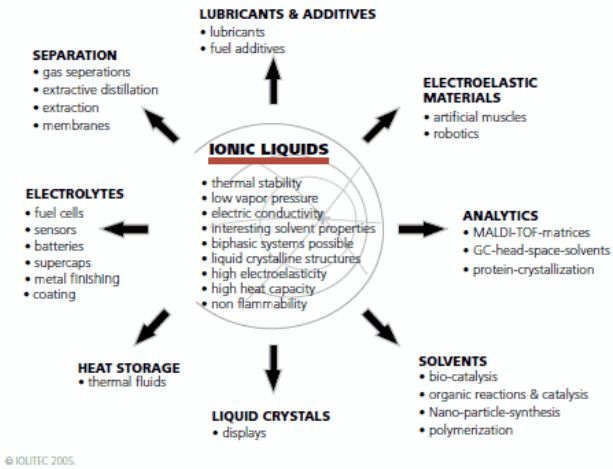


Figure 1.6: A chart of applications and uses of ionic liquids.[55]

In addition to all the prior information presented, a key advantage of ILs is their "green" capabilities. While many ionic liquids are presently created under environmentally unfavourable conditions and taking their whole life cycle into consideration might appear to be toxic, there are ionic liquids available that are not only bio-based, but show green promise.[56−57] Green chemistry is the process of minimising the use and generation of hazardous materials; and the easiest way to determine if a material or process is green, is by confirming that it agrees with the 12 design principles of Green Chemistry. These principles are "prevention, atom economy, less hazardous chemical syntheses, designing safer chemicals, safer solvents and auxiliaries, design for energy efficiency, use of renewable feedstocks, reduce derivatives, catalysis, design for degradation, real-time analysis for pollution prevention and inherently safer chemistry for accident prevention".[4] If a material or process fulfils these principles, it can be deemed to be green. In Figure 1.7 is shown a group of common cations and anions that make up an ionic liquid.

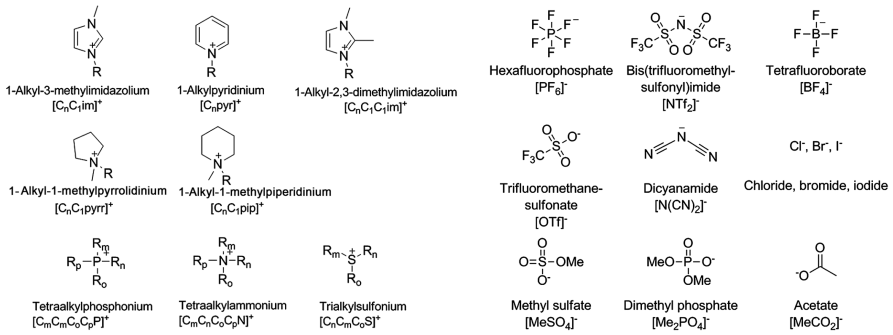


Figure 1.7: The structures of some of the common cations (left) and anions (right) used to make up an ionic liquid.

Despite many favourable properties, ionic liquids are not perfect and over the years disadvantages have been found for this material. As summarised by Wurster and Taylor in their 1965 review article, the relationship between dissolution/dissolution rate and viscosity is typically inverse.[58−61] As such, one can generally expect that as the viscosity of a solution increases, the dissolution and dissolution rate of the solute will decrease. This implies that the highly viscous ionic liquid species might poorly dissolve biopolymers as it (high viscosity) limits the interaction between the solvent and the solute. Formally, viscosity is defined as "the measure of a fluid's resistance to a gradual deformation by shear stress".[62] Informally it is referred to as the "thickness" of the fluid. Viscosity (η) is formulated as $\frac{F}{A}(\frac{v}{\Delta r})^{-1}$ where $\frac{F}{A}$ refers to the shear stress and $\frac{v}{\Delta r}$ is known as the shear velocity.[62] In a viscous solution, the biopolymer will also be transported at a slower rate; this can decrease the degree of dissolution for the biopolymer.[63]

Moreover, it should be noted that some ILs can be quite toxic, are not readily biodegradable, and considering their whole life cycle, have proven to be environmentally unfriendly.[56] Outside of environmental influences and viscosity issues, other disadvantages of ionic liquids are their generally high costs, being 5-20 times more expensive than molecular solvents, their low vapour pressure, which may obstruct distillative solvent separation, their costly multi-step synthesis as well as their high hygroscopicity.[64] Impurities found in the ionic liquid may also negatively affect the chemical properties of the IL,[65] as water contamination in some ILs have shown to drastically decrease the solubility of cellulose, perhaps due to competitive hydrogen bonding to the cellulose microfibrils.[66]

Furthermore, Ding *et al.* showed that the addition of water molecules to a mixture of cellulose and ionic liquid weakens the solute-solvent intermolecular hydrogen bonds, lowering the interaction between the two.[67] Despite these disadvantages, significant benefits remain and a search for cheaper ILs that are less harmful to the environment, but retains otherwise favourable physical and chemical properties of this class of solvents, is of paramount importance.

1.4.1 Biomass Dissolution using Ionic Liquids

In 2002, Swatloski *et al.*[66] reported the first cellulose dissolution study using room temperature dialkylimidazolium-based ILs. They showed that 1-butyl-3-methylimidazolium chloride ($[C_4C_1im]Cl$) could dissolve cellulose with a dissolving capability of 10 wt% (with heating) and up to 25 wt% with microwave irradiation. In 2005 at the conference of the Electrochemical Society, Hanley *et al.*[68] reported the first dissolution study of chitin in 1-ethyl-3-methylimidazolium chloride ($[C_2C_1im]Cl$) and mentioned that dissolved chitin can be re-precipitated from IL solution by adding methanol to the mixture. Heinze *et al.*[69] and Sashina *et al.*[70] reports the dissolution capability of cellulose in 1-butyl-3-methylpyridinium chloride ($[C_4C_1py]Cl$) as 37 wt% at 105 °C, albeit the cellulose strands were significantly depolymerised.

In a 2010 study by Wang *et al.*[71], the dissolution behaviour of biomass chitin was studied in various ionic liquids. The study was conducted on chitin with differing degrees of deacetylation.
1-allyl-3-methylimidazolium chloride ([AMIM]Cl), 1-allyl-3-methylimidazolium acetate ([AMIM]Ac), 1-butyl-3-methylimidazolium chloride ([BMIM]Cl), 1-(2-hydroxyethyl)-3-methylimidazolium chloride ($[C_2OHMIM]Cl$), 1,3-dimethylimidazolium dimethyl phosphate ($[MMIM][Me_2PO_4]$) and 1-ethyl-3-methylimidazolium dimethyl phosphate ($[EMIM][Me_2PO_4]$) were selected as the ionic liquids. From the study, native chitin was found to be insoluble in [BMIM]Cl (it has a dissolution capability of 10 wt% in cellulose for reference) and $[C_2OHMIM]Cl$, while it only had a solubility (%wt) of 0.5%, 1.5%, 1.5% and 5% in [AMIM]Cl, $[MMIM][Me_2PO_4]$, $[EMIM][Me_2PO_4]$ and [AMIM]Ac, respectively. The study also found that as the degree of deacetylation increases for chitin (the number of amino acetyl groups decreases), the solubility of the biopolymer increases. Furthermore, the study illuminated the notion that the stronger the hydrogen bond accepting capabilities of the anion is, the better the biopolymer dissolution should be.

In 2012, Jaworska *et al.*[72] suggested that chitin dissolution studies should be carried out in ionic liquid at elevated temperature rates as all prior processes have been carried out at 80-110 °C or with microwave irradiation. According to Qin *et al.*[73], who performed the dissolution of chitin in 1-ethyl-3-methylimidazolium acetate under microwave heating, a solubility of 20% for chitin was obtained. In a 2018 study by Zhu *et al.*, a one-step dissolution, extrusion and fibre spinning technique was used on chitin together with ionic liquid solvents. The chitin used was isolated from snow crab and added to $[C_2C_1im][CH_3COO]$ and heated to 145 °C for 6 hours with 100 rpm stirring. Hereafter, specifically-designed fibre spinning equipment was used for dry-jet wet fibre spinning of chitin.[74]

In a theoretical study by Ding *et al.*[67], the dissolution of cellulose was studied using density functional theory (DFT). (1,4)-dimethoxy-β-D-glucose (Glc) was the model of choice for cellulose, while $[C_2C_1im][CH_3COO]$ was chosen as the solvent. They showed that the interaction energy (intermolecular) is stronger between Glc and IL molecules than between two Glc molecules. This result suggests there is a thermodynamic preference for the Glc molecules to rather interact with the IL molecules than with other Glc molecules, which might explain why cellulose has a favourable dissolution in $[C_2C_1im][CH_3COO]$. Payal *et al.*[75] studied the interaction between cellulose (using model cellobiose) and various ionic liquids through radial distribution functions and potential of mean force (PMF)

calculations using molecular dynamics simulations. In the study, cellobiose was transferred from vacuum into a solvent system, perpendicular to the solvent surface. From this, the thermodynamic work required in moving the cellobiose monomer into the solvent was calculated. The more negative the resultant PMF for a specific IL, the less work is required to move cellobiose into that IL. Their results showed that cellobiose interacts the strongest with 1-butyl-3-methylimidazolium acetate and the weakest with 1-butyl-3-methylimidazolium hexafluorophosphate.

Gross *et al.*[76] studied the interaction between cellobiose and water as well as cellobiose and 1-butyl-3-methylimidazolium chloride through radial distribution functions (RDFs), spatial distribution functions (SDFs) and interaction energy analyses using molecular dynamics to probe an understanding of the solubility of cellulose in 1-butyl-3-methylimidazolium chloride and insolubility of cellulose in water. The study suggested that the insolubility of cellulose in water is mainly due to its reduction in solvent entropy. Similarly, Brady *et al.*,[77] studied the insolubility of cellulose in water using PMF calculations. Glucose, cellobiose, cellotriose and cellotetraose were all chosen as models for cellulose. The study showed that as the chain length of cellulose increases, so does its PMF well-depth. The well-depth here refers to the first significant local minimum found on a PMF curve. Subsequently, the larger the well-depth is, the more work is required to separate the cellulose dimers far apart in water. This suggests, for example, that glucose should separate out or dissolve better in water than cellobiose does; which is confirmed experimentally by Gray *et al.*[78] For further reading on biomass dissolution using ionic liquids, see the review paper by Mohd *et al.*[27]

While many research papers have reported on the dissolution of biomass in ILs, they have not looked at why specific dissolution results are achieved. Fortunately, research has been done on investigating the biomass dissolution mechanism in ILs.

1.4.2 Biomass Dissolution Mechanism in Ionic Liquids: the Role of the Ions

Ionic liquids' dissolution of chitin, chitosan, cellulose, etc., is mainly due to the anion's interaction with the hydroxyl, acetyl, amino and other hydrogen bond donating substituents on the pyranose rings of these biopolymers.[75] This solvent-biopolymer interaction disrupts the various intermolecular- and intramolecular interactions of the biopolymer, causing its stabilising interactions to occur less frequently and forcing the strands to break apart. Rabideau *et al.*[79] investigated the solvation of cellulose in 15 different imidazolium-based ionic liquids (3 anions and 5 cations) using molecular dynamics simulations. Their results showed that the interaction between the anionic components and cellulose molecules are greater than the interaction between the cationic components and cellulose molecules; supporting the information that the anion is the component that dominantly interacts with the biopolymer.

The dominant interaction from the anion is most likely explained by the hydrogen bond basicity, as the stronger it is, the greater the interaction with the biopolymer and the better it will dissolve the biopolymer.[80] Cations have, however, also been found to influence the solubility of biopolymers. In a recent study, 13 ionic liquids were considered, more specifically, how well they can dissolve microcrystalline cellulose with each IL having a common anion (acetate) but a different cation. The study showed that as the cation of the IL changes, the solubility of cellulose also changes. For example, at 90 °C, the solubility of cellulose was considered when placed in 1-butyl-3-methylimidazolium acetate $[C_4C_1im][CH_3COO]$, N-butyl-N-methylpiperidium acetate $[C_4C_1pip][CH_3COO]$ and N-butyl-N-methylpyrrolidinium acetate $[C_4C_1pyr][CH_3COO]$, respectively. In $[C_4C_1im][CH_3COO]$, the solubility was as high as 49 g of cellulose per mole of ionic liquid used, while the solubility was negligibly small in the other two ILs. All three ILs have a common cationic side chain and anion, but as

can be seen in the above, the solubility results are vastly different. The promising results produced for $[C_4C_1im][CH_3COO]$ was explained by the cation having three acidic protons on the imidazolium ring, where these hydrogen atoms are a part of more favourable hydrogen bonds to the oxygen atoms from cellulose.[80]

Prior to the above, Zhang *et al.*[81] proposed a mechanism for the dissolution of cellulose in ILs; where a synergistic effect of anions and cations is mainly responsible for the dissolution. This was later supported via IR, NMR, dissolution and free energy simulation studies.[76,82–94] Figure 1.8 shows the dissolution mechanism of cellulose in IL and the role of the cation and anion as highlighted by Zhang *et al.*[81]

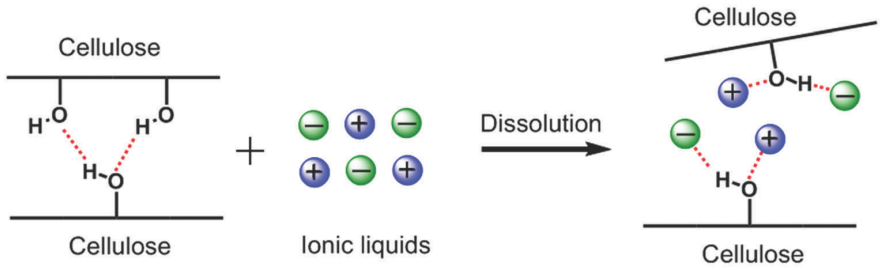


Figure 1.8: The proposed dissolution mechanism of cellulose in ionic liquid.[81]

By considering Figure 1.9, one can also theorise that electrostatic repulsion contributes to the dissolution of cellulose in IL. When the IL gets close enough to the hydrogen bond donating substituents on cellulose, the anion acts as a hydrogen bond acceptor, while the cation interacts as a hydrogen bond donor. As the cellulose strands start separating, cations get close to each other (as do the anions), causing a repulsion of charge, which promotes a further separation of the strands.

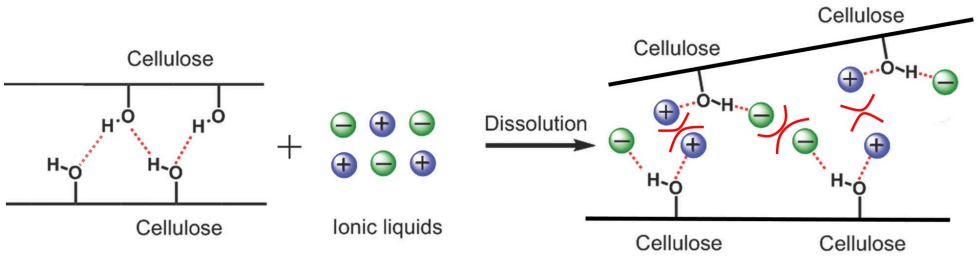


Figure 1.9: An illustration highlighting the repulsive nature of anions and cations during cellulose dissolution.

Gross et al.[76] suggested in their 2011 study that the interaction between 1-butyl-3-methylimidazolium and glucan chains is not only significant enough to be considered important but also showed that the contact distance between the cation and glucose is at its highest from the axial position (above and below the glucose) than equatorial. This suggests a possible interruption of the stacking interactions between glucose molecules, by the cation.

As such, when ionic liquids are introduced to the biopolymer, both the cation and anion serves a vital role in breaking apart the hydrogen bonding network. This polymeric hydrogen bonding network is

broken mainly by the formation of intermolecular hydrogen bonds between the biopolymer and the ionic liquid components. An increase in these observable hydrogen bonds should result in a weaker polymeric hydrogen bonding network, which will allow each polymeric strand to move apart easier and for the whole biomass to dissolve. When this is done, the polymeric strands are still involved in hydrogen bonding; not to each other anymore, but to the ionic liquids. These ionic liquids are mainly removable and separable through filtration and centrifugation.

1.5 A Better Biomass Dissolution with the Addition of Co-Solvents

In 2011, Rinaldi showed that by adding an additional aprotic organic molecular solvent, known henceforth as a co-solvent, to ionic liquids, the dissolution efficiency of cellulose increases.[95] The dissolution efficiency looks at the solubility of a biopolymer per gram of ionic liquid used to dissolve it. The addition of a co-solvent to an ionic liquid lowers the electrostatic interaction between the cation and anion, which not only results in a decrease in the solution's viscosity but should also affect the way the ionic components interact with each other and with the biopolymer. The mixture of a co-solvent and an ionic liquid is known as an organic electrolyte solution (OES).[95]

In 2015, Bioni *et al.*[96] studied the effect of co-solvent addition (DMSO and DMAc) to 1-allyl-3-methylimidazolium chloride. The study showed that as the addition of co-solvent to pure IL increases, the viscosity of the mixture decreases. This viscosity decrease allows the ionic liquid to diffuse quicker, equilibrate faster and increases interaction opportunities to solutes. The study also showed IL conductivity undergoing drastic changes when more co-solvent is introduced, and that at an ionic liquid molar ratio of 0.8 (8:2 IL:DMSO), microcrystalline cellulose has the highest dissolution capability.

In 2016, Gale *et al.*[97] studied the effect of "green"/bio-based co-solvent addition to 1-ethyl-3-methylimidazolium acetate on micro-crystalline cellulose dissolution. For most solvent mixtures, a maximum dissolution efficiency of microcrystalline cellulose was achieved at an ionic liquid molar fraction of 0.2. Additionally, co-solvents have been suggested to accelerate the mass transport of solutes[98], but beyond this, not much is known on the effect of co-solvent addition.

1.6 Research Focus

The computational aspect of this project is to use quantum mechanics (the Gaussian application suite) and classical molecular dynamics (the AMBER MD code) to construct, optimise and parameterise two ionic liquids and three co-solvents to model the dissolution of chitin. While a plethora of studies has been done on the interaction and dissolution of cellulose in ionic liquids (see Chapter 1.4.1), similar studies done on chitin has been minimal. This is the justification of selecting chitin as the biomass in this work. To limit computational expense, a model for chitin will be used in this project, which is a single methylated and acetylated chitobiose monomer, that is, 4-methyl-β-D-N-acetylglucosamine-(1$\rightarrow$4')-1'-methyl-β-D-N'-acetylglucosamine (abbreviated henceforth as $(GlcNAc)_2Me_2$). Figure 1.10 shows the structure of $(GlcNAc)_2Me_2$.

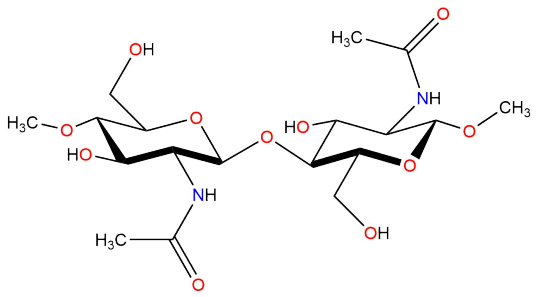


Figure 1.10: The structure of 4-methyl-β-D-N-acetylglucosamine-(1→4')-1'-methyl-β-D-N'-acetylglucosamine ($(GlcNAc)_2Me_2$).

The dissolution of chitin is modelled through the separation of two $(GlcNAc)_2Me_2$ molecules in the various solvent systems by calculating the PMF of separation using umbrella sampling. The PMF work is compared to the radial distribution functions, hydrogen bond analyses and interaction energy results from the solvation of a single $(GlcNAc)_2Me_2$ molecule in an attempt to further support the separation results.

The interaction energy results from the solvation of a single $(GlcNAc)_2Me_2$ molecule is compared to the experimental swelling of chitin, as a good correlation further validates the parameterisation of the solvent molecules. A force field validation is necessary for reliable and accurate MD simulations. The assumption made with the experimental work is that by studying the swelling capacity of chitin, the amount of solvent that has been absorbed can be determined which acts as a representative of how well the solvent interacts with chitin, as the more favourable the interaction is, the larger the absorption and swelling capacity should be.

1.7 Aims and Objectives

1.7.1 Aims of Computational Work

1. Study the solute-solvent interaction and solvation structure of a $(GlcNAc)_2Me_2$ molecule in selected ionic liquid and organic electrolyte solutions.
2. Simulate the separation of a $(GlcNAc)_2Me_2$ dimer in ionic liquid and organic electrolyte solutions.

1.7.2 Aims of Experimental Work

1. Study the (solute-solvent) interaction between chitin in ionic liquid and organic electrolyte solutions.

1.7.3 Objectives of Computational Work

- Validate the force field (GAFF1 or GAFF2) chosen for solvent parameterisation by comparing the computational density, heat of vaporisation or self-diffusion coefficient values of each solvent to that from literature.
- Perform classical molecular dynamics simulations of a $(GlcNAc)_2Me_2$ molecule solvated in ionic liquid and organic electrolyte solutions to obtain the solute-solvent interaction energy, which will be compared to the experimental results, (to further validate the parameterisation of the solvent) and the solvation structure.
- Use umbrella sampling to determine the potential of mean force (PMF) of separating two $(GlcNAc)_2Me_2$ molecules in ionic liquid and organic electrolyte solutions.

- Calculate radial distribution functions (RDFs) and perform hydrogen bond analyses on a $(GlcNAc)_2Me_2$ molecule solvated in ionic liquid and organic electrolyte solutions for supporting the results produced from the separation study.

1.7.4 Objectives of Experimental Work

- Study the swelling capacity of chitin in ionic liquid and organic electrolyte solutions as to quantify the interaction between chitin and solvent.

1.8 Choice of Ionic Liquids, Co-solvents and Organic Electrolyte Solution Ratios

As a result of the popularity of the 1-butyl-3-methylimidazolium cation in ionic liquid research, it was selected as the cation of choice for this project.[66,71,75–76,80] Acetate was selected as the anion as the ionic liquid 1-butyl-3-methylimidazolium acetate (see Figure 1.11 for the IL's structure) has been shown to produce good solubility results for cellulose,[71,76,80] is an ionic liquid regularly used for cellulose dissolution, as well as has literature density and heat of vaporisation results available, which is vital for force field validation. 1-Butyl-3-methylimidazolium methyl sulfate is an additional ionic liquid chosen in this research project due to the availability of its experimental physical properties (density and self-diffusion coeffients). Figure 1.12 shows the structure of 1-butyl-3-methylimidazolium methyl sulfate.

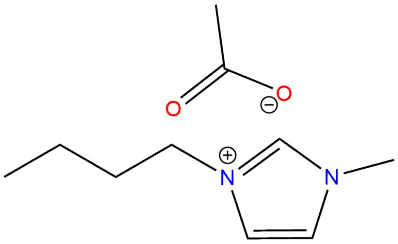


Figure 1.11: The structure of 1-butyl-3-methylimidazolium acetate.

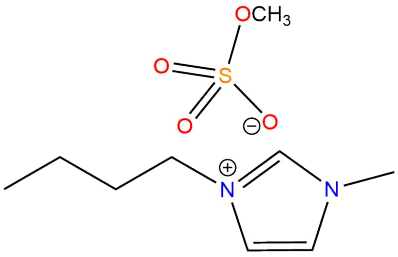


Figure 1.12: The structure of 1-butyl-3-methylimidazolium methyl sulfate.

The co-solvents γ-valerolactone, propylene carbonate and dimethyl carbonate were selected due to them being bio-based, being structurally related (but chemically different), and having experimental physical properties (density and heat of vaporisation) available. The structures of these co-solvents are shown in Figure 1.13.

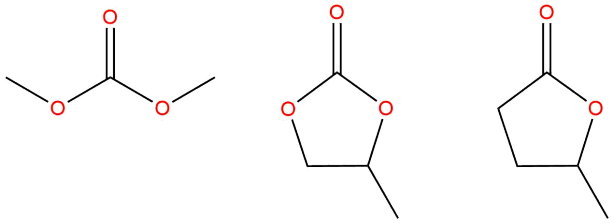


Figure 1.13: The structures of dimethyl carbonate (left), propylene carbonate (middle) and γ-valerolactone (right).

The ratio of 1-butyl-3-methylimidazolium acetate:co-solvent OESs was selected as 2:8 and 8:2 as Bioni and Gale have indicated that these OES ratios can produce favourable dissolution results for microcrystalline cellulose.[96–97] The dissolution of cellulose in 1-ethyl-3-methylimidazolium acetate:propylene carbonate ($[C_2C_1im][CH_3COO]$:PC) and 1-ethyl-3-methylimidazolium acetate:γ-valerolactone ($[C_4C_1im][CH_3COO]$:γV) OESs has previously been studied by Gale *et al.*[98] Their results show that the 8:2 $[C_2C_1im][CH_3COO]$:co-solvent OESs give better dissolution results than the 2:8 $[C_2C_1im][CH_3COO]$:co-solvent OESs, and that the 8:2 $[C_2C_1im][CH_3COO]$:γV OES gives better dissolution results than the 8:2 $[C_2C_1im][CH_3COO]$:PC OES.

Vitz *et al.*[98], Uto *et al.*,[99] and Pinkert *et al.*[100] investigated the dissolution of cellulose in 1-butyl-3-methylimidazolium acetate and 1-ethyl-3-methylimidazolium acetate, respectively, and found the weight percentage results for cellulose to be significantly close. Weight percentages of 12% and 8% were obtained by Vitz *et al.* for cellulose dissolution via $[C_4C_1im][CH_3COO]$ and $[C_2C_1im][CH_3COO]$, respectively; while Pinkert *et al.* obtained weight percentages of 10.7% and 7.4%, respectively. From these results, the change in alkyl chain length will not drastically affect the dissolution results, as such, it is hypothesised that $[C_4C_1im][CH_3COO]$:PC and $[C_4C_1im][CH_3COO]$:γV OESs will give similar dissolution results to the $[C_2C_1im][CH_3COO]$:PC and $[C_2C_1im][CH_3COO]$:γV OESs from above. While the above results are from cellulose dissolution, the trend in results is hypothesised not to change much for chitin. That is, if the 8:2 IL:co-solvent OESs give better dissolution results than the 2:8 IL:co-solvent OESs in cellulose, the same is expected to be true for chitin dissolution. As such, the PMF results produced in the following results will be compared to the general dissolution results from Gale *et al.*[97]

[Page intentionally left blank.]

2 Swelling Capacity of Chitin

This section contains a short overview of some experimental and observable theories, ideas and concepts explored in this project. Many other techniques for quantifying swelling and investigating dissolution exist, but these are not discussed here. See the review papers by Miller-Chou and Hulsey, as well as the book by Semenova for further reading.[101−103]

2.1 Synthesis of $[C_4C_1im][CH_3COO]$

1-Butyl-3-methylimdazolium chloride (in the amount of 30.3906 g/ 0.1740 mol) and 19.0580 g (0.1942 mol) of potassium acetate were added together with 100 ml of ethanol into a 250 ml round bottom flask with a reflux condenser attached. The mixture was kept at 43 °C for 6 hours whilst stirring. Thereafter, the mixture was transferred to centrifuge tubes and centrifuged for 5 minutes at 3500 rpm. This allowed the mixture to separate out according to their densities. The liquid was separated and transferred to a round bottom flask. The remaining solid was washed with ethanol and separated by centrifugation once more. The combined liquid phases were pooled and the solvent removed under reduced pressure (60 mbar) at 40 °C as to remove the ethanol. The resultant oil was then cooled down to -20 °C and kept at this temperature overnight as to precipitate out as much of the leftover potassium acetate and newly formed potassium chloride as possible. Subsequently, the mixture was centrifuged for 5 minutes at 3500 rpm. Once the precipitation of salt stopped, the remaining liquid was thoroughly dried under vacuum. The above synthesis was followed as outlined by Carson and Liu *et al.*[104−105]

The resultant liquid was 1-butyl-3-methylimidazolium acetate $[C_4C_1im][CH_3COO]$, with a percentage yield of 86.44%. Proton ^{1}H NMR data and carbon (^{13}C) NMR data was acquired for the IL using standard pulse sequences on a Bruker 300 MHz nuclear magnetic resonance (NMR) spectrometer. The spectroscopic analysis confirmed the identity of the ionic liquid.

2.1.1 NMR Data Analysis

The results for the experimental NMR analysis is as follows

^{1}H-NMR (300 MHz, $CDCl_3$): δ 10.80 ppm (s, 1H, H4), 7.30 ppm (s, 1H, H5), 7.20 ppm (s, 1H, H6), 4.15 ppm (t, 2H, H1), 3.90 ppm (s, 3H, H8), 1.80 ppm (s, 3H, H9), 1.75 ppm (m, 2H, H2), 1.25 ppm (m, 2H, H3), 0.85 ppm (t, 3H, H7).
^{13}C-NMR (300 MHz, $CDCl_3$): δ 178 ppm (C10), 140 ppm (C4), 123 ppm (C6), 121 ppm (C5), 57 ppm (C1), 49 ppm (C8), 25 ppm (C2), 19 ppm (C9), 18 ppm (C3), 13 ppm (C7).

Most of the peaks observed in the NMR spectra are easy to understand, however, it is noteworthy that the H5 and H6 hydrogen atoms (see Figure 2.1 for a labelled structure of $[C_4C_1im][CH_3COO]$) do not couple to each other in the ^{1}H-NMR spectrum.

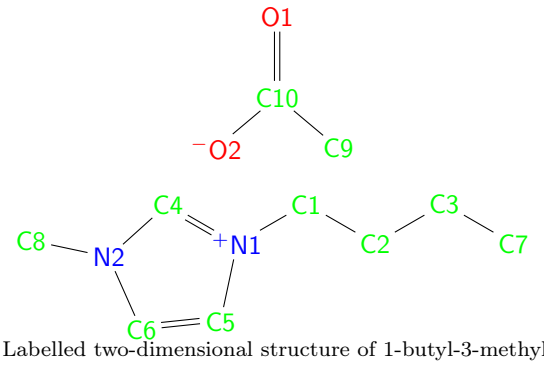


Figure 2.1: Labelled two-dimensional structure of 1-butyl-3-methylimidazolium.

Moreover, additional peaks are observed in the spectra which do not belong to the IL; it being for solvent contamination. The identification of all NMR peaks can be found in figures 2.2 and 2.3.

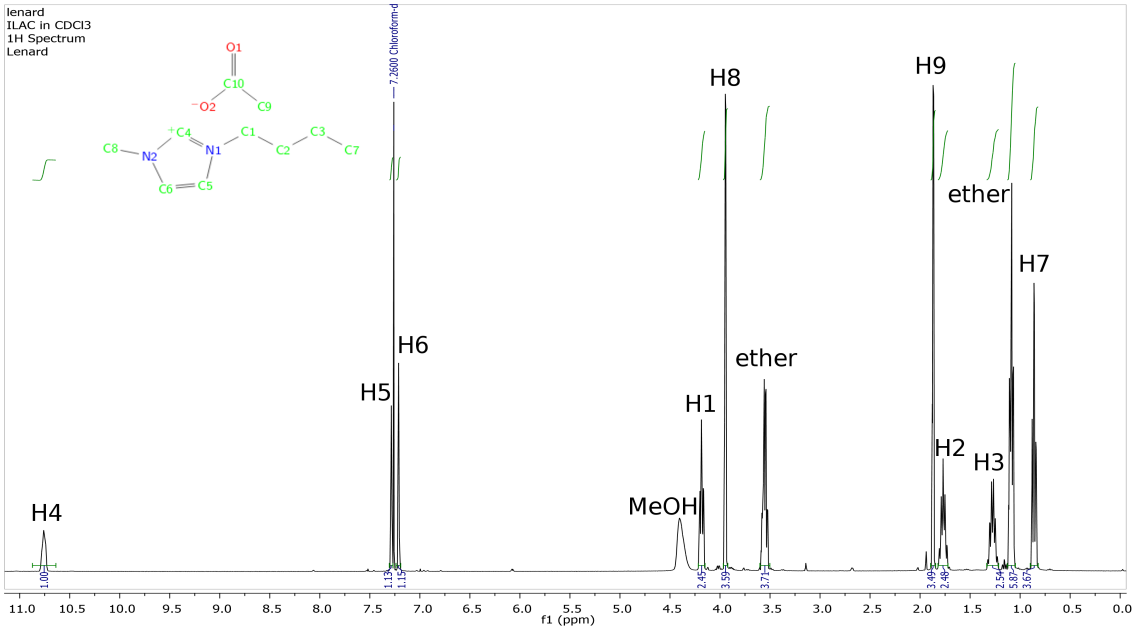


Figure 2.2: ^{1}H-NMR spectrum of 1-butyl-3-methylimidazolium acetate.

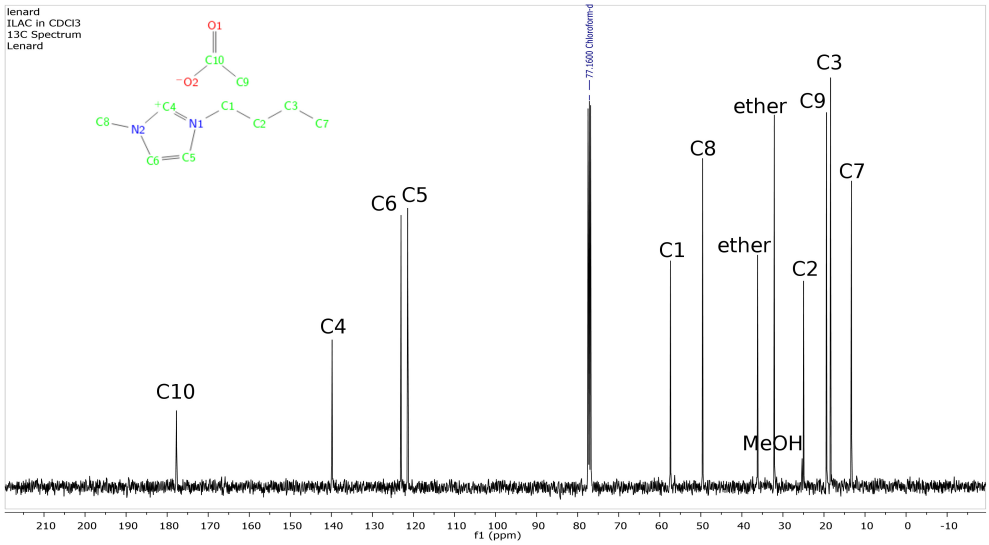


Figure 2.3: ^{13}C-NMR spectrum of 1-butyl-3-methylimidazolium acetate.

2.2 Dissolution and Swelling

Dried (and undissolved) biopolymers that are not solvated have strong hydrogen bonding networks and interlayer dispersion forces to overcome, that keeps all the strands locked together and difficult to break apart.[7] A complete biopolymer dissolution results from the total disruption of all biopolymer-biopolymer hydrogen bonds, which is sought after in the biofuel production process, as the resulting hydrolysis step to create simpler sugars, would be easier and more successful.

Swelling is commonly defined as the increase of volume of a material due to the absorption of a solvent. When a solvent interacts with a biopolymer in such a way that it disrupts some of the biopolymeric hydrogen bonds, replaces them with biopolymer-solvent intermolecular hydrogen bonds and is "locked" into an interaction with the biopolymer, the solvent is said to be absorbed. The more solvent that is absorbed, the more swollen the biopolymer is.[106]

For biopolymers and cross-linked polymers, there does exist a relationship between its dissolution and swelling. In a 1968 paper by Huglin and Pass, the mathematical relationship between the swelling of a polymer and its solubility was identified.[107] This relationship is viewable in terms of Equation 2.1

$$Q = Q_{\mathrm{MAX}} e^{-\mathcal{K} V_M (\delta_S - \delta_P)^2} \tag{2.1}$$

where $Q = \dfrac{S}{\rho}$, S is the swelling capacity of the polymer, ρ is the density of the solvent, Q_{MAX} is the maximum swelling capacity, V_M is the molar volume of the mixture, $\mathcal{K}$ is some factor, and δ_P and δ_S are solubility parameters.[107] Even though a relationship exists, one can only consider swelling as the early stage of trying to dissolve a polymer. If the polymer-polymer forces are stronger than the solvent-polymer forces, the limit of swelling is eventually reached. Conversely, when a structured network of solvent and polymer exists, it could be considered a gel or colloid.

2.2.1 Calculating the Swelling Capacity of Chitin[106]

The swelling capacity of chitin quantifies how swollen the biopolymer is. The larger the swelling capacity is, the more swollen the polymer is. The swelling capacity of chitin was determined by first adding 0.2 g of chitin into a glass vial. This mass of chitin is known as the initial mass (m_i). To each glass vial was then added 5 ml of pure co-solvent (dimethyl carbonate, propylene carbonate or γ-valerolactone), pure 1-butyl-3-methylimidazolium acetate, 2:8 1-butyl-3-methylimidazolium acetate:co-solvent and 8:2 1-butyl-3-methylimidazolium acetate:co-solvent, respectively. All 10 glass vials were then placed in a carousel and heated simultaneously at 80 °C for two hours. Thereafter, the content of each glass vial was transferred into 15 ml centrifuge tubes and centrifuged for 4 minutes at 3500 rpm. The solvent, in each centrifuge tube, was mainly found above the chitin biomass and could be removed with a pipette. Following this, the chitin was subsequently re-weighed, with this new mass being labelled as the mass of chitin absorbent (m_f). The final mass consists of the initial mass of the chitin, as well as the amount of solvent absorbed into chitin. To determine how much solvent has been absorbed as the swelling capacity of chitin, one can use the $\%S_m = \dfrac{m_f - m_i}{m_i} \times 100 = \dfrac{m_s}{m_i} \times 100$ equation, where m_s is the mass of the absorbed solvent. See Figure 2.4 for an illustration of the swelling capacity determination procedure.

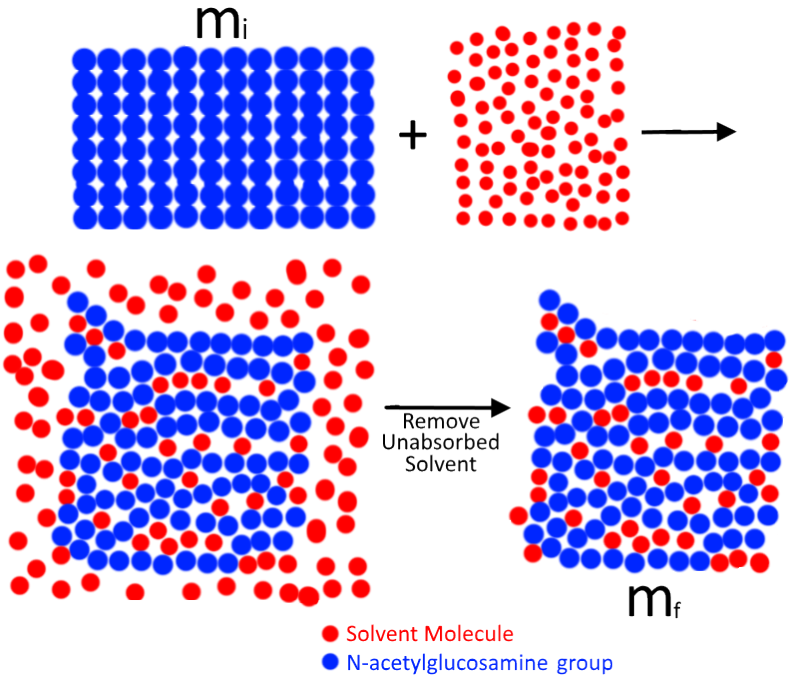


Figure 2.4: An illustration of the swelling capacity determination procedure.

The swelling capacity equation thus considers a ratio of masses. If a ratio of the number of moles is desired then the $\dfrac{n_f - n_i}{n_i} \times 100$ equation should be used. Here, n_f is the number of moles for the chitin absorbent, and n_i is the initial number of moles for dry chitin. The molar swelling capacity can be transformed further to take into consideration masses. See Equation 2.2.

$$\%S_n = \frac{n_f - n_i}{n_i} \times 100 = \frac{n_s}{n_i} \times 100 = \frac{\frac{m_s}{M_s}}{\frac{m_i}{M_i}} \times 100 = \frac{m_s M_i}{m_i M_s} = \%S_m \frac{M_i}{M_s} = \%S_{n/M_i} M_i \tag{2.2}$$

where n_s is the number of moles for the absorbed solvent, M_s is the molar mass of the absorbed solvent and M_i is the molar mass for dry chitin. Since the source of chitin used is the same, M_i is considered as a constant in this study and will not be considered henceforth.

2.2.1.1 Swelling Capacity Results

In Table 2.1 can be found the swelling capacity of chitin in various solvent systems. The higher the percentage is, the more solvent that has been absorbed into the polymer.

Table 2.1: The swelling capacity of chitin in various solvent systems as heated to 80 °C

Absorbed Solvent	Swelling Capacity
Dimethyl carbonate (DC)	212%
γ-Valerolactone (γV)	229%
Propylene carbonate (PC)	237%
1-butyl-3-methylimidazolium acetate (IL)	500%
8:2 IL:DC	398%
8:2 IL:γV	458%
8:2 IL:PC	467%
2:8 IL:DC	261%
2:8 IL:PC	357%
2:8 IL:γV	358%

Table 2.1 shows that a greater mass of pure 1-butyl-3-methylimidazolium acetate is absorbed into chitin than compared to any of the other solvent. For the molar swelling capacity, the following results are observed as per Table 2.2

Table 2.2: The molar swelling capacity of chitin in various solvent systems as heated to 80 °C

Absorbed Solvent	Molar Swelling Capacity ($/M_i$)
γ-Valerolactone (γV)	2.29
Propylene carbonate (PC)	2.32
Dimethyl carbonate (DC)	2.35
1-butyl-3-methylimidazolium acetate (IL)	2.52
8:2 IL:DC	2.25
8:2 IL:γV	2.56
8:2 IL:PC	2.61
2:8 IL:DC	2.33
2:8 IL:PC	2.98
2:8 IL:γV	2.99

The trend in molar swelling capacity is:

2:8 IL:γV > 2:8 IL:PC > 8:2 IL:PC > 8:2 IL:γV > IL > DC > 2:8 IL:DC > PC > γV > 8:2 IL:DC

This trend shows, as an example, that for every mole of chitin used, there would be more 2:8 IL:γV molecules absorbed into the chitin than 2:8 IL:PC molecules. The 8:2 IL:DC OES on the other hand, would have the least number of molecules absorbed by chitin. The hypothesis made now is the solvent that has the most molecules absorbed into chitin (in this case 2:8 IL:γV) interacts the strongest with the biopolymer.

[Page intentionally left blank.]

3 Quantum Mechanics

In computational chemistry, the two main techniques used for calculating the energy necessary to find the interaction strength between two or more systems are quantum mechanics (QM) and molecular mechanics (MM). Quantum mechanics can explain the behaviour of matter and its interactions, with the energy of the matter being on the scale of atoms and subatomic particles through describing the matter as 'wave functions' that characterises all its physical properties.

The starting point of describing what quantum mechanics is, is by considering the Schrödinger equation. The full time-independent form of the equation is (Equation 3.1)

$$\left\{\frac{-\hbar}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V(r,t)\right\}\Psi(r,t)=i\hbar\frac{\partial\Psi(r,t)}{\partial t} \tag{3.1}$$

where Equation 3.1 refers to a single particle with mass m moving through space ($\overrightarrow{r}=\langle x,y,z\rangle$) and time ($t$) under an external field $V(r,t)$. Here, $\hbar$ is the reduced Planck's constant, r is $\sqrt{x^2+y^2+z^2}$, i=$\sqrt{-1}$ and $\Psi(r,t)$ are the time and position dependent wave functions of the particle. When the external field becomes independent of time, the wave function can be written as a product of a spatial and a time part. This gives rise to the time-independent Schrödinger equation (Equation 3.2).

$$\left\{\frac{\hbar}{2m}\nabla^2+V(r)\right\}\Psi(r)=E\Psi(r) \tag{3.2}$$

where ∇^2=$\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$ and E is the energy of the single particle. Typically, the left side of Equation 3.2 gets abbreviated as $\mathcal{H}=\left\{\frac{\hbar}{2m}\nabla^2+V(r)\right\}$ and is known as the Hamiltonian operator. The Schrödinger equation is solved when the wave function and energy of the system is known.

3.1 Operators[108–110]

In quantum mechanics, operators are important for determining the expectation value (or average value) of the energy and position of a system. In QM, the Hamiltonian operator determines the expectation value of the energy of a system. See Equation 3.3

$$E=\frac{\int\Psi^*\mathcal{H}\Psi d\tau}{\int\Psi^*\Psi d\tau} \tag{3.3}$$

where Ψ^* is the complex conjugate of the wave function. For a system with a single electron, a single nucleus and Z protons, the potential energy operator is (Equation 3.4)

$$V(r)=\frac{-Ze^2}{4\pi\varepsilon_0 r} \tag{3.4}$$

Similarly, the expectation value of the linear momentum of a system can be obtained through (Equation 3.5)

$$p_x = -\frac{\int \Psi^* i\hbar \frac{\partial}{\partial x} \Psi d\tau}{\int \Psi^* \Psi d\tau} \tag{3.5}$$

where $-i\hbar\frac{\partial}{\partial x}$ is the linear momentum operator.

3.2 Approximations to Schrödinger's Equation[108–110]

The exact solution to the Schrödinger equation can only be found for a select few problems, such as for a particle in a box, the harmonic oscillator, a particle on a ring, a particle on a sphere and for the hydrogen atom. Any system larger than these problems cannot be solved exactly, including a helium atom with only three particles (two electrons and one nucleus). It is for this reason why approximations have been postulated for solving the Schrödinger equation.

In a many-body problem (any system with multi-electrons), the spin of the electrons must be accounted for as well as the distribution of electron density in space. The Schrödinger equation can incorporate spin and spatial properties, where one-electron wave functions can be re-written in terms of the product of the spin- and spatial function of the electrons. The spin function is constituted of spin orbitals, more specifically, an electron can have a spin of either α or β. Depending on the quantum number of an electron, the spin functions can either have a value of 0 or 1. Thus $\alpha\left(+\frac{1}{2}\right)$=1, $\alpha\left(-\frac{1}{2}\right)$=0, $\beta\left(+\frac{1}{2}\right)$=0 and $\beta\left(-\frac{1}{2}\right)$=1.

Although two electrons with a paired spin can occupy a single spatial orbital, electrons are indistinguishable and exchanging the two electrons in a given orbital keeps the electron density the same. By the Born interpretation, the square of the wave functions should also stay the same, implying that the wave function will remain unchanged or will change in sign. This gives rise to the antisymmetry principle.

One specific approximation for the Schrödinger equation is the Born-Oppenheimer approximation which states that "The electronic wave function depends only on the position of the nuclei of a system and not its momenta."[108]. By this, the total wave function of a molecule is (Equation 3.6)

$$\Psi_{tot}(\text{nuclei}, \text{electrons}) = \Psi(\text{electrons})\Psi(\text{nuclei}) \tag{3.6}$$

From Equation 3.6, the total energy of a molecule becomes E_{tot}=E(electrons)+E(nuclei). According to the Born-Oppenheimer approximation, the nuclei are fixed and the Schrödinger equation is solved for the electrons alone in the field of nuclei.

3.3 Spin Functions, Spatial Functions and Slater Determinants[108−110]

As mentioned prior, wave functions can be written in terms of spatial orbitals, such as 1s(1), 1s(2) and 2s(1). Here, 1s(x) states that electron x can be located in the 1s orbital, while 2s(x) states that electron x can be located in the 2s orbital. Writing 1s(x)α(x) thus means that electron x from the 1s orbital has a spin of α, which is different to 1s(x)β(x) in which electron x is still in the 1s orbital, but it now has a spin of β.

As an example, consider a helium atom which too can be written in terms of spin- and spatial functions. Note that the electrons from the 1s orbital can be excited to the 2s orbital. The spatial functions of a helium atom are (equations 3.7-3.9)

$$\text{1s(1)1s(2) (symmetrical)} \tag{3.7}$$

$$\left(\sqrt{\frac{1}{2}}\right)\text{[1s(1)2s(2)+1s(2)2s(1)] (symmetrical)} \tag{3.8}$$

$$\left(\sqrt{\frac{1}{2}}\right)\text{[1s(1)2s(2)-1s(2)2s(1)] (antisymmetrical)} \tag{3.9}$$

Symmetrical in equations 3.7-3.8 implies that if one were to exchange two electrons in these equations, the original equation will still be obtained. For example (Equation 3.10)

$$1s(1)1s(2) \xrightarrow[\text{electrons}]{\text{swap}} 1s(2)1s(1) = 1s(1)1s(2) \tag{3.10}$$

Antisymmetrical in Equation 3.9 implies that after exchanging two electrons, a change in sign for the equation will be observed. See Equation 3.11

$$\left(\sqrt{\frac{1}{2}}\right)\text{[1s(1)2s(2)-1s(2)2s(1)]} \xrightarrow[\text{electrons}]{\text{swap}} \left(\sqrt{\frac{1}{2}}\right)\text{[1s(2)2s(1)-1s(1)2s(2)]} = \tag{3.11}$$
$$\left(\sqrt{\frac{1}{2}}\right)\text{[-1s(1)2s(2)+1s(2)2s(1)]} = -\left(\sqrt{\frac{1}{2}}\right)\text{[1s(1)2s(2)-1s(2)2s(1)]}$$

Note that in equations 3.8-3.9, $\sqrt{\frac{1}{2}}$ purely serves as a normalisation constant. The spin functions of a helium atom are as per equations 3.12-3.15.

$$\alpha(1)\alpha(2) \text{ (symmetrical)} \tag{3.12}$$

$$\beta(1)\beta(2) \text{ (symmetrical)} \tag{3.13}$$

$$\left(\sqrt{\frac{1}{2}}\right)[\alpha(1)\beta(2) + \alpha(2)\beta(1)] \text{ (symmetrical)} \tag{3.14}$$

$$\left(\sqrt{\frac{1}{2}}\right)[\alpha(1)\beta(2) - \alpha(2)\beta(1)] \text{ (antisymmetrical)} \tag{3.15}$$

There are thus four ways for each of the equations 3.7-3.9 to be combined with the spin wave functions in equations 3.12-3.15. That is a total of 12 combinations, 5 which are antisymmetric. Pauli's exclusion principle, however, states that it is impossible for two electrons to occupy the same orbital with the same spin. In such a case, all the symmetrical combinations must be ignored as it does not obey this principle. For example, the combination 1s(1)1s(2)α(1) α(2) has electrons one and two in the 1s orbitals with the same spin. This is forbidden.[108] One of the linearly combined antisymmetrical wave functions of a helium atom is (Equation 3.16)

$$[\alpha(1)\alpha(2)]\left(\sqrt{\frac{1}{2}}\right)[1\mathrm{s}(1)2\mathrm{s}(2)+1\mathrm{s}(2)2\mathrm{s}(1)] = \tag{3.16}$$

$$\frac{1}{\sqrt{2}}[1s(1)\alpha(1)2s(2)\alpha(2) - 1s(2)\alpha(2)2s(1)\alpha(1)] =$$

$$\frac{1}{\sqrt{2}}\begin{vmatrix} 1\mathrm{s}(1)\alpha(1) & 1\mathrm{s}(2)\alpha(2) \\ 2\mathrm{s}(1)\alpha(1) & 2\mathrm{s}(2)\alpha(2) \end{vmatrix}$$

In Equation 3.16, the antisymmetrical linear combination of spin- and spatial functions produces the determinant of a 2×2 matrix. Wave functions that can be written as a determinant take the form (of Equation 3.17)

$$\Psi(1,2,...,N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1(1) & \chi_1(2) & \cdots & \chi_1(N) \\ \chi_2(1) & \chi_2(2) & \cdots & \chi_2 \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N(1) & \chi_N(2) & \cdots & \chi_N(N) \end{vmatrix} \tag{3.17}$$

where $N!$ is the factorial of N, $\chi_1(1)$=1s(1)β(1), $\chi_2(1)$=1s(1)β(1) and so forth. Equation 3.17 is known as Slater's determinant. Since the MP2 (Møller-Plesset 2$^{\mathrm{nd}}$-order perturbation)- and HF (Hartree-Fock) wave function methods are the two considered in this project, these will be the only two looked at for its approximations made relative to the electronic Schrödinger equation.

3.4 Hartree-Fock Theory[108−110]

According to the variation theorem, the energy calculated from a wave function as per an approximation is always lower than the energy calculated from the true wave function. Consequently, this means that the better the wave function is chosen, the lower the energy is and the closer it will be to the true energy. At a minimum, the first derivative of energy (∂E) will be zero. The Hartree-Fock equations impose this condition on the expression for energy, granted the molecular orbitals remain orthonormal (orthogonal and normalised). This is done in Hartree-Fock theory by using a method of Lagrange multipliers, in which the derivative of the energy (which will resultantly be minimised) is added to the derivatives of the constraint(s) multiplied by a Lagrange multiplier. For the Hartree-Fock equations, the Lagrange multipliers are written as -2ε_{ij} to reflect that they are related to MO energies. Equation 3.18 has to subsequently be solved.

$$\partial E - 2\partial \sum_i \sum_j \varepsilon_{ij} S_{ij} = 0 \tag{3.18}$$

where $S_{ij} = \int \chi_i \chi_j d\tau$. Using Equation 3.18, the core Hamiltonian operator $\mathcal{H}^{\text{core}}(1)$ as per Equation 3.19 is obtained as

$$\mathcal{H}^{\text{core}}(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \tag{3.19}$$

If there were no interelectronic interactions present, Equation 3.19 would be the only operator present. If Coulombic interactions are considered, the Coulomb operator $\mathcal{J}_j(1)$ is used. See Equation 3.20.

$$\mathcal{J}_j(1) = \int \chi_j(2) \frac{1}{r_{12}} \chi_j(2) d\tau_2 \tag{3.20}$$

Similarly, the exchange operator $\mathcal{H}_j(1)$ can be represented according to Equation 3.21 as

$$\mathcal{H}_j(1)\chi_i(1) = \left[\int \chi_j(2) \frac{1}{r_{12}} \chi_i(2) d\tau_2 \right] \chi_j(1) \tag{3.21}$$

Using equations 3.19-3.21, the following form is produced (Equation 3.22)

$$\mathcal{H}^{\text{core}}(1)\chi_i(1) + \sum_{j \neq i}^{N} \mathcal{J}_j(1)\chi_i(1) - \sum_{j \neq i}^{N} \mathcal{H}_j(1)\chi_i(1) = \sum_j \varepsilon_{ij}\chi_j(1) \tag{3.22}$$

Since $\Big[\mathcal{J}_j(1) - \mathcal{H}_1(1)\Big]\chi_i(1) = 0$, Equation 3.22 is altered into Equation 3.23 as

$$\mathcal{H}^{\text{core}}(1)\chi_i(1) + \sum_{j \neq i}^{N} \{\mathcal{J}_j(1) - H_j(1)\}\chi_i(1) = \sum_{j=1}^{N} \varepsilon_{ij}\chi_j(1) \tag{3.23}$$

Equation 3.23 simplified is $\mathcal{F}_i\chi_i = \sum_j \varepsilon_{ij}\chi_i$, in which $\mathcal{F}_i$ is known as the Fock operator and takes the form of (Equation 3.24)

$$\mathcal{F}_j(1) = \mathcal{H}^{\text{core}}(1) + \sum_{j \neq i}^{N} \{\mathcal{J}_j(1) - H_j(1)\} \tag{3.24}$$

From Equation 3.24, the Hartree-Fock equations take up the form of $\mathcal{F}_i\chi_i = \varepsilon_{ij}\chi_i$. A direct solution of Hartree-Fock equations is not practical for molecules; instead, each spin orbital from the molecule is written as a linear combination of single electron orbitals. See Equation 3.25

$$\phi_i = \sum_{v=1}^{K} c_{vi}\psi_v \tag{3.25}$$

where v is a basis function. The coefficients c_{vi} in Equation 3.25 are chosen according to those in which the energy (E) is at a minimum, $\dfrac{\partial E}{\partial c_{vi}} = 0$.

3.5 Møller-Plesset Perturbation Theory[108−110]

In perturbation theory, it is assumed that the perturbed Hamiltonian operator (for a one particle system) $\mathcal{H}=\frac{\hbar}{2m}\nabla^2+V(r)$ can be written in terms of an unperturbed (Hamiltonian) operator $\mathcal{H}_0$ (which the solution is known for) and a small perturbation $\mathcal{V}$ as $\mathcal{H} = \mathcal{H}, + \lambda\mathcal{V}$, where λ is an arbitrary real number.

In general, the theorem further approximates that the wave function and energy of a system can be written as $\psi = \lim\limits_{N\to\infty}\sum\limits_{i=0}^{N}\lambda^i\psi^{(i)}$ and $E = \lim\limits_{N\to\infty}\sum\limits_{i=0}^{N}\lambda^i E^{(i)}$, respectively.

In MP theory, the zeroth-order wave function ($\Psi^{(0)}$) is an exact eigenfunction of the Fock operator. For the first-order perturbation, the wave function and energy can be written as $\Psi = \Psi^{(0)} + \lambda\Psi^{(1)}$ and $E = E^{(0)}+\lambda E^{(1)}$, respectively, in which $E^{(0)}$ is the zeroth-order energy derivative. $\Psi^{(1)}$ and $E^{(1)}$ are the first-order wave function- and -energy derivatives, respectively. Taking these approximations into consideration, the Schrödinger equation simply becomes (Equation 3.26)

$$\{\mathcal{H}_0 + \lambda\mathcal{V}\}[\Psi^{(0)} + \lambda\Psi^{(1)}] = \big(E^{(0)} + \lambda E^{(1)}\big)[\Psi^{(0)} + \lambda\Psi^{(1)}] \tag{3.26}$$

In MP2 theory, the second-order wave function $\psi = \psi^{(0)} + \lambda\psi^{(1)} + \lambda^2\psi^{(2)}$ and second-order energy $E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)}$ is used. Applying it to the Schrödinger equation, the following result as per Equation 3.27 is observed

$$\begin{aligned}&\{\mathcal{H}_0 + \lambda\mathcal{V}\}[\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\psi^{(2)}] = \\ &\big(E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)}\big)[\Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)}]\end{aligned} \tag{3.27}$$

3.6 Basis Sets[108−110]

A basis set is a set of mathematical (basis) functions used to represent molecular orbitals, similar to what is shown in Equation 3.25. Basis functions generally take the form of a Gaussian- or Slater function for molecular calculations as these are two of the simplest forms to compute with. The Gaussian basis function has a functional form of $e^{-\alpha r^2}$. The Slater basis function has a general form of $e^{-\alpha r}$ for its radial contribution. A comparison between Gaussian-type functions and Slater-type functions for the 1s orbital is given in Figure 3.1.

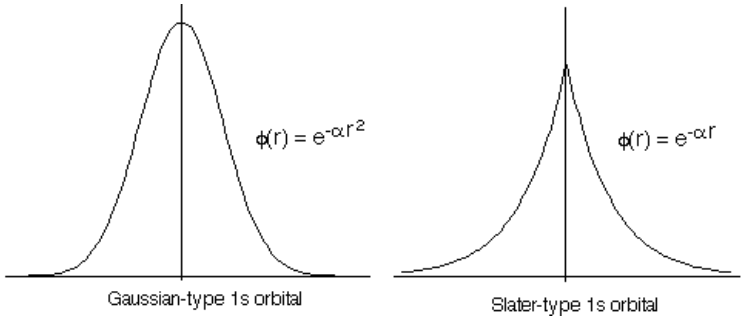


Figure 3.1: The radial dependence of Gaussian- (left) and Slater type functions for the 1s orbital.[111]

The Slater-type function would be the best choice due to its similarity with the eigenfunctions of a hydrogen atom and due to its direct physical interpretation,[108] but because of its expensive computational cost of evaluating two-electron integrals, the Gaussian function is chosen. Two-electron integrals can readily be calculated through the product of two single Gaussian functions, as it becomes a new single Gaussian function. For example $e^{-\alpha r^2}e^{-\beta r^2} = e^{-(\alpha+\beta)r^2}$ For the 1s orbital, in Figure 3.2 a direct comparison of a single Gaussian function with a Slater-type function highlights the shortcoming of the Gaussian basis function.

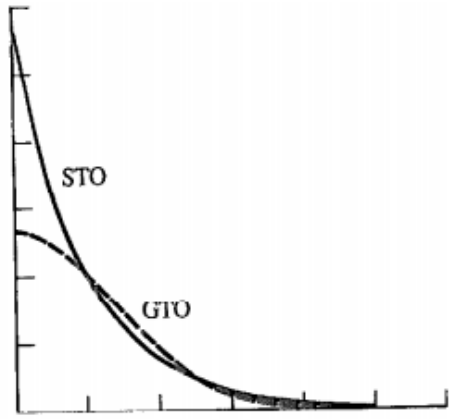


Figure 3.2: A direct comparison of the Gaussian type orbital (GTO) function to the Slater type orbital (STO) function for the 1s orbital.[108]

In Figure 3.2, compared to Slater-type functions, Gaussian-type functions do not have a cusp at the origin and starts to decay to zero faster. This is unwanted as these basis functions are ultimately linearly combined to represent the spin part of the wave function of molecules; thus the better the basis functions chosen are, the better the wave function will be approximated.

Taking the linear combination of Gaussian functions to approximate the Slater basis function will overcome the significant errors produced from the GTOs compared to STOs.[111] The general expression of the linear combination of Gaussian functions is $\phi_\mu = \sum_{i=1}^{L} d_{i\mu}\phi_i(\alpha_{i\mu})$, in which $d_{i\mu}$ is the coefficient of the single Gaussian function ϕ_i. The use of basis functions in wave function methods is

to write each molecular orbital as a linear combination of known single electron functions. As such, the HF (for example) equations can be solved by optimizing the contribution of each basis function to each molecular orbital. The same applied to other wave function methods as well.An example of the approximation of an STO using Gaussian functions is the STO-3G basis set in which 3 primitive Gaussian (G) functions are linearly combined. The larger the number of Gaussian functions used is, the better the approximation will be. These STO type basis sets are overall classified as single-ζ basis sets as each orbital is only described by one Gaussian (basis) function. A minimal basis set for a nitrogen atom will use 5 basis functions, one for each of the 1s, 2s, $2p_x$, $2p_y$ and $2p_z$ orbitals.

Extending basis sets further, consider the one that is denoted 6-31G. This basis set is known as a split-valence basis set and takes the general form of X-YZg, where X refers to the number of primitive Gaussian functions comprising each core atomic orbital basis function. Y and Z, respectively, indicate the number of primitive Gaussian functions that have been linearly combined to produce two individual single Gaussian functions. 'g' indicates that the basis set is made up of Gaussian functions.[112–114] The appearance of two numbers (3 and 1) after the hyphen in 6-31G shows that this is a double-ζ basis set.

To more accurately describe the electronic structure of molecular orbitals of a system, an additional set of functions may be added to the basis set to polarise the molecular orbitals. This addition tends to improve on how the molecular geometry is described as well as improves on the electron distribution. The addition of polarisation functions to heavy atoms for the 6-31G basis set is highlighted by the inclusion of an asterisk [*] or (d), that is, 6-31G* or 6-31G(d). When polarisation functions are added to light atoms as well, the basis set used should now be 6-31G**

To improve on the electron distribution further away from the nucleus such as on electrons in lone-pairs, anions, etc., diffuse functions may be included. The inclusion of diffuse functions will allow for a better description of electrons that are far from the nuclei and are weakly bound. Diffuse functions are represented by small Gaussian functions, in which the value of these Gaussian functions decrease the further away you move from the nucleus, allowing one to produce an electron density far away from the nucleus. The addition of this also allows one to describe certain chemical properties such as HOMO/LUMO energies, electron affinities, etc. more accurately.[112–114] The addition of diffuse functions to heavy atoms for the 6-31G* basis set is represented by the inclusion of a plus [+] between "1" and "G", that is, 6-31+G*. The 6-311+G* basis set would improve on the 6-31+G* basis by describing the valence region with three contracted basis functions, making it a triple-ζ basis set. Triple-ζ basis sets would, however, be significantly more computationally expensive.

3.7 Geometry Optimisation With Quantum Mechanics

Geometry optimisation is the process of finding the atomic arrangement of a system for which it is at its most stable. In quantum mechanics specifically, the geometry of a molecule is found by searching for/creating a potential energy surface (PES) in the vicinity of the initial structure as PESs mathematically relate molecular geometries to their corresponding single point energies.

As such, optimising the geometry of a molecule is necessary as this gives the bond angle, bond distance, energy and other structural parameters of the molecule, which when conformationally searched can confirm if the geometry produced will be the most stable for the molecule. Moreover, geometry optimisations are required as the lowest energy structure is usually the structure that best represents the equilibrium conformation of the molecule.

In quantum mechanics, the geometry optimisation can be performed at various levels of theory. A level of theory is constituted of a wave function method and a basis set. The geometry optimisation and the accuracy thereof are thus directly proportional to the level of theory chosen.

4 Molecular Mechanics and Dynamics

The time-dependent behaviour of atoms, ions and molecules as modelled from molecular mechanics (which combines classical mechanics with force fields and are comprised of potential energy functions) are simulated using molecular dynamics (MD). *Ab initio* MD, Carr-Parrinello MD and other MD methods also exist for such simulation purposes, but will not be discussed here.[118−119]

Molecular mechanics allows one to model the solid, liquid or gas phase behaviour of a system by setting up a simulation box of specific dimension that consists of anywhere between 1 atom to thousands of molecules. Molecular dynamics is dependent on the tools provided by statistical thermodynamics as this makes it possible for properties of the system such as its structure, dynamics and thermodynamics to be quantified. Using a given potential energy function as well as the coordinates of all atoms, the force on an atom can be computed. Newton's laws of motion, specifically $\frac{F_{r_i}}{m_i} = \frac{\partial^2 r_i}{\partial t^2}$, in which F_{r_i} corresponds to the Newtonian force, r_i is the position of particle i and m_i is the mass of the particle i, can be used to obtain the force on the atom and how this force affects its motion.

The force is obtained from the negative derivative of the potential energy $U(r_i^N)$ with respect to the position of the particles r_i, that is, $F_{r_i} = -\frac{\partial U(r_i^N)}{\partial r_i}$. Molecular mechanical force fields compute this force acting on each particle by continuously updating the position and velocity of each particle using a numerical approximation to the integration of Newton's laws of motion.

4.1 Force Fields[108,115−117]

A system's dependence on the coordinates of its particles and the force that is experienced by each particle is described by a force field. A force field is dependent on a potential energy function with a set of parameters built into it. In MM, the bonds in a molecule are viewed as springs. These springs can bend and stretch, but it cannot break. The springs are represented by the simple harmonic potential function of $\frac{1}{2}kx^2$. Simple harmonic forces and force fields thus replace the true quantum mechanical potential by a simplified and classical model.

A common MD code that is used for the molecular dynamics simulation of biomolecules is the AMBER (Assisted Model Building with Energy Refinement) MD code. The AMBER MD simulation engine is very efficient, being able to generate tens of nanoseconds of simulation data per day on either CPU or GPU architectures for systems that have more than 50,000 atoms. The general functional form of a force field is written as the sum of potentials that describes the bond stretching, angle bending, change in torsional angle and non-bonded nature of a system. See Equation 4.1.

$$U(r^N) = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsions}} + E_{\text{non-bonded}} \tag{4.1}$$

The interatomic potential energy $U(r^N)$ is a function of the position r of N particles. For the AMBER MD code the terms in Equation 4.13 expands as follows.

E_{bonds} is typically written as the sum over all bonds, $\sum_{\text{bonds}} k_i(l_i - l_{i,0})^2$, where k_i is the bond force constant, l_i is the equilibrium bond distance and $l_{i,0}$ is the reference bond distance.[108] The change in energy due to the stretching nature of covalently bonded atoms are thus described by this term.

E_{angles} describes an energy change when the angle between any three covalently bonded atoms deforms or rather, changes. E_{angles} is generally written as a sum over all valence angles, $\sum_{\text{angles}} k_i(\theta_i - \theta_{i,0})^2$, where k_i refers to the bond angle force constant, θ_i to the equilibrium bond angle and $\theta_{i,0}$ is the reference bond angle.[108,115–117]

E_{torsions} is written as $\sum_{\text{torsions}} \sum_i V_i[1 + cos(\eta\omega - \gamma)]$, where V_i gives a qualitative indication of the relative barriers of rotation, η (multiplicity) refers to the number of minimum points in the function as the bond is rotated by 360°, ω represents the torsion angle and γ (phase factor) determines where the torsional angle passes through its minimum angle.[108,115–117] The E_{torsions} term arises from the twisting of a bond due to a specific bond order and neighbouring bonds or lone pairs of electrons.

Arguably the more complicated term to describe, $E_{\text{non-bonded}}$ represents the non-bonded energy between all atom pairs. This term excludes atoms involved in bonding pairs and angle triads, and a special scaling factor is applied to atoms separated by three bonds to offset the effects of the torsional energy described in the above. This term combines Lennard-Jones and Coulombic potentials as $\sum_{j=1}^{N-1} \sum_{i=j+1}^{N} [\varepsilon_{ij}((\frac{r_{ij,0}}{r_{ij}})^{12} - 2(\frac{r_{ij,0}}{r_{ij}})^6) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}]$, where ε_{ij} is the well-depth in the van der Waals energy, $r_{ij,0}$ is the reference equilibrium distance between particles i and j, r_{ij} is the actual distance between the systems i and j, and q_i refers to the charge of the particle i.[108,115–117]

4.2 Electrostatic Interactions[108–110]

The electrostatic interaction is the attractive- or repulsive interaction experienced between two objects that have electrical charges. The electrostatic interaction between two molecules (or two ions from the same molecule – think of an ionic liquid) are described (according to Equation 4.2) by the fractional point charges (ζ) of the molecules/ions and by summing the interactions between a pair of these point charges using Coulomb's law

$$\zeta = \sum_{(i,j)=(1,1)}^{(N_A,N_B)} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{4.2}$$

where N_A and N_B refers to the number of point charges in molecular/ionic components A and B respectively. q_i is the partial charge of atom i in component A and r_{ij} corresponds to the distance between atom i from component A and atom j from component B.

For ionic liquids, the electrostatic interaction between cations and anions not only affects how strongly the two components interact but it also has a direct effect on its thermodynamic properties such as density, heat of vaporisation, viscosity and more. When the electrostatic interaction between two components increases, the distance between them will decrease (as the electrostatic energy is inversely proportional to the distance separating the two components), which will also result in a decrease in the total volume of the system as the volume is directly proportional to the distance between the two components. Since density is inversely proportional to volume, the density will increase. This result implies that the stronger the electrostatic interaction is between two components, the higher the system's density will be.

As a result of the neglect of polarisation from partial charges parameterised from gas phase calculations (and to some extent charge transfer), the electrostatic interaction (attraction) between ions are overestimated and can lead to significant errors in thermodynamic- and transport properties.[120−123] The electrostatic interaction between the ions can thus be lowered by scaling down the atomic partial charges of the ions in Equation 4.2.[123−126] The ideal atomic partial charge scaling factor, however, is not easy to find from first principles; and must be found by screening various scaling factors and determining which one describes the various physical properties the best.

4.3 Van der Waals Interactions[108−110]

Permanent electrostatic interactions cannot account for all the non-bonded interactions that exist within a system. It is known that there exist other non-bonded interactions, such as in the case of rare gas atoms where all the permanent multipole moments of these atoms are zero, which implies that there are no dipole-dipole or dipole-induced dipole interactions. This phenomenon is a non-bonded interaction as there must be some interaction present that allows for the rare gas atoms to have a liquid– and solid phase or at the least show deviations from ideal gas behaviour.[108] Various experimental studies, such as the use of molecular beams and spectroscopy, provide evidence of the existence of van der Waals interactions.[127−128]

The most popular potential function that is used to describe the van der Waals interaction is the Lennard-Jones 12-6 function, which takes the functional form as per Equation 4.3 for the interaction between two atoms.[108]

$$u(r) = \varepsilon_{ij}[(\frac{r_{ij,0}}{r_{iJ}})^{12} - 2(\frac{r_{ij,0}}{r_{iJ}})^{6}] \tag{4.3}$$

The functional form in Equation 4.3 is found in the non-bonded interaction potential term in the AMBER force field. Here, ε_{ij} refers to the well-depth, r_{ij} refers to the distance between atoms i and j and $r_{ij,0}$ is the equilibrium distance between the two atoms.

4.4 Force Fields in AMBER

Two popular force fields that are available in the AMBER MD code are the GLYCAM[129] and GAFF (General Amber Force Field)[130] force fields. For the parameterisation of the GLYCAM06 force field, approximately 100 molecules from the chemical families of ethers, alcohols, hydrocarbons, amides, carboxylates, molecules of mixed functional groups and also simple ring systems were tested. If any internal hydrogen bonds formed during the testing, two states were considered by the developers. One in which the hydrogen bond was allowed and the other where it was disallowed[129].

For valence quantum calculations, the B3LYP/6-31++G**//HF/6-31G* level of theory was chosen, as it was expected that the energies produced at this level of theory would not be heavily affected by basis set superposition error(s) (BSSE). This assumption is based on the research paper by Lii *et al.*, who dramatically lowered the BSSE in water dimer hydrogen bonding using the B3LYP/6-31++G** level of theory.[131] A further justification of the choice of level of theory was to accurately determine interaction energies, and the B3LYP functional has been shown to work well for interactions involving carbohydrates.[129] For neutral fragments, the GLYCAM developers made use of the HF/6-31G* level of theory and added diffuse functions when anionic species were considered. Single point energies were computed at the B3LYP/6-31++G** level of theory.[129] The GLYCAM06 force field was developed to not only derive parameter sets for carbohydrates but to also provide property information on systems which are too difficult or impossible to access experimentally. The overall results

produced in the GLYCAM06 work and the resultant set of parameters created agree reasonably well with experimental observations and has been a force field of choice for many researchers.[129,132−134]

GAFF (General AMBER Force Field) was developed by Wang et al. to work with the AMBER MD code to describe most pharmaceutical molecules and to be compatible with the traditional AMBER protein force field. The development of the GAFF force field came at the time where AMBER had a limited set of parameters for organic molecules.[130]

For GAFF, 35 basic atom types were introduced. These include different atom types for sulfur (S), phosphorus (P), hydrogen (H), oxygen (O), nitrogen (N), carbon (C), fluorine (F), chlorine (Cl), bromine (Br) and iodine (I). For GAFF, equilibrium bond lengths and bond angles come from experimental data and high-level *ab initio* calculations, while force constants were estimated through an empirical approach and further optimised. The GAFF parameterisation is based on more than 3000 MP2/6-31G* optimisations and 1260 MP4/6-311G** single-point calculations. Additionally, non-bonded parameters were assigned using a restrained electrostatic potential fit model.[130] GAFF has been widely used in papers by Sprenger *et al.*[135], Jia *et al.*[136], and Peri *et al.*[137] In this project, the GAFF2 force field is used which is an extension of the GAFF force field but is still under active development, and its final form has not been published yet, but is openly available.

4.5 Some Background on Molecular Dynamics

To simulate the intermolecular interactions of a system as the position of a particle changes or as the position of a particle that the tagged particle is interacting with changes, the resultant force must also change. In such a case, the integration of Newton's laws of motion gets divided into stages where each stage is separated by a fixed time of $\delta\tau$. The total force on each particle in a specific configuration at a time t is calculated by using the vector sum of the interaction between particles. This total force is used to calculate the acceleration of particles at time t, which when combined with its position and velocity, can describe its position and velocity at the time $t + \delta\tau$.

Many algorithms exist for integrating the equations of motion using the method described, but they all assume that the positions, velocities, accelerations and other dynamic positions can be approximated by making use of Taylor series expansions as found below in equations 4.4-4.6.[108]

$$r(t+\delta\tau) = \sum_{n=0}^{\infty} \left(\frac{1}{n!}\right) \delta\tau^n \frac{\partial^n r(t)}{\partial t^n} \tag{4.4}$$

$$v(t+\delta\tau) = \sum_{n=0}^{\infty} \left(\frac{1}{n!}\right) \delta\tau^n \frac{\partial^{n+1} r(t)}{\partial t^{n+1}} \tag{4.5}$$

$$a(t+\delta\tau) = \sum_{n=0}^{\infty} \left(\frac{1}{n!}\right) \delta\tau^n \frac{\partial^{n+2} r(t)}{\partial t^{n+2}} \tag{4.6}$$

where $r(t)$ refers to the position of a particle at time t, while v and a refer to its velocity and acceleration, respectively.

Making use of Newton's laws of motion and the equations 4.4-4.6, the force at time $t + \delta\tau$, F_{r_i}, can be derived as per Equation 4.7.

$$F_{r_i}(t + \delta\tau) = m_i a(t + \delta\tau) = m_i \sum_{n=0}^{\infty} \left(\frac{1}{n!}\right) \delta\tau^n \frac{\partial^{n+2} r(t)}{\partial t^{n+2}} \tag{4.7}$$

Two popular algorithms that can describe this force is the Verlet algorithm and the leapfrog algorithm.

4.5.1 Verlet Algorithm[108]

Making use of time t and the position of $r(t - \delta\tau)$, this algorithm can calculate new positions of $r(t + \delta\tau)$ at time $t + \delta\tau$. Taylor series expansions can be used to write these positions as shown in equations 4.8 and 4.9.

$$r(t + \delta\tau) = \sum_{n=0}^{\infty} \left(\frac{1}{n!}\right) \delta\tau^n \frac{\partial^n r(t)}{\partial t^n} \tag{4.8}$$

$$r(t - \delta\tau) = \sum_{n=0}^{\infty} \left(\frac{(-1)^n}{n!}\right) \delta\tau^n \frac{\partial^n r(t)}{\partial t^n} \tag{4.9}$$

Adding Equation 4.8 and 4.9 together, the following equation (Equation 4.10) is derived

$$r(t + \delta\tau) + r(t - \delta\tau) = \sum_{n=0}^{\infty} \left(\frac{2}{(2n)!}\right) \delta\tau^{2n} \frac{\partial^{2n} r(t)}{\partial t^{2n}} \approx 2r(t) + \delta\tau^2 a(t) \tag{4.10}$$

While the Verlet algorithm does not explicitly have velocity terms, it can still be calculated by dividing position differences at time $t + \delta\tau$ and $t - \delta\tau$ by $2\delta\tau$. See Equation 4.11.

$$v(t) = \frac{r(t + \delta\tau) - r(t - \delta\tau)}{2\delta\tau} = \sum_{n=0}^{\infty} \left(\frac{2}{(2n+1)!}\right) \delta\tau^{2n+1} \frac{\partial^{(2n+1)} r(t)}{\partial t^{(2n+1)}} \tag{4.11}$$

Newton's laws of motion and the above equations thus makes calculating the force $F_{r_i}(t)$ possible as per Equation 4.12.

$$F_{r_i}(t) = m_i \left(\frac{\partial v_i(t)}{\partial t}\right) = m_i \sum_{n=0}^{\infty} \left(\frac{2}{(2n+1)!}\right) \delta\tau^{2n+1} \frac{\partial^{(2n+1)} r(t)}{\partial t^{(2n+2)}} \tag{4.12}$$

Applying an approximation via the Verlet algorithm to Equation 4.12, the series is turned into a finite sum, $m_i \sum_{n=0}^{k} \left(\frac{1}{(2n+1)!}\right) \delta\tau^{2n+1} \frac{\partial^{(2n+1)} r(t)}{\partial t^{(2n+2)}}$; where k is a natural number.

The major drawback of the Verlet algorithm is thus some of the approximations and simplifications it makes.

4.5.2 Leapfrog Algorithm[108]

This algorithm is similar to the Verlet algorithm, however, it uses the following two core position and velocity equations (4.13 and 4.14):

$$r(t+\delta\tau) = r(t) + v(t+\frac{1}{2}\delta\tau)\delta\tau \tag{4.13}$$

$$v(t+\frac{1}{2}\delta\tau) = v(t-\frac{1}{2}\delta\tau) + a(t)\delta\tau \tag{4.14}$$

The velocity at time t can then easily be calculated as per Equation 4.15

$$v(t) = \frac{1}{2}\left(v(t+\frac{1}{2}\delta\tau) + v(t-\frac{1}{2}\delta\tau)\right) \tag{4.15}$$

In the leapfrog algorithm the velocity and position terms "leapfrog" over one other and outputs their values at a time $t+\frac{k}{2}\delta\tau$, where k is some natural number. For example, the velocity will "leapfrog" over the position to give a new velocity term at $t+\frac{\delta\tau}{2}$. Hereafter, the positions will "leap" over the velocities and a new position term at $t+\delta\tau$ is outputted. Unlike the Verlet algorithm, the leapfrog algorithm includes a velocity term, as such, the position and velocity terms are not synchronised. This non-synchronicity means that the kinetic energy will not contribute to the total energy at the same time the positions are defined.

4.5.3 Particle Mesh Ewald Method: Ewald Summation[108]

The Ewald (summation) method gives an accurate approximation for calculating long-range interactions such as "the electrostatic energy of a periodic system consisting of an infinitely replicated neutral box of charged particles".[138] In this method, all interacting particles are interacting with all other particles in the simulation cell, including all their images in the infinite periodic cells. The interactions between the charges in the central box with the charges in the image box are given by the following potential function (Equation 4.16)

$$\sum_{|n|=0}^{\infty} \sum_{(i,j)=(1,1)}^{(N,N)} \frac{q_i q_j}{4\pi\epsilon_0 |r_{ij}+n|} \tag{4.16}$$

where n is the position at a cubic lattice point (n_xL;n_yL;n_zL) with n_x, n_y and n_z being integers and L being one n^{th} of the distance between simulation boxes. When n=0, i cannot be equal to j. Unfortunately, the summation in Equation 4.16 converges slowly and ends up being conditionally convergent, in which the order terms are considered affects if the sum converges or diverges. Just a reminder, if a sum converges it means that the sum will tend to a finite number, while a divergent sum is one in which the sum tends to $+\infty$ or -∞. To bypass this in the Ewald method, the summation is divided up into two series which converge, quickly too. Subsequently, the potential function can also be divided into several parts. The first part is as below (Equation 4.17)

$$V_1 = \frac{1}{4\pi\epsilon_0\sqrt{\pi}} \sum_{n=0}^{\infty} \sum_{(i,j)=(1,1)}^{(N,N)} \frac{1}{|r_{ij}+n|} \int_{\alpha|r_{ij}+n|}^{\infty} e^{-t^2} dt \tag{4.17}$$

Here, α is the Ewald convergence parameter and t is some arbitrary parameter. The second part of the potential function is (Equation 4.18)

$$V_2 = \frac{1}{2\mathrm{L}^3\epsilon_0} \sum_{k=\frac{2\pi n}{\mathrm{L}^2}\neq 0} \sum_{(i,j)=(0,0)}^{(N,N)} \frac{q_i q_j}{k^2} e^{-\frac{k^2}{4\alpha^2}} \cos(k \cdot r_{ij}) \tag{4.18}$$

A self-interaction correction is done on Equation 4.18 as the Gaussian functions in Equation 4.18 are subsequently interacting with itself as well. This correction term is (Equation 4.19)

$$V_3 = -\frac{\alpha}{\sqrt{\pi}} \sum_{m=1}^{N} \frac{q_m^2}{4\pi\epsilon_0} \tag{4.19}$$

A correction term is then finally included for surrounding the simulation boxes. Since in this project the medium is a vacuum, the following potential function is introduced (Equation 4.20)

$$V_4 = \frac{2\pi}{3\mathrm{L}^3} \left| \sum_{f=1}^{N} \frac{q_f}{4\pi\epsilon_0} r_f \right|^2 \tag{4.20}$$

Therefore, the Ewald summation can be written as $V_{\text{Ewald}} = V_1 + V_2 + V_3 + V_4$.

[Page intentionally left blank.]

5 Force Field Parameterisation

5.1 Introduction to Theory

5.1.1 Parameterisation of a System

Unlike computational methods such as *ab initio* molecular dynamics and quantum mechanics, (classical) molecular dynamics requires parameterisation. Parameterisation is the process of computing and assigning partial charges, bonded and non-bonded parameters to a system of consideration. These parameters are usually based on experimental data and high-level QM calculations. In molecular mechanics, parameterisation is crucial; not only should a good parameter set reproduce experimental data for test systems, but for any suitable system outside of these.

Parameters can be validated by performing molecular dynamics simulations and comparing the physical properties calculated from this to experimentally known values. The better the correlation, the more accurate the parameterisation is. The most significant bottleneck with parameterisation validation is obtaining literature values for properties such as density and heat of vaporisation. In this project, the parameterisation protocol was divided into three parts. Part 1 was using the R.E.D. (RESP ESP charge Derive) Server to assign atomic partial charges to each component. Part 2 was using `parcmhk2` to obtain all the bond, bond angle, torsional and non-bonded parameters for each component. And part 3 was a torsional fitting (using a genetic algorithm) of the ionic components to improve on their parameters.

Since the ionic compounds in this project was not part of the GAFF(1)/GAFF2 test set, torsional parameters have not been refined for these species. However, quantum mechanics is the best model available to describe and replicate the nature of a system yet is time-consuming; therefore a moulding of MM energies onto QM energies can generate highly accurate results in a short amount of time. Using the process of fitting an MM (AMBER) torsional profile over a QM one, better force constants for the various torsions of the ionic components can be obtained. This process makes use of a genetic algorithm. A generalised description of the genetic algorithm (GA) follows that a population of species are created and will undergo recombination and mutation to produce offspring. This process will continue for several generations of offspring in which each individual is assigned a fitness value. This fitness value will be compared to that of an end state. The "fitter" individual is awarded a higher chance of mating and in some cases, this mating can lead to the mutation of even "fitter" individuals. This process continues until a stopping point is decided on. The longer the process continues, the closer the system will eventually be to the end state.[139]

In this work, a genetic algorithm is employed to improve on the torsional parameters of ionic species as obtained from the GAFF(1)/GAFF2 force field. The process commences by guessing a specific set of torsional parameters. The torsional parameters will be evaluated against the QM torsional parameters and better values will resultantly be assigned. After some pre-selected time, the algorithm will be complete. The longer the algorithm runs, and the more accurate the chosen genetic algorithm functions are, the closer the MM torsional parameters will be to that computed by quantum mechanics. GAs have been used for optimisation in many fields of study, including image processing, complex network design, job scheduling and parameter fitting.[140] Freeman *et al.* reported a GA approach as being quick to achieving acceptable NMR pulse shapes.[141–143] Lucasius *et al.* reported a GA approach to "being accurate at choosing an optimum set of wavelengths to measure the absorptivity of a sample to determine the concentrations of a number of RNA nucleotides".[144–145]

5.1.2 Minimisation, Heating, Equilibration and Production

A typical molecular dynamics protocol (setup) involves first building an initial simulation box (configuration) and minimising the energy of the system. Hereafter, the minimised simulation box is heated to a target temperature, is equilibrated and a production phase is performed during which configurations are saved and energy fluctuations are monitored.

The initial configuration of simulation boxes is performed by `Packmol`. `Packmol` creates an initial point for MD simulations by packing molecules in randomised, yet defined regions of space. The packing is done in such a way that short-range repulsive interactions do not disrupt the simulations. Minimisation calculations are then performed before molecular dynamics to minimise the energy of the simulation box by adjusting the atomic coordinates. This minimisation run gives the lowest energy state of the simulation box as it is generally found to fluctuate around in nature.

The heating of a simulation box (which continues from the minimisation) is required as the minimisation only alters atomic positions to lead to a configuration with lower potential energy relative to the initial configuration and does not take kinetic energy (such as temperature) into account.[146] If the kinetic energy contributions are added all at once to the target temperature (and applied in a Maxwell-Boltzmann distribution across all atoms), this can lead to the system behaving nonphysically. The preferred approach is thus to heat the system up from 0 K to the target temperature (in Kelvin) over several molecular dynamics time steps.

The equilibration phase (NPT ensemble) of a molecular dynamics simulation allows the simulation box to relax to a configuration that best describes the equilibrium state of the system. Equilibration calculations are performed at constant temperature and -pressure where the volume of the entire simulation box is allowed to change to reach a constant density (as to be compared to the experimental value).

The production phase is the final molecular dynamics phase and is dependent on the equilibration data. It is typically performed at constant volume and temperature (that is, it follows an NVT ensemble), with the pressure and energy of the system being monitored throughout the run. During this calculation, the atomic coordinates are saved at set intervals which allow for post-processing and analysis. The production phase should be long enough to give an accurate representation of properties fluctuating around their equilibrium values.

5.1.2.1 Periodic Boundary Conditions and Imaging[146]

As a spherical system increases in size so does its volume ($V = \frac{4}{3}\pi r^3$), and its area ($A = \pi r^2$). For a macroscopic system, this minimally affects the chemistry being studied, but in computational chemistry where the systems are generally small, these system properties can have a large effect on the chemistry. To avoid this, during the molecular dynamics simulations, periodic boundary conditions, henceforth known as PBCs, are used. Under these conditions, the system is assumed to be in a unit cell in some ideal crystal, which implies that for a target simulation box, one can assume there exist 26 identical simulation boxes directly surrounding it. In each of these 27 simulation boxes, similar atoms move at the same speed, to the same position and interact in the same way. As such, when an atom in the target simulation box moves out of the box, it is assumed that an atom from a neighbouring simulation box would also be moving out of its box and would thus be moving into the main box. This process is known as imaging.

PBCs thus preserve mass, particle number, total energy and linear momentum; however, it does not conserve angular momentum, but this should be more-or-less fixed over large simulation time averaging. Another advantage of PBCs is that it allows the Ewald summation to be used more accurately.

5.2 Computational Details

Parameterisation. The structures of 1-butyl-3-methylimidazolium, acetate, methyl sulphate, propylene carbonate, dimethyl carbonate and γ-valerolactone were built in `GaussView` and geometry optimised using the `Gaussian09` application suite at the MP2/6-31G* level of theory, with a tight SCF convergence criteria chosen. Diffuse functions were included for geometry optimising the anions as well, as to correctly describe certain chemical properties (e.g. HOMO/LUMO energies, electron affinities, etc.). Hereafter, frequency calculations at the same level of theory were computed on the different components as the absence of imaginary frequencies confirms the component to be at a minimum. The choice of level of theory for the geometry optimisation in this project was influenced by Wang *et al.*,[130] using it to develop the General AMBER Force Field (GAFF).

Multi conformer partial charge fitting was performed hereafter on each ionic component. Conformers were generated using the `Confab algorithm`[147], as implemented in the `Open Babel` package[148], with an energy cut-off of 50.0 kcal/mol. The input files for the conformational search can be found in Appendix 1. Subsequently, each of the ionic conformers produced from the conformational search was geometry optimised at the MP2/6-31G* level of theory and their energies compared. All unique conformers (based on their energy) were then uploaded to the R.E.D. Server (as well as the geometry optimised structures of the co-solvents), which fits restrained electrostatic potential (RESP) partial charges via the HF/6-31G* level of theory.

The `parmchk2` utility, distributed with the `AmberTools` package, was used to assign initial parameters to the bonded and non-bonded contributions of each component using the GAFF1 and GAFF2 force field.[146] Torsions from the ionic components involving strictly heavy atoms and present in a non-rigid environment were altered by fitting MM torsional profiles to QM profiles using a genetic algorithm (see Appendix 1 for the GA scripts). As part of this procedure, a relaxed scan of a tagged torsion is performed (using MP2/6-31G* or MP2/6-31+G*) starting the rotation about the dihedral angle from 0° up to 360° in increments of 5°, while keeping all other target torsions (if any) in the component fixed at either 0° or 180°. An overlap of MM and QM torsional profiles are desired here as the torsional parameters will be related to highly accurate QM calculations. None of the remaining GAFF-derived parameters (bonded and non-bonded) was altered.

Charge scaling. In the case of 1-butyl-3-methylimidazolium acetate and 1-butyl-3-methylimidazolium methyl sulfate, each component's atomic partial charges were scaled to study how it affects the calculated density, heat of vaporisation and self-diffusion coefficients. Arbitrary scaling factors of 0.7, 0.8 and 0.9 were chosen. At each scaling factor, minimisation, heating, equilibration and production calculations were done as described below. Once completed, the calculated properties were compared to each other and to the experimentally known values, with the scaled ionic liquid producing the best results being selected for further calculations. (The experimental properties were obtained from an ionic liquid database.[150])

Minimisation, heating, equilibration and production. Following the parameterisation step, ionic liquid and co-solvent simulation boxes consisting of 256 molecules (or ion pairs) were constructed using `Packmol`, applying a molecular distance tolerance of 2.5 Å.[151] Volumes selected for the simulation boxes are taken from the equation $V = \frac{NM}{N_A \rho}$, where N is the number of molecules, M is the molar mass, N_A is Avogadro's number and ρ is the experimental density of the solvent at 298.15 K. The `tleap` module (part of the `AmberTools` package) was then used to create parameter- and topology files for the simulation. All MD simulations were performed with periodic boundary conditions in the isothermal-isobaric ensemble using the `PMEMD` program of AMBER16. The particle mesh Ewald method was used to calculate the full electrostatic energy of a unit cell in a macroscopic lattice of repeating images. All bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[152] A 2 fs time step was strictly used for all MD simulations, with the temperature being regulated with Langevin dynamics with a collision frequency of 2 ps^{-1}. A Monte Carlo barostat was used for pressure regulation, with 100 steps between volume change attempts. Simulation box trajectories were written 500 or 1000 times over a span of 1 ps. Heating of the simulation boxes was performed from 0 K to 298.15 K for 10 ns. The equilibration calculations were done for 50 ns at 298.15 K. Long equilibration times are essential for the highly viscous ionic liquids. The production run was done for 100 ns at 298.15 K. Single molecules in the gas phase (i.e. a simulation box containing only one molecule or one ion pair as the latter has been shown by previous work to vaporise as a pair.[153]) were minimised, heated, equilibrated and sent through a production run at the same conditions and using the same procedure as for the liquid phase.

Density and heat of vaporisation. Density values were taken directly from the equilibration output file (making use of the `mdout_analyzer.py` python script as distributed in the `AmberTools` package). Simulation boxes are well equilibrated once the time series of density values plateaus. See Figure 5.1 for the density profiles of $[C_4C_1im][CH_3SO_4]$ and $[C_4C_1im][CH_3COO]$, respectively.

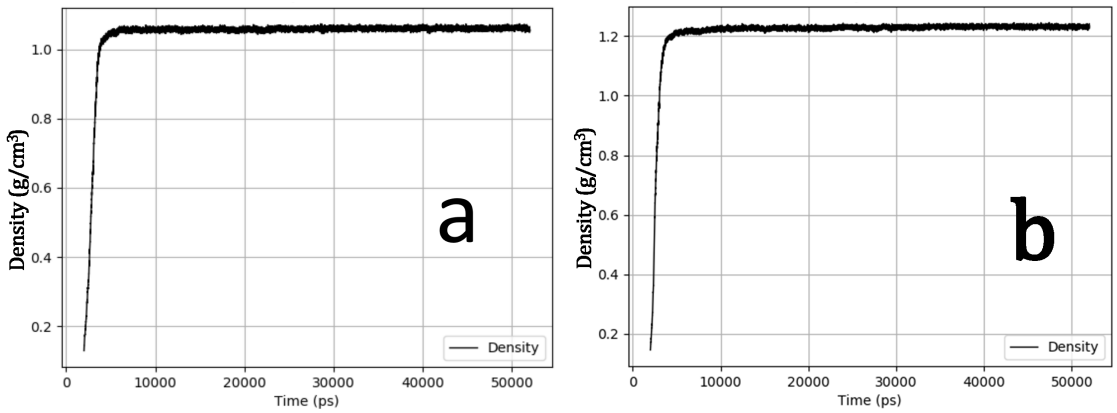


Figure 5.1: The density profiles of the a) $[C_4C_1im][CH_3COO]$ 6-31G* parameter set and b) $[C_4C_1im][CH_3SO_4]$ 6-31+G* parameter set.

Heat of vaporisation values were calculated using Equation 5.1.[154]

$$\begin{aligned}\Delta_{\text{vap}}H = H(p,T)_{\text{gas}} - H(p,T)_{\text{liquid}} &= \frac{\langle E_{\text{gas}}\rangle}{N_{\text{gas}}} - \frac{\langle E_{\text{liquid}}\rangle}{N_{\text{liq}}} + P(V_{\text{gas}} - V_{\text{liquid}}) \\ &= \frac{\sum_{k=1}^{K} E_{\text{gas}}(k)}{KN_{\text{gas}}} - \frac{\sum_{k=1}^{K} E_{\text{liq}}(k)}{KN_{\text{liq}}} + PV_{\text{gas}} - PV_{\text{liquid}} \\ &\approx U_{\text{gas}} - U_{\text{liquid}} + RT\end{aligned} \tag{5.1}$$

where $\langle E_{\text{gas}}\rangle$ and $\langle E_{\text{liq}}\rangle$ (the average ensemble energies for the gas and liquid phase) were taken directly from the production output file using the `mdout_analyzer.py` python script, R refers to the universal gas constant, T is the temperature, N_{liq} and N_{gas} are the number of molecules in each respective physical state, U_{gas} and U_{liquid} are the gas- and liquid-phase internal energies and $\Delta_{\text{vap}}H$ is the heat of vaporisation. In Equation 5.1 the assumption was made that since the gas-phase simulation box consists of only 1 molecule or ion-pair, V_{gas} will be negligibly small. Another assumption made was that the solvents will obey the ideal gas law, transforming PV into nRT, which is just equal to RT for a one mole system.

Self-diffusion. The self-diffusion coefficients (D) of 1-butyl-3-methylimidazolium and methyl sulfate were calculated according to each ionic component's centre of mass using the `Travis Analyzer` package as per the Einstein relation from Equation 5.2.[155]

$$\lim_{\tau\to\infty} \text{MSD} = \lim_{\tau\to\infty} 2n\tau D \tag{5.2}$$

where $\text{MSD} = \dfrac{1}{N}\sum_{i=0}^{N} |r_i(\tau) - r_i(0)|^2$ is the mean square displacement, τ is the time shift, n is the number of dimensions in which the particles diffuse and N is the total number of molecules in the simulation box being studied. Taking the slope of MSD versus τ outputs the self-diffusion coefficient. The early points in the MSD versus τ plot are ignored as these represent ballistic motion in the solvent cage and is not representative of true diffusive motion.

5.3 Building and Parameterising the Pure Solvent Systems

Acetate, methyl sulfate, 1-butyl-3-methylimidazolium, propylene carbonate, dimethyl carbonate and γ-valerolactone were constructed in `GaussView`, with its geometry optimisation being computed at the MP2/6-31G* level of theory. The anions were also geometry optimised at the MP2/6-31+G* level of theory. Frequency calculations at the same levels of theory confirmed the minima on their respective potential energy surfaces, with the appearance of no imaginary frequencies. The geometry optimised structures of these components are shown in figures 5.2-5.5.

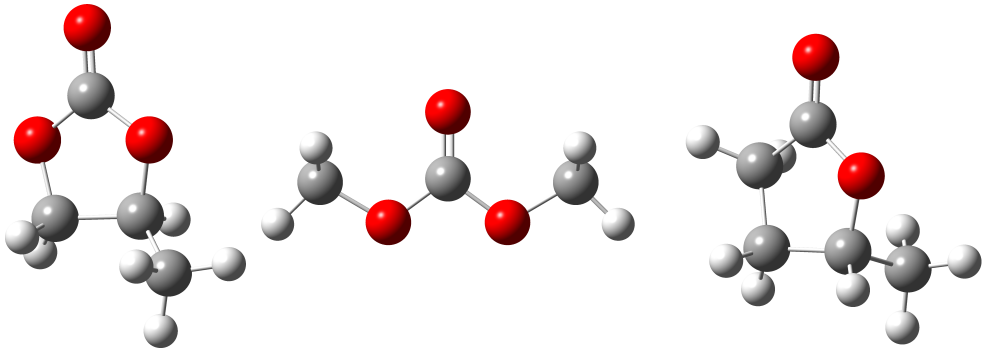


Figure 5.2: The geometry optimised structures of propylene carbonate (left), dimethyl carbonate (middle) and γ-valerolactone (right).

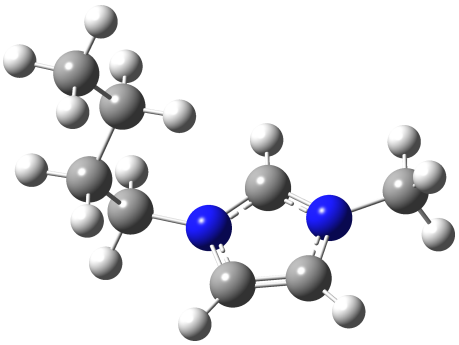


Figure 5.3: The geometry optimised structure of 1-butyl-3-methylimidazolium as computed at the MP2/6-31G* level of theory.

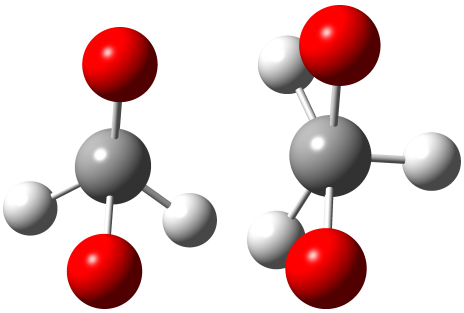


Figure 5.4: A top down view of the geometry optimised structures of the acetate anion at the MP2/6-31+G* (left) and MP2/6-31G* (right) levels of theory.

Notice that for the acetate anion, there is a difference in terms of their structure as the methyl groups have rotated, with the methyl group eclipsed when using the 6-31+G* basis set and staggered when using 6-31G*. Further investigation showed that the alternate structure is a saddle point, depending on the basis set used. However, the difference in energy in structures for Figure 5.4 are negligibly small.

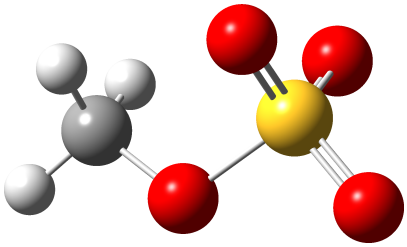


Figure 5.5: The geometry optimised structure of methyl sulfate at the MP2/6-31G* and MP2/6-31+G* levels of theory. [Both were the same.]

Hereafter, atomic partial charges were fitted to the ionic components using multi conformer fitting. This procedure included first performing a conformational search on each component using the `Confab algorithm`[147], as implemented in the `Open Babel` package[148]. Following this, all conformers generated were geometry optimised and the energy of each compared; this was done as to exclude conformers that optimised to the same structure (energy). For propylene carbonate, dimethyl carbonate and γ-valerolactone, their structures from the geometry optimisation were used as is. A conformational search on the acetate anion produced only one conformer, while for the methyl sulfate anions and the 1-butyl-3-methylimidazolium cation, 6 and 26 conformers were generated, respectively. A geometry optimisation of the 26 cationic conformers produced only 8 that are unique. These conformers are shown in Figure 5.6.

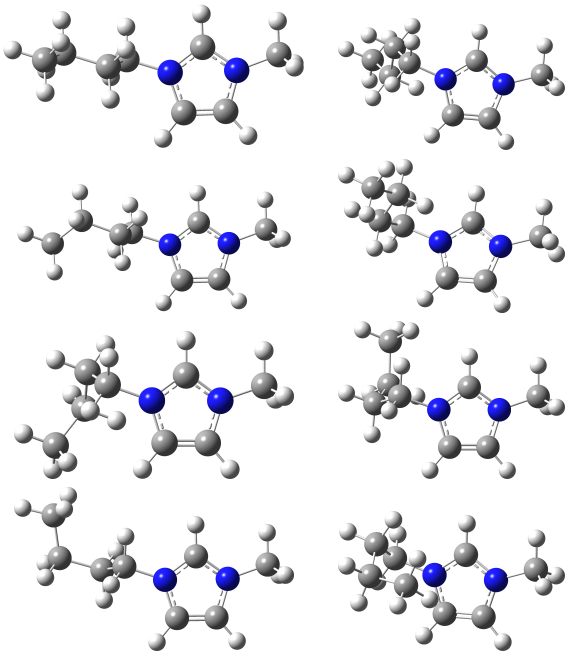


Figure 5.6: The 8 unique conformers of 1-butyl-3-methylimidazolium produced from the geometry optimisation of the 26 conformers obtained from the `Confab`/`Openbabel` conformational search.

It should be noted that the above process was done using a random rotor search, and that other procedures, such as a weighted rotor search, might produce less/more/different unique conformers. Fortunately, the larger number of conformers had very little effect on the partial charges generated, as such, only those in Figure 5.6 were used for charge fitting. Each conformer/geometry optimised structure of the 8 molecules/ions were uploaded to the R.E.D server,[156] where partial charges were fitted using the HF/6-31G* level of theory. HF/6-31+G* charges were also calculated for the anions. Thereafter, `parmchk2` was used to assign bonded- and non-bonded parameters from the GAFF(1) and GAFF2 force fields. Since the target ionic compounds were not in the GAFF/GAFF2 test set, a torsional profile fitting (using a genetic algorithm) was done for the ionic components. The acetate anion, however, does not have any torsions in a non-rigid environment that matches the fitting criteria, as such, was not fitted. The cation has four torsions that can be fitted, but two of them have terminal ends ending in different atom types that are offset by 180 degrees around the sp2 ring nitrogen. The one torsion was thus kept unchanged, while the other was torsional profile fitted. The methyl sulfate anion has three similar torsions which all contribute together, as such only one of them was fitted. The torsions that were fitted for the cation is shown in Figure 5.7, together with their torsional profiles in Figure 5.8. The methyl sulfate torsion and its profile is shown in Figure 5.9.

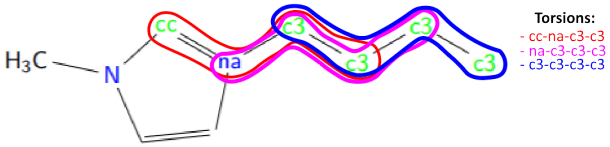


Figure 5.7: The three torsions selected for fitting in 1-butyl-3-methylimidazolium.

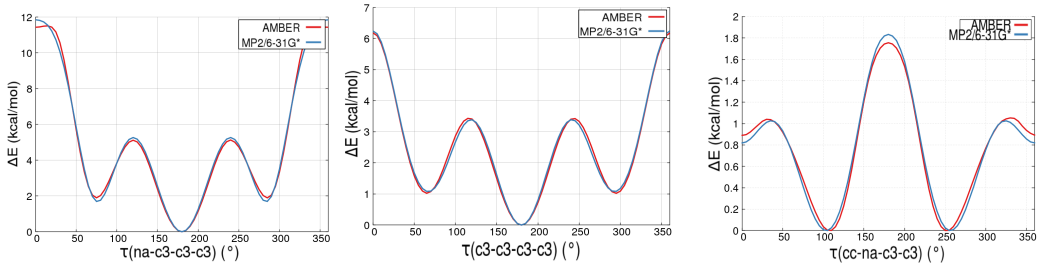


Figure 5.8: The torsional profile fitting of torsions na-c3-c3-c3 (left), c3-c3-c3-c3 (middle) and cc-na-c3-c3 (right). The red profile is for the MM torsional profile and the blue is for QM.

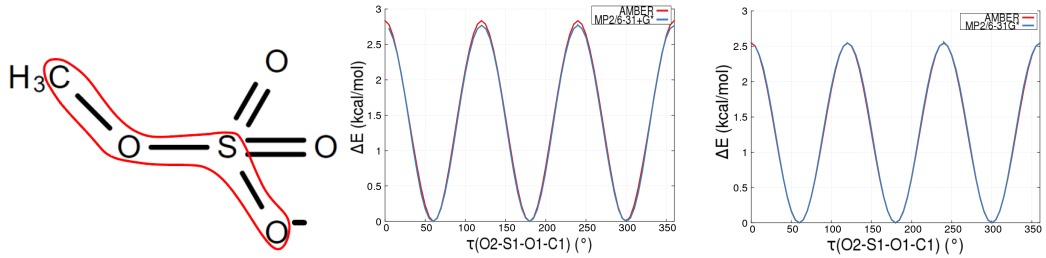


Figure 5.9: The C1-O1-S1-O2 torsion of methyl sulfate (left), the torsional profile fitting of the methyl sulfate anion as computed using the 6-31+G* basis set (middle) and the torsional profile fitting of methyl sulfate as computed using the 6-31G* basis set (right). The red profile is for the MM torsional profile and the blue is for QM.

Hereafter, simulation boxes consisting of 256 solvent molecules/ion-pairs were constructed using `Packmol`; this includes pairing the four sets of anion parameters (with and without diffuse functions) individually with the cation. `Tleap` was then employed to produce parameter- and topology files of each simulation box. The equilibrated simulation box of pure propylene carbonate is shown in Figure 5.10.

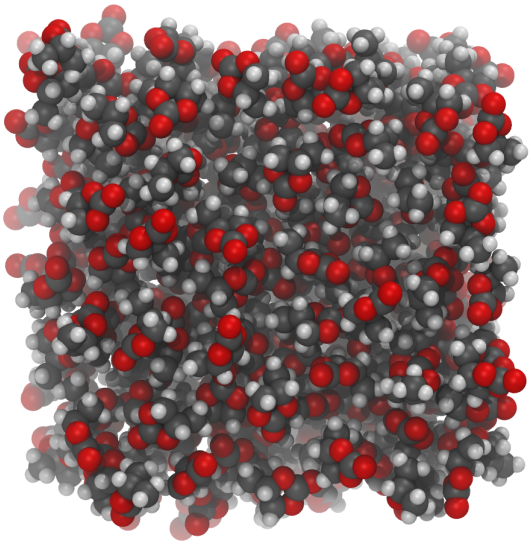


Figure 5.10: A simulation box consisting of 256 propylene carbonate molecules.

Each simulation box was minimised, heated, equilibrated and sent through a production run, after which the density, heat of vaporisation and self-diffusion coefficients were calculated. The percentage error of these properties were calculated using Equation 5.3

$$\frac{\chi_{\text{comp}} - \chi_{\text{lit}}}{\chi_{\text{lit}}} \times 100 \tag{5.3}$$

where χ_{comp} is the value of the property calculated computationally and χ_{lit} is the literature value.

The computational density and -heat of vaporisation are presented in Table 5.1 and 5.2 for the pure co-solvents and the acetate-based IL using the GAFF(1) and GAFF2 force fields; while the

computational density and self-diffusion coefficients of the methyl sulfate-based IL is presented in Table 5.1 and 5.3 using GAFF2 (only). Literature heat of vaporisation values were not available for $[C_4C_1im][CH_3SO_4]$, however self-diffusion coefficients were. Conversely, literature self-diffusion coefficients were not available for all the other solvents, but heat of vaporisation values was. Density and heat of vaporisation values were only calculated for a few of the simulation boxes using GAFF(1) parameters as it became evident that GAFF2 was better for the purposes of this work.

Table 5.1: Comparison of the computational and experimental density of propylene carbonate, dimethyl carbonate, γ-valerolactone, 1-butyl-3-methylimidazolium acetate and 1-butyl-3-methylimidazolium methyl sulfate

Solvent	ρ_{lit} (in g/cm^3)	$\rho_{\text{comp,GAFF}}$ (in g/cm^3)	$\rho_{\text{comp,GAFF2}}$ (in g/cm^3)	GAFF % Error	GAFF2 % Error
Propylene carbonate	1.1951[157]	1.1566 ± 0.0073	1.1839 ± 0.0081	-3.22%	-0.94%
Dimethyl carbonate	1.0693[158]	1.0153 ± 0.0083	1.0448 ± 0.0096	-3.55%	-5.05%
$\gamma -$ Valerolactone	1.05[159]	1.0513 ± 0.0069	1.0357 ± 0.0069	0.12%	-1.05%
$[C_4C_1im][CH_3COO]$ (6-31+G* set)	1.0526[160]	1.0262 ± 0.0040	1.0619 ± 0.0036	-2.51%	0.88%
$[C_4C_1im][CH_3COO]$ (6-31G* set)	1.0526		1.0603 ± 0.0037		0.73%
$[C_4C_1im][CH_3SO_4]$ (6-31+G* set)	1.2122[161]		1.2291 ± 0.0036		1.39%
$[C_4C_1im][CH_3SO_4]$ (6-31G* set)	1.2122		1.2314 ± 0.0040		1.58%

Table 5.2: Comparison of the computational and experimental heat of vaporisation of propylene carbonate, dimethyl carbonate, γ-valerolactone and 1-butyl-3-methylimidazolium acetate

Solvent	$\Delta_{\text{vap}}H_{\text{lit}}$ (in kJ/mol)	$\Delta_{\text{vap}}H_{\text{comp,GAFF}}$ (in kJ/mol)	$\Delta_{\text{vap}}H_{\text{comp,GAFF2}}$ (in kJ/mol)	GAFF % Error	GAFF2 % Error
Propylene carbonate	66.4[162]	76.66 ± 3.31	67.86 ± 3.34	15.45%	2.27%
Dimethyl carbonate	37.93[162]	48.57 ± 3.17	40.66 ± 3.18	28.05%	7.20%
$\gamma -$ Valerolactone	54.35[162]	63.73 ± 3.56	61.90 ± 3.53	17.26%	11.92%
$[C_4C_1im][CH_3COO]$ (6-31+G* set)	134.8[163]	191.42 ± 5.35	196.99 ± 5.36	42.00%	46.14%
$[C_4C_1im][CH_3COO]$ (6-31G* set)	134.8		197.13 ± 5.32		46.24%

Table 5.3: Comparison of the computational and experimental self-diffusion coefficients of the GAFF2 parameterised 1-butyl-3-methylimidazolium methyl sulfate

Solvent	$D_{\text{cation,lit}}$ ($\times 10^{-12}$m^2/s)	$D_{\text{cation,comp}}$ ($\times 10^{-12}$m^2/s)	$D_{\text{anion,lit}}$ ($\times 10^{-12}$m^2/s)	$D_{\text{anion,comp}}$ ($\times 10^{-12}$m^2/s)	Total % Error
$[C_4C_1im][CH_3SO_4]$ (6-31G* set)	7[164]	0.17	6[160]	0.096	-97.95%
$[C_4C_1im][CH_3SO_4]$ (6-31+G* set)	7	0.12	6	0.11	-98.23%

Parameterising the solvents using GAFF2 parameters produces computational results that are closer to the experimental values than using GAFF(1) parameters, with the exceptions being for the density of γ-valerolactone and dimethyl carbonate, as well as the heat of vaporisation of 1-butyl-3-methylimidazolium acetate. Although the dimethyl carbonate and γ-valerolactone GAFF(1)-parameterised systems produce more favourable density values compared to its GAFF2-parameterised counterparts, their heat of vaporisation is much less accurately defined. From a thermodynamic- and molecular viewpoint, $\Delta_{\text{vap}}H$ provides insight into the strength of intermolecular forces in the liquid. This is an important property to describe accurately if the intention is to model the interaction

of the solvent with the solute. Experimentally, vaporisation enthalpies are often used in evaluating transport properties, is an essential parameter for a multi-component multistage vapour–liquid equilibrium process as it is the one process which controls the temperature as well as liquid and vapour profiles in a column. As such, this makes describing heat of vaporisation well, of fundamental importance.[165] As a result of this, obtaining accurate heat of vaporisation results are targetted; this makes the GAFF2-parameterised dimethyl carbonate and γ-valerolactone better for the purposes of this work.

While the acetate IL produces a more favourable heat of vaporisation using GAFF(1) parameters (as per Table 5.2) notice that in Table 5.1, its density has a negative percentage error. To decrease the heat of vaporisation, atomic partial charges are scaled down to decrease the anion-cation electrostatic interaction. This, however, will also give a greater error for the density, since scaling the atomic partial charges down also decreases the density. The same IL in Table 5.2, which is GAFF2-parameterised, has a positive percentage error for the density, implying that a larger scaling can be done and a lower density and heat of vaporisation percentage error calculated. This is yet another justification as to why GAFF2 parameterisation will only be considered in this project, henceforth and why GAFF(1) parameterisation was not implemented on all the solvent systems.

Since the ionic liquids in Table 5.2 and Table 5.3 produce undesirable heat of vaporisation and self-diffusion coefficient results, atomic partial charge scaling was done to improve the results. Scaling factors of 0.7, 0.8 and 0.9 were selected, where the atomic partial charges were scaled according to these factors. For the four ionic liquid parameter sets, the new density values can be found in Table 5.4.

From Table 5.4, it is evident that scaling atomic partial charges down by a factor of 0.9 produces accurate density values for the ionic liquids in comparison to literature values. When comparing the densities obtained from this work for $[C_4C_1im][CH_3COO]$ to Sprenger *et al.*[135]'s results (where GAFF was also used), all scaled density values are significantly improved relative to their results; where a scaling factor of 0.8 was selected. Table 5.5 contains the heat of vaporisation for the scaled acetate-based ILs.

From Table 5.5, it is evident that from the scaling factors chosen, the heat of vaporisation is represented at its best when choosing the 0.7 scaling factor. With a greater focus on accurate $\Delta_{vap}H$ values, the 0.7 scaling factor is the final scaling factor chosen for the acetate-based IL. When choosing this scaling factor, the density of the acetate-based ionic liquid parameter sets is still represented reasonably well. Note that for the parameters in which diffuse functions were added to the basis set, both the density and heat of vaporisation are represented best. As such, these parameters will be used for all further calculations. Table 5.6 contains the self-diffusion coefficients of the anion and cation from 1-butyl-3-methylimidazolium methyl sulfate.

Table 5.4: Comparison of the computational and experimental density of the scaled 1-butyl-3-methylimidazolium acetate and 1-butyl-3-methylimidazolium methyl sulfate ionic liquids

$[C_4C_1im][CH_3COO]$ (6-31G* set) Scaling Factor	ρ_{lit} (in g/cm^3)	ρ_{comp} (in g/cm^3)	Percentage Error
0.7	1.0526	1.0155 ± 0.0043	-3.52%
0.8	1.0526	1.0329 ± 0.0042	-1.87%
0.9	1.0526	1.0496 ± 0.0041	-0.29%
1.0	1.0526	1.0603 ± 0.0037	0.73%
$[C_4C_1im][CH_3COO]$ (6-31+G* set) Scaling Factor	ρ_{lit} (in g/cm^3)	ρ_{comp} (in g/cm^3)	Percentage Error
0.7	1.0526	1.0171 ± 0.0042	-3.37%
0.8	1.0526	1.0339 ± 0.0040	-1.78%
0.9	1.0526	1.0504 ± 0.0038	-0.21%
1.0	1.0526	1.0619 ± 0.0036	0.88%
$[C_4C_1im][CH_3SO_4]$ (6-31G* set) Scaling Factor	ρ_{lit} (in g/cm^3)	ρ_{comp} (in g/cm^3)	Percentage Error
0.7	1.2122	1.1776 ± 0.0045	-2.85%
0.8	1.2122	1.1960 ± 0.0042	-1.34%
0.9	1.2122	1.2136 ± 0.0041	0.12%
1.0	1.2122	1.2283 ± 0.0040	1.33%
$[C_4C_1im][CH_3SO_4]$ (6-31+G* set) Scaling Factor	ρ_{lit} (in g/cm^3)	ρ_{comp} (in g/cm^3)	Percentage Error
0.7	1.2122	1.1771 ± 0.0046	-2.90%
0.8	1.2122	1.1965 ± 0.0043	-1.30%
0.9	1.2122	1.2155 ± 0.0038	0.27%
1.0	1.2122	1.2314 ± 0.0036	1.58%

Table 5.5: Comparison of the computational and experimental heat of vaporisation of the scaled 1-butyl-3-methylimidazolium acetate ionic liquids

$[C_4C_1im][CH_3COO]$ (6-31G* system) Scaling Factor	$\Delta_{vap}H_{lit}$ (in kJ/mol)	$\Delta_{vap}H_{comp}$ (in kJ/mol)	Percentage Error
0.7	134.8	135.83 ± 5.33	0.76%
0.8	134.8	154.59 ± 5.31	14.68%
0.9	134.8	176.31 ± 5.32	30.79%
1.0	134.8	197.13 ± 5.32	46.24%
$[C_4C_1im][CH_3COO]$ (6-31G* system) Scaling Factor	$\Delta_{vap}H_{lit}$ (in kJ/mol)	$\Delta_{vap}H_{comp}$ (in kJ/mol)	Percentage Error
0.7	134.8	135.58 ± 5.30	0.58%
0.8	134.8	154.22 ± 5.34	14.41%
0.9	134.8	175.89 ± 5.31	30.48%
1.0	134.8	196.99 ± 5.36	46.14%

Table 5.6: Comparison of the computational and experimental self-diffusion coefficients of the scaled 1-butyl-3-methylimidazolium methyl sulfate IL

$[C_4C_1im][CH_3SO_4]$ (6-31G* set) Scaling Factor	$D_{cation,lit}$ ($\times 10^{-12} m^2/s$)	$D_{cation,comp}$ ($\times 10^{-12} m^2/s$)	$D_{anion,lit}$ ($\times 10^{-12} m^2/s$)	$D_{anion,comp}$ ($\times 10^{-12} m^2/s$)	Total % Error
0.7	7	15	6	12	107.69%
0.8	7	4	6	2.6	−49.23%
0.9	7	0.4	6	0.27	−94.85%
1.0	7	0.17	6	0.096	−97.95%
$[C_4C_1im][CH_3SO_4]$ (6-31+G* set) Scaling Factor	$D_{cation,lit}$ ($\times 10^{-12} m^2/s$)	$D_{cation,comp}$ ($\times 10^{-12} m^2/s$)	$D_{anion,lit}$ ($\times 10^{-12} m^2/s$)	$D_{anion,comp}$ ($\times 10^{-12} m^2/s$)	Total % Error
0.7	7	13	6	9.3	71.54%
0.8	7	3.1	6	2.2	−59.23%
0.9	7	0.7	6	0.29	−92.38%
1.0	7	0.12	6	0.11	−98.23%

The self-diffusion coefficients calculated in Table 5.6 were obtained from the slope of an MSD versus time change plot. The points corresponding to ballistic motion (0-15 ns) were removed and the slope calculated from the remaining part of the plot. See Figure 5.11 for the original and truncated MSD versus τ plots for the cation from the unscaled 6-31+G*-parameterised 1-butyl-3-methylimidazolium methyl sulfate IL.

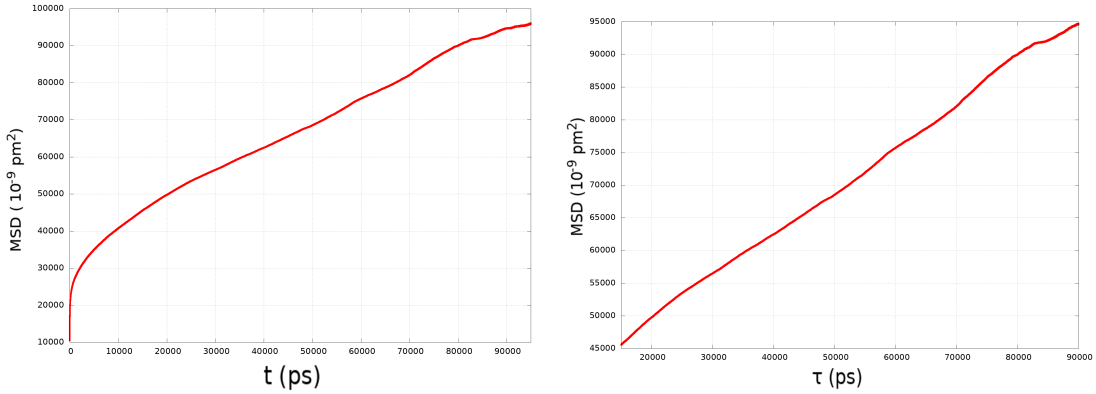


Figure 5.11: The original MSD versus τ plot to calculate the self-diffusion coefficient of 1-butyl-3-methylimidazolium (left), and the truncated MSD versus time change plot for 1-butyl-3-methylimidazolium (right).

As per Table 5.6, the self-diffusion coefficients of this IL were not represented well with unscaled charges, but a scaling factor between 0.7 and 0.8 drastically improves the results. Extrapolating the best scaling factor from a plot of self-diffusion coefficient versus scaling factor, a value of 0.74 was calculated for the 6-31+G*-parameterised set, and a scaling factor of 0.75 for the 6-31G*-parameterised set. For the 6-31+G*-parameterised 1-butyl-3-methylimidazolium methyl sulfate IL, a final self-diffusion coefficient percentage error of 13.85% was calculated, while for the other IL it was calculated as 20.00%. These percentage errors are significantly improved from that in Table 5.6. Since the 6-31+G*-parameterised IL produced the most accurately defined self-diffusion coefficient results with a scaling factor of 0.74 applied, it was selected for all further calculations.

The above results thus show that using a basis set with diffuse functions for the anion can improve IL parameters and that GAFF2 parameterisation produces satisfactory results, especially since the percentage errors produced for solvents in the published work by, for example, Acree, Derecskei and Keller are significantly larger than those produced in this work.[162,166–167].

6 Solvation of $(GlcNAc)_2Me_2$

6.1 Introduction to Theory

Various computational calculations exist which is useful to study the interaction between two molecules as well as the behaviour of the molecules when it is introduced into a particular environment. Some of the more popular post-production calculations that were used in this research project are radial distribution functions, interaction energy, hydrogen bond analysis and potential of mean force.

6.1.1 Radial Distribution Functions[146]

Radial distribution functions (RDFs) are used to represent, on average, the probability of locating an atom b a distance r away from another tagged atom c. As an example, the RDF, $g(r)$, between atoms of atom-type b from molecule B and atoms of atom-type c from molecule C in a spherical volume element can mathematically be defined as (equation 6.1)

$$g_{bc}(r) = \left(\frac{V}{N_b N_c}\right)\left(\sum_{i=1}^{N_b}\sum_{j=1}^{N_c}\delta(r - r(b_i c_j))\right) \tag{6.1}$$

Here, V is defined as the spherical shell's volume, N_b and N_c correlate to the total number of b and c atoms within the shell, δ is the Dirac delta function and $r_{b_i c_j}$ is the distance between atoms b_i and c_j. In Equation 6.2 is the integrated form of the radial distribution functions, which in the case of this project will output how many solvent components are coordinating around the $(GlcNAc)_2Me_2$ monomer at a specific distance.

$$n(r) = \int_{r_0}^{r} \rho g(r) 4\pi r^2 dr \tag{6.2}$$

where $n(r)$ is the coordination number of solvent components, r is the radial distance between two components, $g(r)$ is the radial distribution function and ρ is the density of the simulation box. RDFs were used in this project to describe the solvation structure of a $(GlcNAc)_2Me_2$ molecule in various solvent systems.

6.1.2 Interaction Energy

The interaction energy is calculated between two systems to describe how well they interact with each other when they are introduced to the same surroundings. The total interaction energy can be obtained from the non-bonded part of the potential energy function and gets broken down into electrostatic- and van der Waals energies. The more negative the interaction energy is, the more favourable the interaction between the two systems are. The electrostatic energy component of the IE is calculated in AMBER using the shift function $E_{elec} = k\frac{q_i q_j}{r_{ij}}\left(1 - \frac{{r_{ij}}^2}{{r_{cut}}^2}\right)$, where q_i is the charge on particle i, r_{ij} is the radial distance between particles i and j, and r_{cut} is the cut-off distance chosen for IE calculations.

6.1.3 Hydrogen Bond Analysis[168]

A hydrogen bond can be seen as an essential non-covalent structural force (primarily electrostatic) that involves a hydrogen atom being located between two atoms that have a high affinity for electron

pairs, that are either from the same molecule or two different ones. In a hydrogen bond, one of the atoms "shares" a hydrogen atom while the remaining atom "accepts" the sharing of the hydrogen atom. A hydrogen bond is represented by the expression D-H$\cdots$A, where A refers to the hydrogen atom "acceptor" and D refers to the hydrogen atom "donor". In classical molecular mechanics, however, lone pairs are not used, and the negative charge on the acceptor atom and the positive charge on the donor atom are the only factors that drive hydrogen bond formation.

A typical hydrogen bond analysis outputs the frequency of intermolecular- and intramolecular hydrogen bond formation, that is, over a given number of frames it tells one how often a specific hydrogen bond occurs. The lower this frequency is, the less likely that hydrogen bond will form. In the case of a $(GlcNAc)_2Me_2$ molecule, low intramolecular hydrogen bond formation and high intermolecular hydrogen bond formation is desired.

6.2 Computational Details

The model for chitin. The carbohydrate model for chitin, $(GlcNAc)_2Me_2$, was built with the GLYCAM builder in the `tleap` utility and parameterised through the GLYCAM06j force field. Since Kirschner, Ramraj, Singh and Michel have shown the GLYCAM06 force field to be optimised for sugars and produces accurate results relative to high-end QM computations and experimentation, no further parameterisation or parameter fitting was done.[129,168–170]

Solvating $(GlcNAc)_2Me_2$. Simulation boxes consisting of a single centred $(GlcNAc)_2Me_2$ molecule solvated by 1024 pure solvent molecules or 1020 OES molecules (since using 1024 molecules would have implied partial co-solvent and ionic liquid molecules are needed, which is impossible) were constructed with `Packmol`. The organic electrolyte solution simulation boxes were constructed in ratios of 2:8 and 8:2 for 1-butyl-3-methylimidazolium acetate:co-solvent. These ratios were selected as Bioni and Gale have indicated that they can produce favourable dissolution results for cellulose.[97–98]

Interaction energy. Interaction energy calculations were carried out between the solvent molecules/ions and the solute using the independently written utility, `cpptraj`, as distributed with `AmberTools`,[171] centring the solute in the simulation box to avoid boundary effects. Similarly, it was also calculated between the anion and cation for all OESs. Since a classic Coulomb sum was used, an electrostatic energy cut-off distance of 16.0 Å was selected, while for the van der Waals energy it was 10.0 Å.

Radial distribution functions. Solute-solvent RDFs were calculated in each solvent system using `cpptraj`. The RDF calculations were done in two-fold - between the centre of mass of the solvent components and the centre of mass of the solute, as well as involving the solvent/solute atoms highlighted in Figure 6.1. These atoms were chosen due to its ability to act as a hydrogen bond acceptor or hydrogen bond donor. For all the RDF calculations, a maximum radius of 20.0 Å was selected with a bin size of 0.10 Å.

Hydrogen bond analysis. Hydrogen bond analyses were performed using `cpptraj`, where a D-H$\cdots$A cut-off angle of 120° was selected, and a D$\cdots$A bond cut-off distance of 3.0 Å. The bond cut-off distance is the default value from the AMBER manual, while the bond angle cut-off was influenced by the work by Hunt *et al.*, showing that bond angles of 120° and larger are required for accurate hydrogen bond analyses.[172–173] All OH/NH($(GlcNAc)_2Me_2$)$\cdots$O(solvent) and OH/NH($(GlcNAc)_2Me_2$)$\cdots$O/N($(GlcNAc)_2Me_2$) intermolecular- and intramolecular hydrogen bonds were studied. See Figure 6.1 for the atoms used for this calculation.

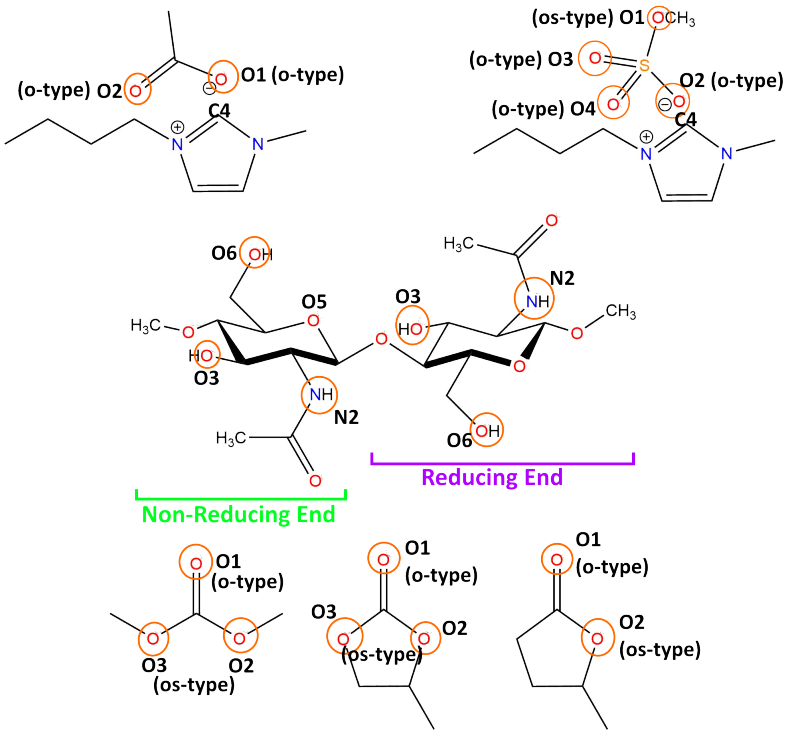


Figure 6.1: The list of atoms from the solvent components and chitin monomer used in the radial distribution function calculations and hydrogen bond analyses.

6.3 Solvation Results

The model for chitin, 4-methyl-β-D-N-acetylglucosamine-(1→4')-1'-methyl-β-D-N'-acetylglucosamine was constructed in `tleap` with the GLYCAM builder, and parameterised with the GLYCAM06j force field. The minimised structure of $(GlcNAc)_2Me_2$ is shown in Figure 6.2.

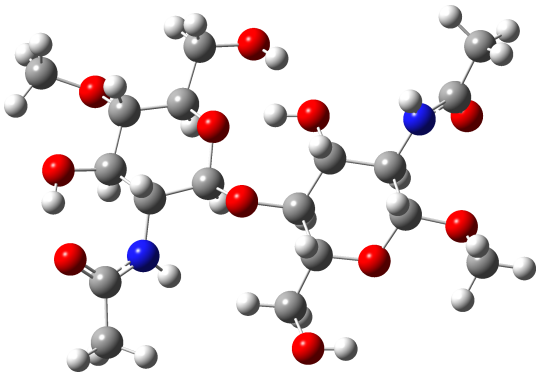


Figure 6.2: The structure of a minimised $(GlcNAc)_2Me_2$ molecule.

Using `Packmol`, a single $(GlcNAc)_2Me_2$ molecule was placed in the centre of simulation boxes containing 1020 organic electrolyte solution molecules/ion pairs or 1024 pure solvent molecules/ion

pairs. `Tleap` was then used to create parameter and topology files for the simulation boxes. An example of such a simulation box is shown in Figure 6.3.

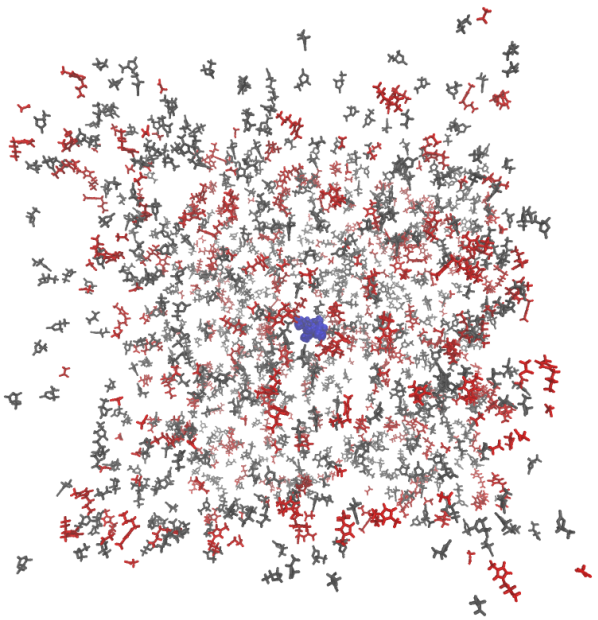


Figure 6.3: A rough initial simulation box built in `Packmol` containing one centred $(GlcNAc)_2Me_2$ (blue) molecule surrounded by 816 propylene carbonate (grey) molecules and 204 1-butyl-3-methylimidazolium acetate (red) ion pairs.

All simulation boxes were minimised, heated, equilibrated and sent through a production phase. From the production phase, radial distribution functions were determined and the interaction energy and hydrogen bonding analysed.

6.3.1 Interaction Energy Results

The interaction energy can only be measured on a purely computational level and consists of an electrostatic- and van der Waals energy contribution. The more negative the interaction energy is between two systems, the more favourably they interact. A time series of the non-bonded energies for some solvent systems is shown in Figure 6.4, while the total solvation energy (this only includes solvent-solute non-bonded interactions) is shown in Table 6.1. The general shape of the non-bonded energy profiles from Figure 6.4 is similar to those in the published work by Pishehvarz *et al.* and Zhang *et al.*, suggesting a sufficient calculation of the average non-bonded energies.[174–175]

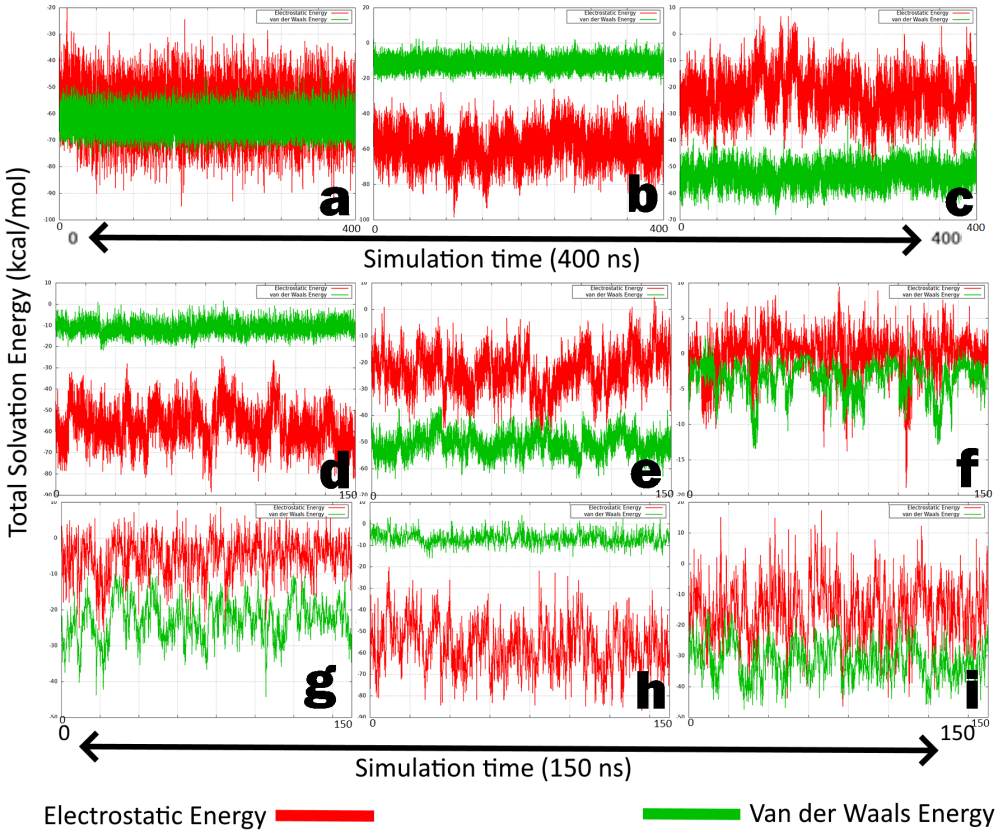


Figure 6.4: Time series of non-bonded energies between the $(GlcNAc)_2Me_2$ monomer and a) pure propylene carbonate, b) pure acetate, c) pure 1-butyl-3-methylimidazolium, d) acetate from 8:2 IL:γV, e) $[C_4C_1im]^+$ from 8:2 IL:γV, f) γ-valerolactone from 8:2 IL:γV, g) dimethyl carbonate from 2:8 IL:DC, h) acetate from 2:8 IL:DC and i) $[C_4C_1im]^+$ from 2:8 IL:DC.

Table 6.1: The total solvation energy between a $(GlcNAc)_2Me_2$ molecule and various solvent components

Solvent	Total Solvation Energy (in kcal/mol)	Swelling Capacity	Molar Swelling Capacity
Dimethyl Carbonate (DC)	-98.17 ± 7.33	212%	2.35
$\gamma -$ Valerolactone (γV)	-116.56 ± 7.96	229%	2.29
Propylene Carbonate (PC)	-122.34 ± 9.10	237%	2.32
$[C_4C_1im][CH_3SO_4]$	-137.98 ± 5.61		
$[C_4C_1im][CH_3COO]$(IL)	-147.26 ± 6.01	500%	2.52
2:8 IL:DC	-138.81 ± 7.41	261%	2.33
2:8 IL:γV	-139.70 ± 8.70	358%	2.99
2:8 IL:PC	-142.08 ± 8.87	357%	2.98
8:2 IL:DC	-145.49 ± 5.79	398%	2.25
8:2 IL:γV	-146.65 ± 5.32	458%	2.56
8:2 IL:PC	-148.94 ± 5.86	467%	2.61

Table 6.1 shows that a $(GlcNAc)_2Me_2$ molecule interacts most favourably with 8:2 IL:PC molecules/ion pairs, and least favourably with pure dimethyl carbonate molecules. Specifically, Table 6.1 shows that the 8:2 IL:co-solvent OESs interact more favourably with a $(GlcNAc)_2Me_2$ molecule than the 2:8 IL:co-solvent OESs do. Note with Table 6.2, that the order of total solvation energy for the pure solvents agree with the order of average atomic partial charges on the oxygen atoms.

Table 6.2: The average partial charge for the oxygen atoms in the pure solvents

Solvent	Average Partial Charge for the Oxygen Atoms	Dipole Moment (Debye)
DC	−0.4941	0.3907
γV	−0.5121	5.0153
PC	−0.5130	6.0835
$[C_4C_1im][CH_3SO_4]$	−0.6131	
IL	−0.9059	

By using Equation 6.3 and Figure 6.5, one can verify if the solvation process can be described by ideal mixing. This is done by calculating the ideal solvation energy of the mixtures using the total solvation energy from the pure solvents, then making a plot of total solvation energy versus ideal total solvation energy for the mixtures. The larger the R^2 value is from the plot, the better the solvation process is described by ideal mixing.

$$E_{\text{solv}}^{\text{ideal}} = \chi_1 E_{\text{solv},1} + \chi_2 E_{\text{solv},2} \tag{6.3}$$

where $E_{\text{solv}}^{\text{ideal}}$ is the ideal total solvation energy for a specific mixture, χ_1 is the mole fraction of the pure co-solvent, $E_{\text{solv, 1}}$ is the total solvation energy of the pure co-solvent, χ_2 is the mole fraction of the pure IL, and $E_{\text{solv, 2}}$ is the total solvation energy of the pure IL.[176]

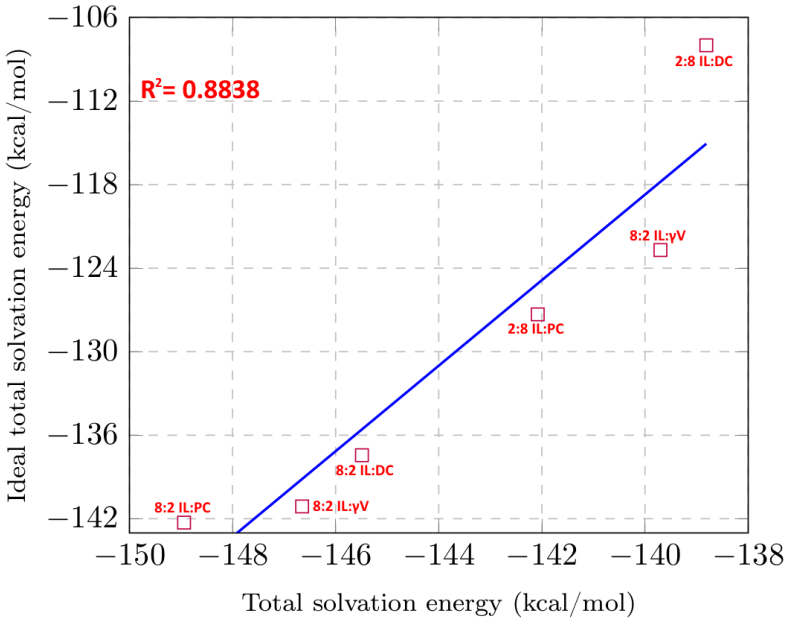


Figure 6.5: Total solvation energy versus ideal total solvation energy plot for the 8:2 and 2:8 IL:co-solvent OESs.

Figure 6.5 shows a good correlation between the total solvation- and ideal solvation energies. From Table 6.1, the solvent absorption in chitin (the experimental swelling capacity per mass) and total solvation energy is in the order of

pure IL > 8:2 IL:PC > 8:2 IL:γV > 8:2 IL:DC > 2:8 IL:PC > 2:8 IL:γV > 2:8 IL:DC > pure PC > pure γV > pure DC (exp.)

8:2 IL:PC > pure IL > 8:2 IL:γV > 8:2 IL:DC > 2:8 IL:PC > 2:8 IL:γV > 2:8 IL:DC > pure PC > pure γV > pure DC (MD)

The relationship between the two can be analysed by making a plot of the swelling capacity versus total solvation energy. See Figure 6.6. While this figure is constructed as a linear plot between swelling capacity and solvation energy, it is plausible that acceleration during the process of swelling can lead to non-linearity. As such, non-linear fitting can also be explored in future work to better express this relationship. However, for the purpose of providing a basic trend between solvation energy and swelling, as is the intention here, a linear fit is sufficient.

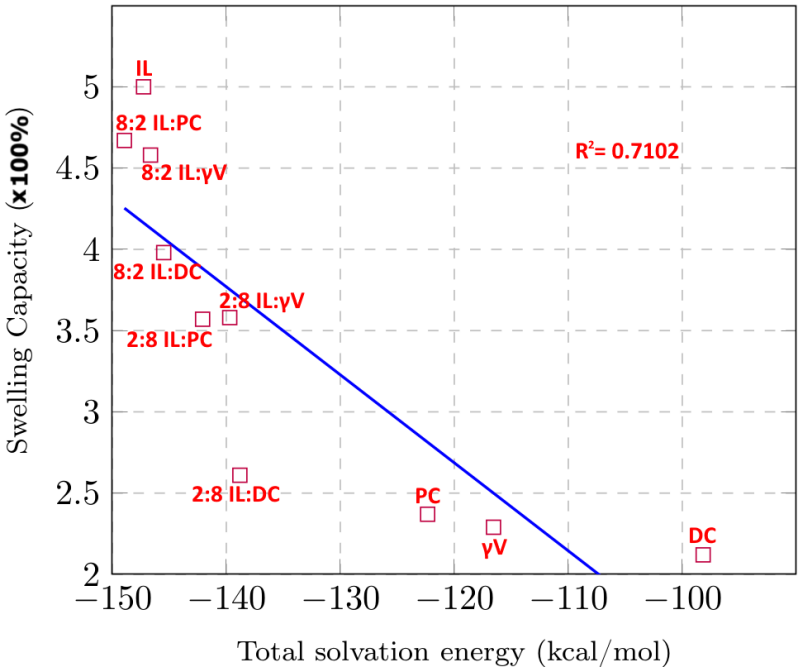


Figure 6.6: Total solvation energy versus swelling capacity (per mass) plot.

From Table 6.1 one can see that a greater quantity of pure 1-butyl-3-methylimidazolium acetate is absorbed compared to all the other solvents; while for pure dimethyl carbonate, the least amount of solvent is absorbed. While the above comparison between swelling capacity and interaction energy is a mass to number of moles comparison, a more appropriate comparison might be with the molar swelling capacity (see Table 6.1). The total solvation energy results, unfortunately, do not agree with the molar swelling capacity results, and this could be either because of an unrepresentative ensemble average or that the interaction with $(GlcNAc)_2Me_2$ is not representative of the absorption into chitin. Moreover, it could be wrong to assume that if n solvent X molecules are absorbed by chitin it will produce a more favourable interaction compared to when only $n-1$ solvent Y molecules are absorbed. This is what the results from the molar swelling capacity infer.

Aside from the 8:2 IL:PC and the 2:8 IL:PC OESs, the total solvation energy results agree with the (mass) swelling capacity result trend. The two OESs being outliers can be as a result of an over-represented average PC-$(GlcNAc)_2Me_2$ solvation energy. One can see that in Table 6.1, these

two OESs give the largest standard deviation for their respective mixture sets. Figure 6.7 shows a time series of the non-bonded energies for PC-$(GlcNAc)_2Me_2$ from the 8:2 IL:PC OES.

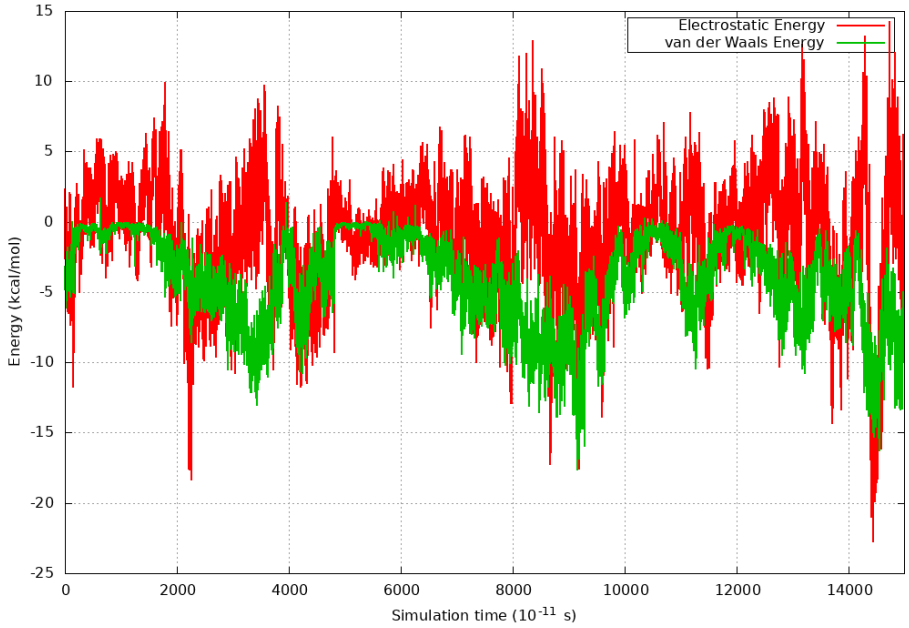


Figure 6.7: The electrostatic- and van der Waals energy between propylene carbonate and the $(GlcNAc)_2Me_2$ molecule in the 8:2 1-butyl-3-methylimdiazolium acetate:propylene carbonate OES.

The average van der Waals interaction energy is -4.23±3.25 kcal/mol. However, as is evident from the large standard deviation (and is also clearly visible in the figure), the value fluctuates wildly. This fluctuation is as a result of slow dynamics, where the true isothermal-isobaric ensemble average in these high viscosity systems can only be represented with simulation times at "reasonable" lengths. At that point, ergodicity would be reached. To improve the sampling, the following can be done

1). Run longer simulations
2). Run multiple simulations with different initial configurations
3). Increase the temperature

Although the 9:1 IL:PC results were not focused on in this project, it was an OES initially considered. The production phase of the solvation of a $(GlcNAc)_2Me_2$ molecule with 9:1 IL:PC was run for 400 ns, yet no improvements on the average non-bonded energies were noticed, suggesting that method 1 will not be sufficient to improve sampling. See Figure 6.8. Moreover, at the time of running the above simulations, the intention was to run all simulations and perform all experimentation at 298.15 K since literature heat of vaporisation and density values were only available for all the pure solvents at this temperature. As a result of this, method 3 was not employed either. Instead what was later investigated was method 2, in which multiple simulation boxes (see Figure 6.9) were considered for the 8:2 IL:PC OES.

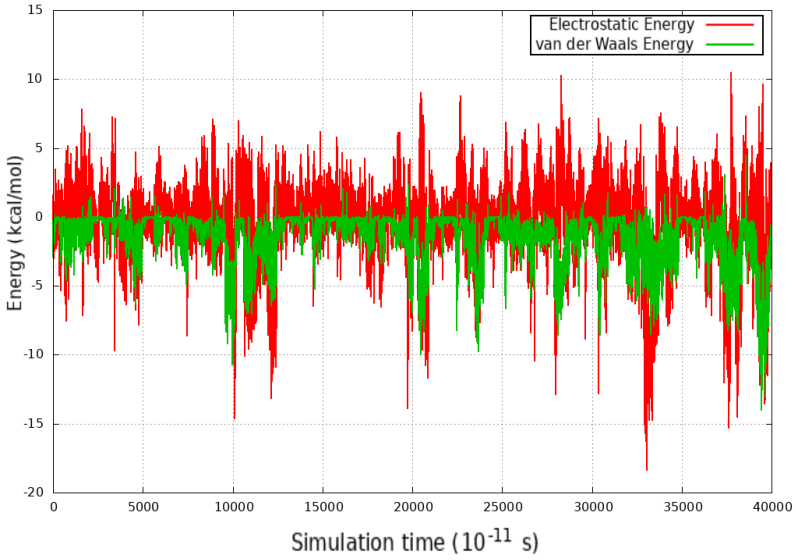


Figure 6.8: The electrostatic- and van der Waals energy between propylene carbonate and the $(GlcNAc)_2Me_2$ molecule in the 9:1 1-butyl-3-methylimdiazolium acetate:propylene carbonate OES.

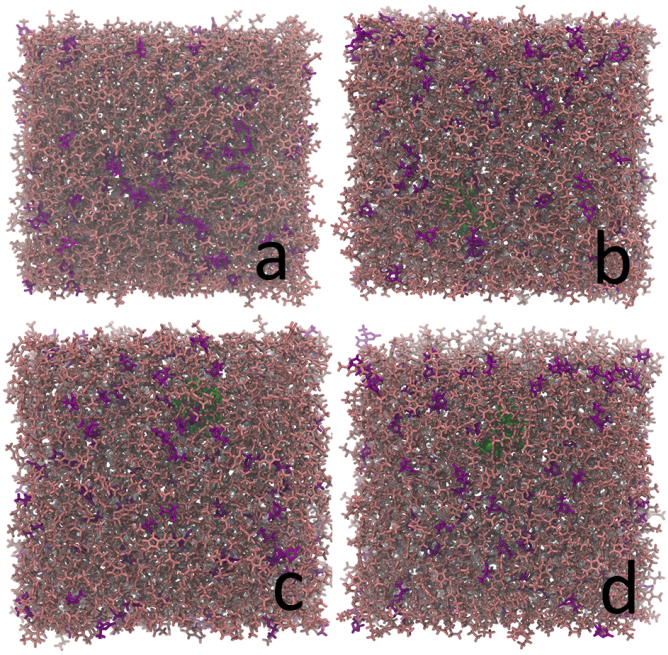


Figure 6.9: The equilibrated simulation boxes used for solvating a $(GlcNAc)_2Me_2$ monomer with 8:2 IL:PC. Here the IL ion pairs are coloured pink, while the propylene carbonate molecules are represented in purple and the carbohydrate as green.

Each of the four simulation boxes from Figure 6.9 was simulated using the same procedure as outlined in Chapter 6.2, meaning that each simulation box had a production run of 150 ns carried out.

The average total solvation energy of each system was calculated, with the time series for the average non-bonded energies shown in Figure 6.10.

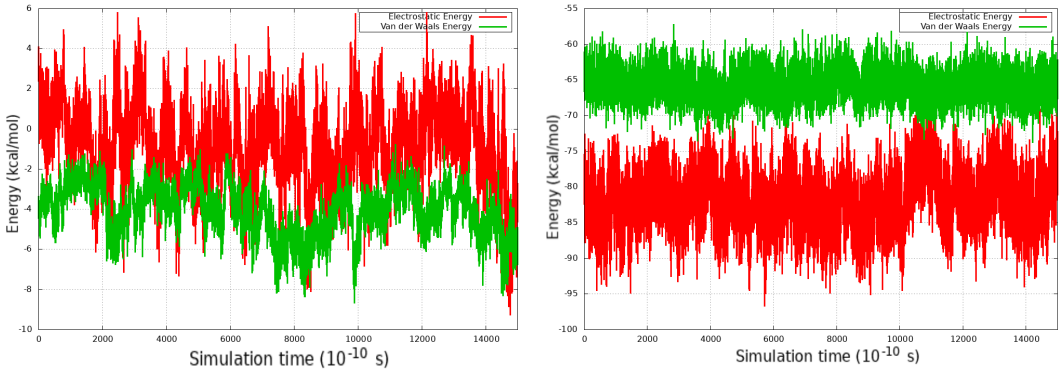


Figure 6.10: A time series of the average non-bonded energies for the interaction between a $(GlcNAc)_2Me_2$ monomer with propylene carbonate molecules (left) and all solvent components (right) from the four 8:2 IL:PC simulation boxes.

The plot appears to be defined better (as compared to Figure 6.7), producing an enhanced average for the energies. The van der Waals interaction energy is now -4.11±1.25 kcal/mol, which is a lower average value and a significantly lower standard deviation value as compared to that from Figure 6.7, implying less fluctuation. Additionally, the total solvation energy for the PC-$(GlcNAc)_2Me_2$ interaction is now -4.93±2.42 kcal/mol, which is improved from the -5.26±5.14 kcal/mol total solvation energy from Figure 6.7. Considering the average total solvation energy of the 8:2 IL:PC OES (see Figure 6.10), a value of -147.60±4.72 kcal/mol is calculated, which improves the 8:2 IL:PC OES result from Table 6.1, showing that method 2 does give better sampling and results that are more representative of ergodicity. Moreover, the R^2 value from Figure 6.5 improves when including this new total solvation energy, to 0.9129.

Averaging over more simulation boxes and increasing the simulation time might improve on the average total solvation energy (including the standard deviation), but this also significantly increases the total calculation time. More specifically, using this method is impractical in this project for all the OESs (2:8 and 8:2 IL:co-solvent), as such, was not implemented further.

Table 6.3 shows the interaction energy between 1-butyl-3-methylimidazolium and acetate from the pure IL and various OESs.

Table 6.3: The total normalised (to 1020 ion pairs) interaction energy between acetate and 1-butyl-3-methylimidazolium from the solvation of a $(GlcNAc)_2Me_2$ molecule by various solvent systems

Solvent	Total Cation-Anion Interaction Energy (kcal/mol)
$[C_4C_1im][CH_3COO]$ (IL)	−237671.95
2:8 IL:PC	−87185
2:8 IL:γV	−92180
2:8 IL:DC	−101720
8:2 IL:γV	−213343.75
8:2 IL:PC	−215613.75
8:2 IL:DC	−216023.75

Table 6.3 suggests that the more co-solvent molecules there are relative to IL ion pairs, the lower the interaction energy becomes between the cation and anion. A further study on this can conclude if a lower interaction energy means a lower density and -viscosity as well.

6.3.2 RDF Results

The radial distribution functions were first calculated between the glycosidic oxygen atom from $(GlcNAc)_2Me_2$ (approximated as the centre of mass) with the carbonyl (carbon) atoms from the co-solvents, the carbonyl (carbon) atom from acetate, the sulfur atom from methyl sulfate and the -N-C-N (C4) carbon atom from 1-butyl-3-methylimidazolium. These RDFs are shown in figures 6.12, 6.14 and 6.16, and also includes the integrated form.

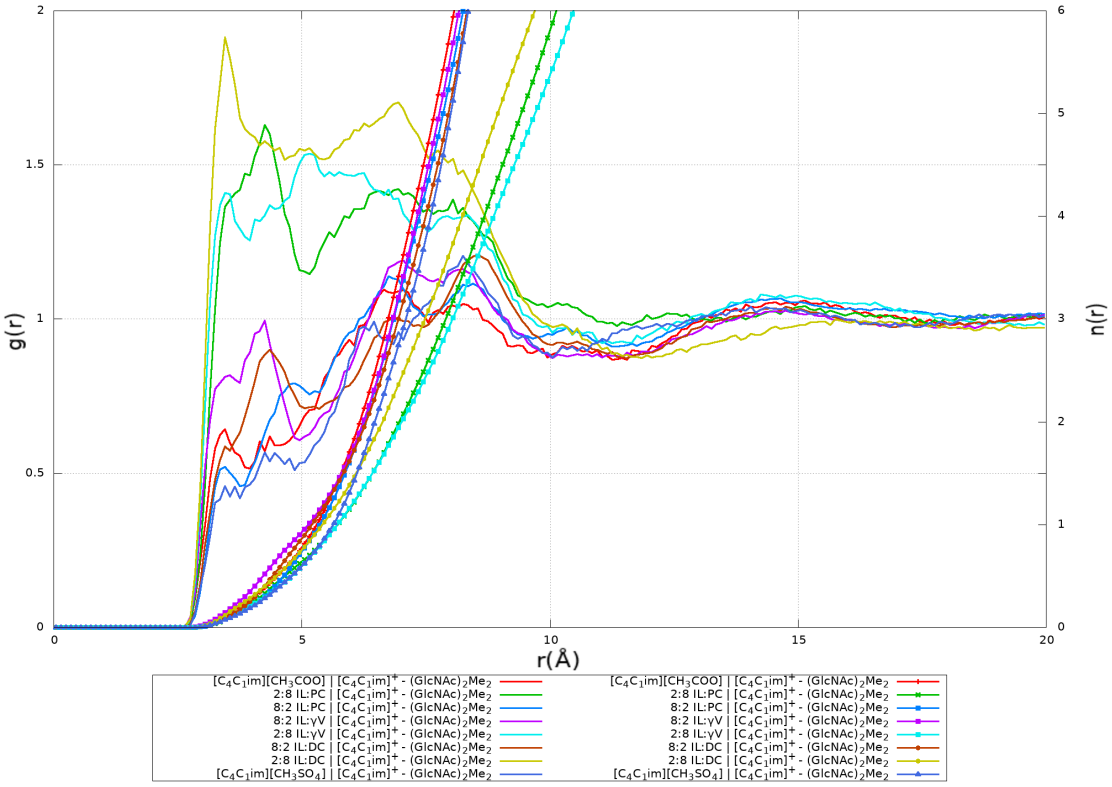


Figure 6.12: Radial distribution functions and integrated RDFs between C4 carbon atom from 1-butyl-3-methylimidazolium and the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer. The shaped-lines (or those which are not purely a solid representation) show the integration of RDFs.

For an RDF plot, a local minimum (present after a local maximum) generally represents a coordination shell around the solute, that is, it represents a shell of solvent coordinating around the solute within a volume sphere of $\frac{4}{3}\pi r^3$ Å^3 where r is the distance corresponding to the local minimum. When integrating the RDF up to this local minimum, the number of solvent components coordinating around the solute (within a volume sphere of $\frac{4}{3}\pi r^3$ Å^3) is outputted. The rough shape of the RDFs in Figure 6.12,6.14 and 6.16 is due to radial averaging on a non-spherical system, however, it

still shows some characteristic features. For example, in Figure 6.12 most of the RDFs have sharp initial peaks around 3-4 Å, followed soon after with a local minimum. Moreover, a secondary local maximum appears around 8 Å (for most of the RDFs), followed by a local minimum at approximately 11 Å. With a small sphere of volume selected, there is a clear difference between the RDFs. The 2:8 IL:co-solvent OESs have sharp and high peaks, followed by the 8:2 IL:co-solvent OESs and then the pure ionic liquids. This suggests that there is a higher probability of locating a C4 carbon atom (from the cation) from the 2:8 IL:co-solvent OESs coordinating around the solute than from the 8:2 IL:co-solvent OESs and the pure ionic liquids. The reason for this is not known but it might be due to these OESs being less viscous, as such, the cations can coordinate easier and more regularly around the solute. For both the pure ionic liquids, the first well-established local maximum occurs at around 8 Å, followed by a local minimum at 11 Å. At the local maximum, the RDF value is higher for the methyl sulfate-based IL (1.20) than the acetate-based IL (1.05), suggesting that there is a higher probability of finding a cationic C4 carbon atom coordinating around the glycosidic oxygen atom in a volume sphere of $\frac{4}{3}\pi(8)^3$ Å^3 than for the acetate-based IL. Integrating the RDFs up to 11 Å (average location of both local minima) outputs coordination numbers of 15.6 and 16.2 for $[C_4C_1im][CH_3SO_4]$ and $[C_4C_1im][CH_3COO]$, respectively. This implies there is slightly more cationic C4 carbon atoms from $[C_4C_1im][CH_3COO]$ coordinating (on average) around the glycosidic oxygen atom in the coordination shell than from $[C_4C_1im][CH_3SO_4]$. This is perhaps due to a more favourable interaction with the monomer as the acetate-based IL has a better solvation energy with the $(GlcNAc)_2Me_2$ monomer than compared to the methyl sulfate-based IL.

When considering the local maxima of the 8:2 IL:co-solvent OESs, the first well-defined maximum (for all three together) occurs close to 8 Å followed by a minimum at 10-11 Å. At the local maximum, the 8:2 IL:DC OES has a higher RDF value than the 8:2 IL:γV OES, followed by the 8:2 IL:PC mixture. This suggests there is a higher probability of locating a cationic C4 carbon atom from the 8:2 IL:DC OES within the volume sphere of $\frac{4}{3}\pi(8)^3$ Å^3 than for the other two OESs. At 11 Å (the general location of the local minima), the integration of RDFs outputs coordination numbers of 15.77, 15.59 and 15.86 for the 8:2 IL:DC, 8:2 IL:γV and 8:2 IL:PC OESs, respectively. This suggests the most number of cationic C4 carbon atoms coordinating around the monomer in this solvation shell (of radius 11 Å) comes from the 8:2 IL:PC OES, which matches the result in Chapter 6.3.1 in which it has the most favourable solvation energy with the $(GlcNAc)_2Me_2$ monomer.

For the 2:8 IL:co-solvent mixtures, it is the 2:8 IL:DC OES that has the highest RDF peak for the local maximum (around 7-8 Å), suggesting that there is a higher probability of finding a cationic C4 carbon atom from this OES coordinating around the glycosidic oxygen atom within a volume sphere of $\frac{4}{3}\pi(8)^3$ Å^3 than compared to the other OESs. The integration of 2:8 IL:co-solvent RDFs up to the local minima close to 11 Å outputs coordination numbers of 8.29, 7.13 and 7.84, respectively for the 2:8 IL:DC, 2:8 IL:γV and 2:8 IL:PC OESs. This implies there is more cationic C4 carbon atoms from the 2:8 IL:DC OES coordinating around the solute in the solvation shell (of 11 Å radius) than from the 2:8 IL:γV and 2:8 IL:PC OESs. Overall, Figure 6.12 shows that the more co-solvent is added to the IL, the fewer cations will be found (on average) coordinating around the solute, however, these cations will more regularly coordinate around it (since it has higher RDFs).

For pure $[C_4C_1im][CH_3COO]$, 2 cations are found (on average) coordinating around the $(GlcNAc)_2Me_2$ monomer within volume sphere of $\frac{4}{3}\pi 6^3$ Å^3. This is supported by Figure 6.13, in which (on average)

only two C4 carbon atoms from the cation are found solvating close to the glycosidic oxygen atom.

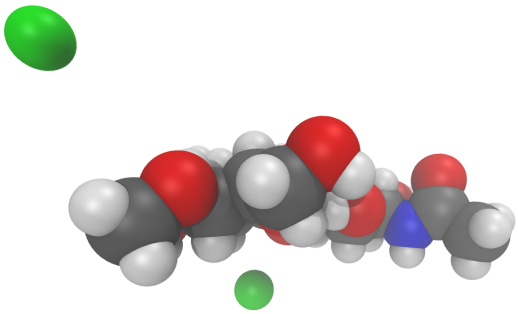


Figure 6.13: A snapshot of the 1-butyl-3-methylimdazolium C4 carbon atoms (green spheres) solvating the $(GlcNAc)_2Me_2$ monomer within 6 Å (away from it) for the $[C_4C_1im][CH_3COO]$ IL.

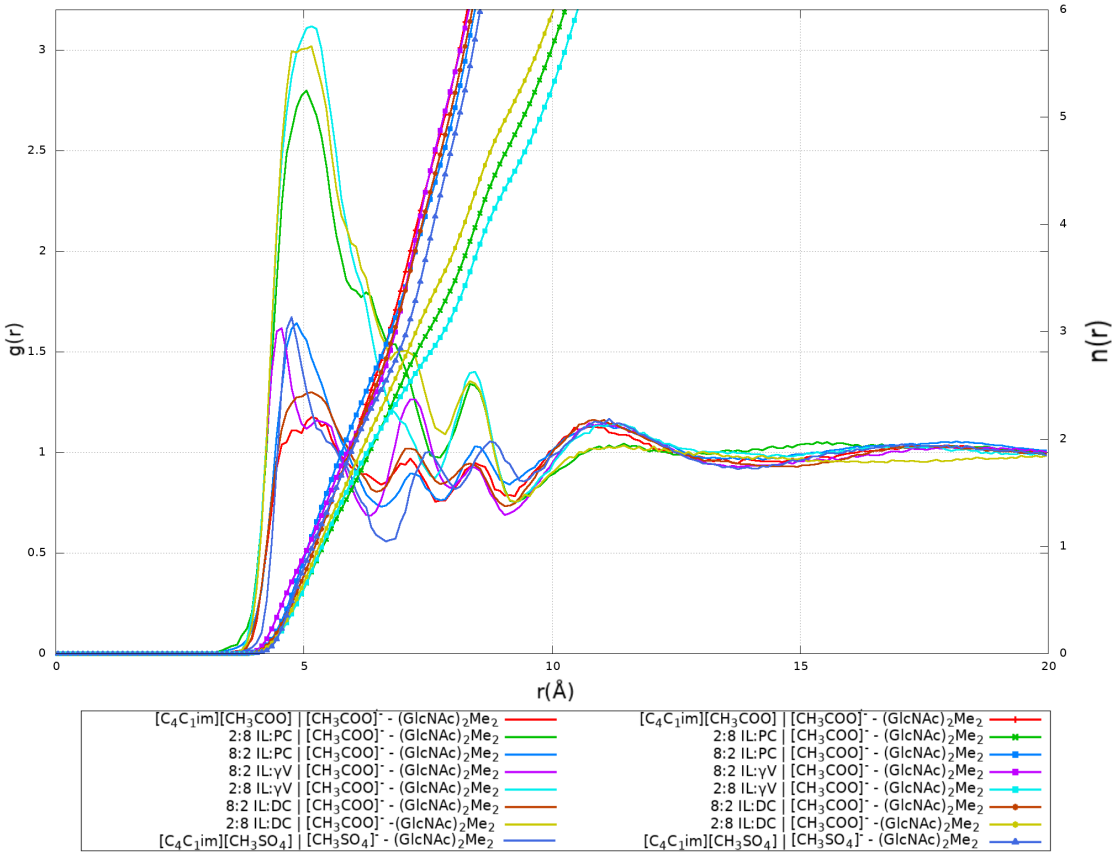


Figure 6.14: Radial distribution functions and integrated RDFs between the carbonyl (carbon) atom from acetate or sulfur atom from methyl sulfate and the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer. The shaped-lines (or those which are not purely a solid representation) show the integration of RDFs.

All RDFs in Figure 6.14 have large initial peaks (local maxima) close to 5 Å, followed by local minima between 6-8 Å. For the pure ionic liquids, it is the methyl sulfate-based IL that has the larger local maximum at (approximately) 5 Å, suggesting there is a higher probability of finding sulfur atoms from the methyl sulfate anion coordinating in the volume sphere of $\frac{4}{3}\pi 5^3$ than carbon atoms from the acetate anion. When integrating the RDFs up to the local minima, coordination numbers of 2.46 and 2.74 are produced for the $[C_4C_1im][CH_3SO_4]$ and $[C_4C_1im][CH_3COO]$ ionic liquids, respectively. This means more acetate carbonyl carbon atoms will coordinate around the glycosidic oxygen atom from $(GlcNAc)_2Me_2$ than sulfur atoms (form methyl sulfate) in the coordination shell. This matches the results from Figure 6.12 and the solvation energy results from Chapter 6.3.1.

For the 8:2 IL:co-solvent OESs, it is the 8:2 IL:PC mixture that shows the biggest RDF peak at its global maximum, followed by the 8:2 IL:γV and then 8:2 IL:DC OESs. This suggests that there is a higher probability that acetate carbonyl (carbon) atoms from 8:2 IL:PC will be found coordinating around the glycosidic oxygen atom from $(GlcNAc)_2Me_2$ in a volume sphere of $\frac{4}{3}\pi 5^3$ Å^3 than from the other 8:2 IL:co-solvent mixtures, matching the solvation energy results. When considering the integration of the RDFs up to the first local minimum (average location chosen as 6.5 Å) for the 2:8 IL:co-solvent OESs, coordination numbers of 2.77, 2.55 and 2.62 are produced for the 2:8 IL:PC, 2:8 IL:γV and 2:8 IL:DC mixtures, respectively. This implies that (on average) more acetate carbonyl (carbon) atoms from the 2:8 IL:PC OES will coordinate around the glycosidic oxygen atom than from the 2:8 IL:DC and 2:8 IL:γV OESs, supporting the solvation energy results. Adding more co-solvent to the ionic liquid thus seems to lower the number of anions that will coordinate around the $(GlcNAc)_2Me_2$ monomer, but these anions will coordinate more frequently, matching the results from Figure 6.12.

For pure $[C_4C_1im][CH_3COO]$, approximately 2 (carbonyl) carbon atoms from the acetate anion are found (on average) coordinating around the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer within a volume sphere of $\frac{4}{3}\pi 6^3$ Å^3. This is supported by Figure 6.15, in which (on average) only two carbonyl (carbon) atoms from the anion are found solvating close to the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer.

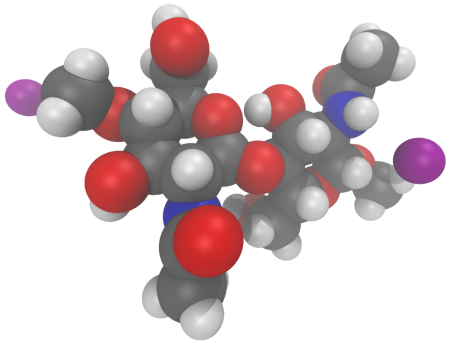


Figure 6.15: A snapshot of the acetate carbonyl (carbon) atoms (purple spheres) solvating the $(GlcNAc)_2Me_2$ monomer within 6 Å (away from it) for the $[C_4C_1im][CH_3COO]$ IL.

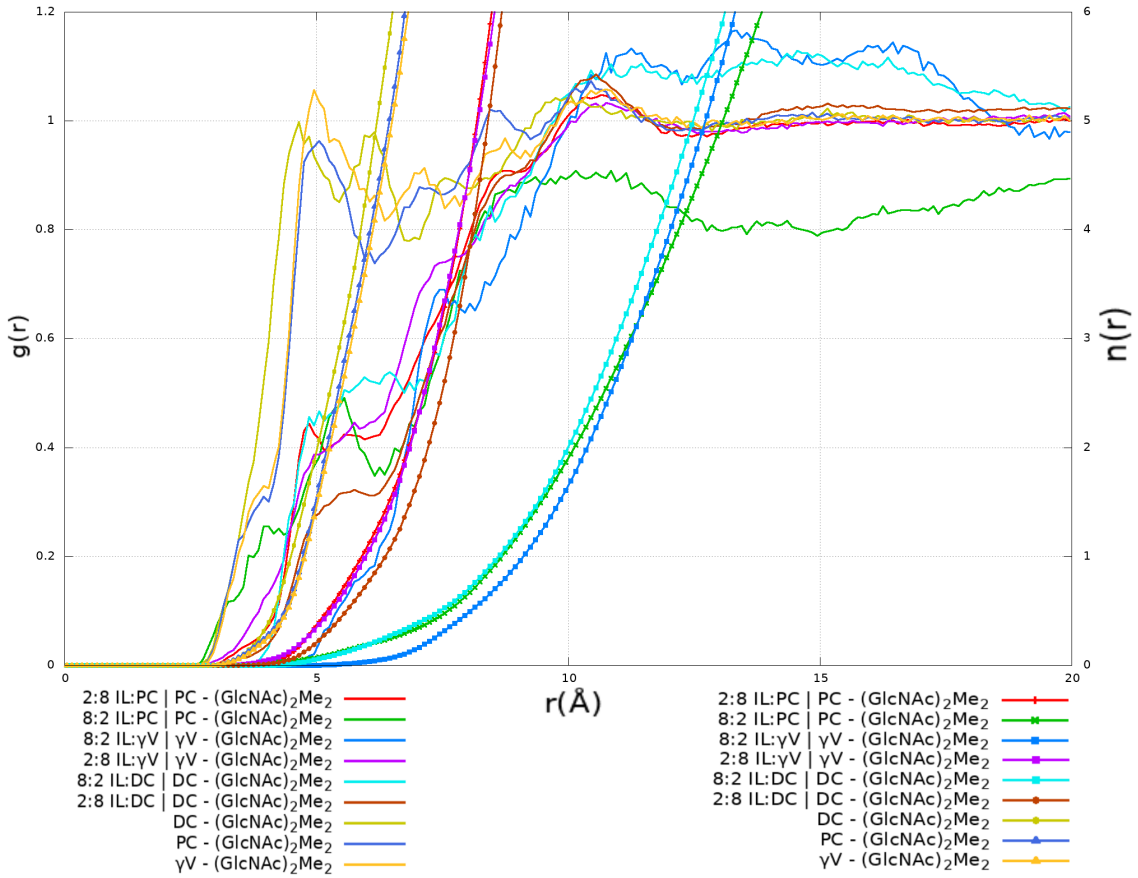


Figure 6.16: Radial distribution functions and integrated RDFs between the carbonyl (carbon) atoms from the co-solvents and the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer. The shaped-lines (or those which are not purely a solid representation) show the integration of RDFs.

For the RDFs in Figure 6.16, no well-defined local minima and -maxima are easily observed, but there appears to be a local maximum at 10 Å and a local minimum between 12-14 Å for the pure co-solvents and 2:8 IL:co-solvent OESs. For the pure co-solvents, the local maxima at (approximately) 10 Åare very close, implying that there is an almost equal probability of locating a glycosidic oxygen and carbonyl (carbon) atom pair within a volume sphere of $\frac{4}{3}\pi 10^3$ Å^3 for the propylene carbonate, γ-valerolactone and dimethyl carbonate simulation boxes. Integrating the RDFs up to 12 Å (average location of the local minima), coordination numbers of 50.8, 45.3 and 50.2 are produced for the propylene carbonate, γ-valerolactone and dimethyl carbonate solvent systems, respectively. This means that more carbonyl (carbon) atoms from propylene carbonate will be found coordinating around the glycosidic oxygen atom from $(GlcNAc)_2Me_2$ in the coordination shell (of radius 12 Å), than from dimethyl carbonate and γ-valerolactone. This can possibly be due to propylene carbonate interacting the most favourably with the $(GlcNAc)_2Me_2$ monomer according to the solvation energy results.

For the 2:8 IL:co-solvent mixtures, the RDF value for the 2:8 IL:DC OES at the 10 Å local maximum is the highest, followed by the 2:8 IL:PC OES and then the 2:8 IL:γV mixture. Integrating the RDFs up to the local minimum (close to 12 Å) produces coordination numbers of 28.5, 28.9 and 26.9 for the 2:8 IL:DC, 2:8 IL:PC and 2:8 IL:γV OESs, respectively. This suggests that more carbonyl (carbon) carbon atoms from propylene carbonate will coordinate around the glycosidic oxygen atom in the coordination shell of volume $\frac{4}{3}\pi 12^3$ Å^3 than from dimethyl carbonate or γ-valerolactone, supporting the solvation energy results from Chapter 6.3.1. The 8:2 IL:co-solvent OESs have lower coordination numbers than the pure co-solvents and 2:8 IL:co-solvents which shows that the fewer co-solvents are present, the less of them can solvate the monomer.

For pure propylene carbonate, approximately 4 (carbonyl) carbon atoms are found (on average) coordinating around the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer within a volume sphere of $\frac{4}{3}\pi 6^3$Å^3. This is supported by Figure 6.17, in which (on average) only four carbonyl (carbon) atoms from the anion are found solvating close to the glycosidic oxygen atom from the $(GlcNAc)_2Me_2$ monomer.

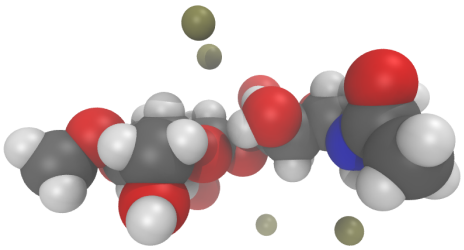


Figure 6.17: A snapshot of the propylene carbonate carbonyl (carbon) atoms (brown spheres) solvating the $(GlcNAc)_2Me_2$ monomer within 6 Å (away from it) for the pure propylene carbonate simulation box.

For the same volume sphere of $\frac{4}{3}\pi 6^3$ Å^3, the same number of propylene carbonate molecules are found coordinating around the $(GlcNAc)_2Me_2$ monomer than $[C_4C_1im][CH_3COO]$ ion pairs. In Figure 6.18 is shown the ratio of propylene carbonate molecules to IL ion pairs solvating the monomer at various distances. After approximately 8 Å, the number of propylene carbonate molecules solvating the monomer is more than the number of IL ion pairs. One would expect that since more propylene carbonate molecules solvate the carbohydrate monomer that it interacts more favourably, but by Chapter 6.3.1 one can see that it is the mixtures and ILs that interact more favourably. This supports the suggestion made in Chapter 6.3.1 that if more solvent X molecules solvate a monomer compared to solvent Y, it does not mean solvent X interacts more favourably.

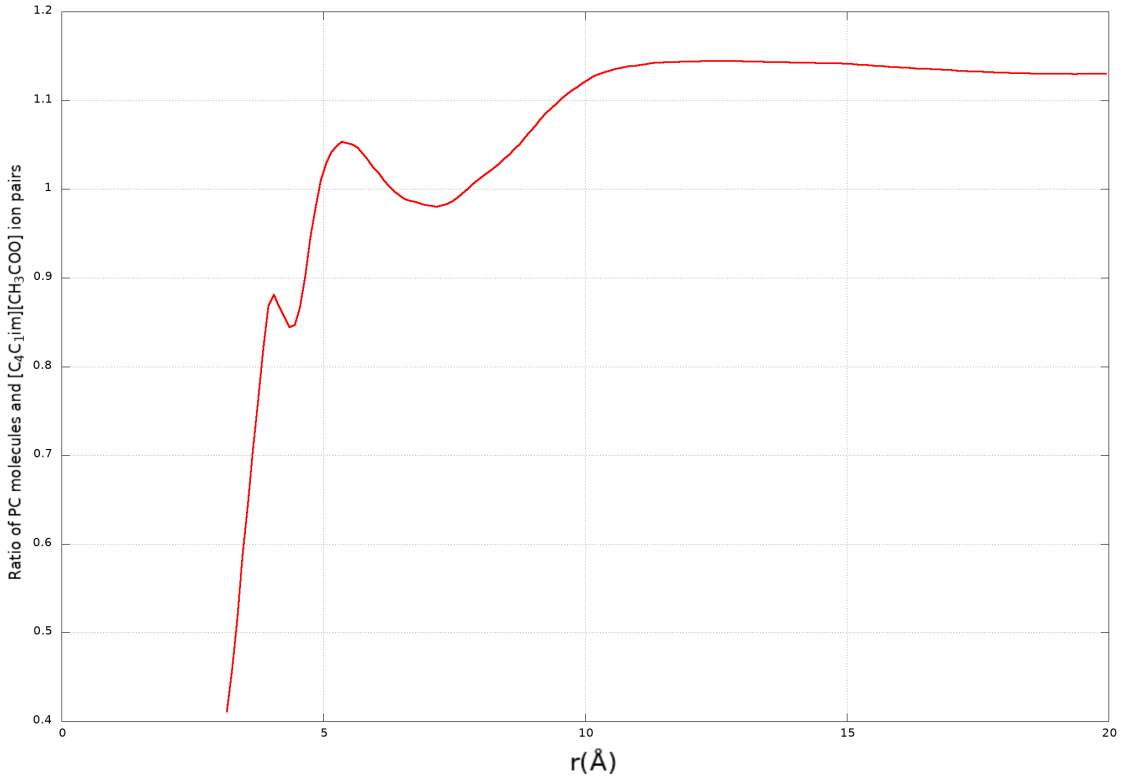


Figure 6.18: Ratio of propylene carbonate carbonyl (carbon) atoms and $[C_4C_1im][CH_3COO]$ carbonyl- and -N-C-N- carbon atoms coordinating around the glycosidic oxygen atom from $(GlcNAc)_2Me_2$.

In Figure 6.19 is shown the charge of ion pairs solvating the $(GlcNAc)_2Me_2$ monomer at various separations. If the ions from the pure ionic liquids solvated the monomer as pairs, the plot in Figure 6.16 would remain flat at 0, but this is not observed. Instead, it starts off at zero, but increases (positively) after 3 Å, suggesting that at this point more cations are solvating the monomer than anions. Close to 5 Å, the charge becomes negative, showing that more anions than cations solvate the monomer in the volume sphere of $\frac{4}{3}\pi 5^3$ Å^3. Figure 6.16 supports Lu et al.[80] and Zhang et al.[81]'s results on how important the role of the cation is in biomass interaction/dissolution, as one can see here that it is the cation that first solvates the monomer (at least in a greater quantity). Analysing the solvation energy results further, the cation is found to interact stronger with the $(GlcNAc)_2Me_2$ monomer than the anion, in which for the 1-butyl-3-methylimidazolium acetate ionic liquid, the cation has a -75.94 kcal/mol solvation energy with the monomer, while the anion has a solvation energy of -71.32 kcal/mol. As such, for these ionic liquids, the cation appears to be more important than the anions in terms of biomass interaction.

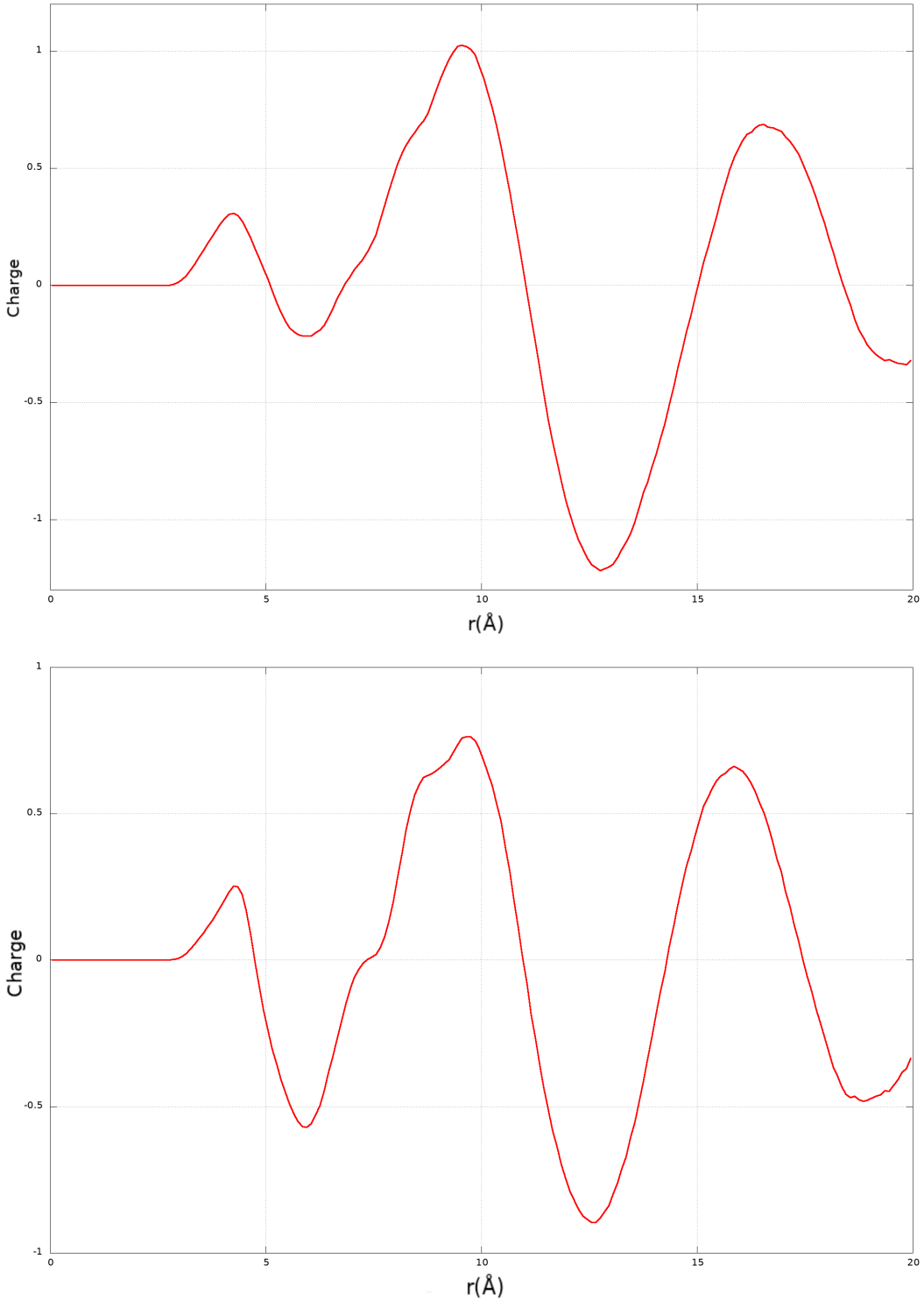


Figure 6.19: Charge of the solvent species coordinating around the $(GlcNAc)_2Me_2$ monomer at various separations. In the top figure $[C_4C_1im][CH_3COO]$ is used, while for the bottom figure $[C_4C_1im][CH_3SO_4]$ was chosen.

In Figure 6.20 is more RDFs, now between the hydrogen bond acceptors (of the solvents) and the hydrogen bond donors (of the solute), as this will give more information about the solvation structure of $(GlcNAc)_2Me_2$. The radial distributions produced in Figure 6.20 are similar in shape to those produced by Payal *et al.*[75], suggesting a satisfactory RDF calculation.

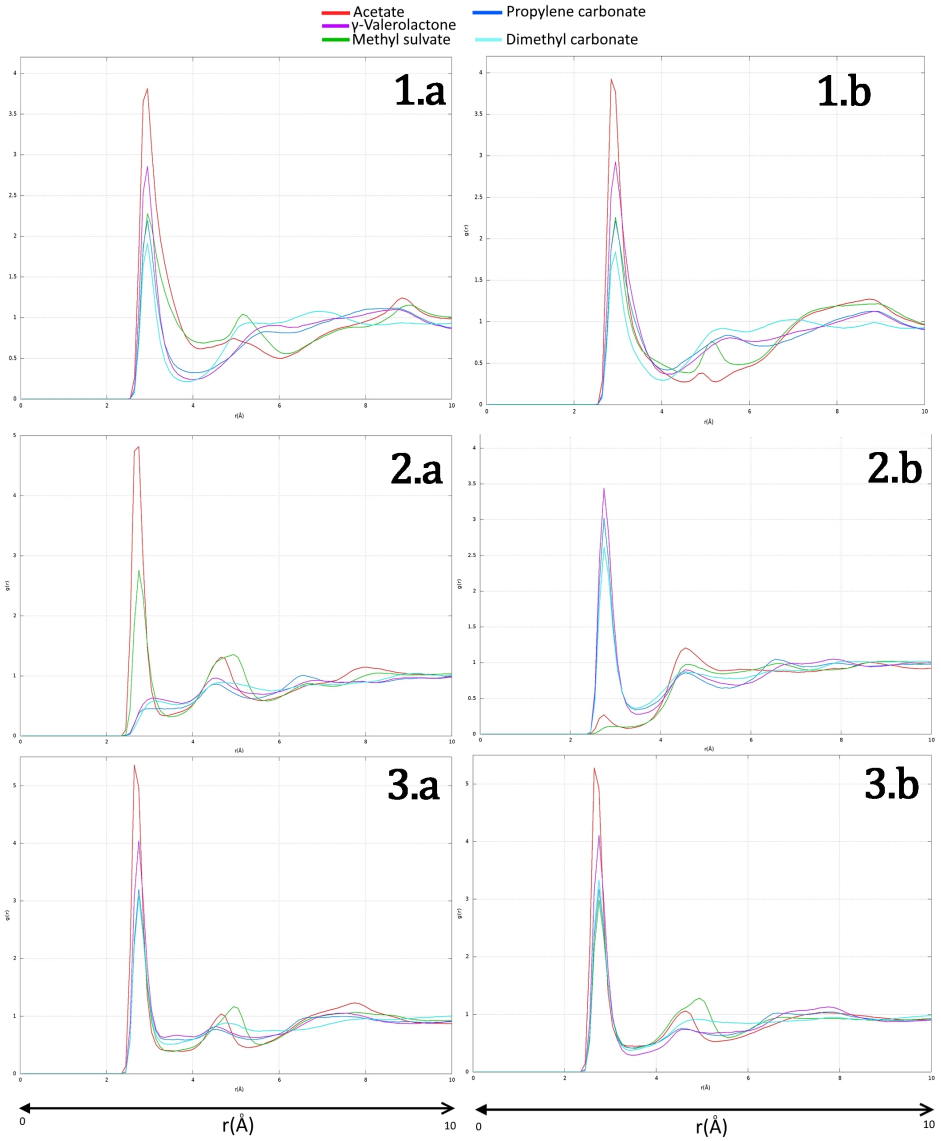


Figure 6.20: Radial distribution functions involving the carbonyl oxygen atoms (or those of o-type) with 1.a) the N2 non-reducing nitrogen atom, 1.b) the N2 reducing nitrogen atom, 2.a) the O3 non-reducing oxygen atom, 2.b) O3 reducing oxygen atom, 3.a) O6 non-reducing oxygen atom and 3.b) O6 reducing oxygen atom from the $(GlcNAc)_2Me_2$ monomer.

When performing the radial distribution functions with discrete atoms, a bit more information becomes available on the solvation structure of $(GlcNAc)_2Me_2$. In terms of the pure solvent systems and their carbonyl oxygen atoms (or those of o-type), the acetate RDFs have the biggest and broadest initial peaks (close to 3 Å) with the N2 nitrogen atoms, the O3 non-reducing oxygen atom and the O6 oxygen atoms from $(GlcNAc)_2Me_2$. This means that acetate is densely situated around these atoms at close distances and might be interacting the most favourably here. γ-Valerolactone has the second largest and broadest initial RDF peaks with the N2 nitrogen atoms and the O6 oxygen atoms. Methyl sulfate has the second largest initial peak with the O3 non-reducing oxygen atom. Propylene carbonate and dimethyl carbonate produce the smallest initial RDF peaks with all the hydrogen bond donating substituents (suggesting that it might be interacting the weakest here), except with the O3 reducing oxygen atom from $(GlcNAc)_2Me_2$. In terms of the R-O-R' oxygen atoms (or those of os-type) from the solvent components, methyl sulfate and propylene carbonate produce the biggest initial RDF peaks with all the hydrogen bond donating substituents from $(GlcNAc)_2Me_2$, except for the O3 reducing oxygen atom. For the mixtures, the acetate oxygen atoms from the 2:8 IL:co-solvent OESs produce the broadest and biggest initial RDF peaks with all the hydrogen bond donating substituents, followed by the 8:2 IL:co-solvent OESs and then pure IL. Finally, when considering the RDF peaks in terms of the co-solvent oxygen atoms for the OESs, all peaks are significantly smaller than that from the pure co-solvent; however, not much more information is available from these RDFs. When considering the RDFs involving the C4 carbon atom (-N-C-N-) from 1-butyl-3-methylimidazolium, broad and large RDFs peaks are observed only at 3.25 Å as compared to 2.65 Å for the methyl sulfate and acetate anions. This means that there is a larger probability of locating anions close to the monomer, but more cations are expected to coordinate very close to the monomer (according to the Figure 6.16).

The RDF results thus shows that for all the solvent systems, the broadest and tallest initial RDF peaks are observed for the O6 oxygen atoms from the $(GlcNAc)_2Me_2$ monomer, suggesting that the solvent components might interact the most in these regions. This is sensible as the O6 oxygen atoms are more accessible than the rest of the hydrogen bond donating substituents. Additionally, the smallest initial RDF peaks are observed with the O3 (reducing) oxygen atom. With the hydrogen bond analysis, the interaction between atoms (solvent and solute) is further investigated.

6.3.3 Hydrogen Bond Analysis Results

For the HB analysis, intramolecular- as well as intermolecular hydrogen bonds were investigated. The frequency of occurrence of the intermolecular hydrogen bonds are shown in Table 6.4, while other hydrogen bonding results are found in Table 6.5. The results in Table 6.4 shows that the least number of intermolecular hydrogen bonds are formed with the O3 (reducing) oxygen atom (from $(GlcNAc)_2Me_2$) which supports the RDF results, while the most number of intermolecular hydrogen bonds are formed with the O6 oxygen atoms from $(GlcNAc)_2Me_2$ molecule, also supporting the RDF results. In terms of the intermolecular hydrogen bonds formed with the pure solvents, pure $[C_4C_1im][CH_3COO]$ forms the most with the N2 nitrogen atoms (from $(GlcNAc)_2Me_2$), followed by γ-valerolactone, $[C_4C_1im][CH_3SO_4]$, propylene carbonate and then dimethyl carbonate, supporting the RDF results. For the O3 (reducing) oxygen atom, the most intermolecular hydrogen bonds are formed with the co-solvents, followed by the pure ILs, supporting the RDF results. With the O3 (non-reducing) oxygen atom, the most intermolecular hydrogen bonds are formed with the pure ILs, then the co-solvents, also supporting the RDF results. For the mixtures, it appears that the 8:2 IL:co-solvent OESs forms the most intermolecular hydrogen bonds with the $(GlcNAc)_2Me_2$ hydrogen bond donating substituents using their carbonyl oxygen atoms, while the 2:8 IL:co-solvent OESs follow closely.

Table 6.4: Number of average intermolecular hydrogen bonds formed between the carbonyl oxygen atoms from the solvent components and hydrogen bond donating substituents of $(GlcNAc)_2Me_2$

Solvent System	N2 (non-reducing)	N2 (reducing)	O3 (non-reducing)	O3 (reducing)	O6 (non-reducing)	O6 (reducing)
DC	0.3389	0.3571	0.5871	0.0566	0.7344	0.7143
PC	0.3914	0.4032	0.6818	0.0976	0.7185	0.7361
γV	0.4387	0.4511	0.6780	0.0790	0.7881	0.7860
$[C_4C_1im][CH_3SO_4]$	0.4327	0.4396	0.7425	0.0272	0.8156	0.7966
IL	0.5864	0.6352	0.9142	0.0520	0.9570	0.9545
IL:PC (2:8)	0.5549	0.5664	0.8524	0.0580	0.9269	0.9375
IL:γV (2:8)	0.6019	0.5910	0.8241	0.0646	0.9328	0.9086
IL:DC (2:8)	0.5826	0.6138	0.8783	0.0316	0.9386	0.9215
IL:PC (8:2)	0.5851	0.6259	0.8863	0.0849	0.9494	0.9597
IL:γV (8:2)	0.6025	0.6300	0.9093	0.0159	0.9537	0.9267
IL:DC (8:2)	0.5677	0.6229	0.9062	0.0211	0.9627	0.9529

Table 6.5: Frequency of occurrence (%) of the intraring O3 (Reducing)-H$\cdots$O5 (Non-Reducing) intramolecular hydrogen bond from $(GlcNAc)_2Me_2$, the number of average intramolecular hydrogen bonds formed in a $(GlcNAc)_2Me_2$ molecule, the number of average intermolecular hydrogen bonds formed between a $(GlcNAc)_2Me_2$ molecule and various solvent components, and the frequency of occurrence of all cation (using the -N-C-N- carbon atom) to $(GlcNAc)_2Me_2$ intermolecular hydrogen bonds

Solvent System	Frequency of occurrence of the intraring hydrogen bond	Number of average intramolecular hydrogen bonds	Number of average intermolecular hydrogen bonds formed	Frequency of occurrence of all cationic C4-$(GlcNAc)_2Me_2$ hydrogen bonds
DC	82.05%	0.9882	2.79	
PC	74.51%	0.9431	3.03	
γV	77.38%	0.9119	3.22	
$[C_4C_1im][CH_3SO_4]$	88.84%	0.9323	3.25	0.52%
IL	87.05%	0.8857	4.10	0.34%
IL:PC (2:8)	85.40%	0.8879	3.90	0.46%
IL:γV (2:8)	82.10%	0.8793	3.92	0.47%
IL:DC (2:8)	87.33%	0.9032	3.97	0.56%
IL:DC (8:2)	90.84%	0.9319	4.03	0.50%
IL:γV (8:2)	90.65%	0.9223	4.04	0.56%
IL:PC (8:2)	80.81%	0.8335	4.09	0.63%

Table 6.5 suggests that pure 1-butyl-3-methylimidazolium acetate forms the most intermolecular hydrogen bonds with a $(GlcNAc)_2Me_2$ molecule, and pure dimethyl carbonate the least. In terms of the pure solvents, it can be seen that the more intermolecular hydrogen bonds are formed between the $(GlcNAc)_2Me_2$ molecule, the less number of $(GlcNAc)_2Me_2$ intramolecular hydrogen bonds are available. This trend supports the RDF results. Table 6.5 also suggests that pure propylene carbonate disrupts the O3 (Reducing)-H$\cdots$O5 (Non-Reducing) intramolecular hydrogen bond the most, and 8:2 1-butyl-3-methylimidazolium acetate:dimethyl carbonate the least. As is evident from the second and third column in Table 6.5, the intraring hydrogen bond contributes the most to the number of average intramolecular hydrogen bonds, making it an important interaction to disrupt. Table 6.5 further suggests that as more co-solvents are added to the ionic liquid, the total number of intermolecular hydrogen bonds to the $(GlcNAc)_2Me_2$ molecule decreases, which agrees with the interaction energy results from Table 6.1 as well as the order of results in Table 6.3. Additionally, while literature has shown that a cation (from an ionic liquid) does have an effect on hydrogen bond formation and can contribute significantly to biomass-solvent interaction, Table 6.5 shows that the

cation forms negligibly few intermolecular hydrogen bonds with the $(GlcNAc)_2Me_2$ monomer, although this only due to the cut-off distance and cut-off angle chosen for the HB analysis. Changing the HB variables increases the frequency of occurrence results for the cation.

Considering the structure of the $(GlcNAc)_2Me_2$ monomer, there are 6 hydrogen bond acceptors available, while the solvent components (anion and co-solvent) do not have any hydrogen bond donors. This means that at most 6 intermolecular hydrogen bonds can be formed with the $(GlcNAc)_2Me_2$ monomer (ignoring the contribution from the cation). The results from Table 6.5 show intermolecular hydrogen bonds below this value, with a maximum of 4.10 produced. From Table 6.5, $[C_4C_1im][CH_3SO_4]$ is shown to form approximately 3 intermolecular hydrogen bonds with the monomer, with this being visually confirmed in Figure 6.21. In the same figure only one intramolecular hydrogen bond can be seen, the O3 (Reducing)-H$\cdots$O5 (Non-reducing) intramolecular hydrogen bond, which matches the number of average intramolecular hydrogen bonds from Table 6.5. It also highlights this as the most difficult hydrogen bond to disrupt for the monomer. Figure 6.22 shows pure propylene carbonate forming 3 intermolecular hydrogen bonds with the $(GlcNAc)_2Me_2$ monomer, further supporting the results from Table 6.5. Furthermore, there are 4 intermolecular hydrogen bonds available between the 8:2 IL:γV OES and the $(GlcNAc)_2Me_2$ monomer in Figure 6.23, 3 with the acetate anions and 1 with the co-solvent. This too supports the results from Table 6.5.

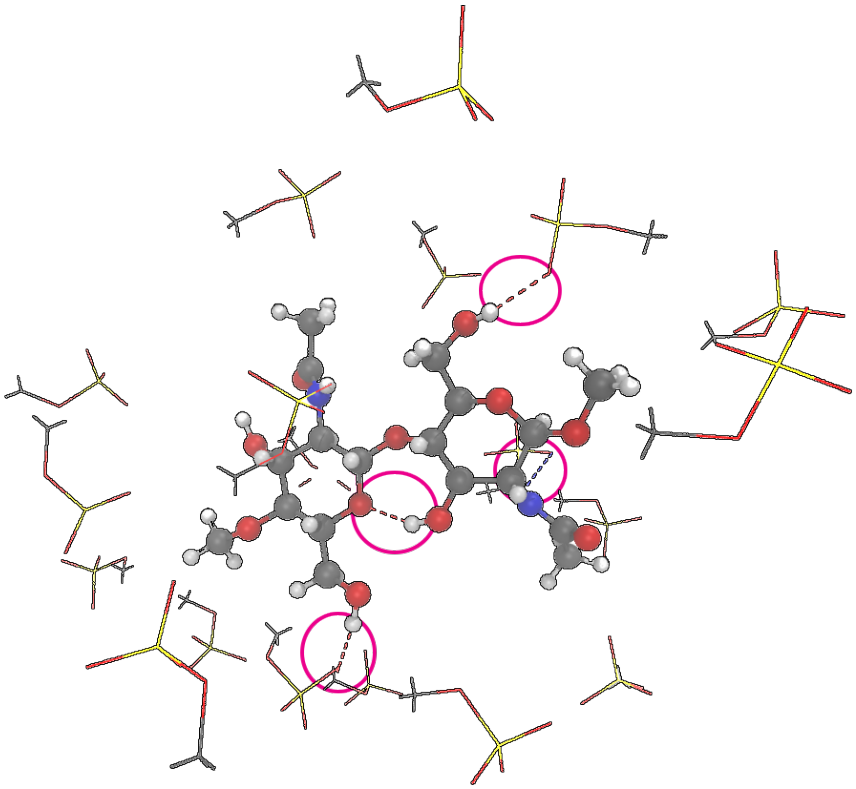


Figure 6.21: A snapshot of the methyl sulfate anions solvating around the $(GlcNAc)_2Me_2$ 6 Å away from it.

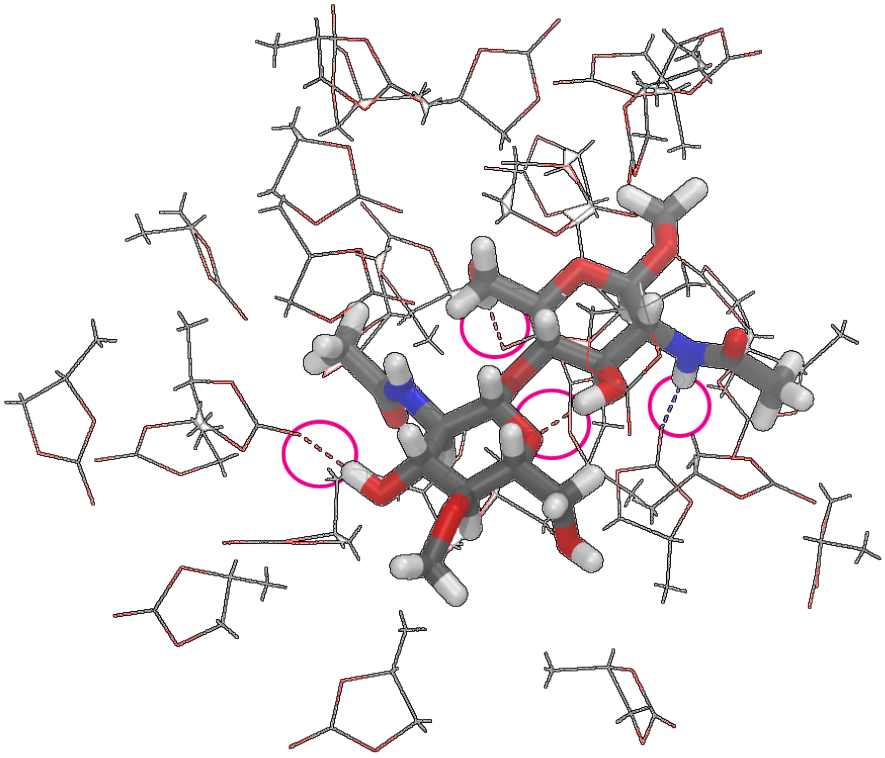


Figure 6.22: A snapshot of the methyl sulfate anions solvating around the $(GlcNAc)_2Me_2$ 6 Å away from it.

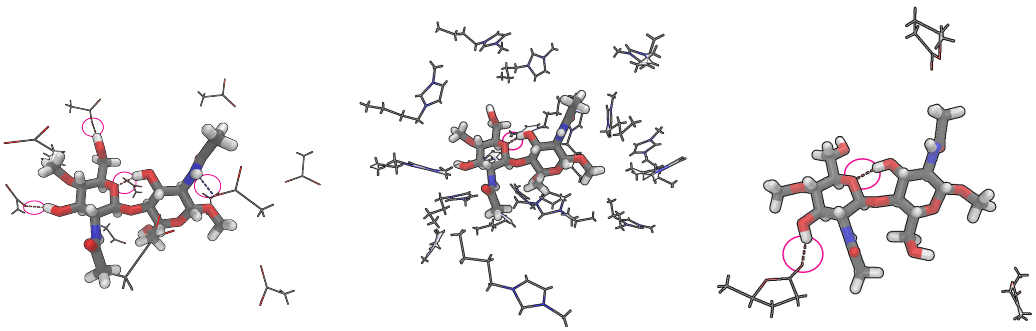


Figure 6.23: Snapshots of all the (left) acetate ions, (middle) 1-butyl-3-methylimidazolium cations and (right) γ-valerolactone molecules from 8:2 IL:γV solvating the $(GlcNAc)_2Me_2$ monomer 6 Å away from it.

From the above results (interaction energy, RDF and HB) it appears that the pure acetate-based ionic liquid will interact most favourably with the $(GlcNAc)_2Me_2$ monomer, followed by the 8:2 IL:co-solvent OESs, the 2:8 IL:co-solvent OESs and then the pure co-solvents. This correlates well with the cellulose dissolution results from Gale *et al.*[97] The results also show that although the solvation structure is important, a greater number of solvating molecules does not infer a more favourable interaction. It does show how important the cation is in interacting with and perhaps dissolving

biopolymers, as the cation interacts the strongest with the monomer and solvates it in a greater quantity (small volume spheres) compared to the anions. The RDF- and HB results suggests that the O6 oxygen atoms from $(GlcNAc)_2Me_2$ contribute significantly to the solute-solvent intermolecular interactions, while the O3 (reducing) oxygen atom contributes the least. The HB results answers why the latter contributes the least by suggesting that the O3-H$\cdots$O5 intramolecular hydrogen bond is the most difficult hydrogen bond to disrupt.

7 Potential of Mean Force Calculations Using Umbrella Sampling

7.1 Introduction to Theory

7.1.1 Potential of Mean Force[141]

The potential of mean force (PMF) studies how the free energy of a system changes as a function of some reaction coordinate parameter (e.g. the distance between two atoms or a bond angle); or rather the work required in altering/achieving a reaction coordinate parameter is calculated. In the case of this project, the work required in separating two $(GlcNAc)_2Me_2$ molecules is investigated, where a radial distance reaction coordinate parameter is selected. The PMF (W) can mathematically (Equation 7.1) be determined as

$$W(\vec{r}) = -\ln[\Pi(\vec{r})]k_BT = \ln[\Pi(\vec{r})^{-1/\beta}];\ \text{where } \beta = \frac{1}{k_BT} \tag{7.1}$$

Here, $\vec{r}$ is the coordinate, k_B is the Boltzmann constant, T is the temperature of the system and $\Pi(\vec{r})$ is the probability of the coordinate $\vec{r}$ taking on a specific value. $\Pi(\vec{r})$ is defined as (Equation 7.2)

$$\Pi(\vec{r}) = Q^{-1}\int\int \delta[r'(\vec{r}) - \vec{r}]e^{\frac{-E(\vec{r},\vec{s})}{k_BT}}d\vec{r}d\vec{s} \tag{7.2}$$

where Q is the normalising full partition function, δ is the Dirac delta function, r' is the value of the PMF coordinate for any arbitrary point in phase space and $E(\vec{r},\vec{s})$ is the sum of potential- and kinetic energies depending on parameters $\vec{r}$ and $\vec{s}$.

In order to exactly equate PMF with free-energy profile, one must consider the Jacobian contribution[77,177–179] As is customary with the reaction coordinate (r) chosen in this work, $2k_BT\ln(r)$ was used as the Jacobian contribution term ($\mathcal{J}$).[177] Here, k_B is the Boltzmann constant, T is the temperature and $\ln(r)$ is the natural logarithm of $(GlcNAc)_2Me_2$ separation. A second correction term has been suggested by Wong and York, but because this term cannot be trivially solved, is not included here.[178]

7.1.2 Umbrella Sampling[108]

The calculation of a PMF is one of the biggest challenges in computational chemistry. Calculating a fully-converged PMF requires each point along phase space to be equally sampled, and this cannot be done in simple equilibrium molecular dynamics. Umbrella sampling provides an alternative way to calculate the work along a reaction coordinate by forcing the simulation into regions of phase space that is not normally visited in equilibrium simulations.

In this method, bias potentials along a pre-defined reaction coordinate drive a system from the reactants to the products. The intermediate from the one state to the other is generally covered by harmonic potential windows at which constrained MD simulations are performed. From the sampled distribution of the system along the reaction coordinate, the work required in restraining the two molecules at each window can be calculated. Combining these windows together, the work done (on the system/surroundings) across the entire reaction coordinate can be calculated. Mathematically, the potential of mean force from the umbrella sampling is defined by (Equation 7.3)

$$\Delta W = W(r_2) - W(r_1) = \int_{r_1}^{r_2} \langle f_c(r) \rangle_c dr \tag{7.3}$$

where ΔW is the work done in moving the system from a radial distance r_1 to a radial distance r_2, $f_c(r)$ is the force of constraint and $\langle ... \rangle$ is the ensemble average over a set of separation windows.[175] Alternative methods are available for calculating potential of mean force, such as the Jarzynski relation, thermodynamic integration (TI), metadynamics and slow growth, but are not discussed here.[180–182].

7.1.3 Using the Weighted Histogram Analysis Method to Obtain Unbiased PMFs[183–184]

While umbrella sampling can be used to obtain a potential of mean force, note that the resultant PMFs outputted will be biased. Moreover, the PMFs are only available at discrete points, which is undesirable if a continuous PMF curve is required. The Weighted Histogram Analysis Method (`WHAM`) provides a way to use biased distributions to obtain unbiased PMFs, while also approximating intermediate PMF values between discrete states. When plotting histograms of the biased distributions, the more the histograms overlap the smaller the square error becomes on each histogram $\text{var}(\vec{r_i})$ as well as for the resultant PMF $\text{var}(W(\vec{r_i}))$, as is evident by Equation 7.4

$$\text{var}(W(\vec{r_i})) \propto \sum_{i=1} \text{var}(\vec{r_i}) \tag{7.4}$$

From Equation 7.4, the smaller the square error of the biased distributions becomes, the more accurately the unbiased PMFs and intermediate PMFs will be defined. Figure 7.1 shows an example of poor and good histogram overlap.

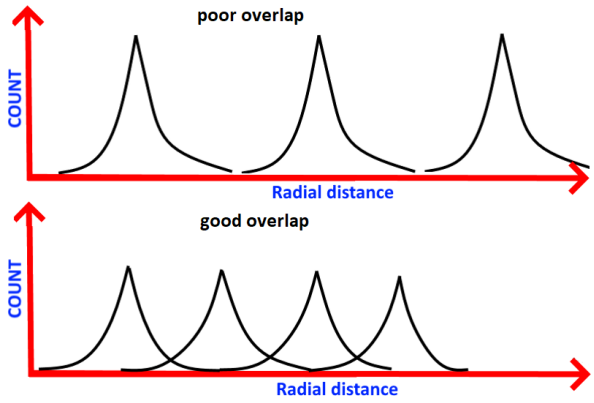


Figure 7.1: An example of poor- and good histogram overlap.

The `WHAM` method specifically makes use of the biased histograms to get a probability distribution function $P_i^{\text{unbias}}(r) = \langle e^{-2\beta U_i(r)} \rangle_i$, where $\langle ... \rangle$ is an ensemble average, $\beta = \dfrac{1}{k_B T}$ and $U_i(r)$ is the potential function; which is then used to obtain the unbiased PMF. The unbiased PMF is approximated from $W_i^{\text{unbias}}(r) = -k_B T \ln\left(\dfrac{P_i^{\text{unbias}}(r)}{\Delta x_{\text{bin}}}\right)$, where Δx_{bin} is the width of bin i in $P_i^{\text{unbias}}(r)$.

As such, from the unbiased probability distribution function can be approximated not only unbiased PMFs at discrete states, but also intermediate PMF values as the probability distribution function (PDF) is a continuous function. Once the WHAM method has been employed, the Jacobian contribution term $2k_BT\ln(r)$ is added to the PMF functions.[77] If this term is not added at this point, the PMF curves are found to never plateau.

7.2 Computational Details

Potential of Mean Force calculations. A potential of mean force calculation for the radial separation of two $(GlcNAc)_2Me_2$ molecules in the gas-phase was performed at 298.15 K using umbrella sampling. The starting arrangement of the two molecules is based on the face-to-face arrangement of the crystal structure of chitin. (See Figure 7.2)

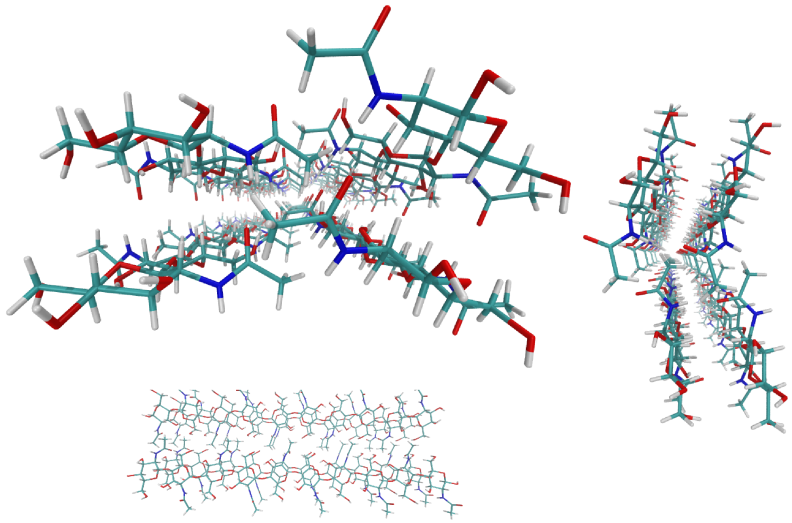


Figure 7.2: The crystal structure of chitin.

The separation of the $(GlcNAc)_2Me_2$ dimer was not continuous, instead the umbrella sampling procedure uses frames of separation. In each frame, a separation restraint was implemented with a harmonic potential and an AMBER force constant of k=2.5 kcal/mol (as used by Dickson *et al.*[185]) to restrain the two $(GlcNAc)_2Me_2$ molecules at various radial distances. The radial distances selected were $r \in \{2.75 + 0.75x; x \in \{1, 2, 3, ..., 36\}\}$ Å with respect to the centre of mass (which is calculated according to the heavy atoms only) of each $(GlcNAc)_2Me_2$ molecule. During the minimisation phase, the molecules were moved to each of the 36 restraint positions, thus creating 36 frames in which the molecules were separated at different radial distances. These are called windows. For the heating, equilibration and production phases, the two molecules were restrained. The production phase, in particular, was performed over 50 ns for each of the 36 windows, equating to 1.8 µs of simulation time. Figure 7.3 shows the starting- and end separation frames of the two $(GlcNAc)_2Me_2$ molecules.

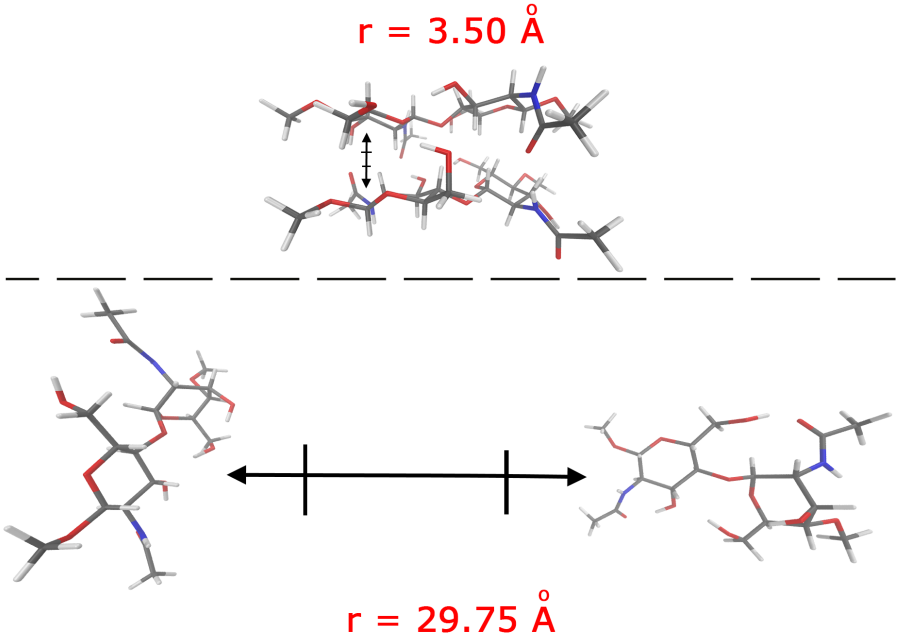


Figure 7.3 The gas-phase separation of two $(GlcNAc)_2Me_2$ molecules at 3.50 Å and 29.75 Å, respectively.

Once the MD simulation of all 36 windows were completed, a histogram plot was made of the distributions as to evaluate their overlap and decide if satisfactory intermediate values would be obtained between the windows. The better the overlap, the smaller the square error of each distribution and the better the eventual intermediate PMF values will be.

Hereafter, the `WHAM` (Weighted Histogram Analysis Method) program used the biased distributions to calculate a probability distribution, which in turn is used to output the unbiased PMF curve of the gas-phase separation of the two $(GlcNAc)_2Me_2$ molecules. This includes the PMF values at intermediate states. With `WHAM`, a tolerance of 0.000001 kcal was selected and a force constant of 5.0 kcal/mol. Note that AMBER uses the kx^2 harmonic potential, while `WHAM` uses $\frac{1}{2}kx^2$, hence the force constant used in `WHAM` is double that used in AMBER. The appearance of a plateau in the PMF curve suggests that the two $(GlcNAc)_2Me_2$ molecules are not interacting much anymore and at this point, the separation is stopped.

Following this, 13 of the 36 gas-phase windows were solvated with the various solvent systems using `Packmol`, specifically those in which the radial distance between two $(GlcNAc)_2Me_2$ molecules are $r \in \{2.00 + 1.50x; x \in \{1, 2, 3, ..., 13\}\}$ Å. All windows were then minimised, heated (to 353.15 K) and equilibrated (for 35 ns) using the same procedure as for the gas-phase simulation. For the production run of 50 ns on each window, a step size of 10 (here the distance and force data is outputted every $10 \times 0.02/1000 = 20$ fs) and a restraint force constant of 2.5 kcal/mol was chosen.

As with the gas-phase simulation, the `WHAM` program was also used to produce PMF curves for the separation of the two $(GlcNAc)_2Me_2$ molecules in the various solvent systems. These PMF curves were then compared to each other and the best solvent for $(GlcNAc)_2Me_2$ separation chosen.

7.3 Potential of Mean Force Calculation Analysis

Multiple methods were tested for calculating the PMF curves in this project, including steered molecular dynamics (SMD) to determine the reaction coordination in solution, followed by umbrella sampling of each window; and SMD alone but using the Jarzynski equation directly to calculate the PMF from an integration of the force, but the curves produced from these were not as desired (did not plateau satisfactorily). Instead, umbrella sampling is used here with a specific path of separation; this path is that of $(GlcNAc)_2Me_2$ gas-phase separation. At a far enough separation, the two molecules are not expected to interact any more, and at this point, the work required to separate them further should not change. As a result of this, the PMF curves were subsequently shifted to represent a 0 kcal/mol PMF (or energy change) at the point the PMF plateaus. Using the 36 gas-phase windows, `WHAM` was employed to create a distribution of histograms to ultimately approximate the PMF values between discrete points. The histograms for the gas-phase separation are shown in Figure 7.4, while the corresponding gas-phase PMF curve at 298.15 K is shown in Figure 7.5

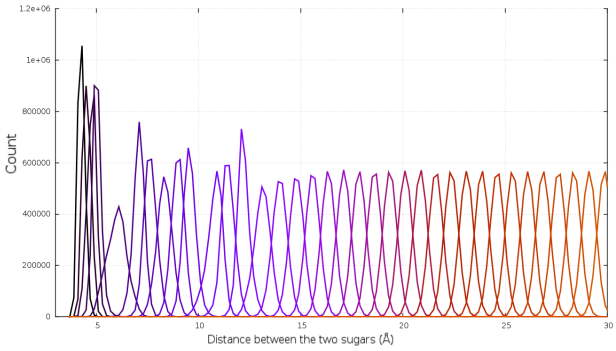


Figure 7.4: The histograms created by `WHAM` for the gas-phase separation of two $(GlcNAc)_2Me_2$ molecules.

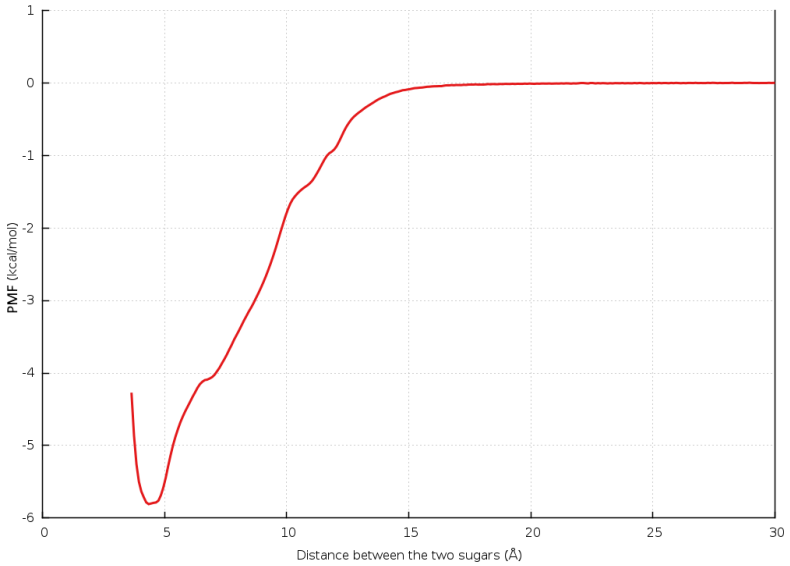


Figure 7.5: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in the gas-phase.

The greater the number of windows used are, the better the overlap of histograms will be (as seen in Figure 7.4) and the more accurate the PMF values. After a separation of 16 Å, the PMF curve in Figure 7.5 plateaus to (the shifted) 0 kcal/mol. It is thus expected that at a separation larger than 16 Å, the interaction between the two $(GlcNAc)_2Me_2$ molecules should be negligible (in the gas phase). Notice also the highly negative global minimum close to 5 Å (this is the well-depth), which suggests the two $(GlcNAc)_2Me_2$ molecules require a substantial driving force to separate beyond this point.

Since a path has been constructed from the gas-phase separation, 13 of the 36 windows were selected, solvated with TIP3P water[186] and each window was minimised, heated (298.15 K), equilibrated and sent through a production run with the two $(GlcNAc)_2Me_2$ molecules being restrained in each window. The separation of the two molecules was done from 3.50 Å to 21.50 Å in increments of 1.50 Å. TIP3P water was selected as the solvent as to validate the PMF calculation procedure since this water model is a common solvent used in cellulose dissolution studies.[76–77] The PMF curve produced here is specifically compared to the PMF curve produced by Brady *et al.*[77], in which the separation of a cellobiose dimer was studied in the TIP3P water model (see Figure 7.6).

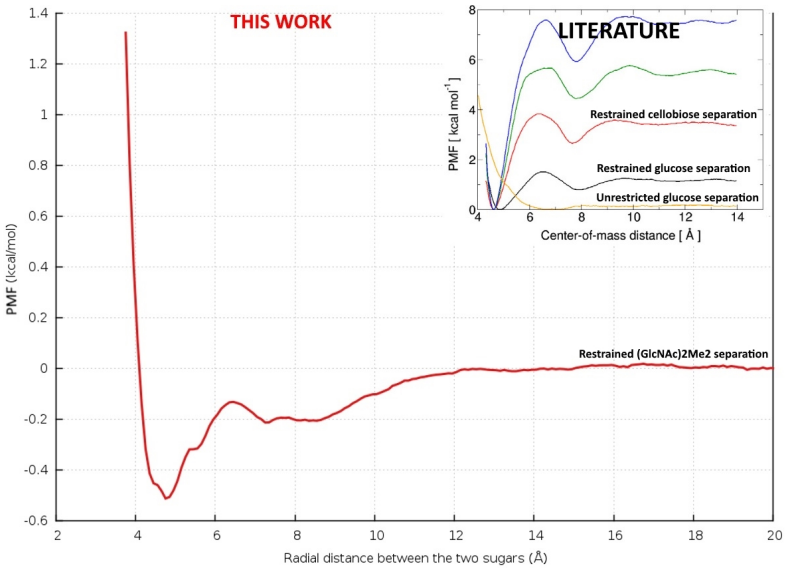


Figure 7.6: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in TIP3P water.

Just as with the gas-phase separation, a (shifted) 0 kcal/mol plateau is observed as well as a global minimum (well-depth) between 4-5 Å, suggesting a significant (yet lower than in vacuo) driving force will be required to separate the two molecules beyond this point. When comparing the separation of a $(GlcNAc)_2Me_2$ dimer in water to the separation of a cellobiose dimer in TIP3P water from Brady *et al.* (see the red curves in Figure 7.6), similarities in terms of well-depth, two large local minima (one at 5 Å and the other at 7 Å), as well as an energy barrier at 6.5 Å are observed. Brady *et al.* suggests the energy barrier between two local minima to be "as a result of artifacts of the artificial reaction coordinates chosen. This results either from the energy needed to pull a vacuum between the dimer as they separate or to squeeze out the last solvation layer between the dimer as the monomers are brought together."[77]. Moreover, the yellow plot in Figure 7.6 (literature) shows the unrestricted simulation of separating a glucose dimer in TIP3P water, while the black

profile shows the restrained PMF of this separation. Bergenstråhle suggests that the shape of the unrestricted PMF curve for separating two glucose monomers in water is most likely more accurate than the shape of the restrained PMF curve, since glucose does not show a tendency to associate in water (rather dissociate) and thus spontaneously separates. Since the well-depth from the restrained PMF curve is less than k_BT (implying spontaneous separation), it is unclear if the appearance of this barrier is statistically significant..[77,187]

Since a PMF curve was produced for separating two $(GlcNAc)_2Me_2$ molecules in water that agrees well with the published PMF curve by Brady *et al.*[77], attention was shifted to solvating the gas-phase windows with 1-butyl-3-methylimidazolium acetate ion pairs. However, it is important to consider the greater viscosity of the IL and hence, production runs of 50 ns, as well as 100 ns, were considered. The resultant PMF curves at 298.15 K are shown in Figure 7.7.

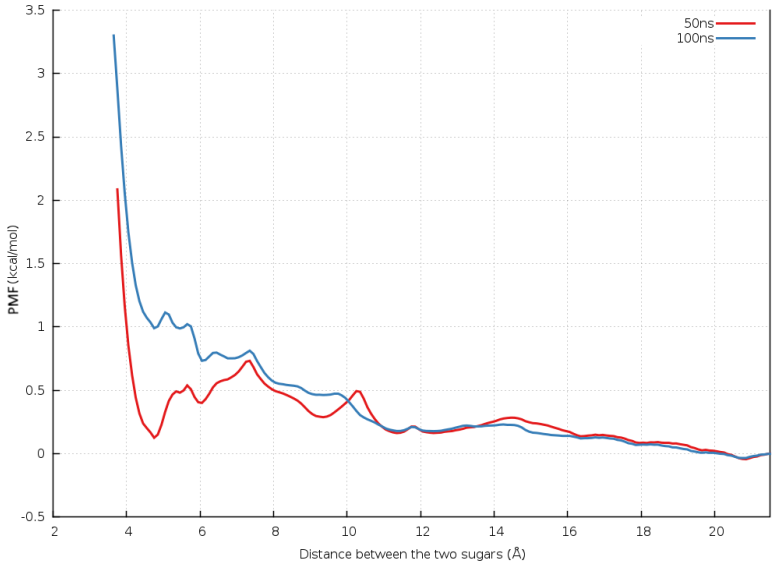


Figure 7.7: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in 1-butyl-3-methylimidazolium acetate at 298.15 K, using production runs of 50 ns (red) and 100 ns (blue).

After running the production phase for each window for 50 and 100 ns, the PMF curves in Figure 7.7 are still not plateauing satisfactorily, while also having a bumpy shape, producing an energy barrier for the 50 ns curve close to 5.5 Å. Longer simulation times and more windows being used with umbrella sampling could improve the shape of the PMF curve, but this quickly becomes impractical.[†] Rather than pursue longer simulation times, an alternative procedure is running the MD simulations at a higher temperature, namely 353.15 K. The viscosity of the ionic liquid will decrease allowing the IL molecules to diffuse quicker, equilibrate faster and give a better production plateau.

[†]If all 36 gas-phase windows were solvated with each solvent system and restrained MD simulations performed at 298.15 K, it would take on average 1.5 days to complete just one window based on the computational resources available. This equates to 18 days of calculation time for one simulation box or 324 days for all solvent systems. This is impractical for a master's research project.

Note also the significant change between the two curves in Figure 7.7 at radial separations of 4-6 Å, where the well-depth "disappears" when running the simulations for longer, as well as the energy barrier at 5.5 Å. The energy barriers (although significantly smaller for the 100 ns curve) has values of less than RT, meaning that at 80 °C, it can easily be crossed (a spontaneous separation). To see if the separation will work sufficiently well at a higher temperature (specifically since the molecules separate spontaneously in IL), a rough PMF is produced from the radial distribution function between the $(GlcNAc)_2Me_2$ monomers from an unrestrained MD simulation, followed by umbrella sampling restrained MD simulations. Since the IL is highly viscous, the production phase was carried out at 150 ns for the unrestrained simulation (see Figure 7.8). The rough PMF is approximated from the RDF as follows.[188]

$$W(r_i) = -k_B T \ln\, g(r_i) \tag{7.5}$$

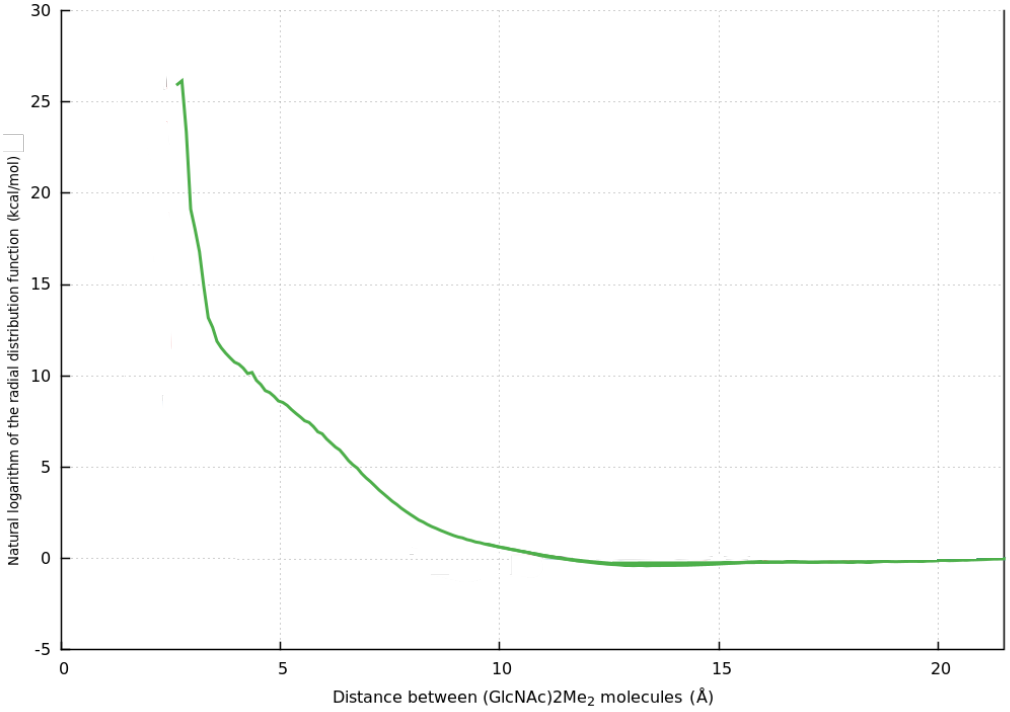


Figure 7.8: Natural logarithm of the RDF between two $(GlcNAc)_2Me_2$ molecules in 1-butyl-3-methylimidazolium acetate at 353.15 K, after 150 ns of production simulation time.

The rough PMF derived from the RDF suggests a higher temperature will give reasonable results, such as a spontaneous separation. In terms of the PMF curve calculated via umbrella sampling, see Figure 7.9, where the production run was done for 50 ns per window.

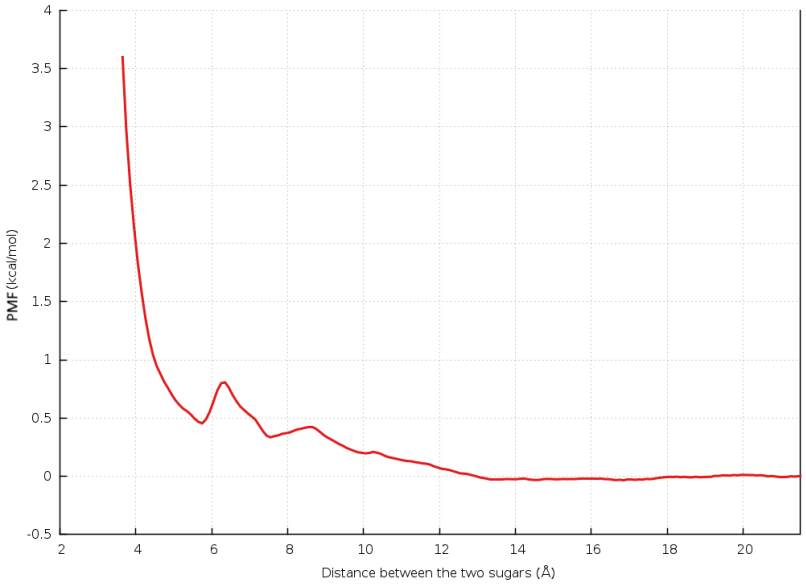


Figure 7.9: The potential of mean force curve for the separation of two $(GlcNAc)_2Me_2$ molecules in 1-butyl-3-methylimidazolium acetate at 353.15 K for a production run of 50 ns.

The energy barrier present on the PMF curve in Figure 7.9 at 6.5 Å is suggested to be due to artifacts of the reaction coordinate chosen, in which a solvent component is "squeezing" in-between the two monomers.[77] The barrier occurs at the same separation distance as in Brady's study of the separation of cellobiose in TIP3P water, as well as for the separation of a $(GlcNAc)_2Me_2$ dimer in TIP3P water (see Figure 7.6) and in pure ionic liquid (at 298.15 K in Figure 7.7). To confirm that the barrier is indeed due to solvent "squeezing" in-between the dimer, the immediate solvation of the dimer should be visually investigated. If no solvent is present in-between the dimer at a radial solute-solute separation of 5 Å (before the barrier is observed), but at 6.5 Å (at the peak of the barrier), then the energy barrier is due to the solvent. See figures 7.10-7.11.

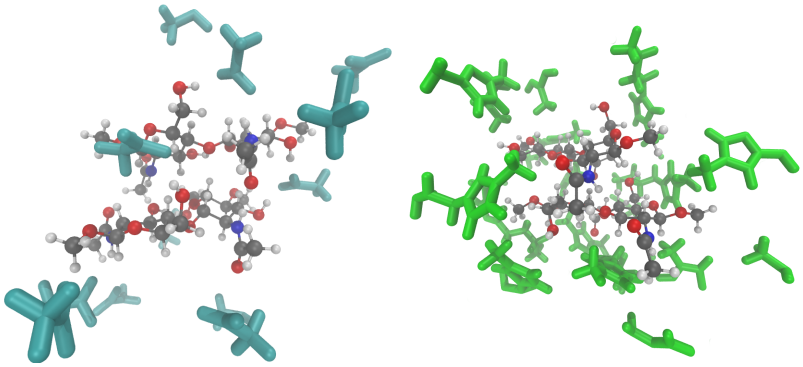


Figure 7.10: A snapshot of the immediate solvation of the $(GlcNAc)_2Me_2$ dimer in $[C_4C_1im][CH_3SO_4]$ separated at 5.00 Å.

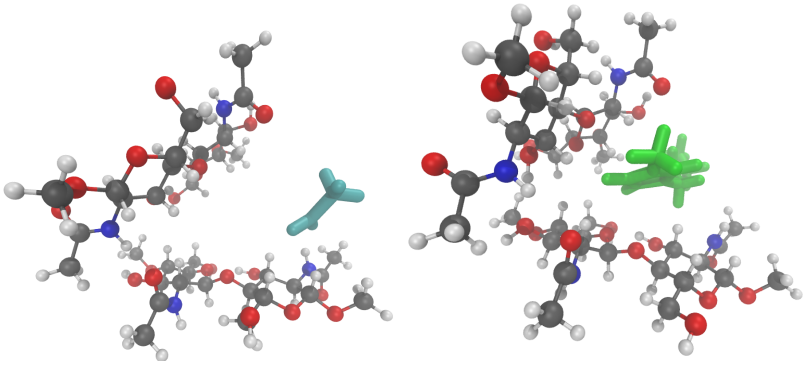


Figure 7.11: A snapshot of the immediate solvation of the $(GlcNAc)_2Me_2$ dimer in $[C_4C_1im][CH_3SO_4]$ separated at 6.50 Å.

In Figure 7.10, no solvent is present between the dimer, but there is solvent present between the dimer in Figure 7.11. This supports the suggestion made by Brady et al.[77], that the energy barrier is due to solvent. It is worth noting that it is the cation that "squeezes" in-between the dimer first, further supporting the results from Chapter 6 on how important the role of the cation is in interacting with or dissolving biomass. Nonetheless, the barrier is less than RT and should easily be crossed at 80 °C. Moreover, as can be seen in Figure 7.7, the barrier is expected to decrease with a longer simulation time. This was confirmed with a longer simulation run of 150 ns, see Appendix 2. From this point onward, all restrained MD simulations were performed at 353.15 K using the same PMF umbrella sampling procedure as in the above.

7.3.1 Pure Solvent PMF Results

In Figure 7.12 and 7.13 is the histogram plot of distributions for $(GlcNAc)_2Me_2$ separation in pure γ-valerolactone, as well as the PMF curves for the dimer separation in pure solvent.

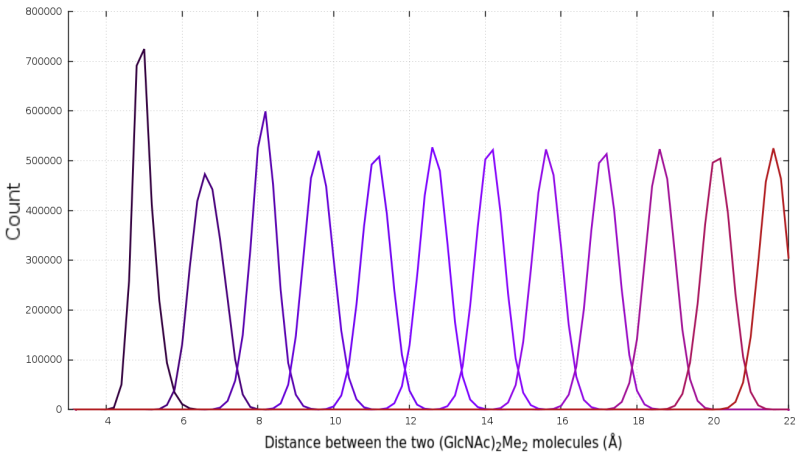


Figure 7.12: The histograms created by `WHAM` for the separation of two $(GlcNAc)_2Me_2$ molecules in pure γ-valerolactone (at 353.15 K.)

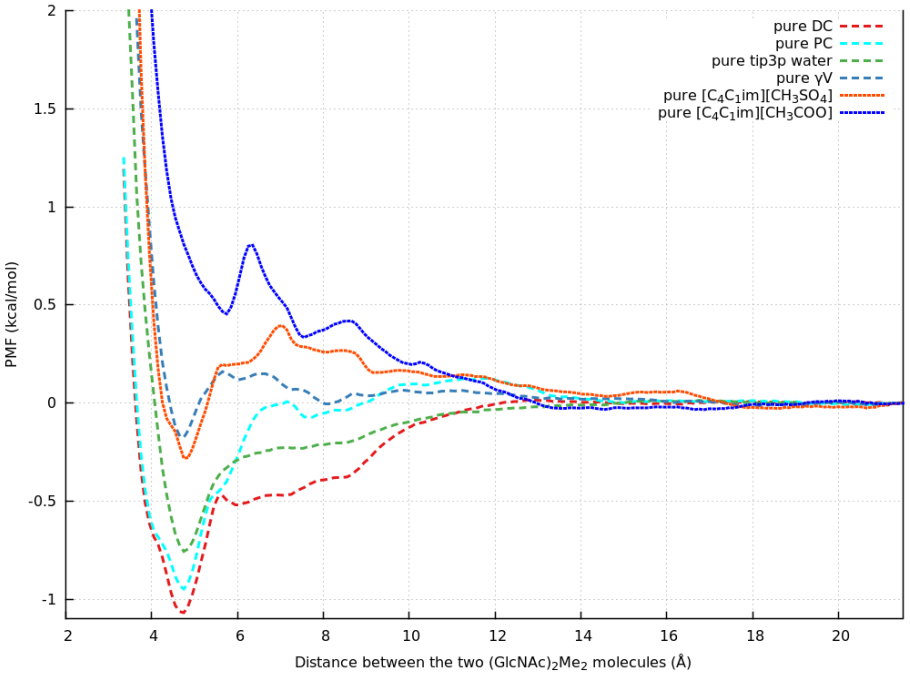


Figure 7.13: The PMF curves of separating a $(GlcNAc)_2Me_2$ dimer in pure $[C_4C_1im][CH_3COO$, $[C_4C_1im][CH_3SO_4]$, γ-valerolactone, propylene carbonate and dimethyl carbonate.

The methyl sulfate-based IL and pure co-solvents all share a well-depth at a common radial separation, that being 4.75 Å. The values of the well-depths can be found in Table 7.1. All PMF curves also have (initial) energy barriers (between 4-6 Å) to cross to start their dissociation. The values for the (initial) energy barriers is shown in Table 7.1.

Table 7.1: The well-depth and energy barriers of $(GlcNAc)_2Me_2$ separation in pure solvent

Solvent	Well-depth* (kcal/mol)	Energy Barrier** (kcal/mol)
DC	−1.07	0.60
PC	−0.95	0.95
γV	−0.17	N/A
TIP3P Water	−0.76	0.41
$[C_4C_1im][CH_3SO_4]$	−0.28	0.48
$[C_4C_1im][CH_3COO]$	0.45	0.35

* - The well-depth is calculated as the difference between the first local minima (or global minima) and the infinite separation (0 kcal/mol).
** - The energy barrier is calculated by taking the difference between the energy from the local maxima on the barrier and the energy for the local minima before the barrier.

Table 7.1 shows that the PMF curve of the acetate-based IL is significantly different (higher in energy) compared to the rest of the solvents, producing a well-depth (which is expected to decrease with a longer simulation time) of 0.45 kcal/ion pair. The first energy barrier on the $[C_4C_1im][CH_3COO]$ PMF curve is also quite low, with a value of 0.35 kcal/ion pair being produced - one of the weakest in this work. In pure dimethyl carbonate, on the other hand, the biggest well-depth is produced, corresponding to a value of -1.07 kcal/mol. Additionally, for the propylene carbonate PMF curve,

the largest initial energy barrier is observed, with a value of 0.95 kcal/mol. Overall, based on the well-depth results, the most amount of thermodynamic work will be required to separate the dimer in dimethyl carbonate, and the least in $[C_4C_1im][CH_3COO]$. Moreover, the dimer will require more energy to cross the first barrier of separation in propylene carbonate, and the least in γ-valerolactone and $[C_4C_1im][CH_3COO]$. Note however, that all the above energy values are significantly lower than RT, which means that at 80 °C, without much/any external work required, the dimer can easily cross the energy barriers and will spontaneously separate in all the solvents. Longer simulation times should improve the shape of the PMF curves and the values produced in Table 7.1 but this was not computationally feasible at the time of running the simulations. Furthermore, the above well-depth results for Table 7.1 matches the results for the total solvation energy (Chapter 6.3.1) and average intermolecular hydrogen bonds (Chapter 6.3.3) between the pure solvent and $(GlcNAc)_2Me_2$ monomer well, a relationship that should be investigated further.

7.3.2 OES PMF Results

The PMF calculations for the organic electrolyte solutions were run in two parts. This was to investigate how different configurations can affect the PMF calculations. The first part had the co-solvent molecules start the simulations off by closely surrounding the $(GlcNAc)_2Me_2$ molecules, while for the second part, the ion pairs directly surrounded the dimer. See Figure 7.14 and 7.15 for how these simulation boxes were constructed, and figures 7.16 and 7.17 for a look at the PMF curves.

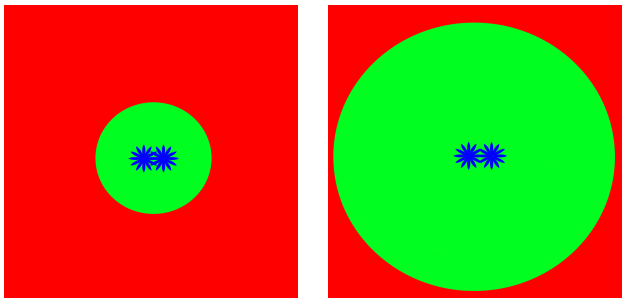


Figure 7.14: An illustration of simulation boxes created for the 8:2 and 2:8 IL:co-solvent OESs, in which the (blue) $(GlcNAc)_2Me_2$ dimer (separated at 3.50 Å) is directly surrounded by co-solvent (green) molecules and then ionic liquid (red) ion pairs.

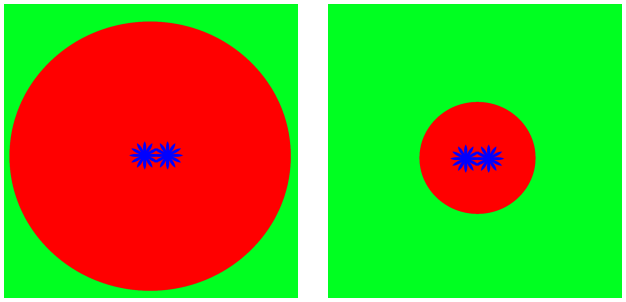


Figure 7.15: An illustration of simulation boxes created for the 8:2 and 2:8 IL:co-solvent OESs, in which the (blue) $(GlcNAc)_2Me_2$ dimer (separated at 3.50 Å) is directly surrounded by ion pairs (red) and then co-solvent (green) molecules.

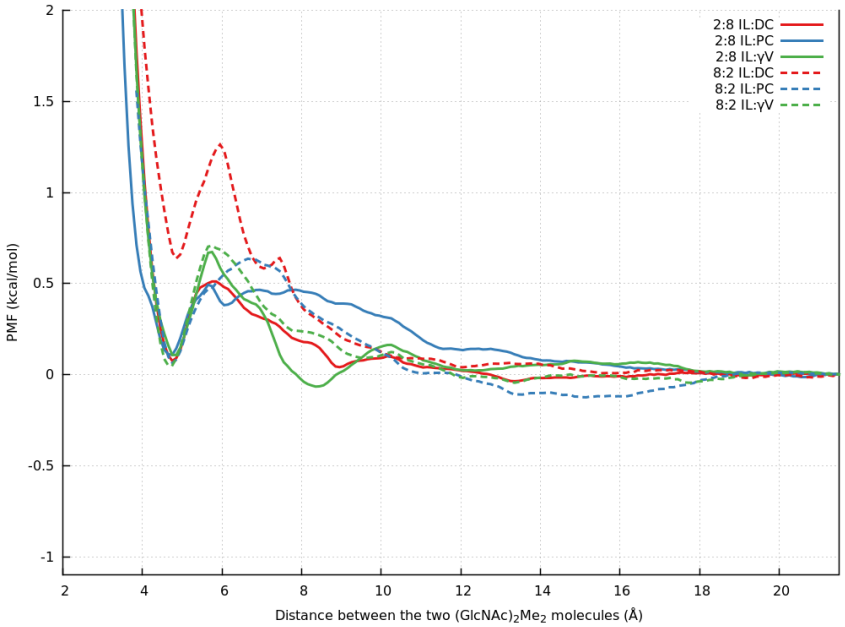


Figure 7.16: The potential of mean force curves for the separation of two $(GlcNAc)_2Me_2$ monomers in the 8:2 and 2:8 IL:co-solvent OESs. The simulation boxes used here was for the co-solvent starting close to the dimer.

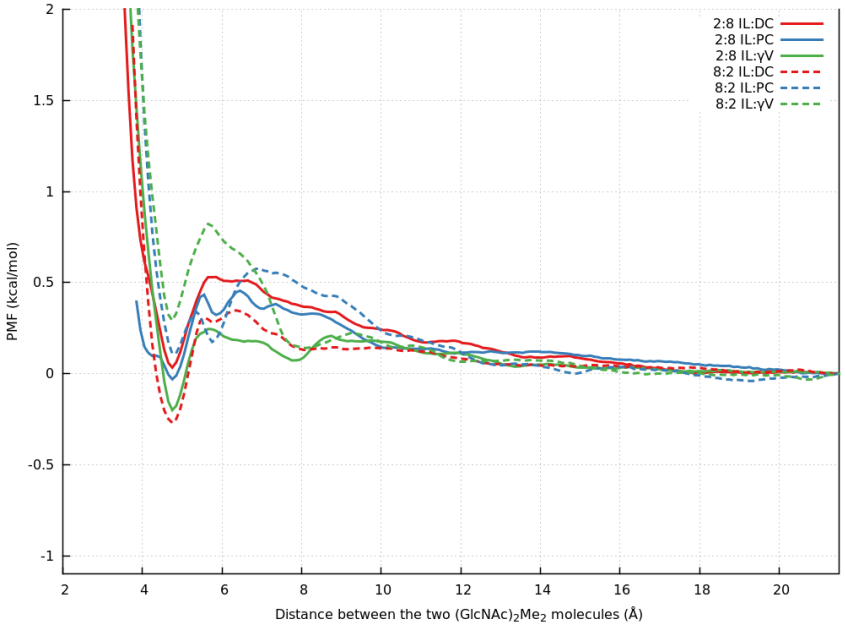


Figure 7.17: The potential of mean force curves for the separation of two $(GlcNAc)_2Me_2$ monomers in the 8:2 and 2:8 IL:co-solvent OESs. The simulation boxes used here was for the ion pairs starting close to the dimer.

All the PMF curves in Figure 7.16 have well-depths at around 4.75 Å (same as in Figure 7.13) and also have energy barriers between 5-7 Å, with 8:2 IL:DC's being the most significant. The 8:2 IL:DC PMF curve has a well-depth of 0.68 kcal/mol, while it has an initial energy barrier of 0.62 kcal/mol. The rest of the OESs have similar well-depths (close to 0.05 kcal/mol), while the 2:8 IL:PC PMF curve has the lowest initial energy barrier (0.17 kcal/mol). The energy barriers and well-depths are, however, lower than RT in value, which implies that dimer separation should happen spontaneously in all the mixtures.

The PMF curve with the highest well-depth in Figure 7.17 now corresponds to the 8:2 IL:γV one (0.29 kcal/mol), while the 8:2 IL:DC curve produces the lowest well-depth (-0.27 kcal/mol). All PMF curves too have well-depth locations close to 4.75 Å (just as in Figure 7.14) and have initial energy barriers to cross (between 5-7 Å). The 8:2 IL:γV OES PMF curve has the largest initial energy barrier to cross (0.53 kcal/mol), while 8:2 IL:PC has the lowest (0.23 kcal/mol). The well-depths and energy barriers, however, have values below RT, meaning that spontaneous separation should occur in these OESs.

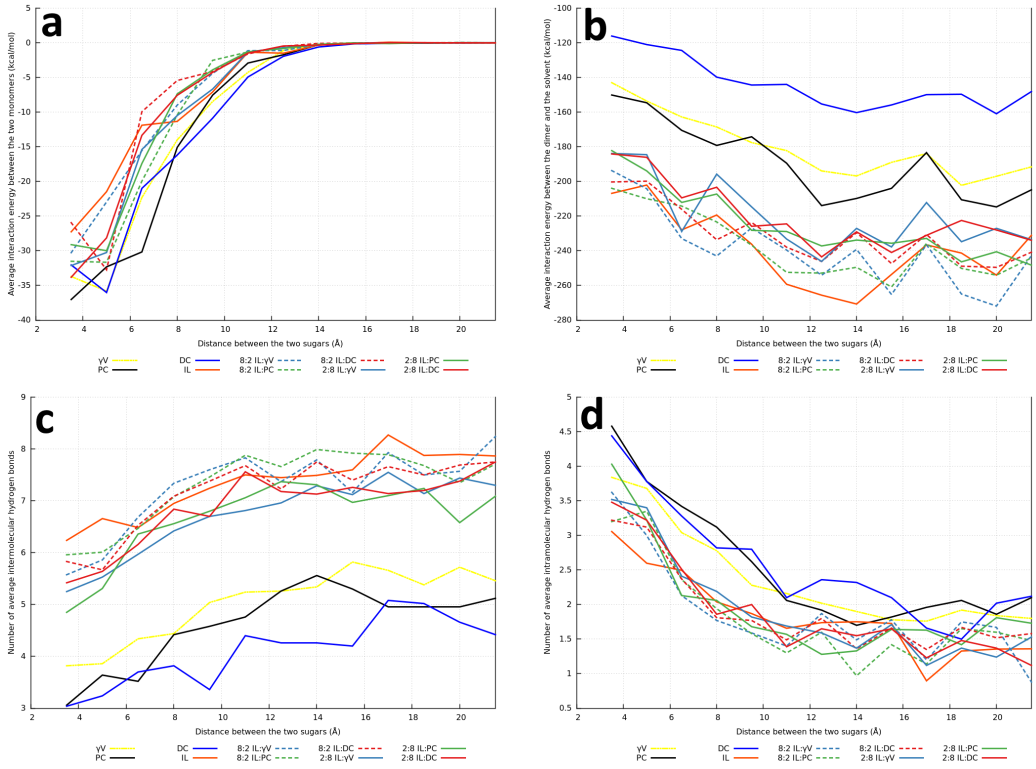


Figure 7.18: a) Interaction energy between each $(GlcNAc)_2Me_2$ monomer at various radial separations. b) Solvation energy between the solvent and the dimer (single entity) at various monomer-monomer radial separations. c) Number of average intermolecular hydrogen bonds between the two separate monomers at various radial separations. d) Number of average intramolecular hydrogen bonds in the dimer (single entity) at various monomer-monomer radial separations. The mixtures chosen here are those fitting the description of Figure 7.14.

Aside from using the restrained windows for PMF calculations, hydrogen bonding and interaction energies can also be calculated. In Figure 7.18.a the interaction energies between the two $(GlcNAc)_2Me_2$ monomers were calculated across various radial separations. It shows that as the separation increases, so does the interaction energy, to the point where it plateaus off to 0 kcal/mol. In Figure 7.18.b was calculated the solvation energy between the dimer and the solvent across various monomer-monomer separations. As the separation between the dimer increases, the solvation energy between the solvent and dimer decreases. This is since the more the dimer separates, the more it allows the solvent to directly solvate around it and increases the interaction. In Figure 7.18.c the number of average intermolecular hydrogen bonds between the solvent and dimer was calculated, with a general increase being observed as the dimer separates apart. The reasoning follows the same as for Figure 7.18.b in which more solvent gets to directly solvate the monomers as they separate apart. Lastly, in Figure 7.18.d the number of intramolecular hydrogen bonds was calculated for the dimer as they separate, with the number generally decreasing as the two monomers move apart. This is expected since the more intermolecular hydrogen bonds that are formed, the less intramolecular hydrogen bonds can remain.

Longer simulations times and performing a Boltzmann distribution over multiple PMF calculations can improve the results in Figure 7.16-7.17, but this was impractical in the project. What one can conclude from the PMF curves in Figure 7.16-7.17 is that the initial configuration of simulation boxes do have an effect on the PMF calculations, and averaging over multiple states (similar to what was done in Chapter 6.3.1) is necessary for improved results. While the PMF calculation procedure proved to be insufficient in predicting the separation trend for a $(GlcNAc)_2Me_2$ dimer in OES, it worked significantly well with the pure solvents. Using longer simulation times, more restraint windows, multiple simulation boxes for one window, higher temperatures and even other PMF calculation methods can shed more light on the separation of two $(GlcNAc)_2Me_2$ monomers in various solvent systems.

[Page intentionally left blank.]

8 Conclusion

The aim of this project was to model the dissolution of chitin, as well as investigate the interaction between it and various solvent systems using molecular dynamics. A force field validation and the inclusion of experimentation was done to validate the computational results produced and accuracy of the simulations. Radial distributions, interaction energies and hydrogen bond analyses were calculated to investigate the solvation structure and interaction of a $(GlcNAc)_2Me_2$ monomer with the solvent, while PMF calculations were used for the separation of two monomers in the solvent.

Parameterisation via the GAFF2 force field was found to be a better fit than the GAFF(1) force field in this project as it produced accurate density, heat of vaporisation and self-diffusion coefficients relative to literature values. Atomic partial charge scaling of the ionic liquids were also shown to drastically change the thermodynamic properties of ILs, in which a 0.9 scaling factor best represented the density property for the ILs and the 0.7 scaling factor best described the heat of vaporisation property. The addition of diffuse functions to anions also showed to improve on the parameterisation of the ILs. A satisfactory correlation was achieved between the interaction energy and (mass) swelling capacity, further validating the force field chosen. An exact correlation was not achieved, however, and this was suggested to be due to an over-represented co-solvent-$(GlcNAc)_2Me_2$ solvation energy. Averaging over multiple simulation boxes (with different initial configurations) was found to significantly improve the solvation energy, however.

The radial distribution functions showed that there are more $[C_4C_1im][CH_3COO]$ ions pairs that coordinate close to the $(GlcNAc)_2Me_2$ monomer than 8:2 IL:co-solvent OES components, followed by the 2:8 IL:co-solvent OES and then the pure co-solvents. It further highlighted how important the cation is in biomass interaction and potentially dissolution, since more cations coordinate close to the monomer than anions, although the anions solvate within a closer radius. The RDF results also showed the various solvent components to be the most densely packed close to the O6 oxygen atoms from the sugar, while coordinating the least around the O3 (reducing) oxygen atom. The hydrogen bond analysis results support the above, while also producing the most number of average intermolecular hydrogen bonds between $(GlcNAc)_2Me_2$ and $[C_4C_1im][CH_3COO]$, followed by the 8:2 IL:co-solvent OESs, the 2:8 IL:co-solvent OESs and then the pure co-solvents.

When analysing the PMF curves of separating two $(GlcNAc)_2Me_2$ monomers in the pure solvent, plenty of results can be extracted. In $[C_4C_1im][CH_3COO]$ the dimer requires the least amount of thermodynamic work to separate, while in dimethyl carbonate it requires the most. We also see the monomers having a high initial energy barrier to cross in pure propylene carbonate. For the OESs, however, the PMF curves are very similar in shape. We see that using various initial configurations for the mixtures will produce PMF results that are significantly different (in terms of well-depth), thus suggesting that an averaging over multiple states are required. Since the well-depths and energy barriers for all PMF curves are smaller than RT, the dimer should nonetheless separate spontaneously in each of the solvent systems.

Overall, the suggestion is that $(GlcNAc)_2Me_2$ interacts more favourably with $[C_4C_1im][CH_3COO]$ than with the 8:2 IL:co-solvent OESs, followed by the 2:8 IL:co-solvent OESs and then the pure co-solvents. This agrees with the (cellulose) dissolution results from Gale *et al.*[97] Determining the exact order of solute-solvent interaction was unfortunately not possible, but longer simulation time and applying the improvements as discussed in Chapter 9, might make it feasible in the future.

[Page intentionally left blank.]

9 Future Work

While PMF calculations are highly accurate and useful in getting information on a system, PMF calculations have to be well-sampled and well-defined. With umbrella sampling, this requires a good overlap between windows and sufficiently long simulation times. With a GPU cluster being available at the CHPC now, future work can include using more restraint windows and longer simulation times.

Moreover, in terms of the mixtures, better averaging should be done with the PMF curves. For each mixture, multiple simulation boxes should be considered, each with different initial configurations. A Boltzmann distribution can then be used to obtain a better sampled and well-defined PMF curve. Alternative methods for calculating the PMF should also be investigated, including using metadynamics, alchemical free energy analysis, MBAR, thermodynamic integration, etc. For the solvation of a $(GlcNAc)_2Me_2$ monomer, better sampling should be done for the mixtures. Multiple simulation boxes should be considered for the mixtures and a Boltzmann distribution used to obtain an average result for the solvation energies, radial distribution functions and hydrogen bond analyses. This was shown to work well in Chapter 6.3.1. Furthermore, longer N-acetylglucosamine chains should be used for future MD studies as it will be more representative of the full chitin crystal structure. Moreover, when modelling the dissolution of chitin, stacking (interactions) should also be factored into the simulation as it has been shown to significantly contribute to why biopolymers are difficult to dissolve.

Experimentally, the dissolution of chitin should be investigated as to compare it to the PMF results; while other methods should be explored for further validating the interaction energy results. Isothermal scanning calorimetry (ISC) and time-domain NMR (TD-NMR) are two methods suggested here. A recent study has also considered the effect of an electrical field on the interaction between cellulose and LiCl (in DMAc) and has produced promising results. It has shown that in the presence of an electrical field, anions interact stronger with cellulose. Future work can be to study the electrical field effect on chitin in various organic electrolyte solutions and to investigate the effect of anion interaction. This study can be done on a computational- and experimental level.

[Page intentionally left blank.]

10 References

1. Höök, M.; Tang, X. *Energy Policy* **2013**, *52*, 797-809.
2. Is Solar Energy Expensive Energy http://www.soventix.co.za/news/solar-energy-expensive-energy (accessed on Mar 11, 2019).
3. Fatehi, P. In *Cellulose - Biomass Conversion*; 2019.
4. 12 Design Principles of Green Chemistry https://www.acs.org/content/acs/en/greenchemistry/principles/12-principles-of-green-chfemistry.html (accessed Feb. 22, 2019).
5. Biofuel Facts and Information | National Geographic https://www.nationalgeographic.com/environment/global-warming/biofuel/ (accessed Apr. 30, 2019.
6. Alonso, D.; Wettstein, S.; Bond, J.; Root, T.; Dumesic, J. *ChemSusChem* **2011**, *4*, 1078-1081.
7. Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2581-2583.
8. Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802-11808.
9. Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143-168.
10. Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247-3276.
11. Klemm D.; Heublein, B.; Fink, H.; Bohn, A. *J. ChemInform.* **2005**, *44*, 3358-3393.
12. Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.
13. Parthasarathi, R.; Bellesia, G.; Chundawat, S.; Dale, B.; Langan, P.; Gnanakaran, S. *J. Phys. Chem. A* **2011**, *115*, 14191-14202.
14. The Many Forms of Cellulose https://biocyclopedia.com/index/plant_pathways/the_many_forms_of_cellulose_a_brief_introduction.php (accessed Mar. 11, 2019).
15. Jafari, S.; Fathi, M.; Mandala, I. Emerging product formation. (accessed Jun. 17, 2019).
16. Li, J.; Revol, J.; Marchessault, R. *J. Appl. Polym. Sci.* **1997**, *65*, 373-380.
17. Cellulose http://www1.lsbu.ac.uk/water/cellulose.html (accessed Apr. 30, 2019).
18. Kaya, M.; Mujtaba, M.; Ehrlich, H.; Salaberria, A.; Baran, T.; Amemiya, C.; Galli, R.; Akyuz, L.; Sargin, I.; Labidi, J. *Carbohydr. Polym.* **2017**, *176*, 177-186.
19. Sikorski, P.; Hori, R.; Wada, M. *Biomacromolecules* **2009**, *10*, 1100-1105.
20. Seddiqi, H.; Honarkar, H. *Iran Polym. J.* **2014**, *23*, 307-326.
21. Somorin, O.; Nishi, N.; Tokura, S.; Noguchi, J. *Polym. J.* **1979**, *11*, 391-396.
22. Roberts, G. (1992). Roberts, G. In *Chitin Chemistry*; Roberts, G., 1st Ed.; Macmillan Education UK: London, 1992; pp. 274-329.
23. Roy, J.; Salaün, F.; Giraud, S.; Ferri, A. In *Solubility of Polysaccharides*; Xu, Z.; Guo, M.; Hu, X.; Wang, C.; Ai, L.; Joanna, K.; Pandithavidana, D.; Sameera, W.; von Borries-Medrano, E.; Jaime-Fonseca, M.; Aguilar-Mẽndez, M.; Wu, S.; Roy, J.; Salaün, F.; Giraud, S.; Ferri, A.; 2017; pp. 110-127.
24. Igarashi, K.; Wada, M.; Samejima, M. *FEBS J.* **2007**, *274*, 1785-1792.
25. Yoneda, Y.; Mereiter, K.; Jaeger, C.; Brecker, L.; Kosma, P.; Rosenau, T.; French, A. *J. Am. Chem. Soc.* **2008**, *130*, 16678–16690.
26. Isik, M.; Sardon, H.; Mecerreyes, D. *Int. J. Mol. Sci.* **2014**, *15*, 11922-11940.
27. Mohd, N.; Draman, S.; Salleh, M.; Yusof, N. *AIP Conf. Proc.* **2017**, *1809*, 020035.
28. Hydrothermal liquefaction of biomass: Developments from batch to continuous process http://my.rsc.org/blogs/84/1168 (accessed Apr. 2, 2019).
29. Basic Information about Anaerobic Digestion (AD) https://www.epa.gov/anaerobic-digestion/basic-information-about-anaerobic-digestion-ad (accessed Apr. 16, 2019).
30. Anaerobic Digestion https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/anaerobic-digestion (accessed Apr. 16, 2019).
31. Seddon, K.; Rogers, R. *Science* **2003**, *302*, 792-793.

32. Freemantle, M. An Introduction to Ionic Liquids; RSC Publishing: Cambridge, 2010.
33. Seddon, K. *J. Chem. Tech. Biotechnol.* **1997**, *68*, 351-356.
34. Seddon, K.; Plechkova, N. *Chem. Soc. Rev.* **2008**, *37*, 123-150.
35. Liu, J.; Jiang, G.; Jonsson, J. *TrAC.* **2005**, *24*, 20-27.
36. Kubisa, P. *Prog. Polym. Sci.* **2004**, *29*, 3-12.
37. Zhu, S.; Wu, Y.; Chen, Q.; Yu, Z.; Wang, C.; Jin, S.; Ding, Y.; Wu, G. *GreenChem.* **2006**, *8*, 325-327.
38. Brennecke, J.; Maginn, E. *AIChE Journal* **2001**, *47*, 2384-2389.
39. Welton, T. *Biophys. Rev.* **2018**, *10*, 691-706.
40. Walden, P. *Bull. Acad. Imper. Sci.* **1914**, *8*, 405-422.
41. Parshall, G. *J. Am. Chem. Soc.* **1972**, *94*, 8716-8719.
42. Rosenthal, K.; Storm, D.; Ford, W. *Antimicrob. Agents Chemother.* **1975**, *8*, 510-512.
43. Magnuson, D.; Bodley, J.; Evans, D. *J. Solut. Chem.* **1984**, *13*, 583-587.
44. Poole, C.; Kersten, B.; Ho, S.; Coddens, M.; Furton, K. *J. Chromatogr.* **1986**, *352*, 407.
45. Wilkes, J.; Levisky, J.; Wilson, R.; Hussey, C. *Inorg. Chem.* **1982**, *21*, 1263-1264.
46. Hunt, P. *Top. Curr. Chem.* **2017**, *375*, 59.
47. Boon, J.; Levisky, J.; Pflug, J.; Wilkes, J. *J. Org. Chem.* **1986**, *51*, 480-483.
48. Jaeger, D.; Tucker, C. *Tetrahedro Lett.* **1989**, *30*, 1785-1788.
49. Welton, T. *Chem. Rev.* **1999**, *99*, 2071-2083.
50. Sheldon, R. *Chem. Commun.* **2001**, *0*, 2399-2407.
51. Huddleston, J.; Rogers, R. *Chem. Commun.* **1998**, *0*, 1765-1766.
52. Visser, A.; Swatloski, R.; Reichert, W.; Mayton, R.; She, S.; Wierzbicki, A.; Davis, J.; Rogers, R. *Chem. Commun.* **2001**, *0*, 135-136.
53. Galiński, M.; Lewandowski, A.; Stpniak, I.; *Electrochim. Acta* **2006**, *51*, 5567-5580.
54. Blanchard, L.; Hancu, D.; Beckman, E.; Brennecke, J. *Nature* **1999**, *399*, 28-29.
55. Introduction to Ionic Liquids https://www.sigmaaldrich.com/technical-documents/articles/chemfiles/ionic-liquids0.html (accessed Apr. 16, 2019).
56. Hulsbosch, J.; De Vos, D.; Binnemans, K.; Ameloot, R. A*CS Sustainable Chem. Eng.* **2016**, *4*, 2917-2931.
57. Mallakpour, S.; Rafiee, Z. *J. Polym. Environ.* **2011**, *19*, 447-484.
58. Roller, P. *J. Phys. Chem.* **1935**, *39*, 221.
59. Colburn, A. *Trans. Inst. Chem. Engrs.* London 1933, 29, 174.
60. Wagner, C. *J. Phys. Chem.* **1949**, *53*, 1030.
61. Wurster, D.; Taylor, P. *J. Pharm. Sci.* **1965**, *54*, 169-175.
62. Viscosity | Definition of Viscosity by Mirriam-Webster https://www.merriam-webster.com/dictionary/viscosity (accessed on Jun. 28, 2018).
63. Mass Transport and Dissolution http://rmnj.blogspot.com/2012/12/mass-transport-and-dissolution.html (accessed Feb. 25, 2019.)
64. Kunz, W.; Häckl, K. *Chem. Phys. Lett.* **2016**, *661*, 6-12.
65. Seddon, K.; Stark, A.; Torres, M. *Pure Appl. Chem.* **2000**, *72*, 2275–2287.
66. Swatloski, R.; Spear, S.; Holbrey, J.; Rogers, R. J. *Am. Chem. Soc.* **2002**, *124*, 4974-4975.
67. Ding, Z.; Chi, Z.; Gu, W.; Gu, S.; Liu, J.; Wang, H. *Carbohydr. Polym.* **2012**, *89*, 7-16.
68. Hanley, M.; Green, J.; Fylstra, P.; Fox, D.; Henderson W.; de Long, H.; Trulove, P. *In 210th Meeting of the Electrochemical Society*, Cancun, Mexico, Conference Materials, Abstract **2006**.
69. Heinze, T.; Schwikal, K.; Barthel, S. *Macromol. Biosci.* **2005**, *5*, 520–525.
70. Sashina, E.; Kashirskii, D.; Zaborski, M.; Jankowski, S. *Russ. J. Gen. Chem.* **2012**, *82*, 1994–1998.
71. Wang, W.; Zhu, J.; Wang, X.; Huang, Y.; Wang, Y. *J. Macromol. Sci. B* **2010**, *39*, 528-541.

72. Jaworska, M.l Kozlecki, T.; Gorak, A. J. *Polym. Eng.* **2012**, *32*, 67-69.
73. Qin Y, Lu X, Sun N, Rogers RD. *Green Chem.* **2010**, *12*, 968–971.
74. Zhu, C.; Richardson, R.; Song, Y.; Rahatekar, S. One Step Dissolution, Extrusion, And Fiber Spinning Of Chitin Using Ionic Liquid Solvents. *Polysaccharide-based Fibers and Composites* 2017, 13-27.
75. Payal, R.; Bejagam, K.; Mondal, A.; Balasubramanian, S. *Phys. Chem. B* **2015**, *119*, 1654-1659.
76. Gross, A.; Bell, A.; Chu, J. *J. Phys. Chem. B*, **2011**, *115*, 13433–13440.
77. Bergenstråhle, M.; Wohlert, J.; Himmel, M.; Brady, J. *Carbohydr. Res.* **2010**, *14*, 2060-2066.
78. Gray, M.; Converse, A. *Appl. Biochem. Biotechnol.* **2003**, *105-108*, 179-193.
79. Rabideau, B.; Agarwal, A.; Ismail, A. *J. Phys. Chem. B* **2014**, *118*, 1621-1629.
80. Lu, B.; Xu, A.; Wang, J. *Green Chem.* **2014**, *16*, 1326-1335.
81. Zhang, J.; Wu, J.; Yu, J.; Zhang, X.; He, J.; Zhang, J. *Mater. Chem. Front.* **2017**, *1*, 1273-1290.
82. Liu, Z.; Wang, H.; Li, Z.; Lu, X.; Zhang, X.; Zhang, S.; Zhou, K. *Mater. Chem. Phys.* **2011**, *128*, 220–227.
83. Payal, R.; Bharath, R.; Periyasamy, G.; Balasubramanian, S. *J. Phys. Chem. B* **2012**, *116*, 833–840.
84. Li, J.; Lu, Y.; Yang, D.; Sun, Q.; Liu, Y.; Zhao, H. *Biomacromolecules* **2011**, *12*, 1860–1867.
85. Stark, A.; Sellin, M.; Ondruschka, B.; Massonne, K.; *Sci. China: Chem.* **2012**, *55*, 1663–1670.
86. Pinkert, A.; Marsh, K.; Pang, S. *Ind. Eng. Chem. Res.* **2010**, *49*, 11121–11130.
87. Muraoka, J.; Kamiya, N.; Ito, Y. *J. Mol. Liq.* **2013**, *182*, 76–78.
88. Lovell, C.; Walker, A.; Damion, R.; Radhi, A.; Tanner, S.; Budtova, T.; Ries, M. *Biomacromolecules* **2010**, *11*, 2927–2935.
89. Ries, M.; Radhi, A.; Keating, A.; Parker, O.; Budtova, T. *Biomacromolecules* **2014**, *15*, 609–617.
90. Xu, A.; Cao, L.; Wang, B. *Carbohydr. Polym.* **2015**, *125*, 249–254.
91. Lu, B.; Xu, A.; Wang, J. *Green Chem.* **2014**, *16*, 1326–1335.
92. Okushita, K.; Chikayama, E.; Kikuchi, J. *Biomacromolecules* **2012**, *13*, 1323–1330.
93. Chang, H.; Zhang, R.; Hsu, D. *Phys. Chem. Chem. Phys.* **2015**, *17*, 27573–27578.
94. Cho, H.; Gross, A.; Chu, J. *J. Am. Chem. Soc.* **2011**, *133*, 14033–14041.
95. Rinaldi, R. *Chem. Commun.* **2011**, *47*, 511-513.
96. Bioni, T.; Arêas, E.; Couto, L.; Favarin, G.; El Seoud, O. *Nord. Pulp Pap. Res. J.* **2015**, *30*, 105-111.
97. Gale, E; Wirawan, R.; Silveira, R.; Pereira, C.; Johns, M.; Skaf, M.; Scott, *J. ACS Sustain. Chem. Eng.* **2016**, *4*, 6200-6207.
98. Vitz, J.; Erdmenger, T.; Haensch, C.; Schubert, U. *Green Chem.* **2009**, *11*, 417-424.
99. Uto, T.; Yamamoto, K.; Kadokawa, J. *J. Phys. Chem. B* **2018**, *122*, 258-266.
100. Pinkert, A.; Marsh, K.; Pang, S.; Staiger, M. *Chem. Rev.* **2009**, *109*, 6712-6728.
101. Miller-Chou, B.; Koenig, J. *Prog. Polym. Sci.* **2003**, *28*, 1223-1270.
102. Hulsey, S.; Absar, S.; Choi, H. *Procedia Manuf.* **2017**, *10*, 652-661.
103. Semenova, M.; Dickinson, E. *Biopolymers in Food Colloids*; 1st ed.; Brill, 2014.
104. Carson, L.; Chau, P.; Earle, M.; Gilea, M.; Gilmore, B.; Gorman, S.; McCann, M.; Seddon, K. *Green Chem.* **2009**, *11*, 492–497.
105. Liu, F.; Li, L.; Yu, S.; Lv, Z.; Ge, X.; *J. Hazard. Mater.* **2011**, *189*, 249–254.
106. El Seoud, O.; Fidale, L.; Ruiz, N.; D' Almeida, M.; Frollini, E. *Cellulose* **2008**, *15*, 371-392.
107. Huglin, M.; Pass, D. *J. Appl. Polym. Sci.* **1968**, *12*, 473-485.
108. Leach, A. *Molecular modelling; 2nd ed.*; Pearson/Prentice Hall: Harlow [u.a.], 2009.

109. Phillips, A. *Introduction to Quantum Mechanics*; Wiley: Hoboken, 2013.
110. Griffiths, D. *Introduction to quantum mechanics*; Prentice Hall: Englewood Cliffs, N.J., 1995.
111. Chapters 11: Computational Quantum Chemistry 11.1: Gaussian Basis Sets https://chem.libretexts.org/Courses/Colorado_State_University/Chem_476%3A_Physical_Chemistry_II_(Levinger)/Chapters/11%3A_Computational_Quantum_Chemistry/11.1%3A_Gaussian_Basis_Sets (accessed Apr. 9, 2019).
112. Stewart, R.; Hehre, W.; Pople, J. *J. Chem. Phys.* **1969**, *51*.
113. Ditchfield, R.; Hehre, W.; Pople, J. *J. Chem. Phys.* **1971**, *54*, 724-728.
114. Bauzà, A.; Quiñonero, D.; Deyá, P.; Frontera, A. *J. Phys. Chem. A* **2013**, *117*, 2651-2655.
115. Hoower, W. *Molecular dynamics*, 1st ed.; Springer: Berlin, 1986.
116. Jensen, F. *Introduction to computational chemistry*, 2nd ed.; Wiley: Chichester [u.a.], 2007.
117. Young, D. *Computational chemistry*, 1st ed.; Wiley: New York, 2001.
118. Kühne, T. *Chem. Phys.* **2013**, *0*, 1-13.
119. Car-Parrinello molecular dynamics http://www.acmm.nl/ensing/thesis/node10.html (accessed Apr. 10, 2019).
120. Müller-Plathe, W.; van Gunsteren, W. *J. Chem. Phys.* **1995**, *103*, 4745–4756.
121. Zhao, W.; Eslami, H.; Cavalcanti, W.; MJller-Plathe, F.; *Z. Phys. Chem.* **2007**, *221*, 1647–1662.
122. Sieffert, N.; Wipff, G. *J. Phys. Chem. B* **2006**, *110*, 13076–13085.
123. Chevrot, G.; Schurhammer, R.; Wipff, G. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4166–4174.
124. Schröder, C. *Phys. Chem. Chem. Phys.* **2012**, *14*, 3089-3102.
125. Chaudhari, M.; Nair, J.; Pratt, L.; Soto, F.; Balbuena, P.; Rempe, S. *J. Chem. Theory Comput.* **2016**, 12, 5709-5718.
126. Quevillon, M.; Whitmer, J. *Materials* **2018**, *11*, 64.
127. Long, C.; Ewing, G. *J. Chem. Phys.* **1973**, *58*, 4824.
128. Howard, B. *Faraday Discuss Chem. Soc.* **1981**, *71*, 23-29.
129. Kirschner, K.; Yongye, A.; Tschampel, S.; González-Outeiriño, J.; Daniels, C.; Foley, B.; Woods, R. *J. Comput. Chem.* **2008**, *29*, 622-655.
130. Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
131. Lii, J.; Ma, B.; Allinger, N. *J. Comput Chem.* **1999**, *20(15)*, 1593–1603.
132. Bengtsson, J.; Olsson, C.; Hedlund, A.; Köhnke, T.; Bialik, E. *J. Phys. Chem. B* **2017**, *121*, 11241-11248.
133. Zhao, Y.; Liu, X.; Wang, J.; Zhang, S. *ChemPhysChem.* **2012**, *13*, 3126-3133.
134. Strelcová, Z; Kulhánek, P.; Friák, M; Febritius, O.; Petrov, M.; Neugebauer, J.; Koca, J. *RSC Adv.* **2016**, *6*, 30710-30721.
135. Sprenger, K.; Jaeger, V.; Pfaendtner, J. *J. Phys. Chem. B* **2015**, *119*, 5882-5895.
136. Jia, X.; Zhang, J.; Mei, Y. *J. Mol. Model.* **2013**, *19*, 2355-2361.
137. Peri, S.; Karim, N.; Khare, R. *Carbohydr. Res.* **2011**, *346*, 867-871.
138. The smooth particle mesh Ewald method http://www.chim.unifi.it/orac/MAN/node20.html (accessed Nov. 5, 2018).
139. Genetic Algorithms Introduction https://www.tutorialspoint.com/genetic_algorithms/genetic_algorithms_introduction.htm (accessed Nov. 9, 2018).
140. Davis, L. *Handbook of genetic algorithms*; Van Nostrand Reinhold: New York, 1991.
141. Freeman, R.; Wu, X. *J. Magn. Reson.* **1987**, *75*, 184-189.
142. Wu, L.; Freeman, R. *J. Magn. Reson.* **1989**, *85*, 414-420.
143. Wu, L.; Xu, P.; Freeman, R. *Magn. Reson. Med.* **1991**, *20*, 165-170.
144. Lucasius, C.; Kateman, G. *Trends Analyt. Chem.* **1991**, *10*, 254-261.

145. Hibbert, B. *Chemom. Intell. Lab. Syst.* **1993**, *19*, 277-293.
146. Cramer, C. Essentials of computational chemistry; 2nd ed.; Wiley: Chichester [u.a.], 2014.
147. O'Boyle, N.; Vandermeersch, T.; Flynn, C.; Maguire, A.; Hutchison, G. *J. Cheminform.* **2011**, *3*, 1-9.
148. O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. *J. Cheminform.* **2011**, *3*, 1-14.
149. Development of the Second Generation of the General AMBER Force Field https://www.researchgate.net/project/Development-of-the-Second-Generation-of-the-General-AMBER-Force-Field (accessed Jun. 21, 2018).
150. Ionic Liquids Databse - ILThermo https://ilthermo.boulder.nist.gov (accessed Jun. 28, 2018).
151. Martìnez, L; Andrade, R; Birgin, E.; Martìnez, J. *J. Comput. Chem.* **2009**, *30*, 2157-2164.
152. Coccotti, R.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327-341.
153. Malberg, F.; Hollóczki, O.; Thomas, M.; Kirchner, B. *Nature* **2015**, *26*, 1343-1349.
154. Liu, H.; Sale, K.; Holmes, B.; Simmons, B.; Singh, S. *J. Phys. Chem. B* **2010**, *114*, 4293-4301.
155. Brehm, M.; Kirchner, B. *J. Chem. Inf. Model.* **2011**, *51*, 2007-2023.
156. RED Server Development https://upjv.q4md-forcefieldtools.org/REDServer-Development/ (accessed Jun. 22, 2019).
157. Propylene Carbonate Solvent Properties http://macro.lsu.edu/HowTo/solvents/Propylene%20Carbonate.htm (accessed Jun. 28, 2018).
158. Dimethyl Carbonate Chemical Properties https://www.chemicalbook.com/ProductChemicalPropertiesCB8853983_EN.htm (accessed Jun. 28, 2018).
159. gamma-Valerolactone CAS#: 108-29-2 https://www.chemicalbook.com/ProductChemicalPropertiesCB4285373_EN.htm (accessed on Jun. 28, 2018).
160. Safarov, J.; Geppert-Rybczynska, M.; Kul, I.; Hassel, E. *Fluid Phase Equilib.* **2014**, *383*, 144-155.
161. Pereiro, A.; Verdia, P.; Tojo, E.; Rodriguez, A. *J. Chem. Eng. Data* **2007**, *52*, 377-380.
162. Acree, W.; Chickos, J. *J. Phys. Chem. Ref. Data* **2016**, *45*, 033101.
163. Wei, J.; Bu, X.; Guan, W.; Xing, N.; Fang, D.; Wu, Y. *RSC Adv.* **2015**, *5*, 70333.
164. Andanson, J.-M.; Traikia, M.; Husson, P. *J. Chem. Thermodyn.* **2014**, *77,* 214-221.
165. Mehmandoust, B.; Sanjari, E.; Vatani, M. *J. Adv. Res.* **2013**, *5*, 261-269.
166. Derecskei, B.; Derecskei-Kovacs, A. *Mol. Simul.* **2007**, *32*, 109-115.
167. Keller, M.; Maginn, E. *J. Phys. Chem. B Lett.* **2007**, *111*, 9424-9427.
168. Ramraj, A.; Raju, R.; Wang, Q.; Hillier, I.; Bryce, R.; Vincent, M. *J. Mol. Graph. Model.* **2010**, *29*, 321-325.
169. Singh, A.; Tessier, M.; Pederson, K.; Wang, X.; Venot, A.; Boons, G.; Prestegard, J.; Woods, R. *Can. J. Chem.* **2016**, *94*, 927-935.
170. Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate complexes www.julienmichel.net/lab/assets/docs/JCTC-2015-Glycam-lectin-OA.pdf (accessed Apr. 12, 2019).
171. Roe, D.; Cheatham, T. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.
172. Skarmoutsos, I.; Dellis, D.; Matthews, R.; Welton, T.; Hunt, P. *J. Phys. Chem. B* **2012**, *116*, 4921-4933.
173. CPPTRAJ Hydrogen Bond Analysis Tutorial http://www.amber.utah.edu/AMBER-workshop/London-2015/Hbond/ (accessed Oct. 27, 2017).
174. Pishehvarz, G.; Erfan-Niya, H.; Zaminpayma, E. *Comput. Mater. Sci.* **2018**, *155*, 499-523.
175. Zhang, J.; Zhao, X.; Huo Liu, Q. *RSC Adv.* **2016** *6*, 90388.
176. Chatel, G.; Pereira, J.; Debbeti, V.; Wang, H.; Rogers, R. *Green Chemistry* **2014**, *16*, 2051.
177. Trzesniak, D.; Kunz, A.; van Gunsteren, W. *ChemPhysChem.* **2007**, *8*, 162– 169.

178. Wong, K.; York, D. *J. Chem. Theory Comput.* **2012**, *8*, 3998–4003.
179. van Gunsteren, W.; Daura, X.; Mark, A. *Helvetica Chimica Acta* **2002**, *85*, 3113-3129.
180. Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690-2693.
181. Thermodynamic integration method - Apte.pdf http://chemeng.iisc.ac.in/molsim/Apte.pdf (accessed Jan 28, 2018).
182. Kollman, P. *Chem. Rev.* **1993**, *93*, 2395-2417.
183. Zhu, F.; Hummer, G. *J. Comput. Chem.* **2011**, *33*, 453-465.
184. Shankar, K.; Rosenberg, J.; Bouzida, D.; Swendsen, R.; Kollman, P. *J. Comput. Chem.* **1992**, *13*, 1011-1021.
185. AMBER-Umbrella_COM_restraint_tutorial https://github.com/callumjd/AMBER-Umbrella_COM_restraint_tutorial (accessed Jun 12, 2019).
186. Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954-9960.
187. Mason, P.; Neilson, G.; Enderby, J.; Saboungi, M.; Brady, J. *J. Phys. Chem. B* **2005**, *109*, 13104–13111.
188. Chandler, D. *Introduction to modern statistical mechanics*; Oxford Univ. Press: New York, 1987.

[Page intentionally left blank.]

11 Appendix 1: Conformer- and Genetic Algorithm Scripts

The following chapter contains the scripts used for the generating- and analysing conformers needed for multi-conformer partial charge fitting, as well as the scripts used for the torsional profile fitting.

11.1 Generating Conformers

Once the geometry optimisation of the ionic components finished, the output file (`.log`) was converted to a `.mol` file.

```
obabel -ig09 MOLECULE.log -omol -OMOLECULE.mol #Convert .log file to .mol file

#Correct Charges in .mol file in acetate for example
#M  CHG  1   1   1 #Put a positive charge on the carbonyl carbon atom
#M  CHG  1   2  -1 #Put a negative charge on the oxygen atom
#M  CHG  1   3  -1 #Put a negative charge on the oxygen atom
#M  END
```

In the `.mol` file, with respect to the acetate anion, the charge sign on the carbonyl carbon and oxygen atoms are corrected. As such, by resonance, each oxygen atom will have a negative charge, while the carbon atom will have a positive charge. Overall this sums to a negative charge for the acetate anion. If the charge correction is not implemented, in the next step where the `Confab` algorithm is run, extra hydrogen atoms will be places on these atoms. This is unwanted. The conformers were generated next for the ionic components, in which from 5000 possible conformers were searched those that are below an energy of 50.0 kcal/mol. All the conformers that lie below this energy are outputted to a `.sdf` file.

```
obabel MOLECULE.mol --confab -OMOLECULE_conformers.sdf --ecutoff 50.0 --conf 5000
#Use confab to search from 5000 conformers,  those that are below energy  cutoff
↪  of 50.0 kcal/mol
```

Hereafter, each of the conformers outputted in the `.sdf` file that falls below an energy cut-off of 50 kcal/mol, were geometry optimised at the MP2/6-31G* level of theory. All the `.log` files from the geometry optimisation were then placed into the same directory and analysed through the following `Python` script. This script sorts the conformers in terms of their energy, weighs them based on a global minimum, and outputs a Boltzmann distribution of the unique conformers, confirming which conformers and how many conformers have similar energies.

```
###Analyze Conformers
import sys,  os
sys.path.append('/home/people/USER/miniconda2/lib/python2.7/site-packages')
from file_read_backwards import FileReadBackwards
from operator import itemgetter
```

```
from math import exp

# setup & constants
output = file('energies.txt', 'w') #The structural energies are all written to
↪  this file
energies = []
seen = set()
RT = (8.314/1000.0)*298.15

# numerical counter over files
lst = range(1, 29)

# cut-off for inclusion in list
cutoff = 5.0

# rounding for energies (in au)
rnd = 6

# read energies
for n in lst:
    f = FileReadBackwards("conf-{:d}.log".format(n, n),  encoding="utf-8")
    for line in f: #the energies of the conformers are read here
        if line.find('EUMP2 =') > -1:
            energy = round(float(line.split()[5].replace("D", "E")), rnd) * 2625.5
            energies.append((n,  energy))
            break
        if line.find('SCF Done:') > -1:
            energy = round(float(line.split()[4]), rnd) * 2625.5
            energies.append((n,  energy))
            break

# get global minimum energy
global_min = map(min,  zip(*energies))[1]

# sort energies
sorted_energies = sorted(energies,  key=itemgetter(1))

# calculate Boltzmann probabilities of unique elements (energies)
unique_energies = [(n,  e - global_min) for (n,  e) in sorted_energies if not (e
↪  in seen or seen.add(e))]
bf = [exp(-e/RT) for (n,  e) in unique_energies]
bsum = sum(bf)
p = [x/bsum for x in bf]

# print the conformers with their relative energies
for (n,  energy) in sorted_energies:
    delta_energy = energy - global_min
    x = ''
```

```
    if delta_energy < cutoff:
        x = '<'
    if delta_energy == 0.0:
        x = '0'
    if energies[1].count(energy) >= 2:
        idx = energies.index(energy)
        x = '={:d}'.format(idx + 1)
    output.write('{:2d},  {:6.2f},  {:s}\n'.format(n,  delta_energy,  x))

output.write('############################\n')
output.write('## Boltzmann distribution ##\n')
output.write('############################\n')

i = 0
for (n,  energy) in unique_energies:
    output.write('{:2d},  {:6.2f},  {:6.1f} %\n'.format(n,  energy,  100*p[i]))
    i += 1 #The relative energies for all conformers are produced here,  and which
    ↪  ones are unique.
```

In the `energies.txt` file can now be viewed the relative energies of all conformers, and which ones are unique.

11.2 The Genetic Algorithm

After performing the scan of a target dihedral angle (as mentioned in Chapter 5), the relative energies as well as the resultant structures from the output are saved using the following `python` script.

```
import argparse
import openbabel as ob
import os
from cclib.parser import Gaussian
from cclib.parser import utils

#Here the parameter that needs to analysed is chosen
def get_parameters(show = True):
    parser = argparse.ArgumentParser()
    parser.add_argument('-b',  '--bonds',  action='append',  nargs=2,  type=int)
    parser.add_argument('-a',  '--angles',  action='append',  nargs=3,  type=int)
    parser.add_argument('-d',  '--dihedrals',  action='append',  nargs=4,
    ↪  type=int)
    parser.add_argument('-i',  '--input',  action='store',  type=str)
    args = parser.parse_args()
    if show:
        print("Bonds: {}".format(args.bonds))
        print("Angles: {}".format(args.angles))
        print("Dihedrals: {}".format(args.dihedrals))
        print("Log file: {}".format(args.input))
```

```
    parameters = []
    if args.bonds != None:
        for bond in args.bonds:
            bond.extend([0]  * (4 - len(bond)))
            parameters.append(bond)
    if args.angles != None:
        for angle in args.angles:
            angle.extend([0]  * (4 - len(angle)))
            parameters.append(angle)
    if args.dihedrals != None: #We are considering dihedrals,  so only this one
    ↪  will be used.
        for dihedral in args.dihedrals:
            parameters.append(dihedral)
    return (args,  parameters)

def make_obmol(atomcoords,  atomnos):
    obmol = ob.OBMol()
    for n in range(len(atomnos)):
        coords = atomcoords[n].tolist()
        atomno = int(atomnos[n])
        obatom = ob.OBAtom()
        obatom.SetAtomicNum(atomno)
        obatom.SetVector(*coords)
        obmol.AddAtom(obatom)
    return obmol

def measure_parameter(obmol,  atomno_1=0,  atomno_2=0,  atomno_3=0,  atomno_4=0):
    if atomno_3 == 0:
        bond = obmol.GetAtom(atomno_1).GetDistance(atomno_2)
        return bond
    if atomno_4 == 0:
        angle = obmol.GetAtom(atomno_1).GetAngle(atomno_2,  atomno_3)
        return angle
    else:
        dihedral = obmol.GetTorsion(atomno_1,  atomno_2,  atomno_3,  atomno_4)
        if dihedral < 0.0:
            dihedral += 360.0
        return dihedral #the dihedral angles from the .log file is outputted.

if __name__ == '__main__':
    obConversion = ob.OBConversion()
    obConversion.SetInAndOutFormats("pdb",  "pdb") #structures with different
    ↪  torsionals are saved as .pdb files
    (args,  parameters) = get_parameters(show=True)
    p = Gaussian(args.input)
    f = file("gaussian_energies.dat", "w") #relative energies (and torsional
    ↪  angles) are written to this file
    if len(parameters) > 1:
```

```
    multi_param = True
else:
    multi_param = False
data = p.parse()
idx = [ i for i,  x in enumerate(data.optstatus) if x == 2 ]
scf = [ data.scfenergies[i] for i in idx ]
rel_scf_kcal = [ utils.convertor(i - min(scf),  "eV",  "kcal") for i in scf ]
 ↪  # energies are converted to kcal
rel_scf_kj = [ utils.convertor(i - min(scf),  "eV",  "kJmol-1") for i in scf ]
molecules = [ make_obmol(data.atomcoords[i],  data.atomnos) for i in idx ]
if not os.path.exists('qm_data'):
    os.makedirs('qm_data')
f.write("#{:^12s}{:^12s}".format('kcal/mol', 'kJ/mol')) #converted energies
 ↪  are written
for parameter in parameters:
    if parameter[2] == 0:
        f.write('{:10d}-{:d}'.format(parameter[0],  parameter[1]))
    elif parameter[3] == 0:
        f.write('{:8d}-{:d}-{:d}'.format(parameter[0],  parameter[1],
         ↪  parameter[2]))
    else:
        f.write('{:6d}-{:d}-{:d}-{:d}'.format(parameter[0],  parameter[1],
         ↪  parameter[2],  parameter[3]))
f.write("\n")
if multi_param:
    parameter_values = [ [ measure_parameter(mol,  parameter[0],
     ↪  parameter[1],  parameter[2],  parameter[3]) for mol in molecules ]
                         for parameter in parameters ]
    parameter_values = map(list, map(None, *parameter_values))
    for index,  values in enumerate(parameter_values):
        f.write("{:12.7f}".format(rel_scf_kcal[index]))
        f.write("{:12.7f}".format(rel_scf_kj[index]))
        for value in values:
            f.write("{:12.2f}".format(value))
        f.write("\n")
        obConversion.WriteFile(molecules[index],
         ↪  './qm_data/scan_{:03d}.pdb'.format(index))
else:
    parameter = parameters[0]
    parameter_values = [ measure_parameter(mol,  parameter[0],  parameter[1],
     ↪  parameter[2],  parameter[3]) for mol in molecules ]
    for index,  value in enumerate(parameter_values):
        f.write("{:12.7f}".format(rel_scf_kcal[index]))
        f.write("{:12.7f}".format(rel_scf_kj[index]))
        f.write("{:12.2f}\n".format(value))
        obConversion.WriteFile(molecules[index],
         ↪  './qm_data/scan_{:03d}.pdb'.format(index))
```

Hereafter, the `.mol2` file with atomic partial charges as obtained by the R.E.D. Server was copied over to the working directory and initial GAFF2 parameters generated (`.frcmod` file). For each tagged dihedral, the dihedral is "zeroed" in the `.frcmod` file as to start the fitting process. For example, if the initial tagged dihedral was `X1-X2-X3-X4 6 6.800 180.000 2.000`, it is now written as `X1-X2-X3-X4 1 0.000 0.000 3.000`. Here, `X1-X2-X3-X4` is the torison, `1` is the divider term, the first `0.000` term is the barrier term, the second `0.000` term is the phase and `3.000` is the periodicity. The following `bash` script is run to output energies from a MM (implemented with SANDER) torsional scan.

```
#torsion is set from 0 to 360 in increments of 5 degrees
number=`seq 0 5 360`
rank=5000 #used in dihedral restraint (force)

#define root name for all files
rootname='BMI' #BMI = 1-Butyl-3-MethylImidazolium
#define dihedral atoms
dihedral_atom1=1 #Here,  the atoms used for the scanning computation is entered.
dihedral_atom2=3
dihedral_atom3=4
dihedral_atom4=5

# template for tleap file,  will write $root.tleap
# requires a .mol2 file with the structure & charges
# and a frcmod file with the parameters
tleapfile="loadAmberParams $rootname.frcmod\n
# the .frcmod file is read here
$rootname = loadMol2 $rootname.mol2\n
# the .mol2 file is read here
check $rootname\n
# the charge of the .mol2 is checked here to see if the compontent/molecule is
↪  correctly charged
charge $rootname\n
saveAmberParm $rootname $rootname.prmtop $rootname.inpcrd\n
# the component's topology and parameter files are written out here
saveOff $rootname $rootname.lib\n
quit\n";
echo -e $tleapfile > $rootname.tleap;

#run tleap as defined above
tleap -f $rootname.tleap;

#template for dihedral restraints file,  save it in $dih_file
dih_file="&rst iat= DIH_1,  DIH_2,  DIH_3,  DIH_4, \n
# these are the atoms initials selected for the scanning
r1=AN_PREV00000,  r2=ANG.000000,  r3=ANG.000000,  r4=ANG.100000,  rk2=RANK.0,
↪  rk3=RANK.0,  /\n
# initial scanning parameters
```

```
&rst iat= 6,  1,  3,  4, \n
# this is the torsion that is not been scanned,  but has to be fixed during the
 ↪  computation as to avoid atom collision from free rotation.
r1=-0.100000,  r2=0.000000,  r3=0.000000,  r4=0.100000,  rk2=5000.0,  rk3=5000.0,
 ↪  /\n
&rst iat= 3,  4,  5,  9, \n
r1=179.900000,  r2=180.000000,  r3=180.000000,  r4=180.100000,  rk2=5000.0,
 ↪  rk3=5000.0,  /"

echo -e $dih_file > dihedral_TEMPLATE.f;

# for each angle,  use dihedral_TEMPLATE.f to create the restraint files for
 ↪  sander
if [ -e dih_tmp ]; then rm dih_tmp/*; else mkdir dih_tmp; fi
for N in $number;
do
  angprev_num=`echo $N-1 | bc`
  if [ $N == 0 ]; then
    ang_prev=-0.1
  else
    ang_prev=$angprev_num.9
  fi
   sed -e 's/DIH_1/'$dihedral_atom1'/g' dihedral_TEMPLATE.f | sed -e
    ↪  's/DIH_2/'$dihedral_atom2'/g' | sed -e 's/DIH_3/'$dihedral_atom3'/g' | sed
    ↪  -e 's/DIH_4/'$dihedral_atom4'/g' | sed -e 's/AN_PREV/'$ang_prev'/g' | sed
    ↪  -e 's/ANG/'$N'/g' | sed -e 's/RANK/'$rank'/g' > dih_tmp/dihedral_$N.f
done

#run minimisations
#template file for min.in (will be changed for each minimisation step)
min_file="Minimization of BMI with dihedral constraints\n
&cntrl\n
imin=1, \n
ntb=0, \n
cut=10, \n
maxcyc=5000, \n
ntpr=1, \n
ioutfm=0, \n
nmropt=1\n
/\n
&wt type='END'\n
/\n
DISANG=dih_tmp/dihedral_XXANGLEXX.f\n"
echo -e $min_file > min.in;

#for each angle,  create input.in with correct dihedral restraint file,  and run
 ↪  sander to minimize the structure
if [ -e out_$rootname ]; then rm out_$rootname/*; else mkdir out_$rootname; fi
```

```
for N in $number;
do
 sed -e 's/XXANGLEXX/'$N'/g' min.in > input.in
 if [ $N -lt 10 ]; then N=00$N; fi
 if [ $N -lt 100 -a $N -gt 9 ]; then N=0$N; fi
sanderexe=`which sander`
$sanderexe -O -i input.in -o out_$rootname/${rootname}_$N.out -p
↪  ${rootname}.prmtop -c ${rootname}.inpcrd -r out_$rootname/${rootname}_$N.ncrst
done

#remove sander_energies if exists
if [ -e sander_energies.txt ]; then rm sander_energies.txt; fi
#get FINAL RESULTS line and retrieve last step energies
for file in out_$rootname/*.out;
do
  echo -n $file >> sander_energies.txt;
  sed -n $(echo `grep -n " FINAL " $file | awk '{print $1}' | sed 's/://'`+5 | bc
  ↪  )p $file >> sander_energies.txt
done
#save energies in a proper format (1 column with the energies sorted by angle)
min=$(awk 'NR == 1 || $3 < min {min = $3}END{print min}' sander_energies.txt)
awk -v min="$min" '{printf "%.7f\n", $3 - min}' sander_energies.txt >
↪  sander_energies.dat

##########CLEAN FILES#######
rm input.in
rm min.in
rm mdinfo
rm sander_energies.txt
rm dihedral_TEMPLATE.f
############################

#########OPTIONAL################
#create a sd file for visualization of all sander process in min_pdbs
#AMBER16 write restart files in netcdf format,  so we use cpptraj to convert to
↪  ASCII
if [ -e min_pdbs ]; then rm min_pdbs/*; else mkdir min_pdbs; fi
for N in $number;
do
  if [ $N -lt 10 ]; then N=00$N; fi
  if [ $N -lt 100 -a $N -gt 9 ]; then N=0$N; fi
  cpptraj -p ${rootname}.prmtop <<-EOF >/dev/null
  trajin out_${rootname}/${rootname}_$N.ncrst ncrestart
  trajout out_${rootname}/${rootname}_$N.rst restart
  EOF
  ambpdb -p ${rootname}.prmtop < out_${rootname}/${rootname}_$N.rst >
  ↪  min_pdbs/min_pdb_${rootname}_$N.pdb 2> /dev/null #create .pdb files from the
  ↪  parameter and topology file
```

```
  babel -ipdb min_pdbs/min_pdb_${rootname}_$N.pdb -osd
  ↪  min_pdbs/min_pdb_${rootname}_$N.sd 2> /dev/null
  cat min_pdbs/min_pdb_${rootname}_$N.sd >> min_pdbs/min_pdb_all.sd
done
################################
```

From here an initial comparison is made between the MM torsional profile and the QM torsional profile. The goal is to refine the MM torsional as to best overlap with the QM profile. The following genetic algorithm `Python` script is run to get new parameters for the tagged torsion. The `Python` script was written by Sergio Ruiz and can be found at http://www.ub.edu/cbdd/?q=content/small-molecule-dihedrals-parametrization (accessed Jun. 24, 2019).

```
python << END
import numpy as npy
import sys
from pyevolve import G1DList
from pyevolve import GSimpleGA
from pyevolve import Selectors
from pyevolve import Mutators
from pyevolve import Initializators
from pyevolve import GAllele
import math

#gaussian_file = sys.argv[1]
#sander_file = sys.argv[2]

gaussian_file = 'gaussian_energies.dat' #the QM energies file is defined
sander_file = 'sander_energies.dat' #the MM energies file is defined

qm = [ float(i.split()[0]) for i in open(gaussian_file).readlines() if
↪  i.split()[0].find("#")]  # Read QM energies and angles
sander = map(float, open(sander_file).read().split()) #read sander energies
if len(open(gaussian_file).readline().split()) > 4: #read sander angles
    angle_list = [ map(float, i.split()[2:]) for i in
    ↪  open(gaussian_file).readlines() if i.split()[0].find("#")]
    multi_angle = True
else:
    angle_list = [ float(i.split()[2]) for i in open(gaussian_file).readlines() if
    ↪  i.split()[0].find("#")]
    multi_angle = False

############################################################
##  THREE CURVE FITTING
############################################################
def calc_fitted_3graphs(pkidvf1=0.0, pn1=0.0, phase1=0.0, pkidvf2=0.0,  pn2=0.0,
↪  phase2=0.0, pkidvf3=0.0, pn3=0.0, phase3=0.0):
```

```
    values=[]
    if multi_angle: #continue if there are mutliple angles
        for i,  angles in enumerate(angle_list):
            edihedral1 = 0.0
            edihedral2 = 0.0
            edihedral3 = 0.0
            for angle in angles: #determine mm potential energies
                edihedral1 += pkidvf1 * (1 + math.cos(math.radians(pn1 * angle -
                ↪  phase1)))
                edihedral2 += pkidvf2 * (1 + math.cos(math.radians(pn2 * angle -
                ↪  phase2)))
                edihedral3 += pkidvf3 * (1 + math.cos(math.radians(pn3 * angle -
                ↪  phase3)))
            values.append(edihedral1 + edihedral2 + edihedral3 + sander[i])
    else:
        for i,  angle in enumerate(angle_list):
            edihedral1 = pkidvf1 * (1 + math.cos(math.radians(pn1 * angle -
            ↪  phase1)))
            edihedral2 = pkidvf2 * (1 + math.cos(math.radians(pn2 * angle -
            ↪  phase2)))
            edihedral3 = pkidvf3 * (1 + math.cos(math.radians(pn3 * angle -
            ↪  phase3)))
            values.append(edihedral1 + edihedral2 + edihedral3 + sander[i])
    values=npy.array(values)
    values=values - min(values)
    err=[]
    for i, value in enumerate(values):
        err.append(value - qm[i])
    err=npy.array(err)
    rmsd=npy.sqrt(npy.sum(err**2)/len(err))
    return rmsd #RMSD between the QM and MM profiles as per the estimator are
    ↪  determined.

#function for ranking chromosomes
def eval_func_3graphs(chromosome): # Chromosome is a set of parameters which
↪  define a proposed solution to the problem that the genetic algorithm is trying
↪  to solve
  score=0.0 # This function is the evaluation function, we want to give high score
   ↪  to more zero'ed chromosomes
        if chromosome[9] == 1: #iterate over the chromosome
      rmsd=calc_fitted_3graphs(pkidvf1=chromosome[0], pn1=chromosome[1],
       ↪  phase1=chromosome[2])
        elif chromosome[9] == 2:
            rmsd=calc_fitted_3graphs(pkidvf1=chromosome[0], pn1=chromosome[1],
            ↪  phase1=chromosome[2],pkidvf2=chromosome[3], pn2=chromosome[4],
            ↪  phase2=chromosome[5])
        else:
```

```
            rmsd=calc_fitted_3graphs(pkidvf1=chromosome[0], pn1=chromosome[1],
            ↪  phase1=chromosome[2],pkidvf2=chromosome[3], pn2=chromosome[4],
            ↪  phase2=chromosome[5],pkidvf3=chromosome[6], pn3=chromosome[7],
            ↪  phase3=chromosome[8])
  score=1/rmsd #makes sure that with the highest RMSD is the most favoured for GA
  return score

#initialize alleles; a specialised set of alternatives to the strings of
↪  choromosomes.
setOfAlleles=GAllele.GAlleles()

#first variable
pkidvf=npy.arange(0.01, 10.010, 0.010) #first class of genes from 0.01 ti 10.010
↪  in increments of 0.01
setOfAlleles.add(GAllele.GAlleleList(pkidvf))
#second variable
pn=npy.arange(1, 7, 1) #second class of genes, 1 to 7 in steps of 1.
setOfAlleles.add(GAllele.GAlleleList(pn))
#third variable
phase=[0.0, 180.0] #3rd class of genes, phases 0 and 180.
setOfAlleles.add(GAllele.GAlleleList(phase))

#repeat assignment of variables for second curve
setOfAlleles.add(GAllele.GAlleleList(pkidvf))
setOfAlleles.add(GAllele.GAlleleList(pn))
setOfAlleles.add(GAllele.GAlleleList(phase))

#repeat assignment of variables for third curve
setOfAlleles.add(GAllele.GAlleleList(pkidvf))
setOfAlleles.add(GAllele.GAlleleList(pn))
setOfAlleles.add(GAllele.GAlleleList(phase))

#Selector for 1,  2 or 3 functions
numfunc=[1, 2, 3]
setOfAlleles.add(GAllele.GAlleleList(numfunc))

#initialize genome with defined alleles
genome = G1DList.G1DList(10)
genome.setParams(allele=setOfAlleles)

#define evaluator function
genome.evaluator.set(eval_func_3graphs)
genome.mutator.set(Mutators.G1DListMutatorAllele)
genome.initializator.set(Initializators.G1DListInitializatorAllele)

ga = GSimpleGA.GSimpleGA(genome)
ga.selector.set(Selectors.GRouletteWheel) #some random individuals chosen to cross
↪  over.
```

```
ga.setCrossoverRate(0.85) #There is a 85% chance that the systems that move over
↪  will mate.
ga.setMutationRate(0.02) #2% chance that systems that move over will have slight
↪  mutations
ga.setPopulationSize(500) #starting population size is 500 individuals
ga.setGenerations(10000) #10000 generations of mating occurs
ga.setElitism(True) #best system goes through immediately
ga.evolve(freq_stats=1000)

best=ga.bestIndividual()

############################################################
##  PRINT RESULTS
############################################################
if best[9]==1:
  print "The optimized values for only 1 function are:"
  print "\tPK/IDIVF = \t"+str(best[0])
  print "\tPHASE = \t"+str(best[2])
  print "\tPN = \t\t"+str(best[1])
elif best[9]==2:
  print "The optimized values for 2 functions are:"
  print "Function 1:"
  print "\tPK/IDIVF = \t"+str(best[0])
  print "\tPHASE = \t"+str(best[2])
  print "\tPN = \t\t"+str(best[1])
  print "Function 2:"
  print "\tPK/IDIVF = \t"+str(best[3])
  print "\tPHASE = \t"+str(best[5])
  print "\tPN = \t\t"+str(best[4])
else:
  print "The optimized values for 3 functions are:"
  print "Function 1:"
  print "\tPK/IDIVF = \t"+str(best[0])
  print "\tPHASE = \t"+str(best[2])
  print "\tPN = \t\t"+str(best[1])
  print "Function 2:"
  print "\tPK/IDIVF = \t"+str(best[3])
  print "\tPHASE = \t"+str(best[5])
  print "\tPN = \t\t"+str(best[4])
  print "Function 3:"
  print "\tPK/IDIVF = \t"+str(best[6])
  print "\tPHASE = \t"+str(best[8])
  print "\tPN = \t\t"+str(best[7])
END
```

Finally, new parameters are produced for the tagged torsion and these are edited into the original `.frcmod` file. The `bash` script from above is run again and a new comparison is made between the MM and QM torsional profiles.

[Page intentionally left blank.]

12 Appendix 2: Additional Results

Below are the partial charges of propylene carbonate (Figure 12.1), dimethyl carbonate (Figure 12.2), γ-valerolactone (Figure 12.3), 1-butyl-3-methylimidazolium (Figure 12.4), acetate (Figure 12.5) and methyl sulfate (Figure 12.6); as well as a comparison of literature and computational bond lengths for the 1-butyl-3-methylimidazolium cation (Table 12.1) and the PMF curve of separation of two $(GlcNAc)_2Me_2$ monomers in pure $[C_4C_1im][CH_3COO]$ at 353.15 K (Figure 12.7).

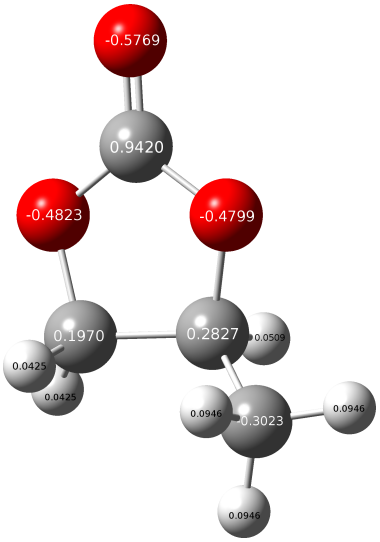


Figure 12.1: Partial charges of propylene carbonate as calculated by the R.E.D. Server at the HF/6-31G* level of theory.

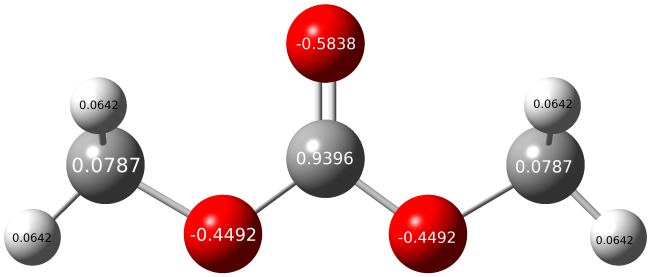


Figure 12.2: Partial charges of dimethyl carbonate as calculated by the R.E.D. Server at the HF/6-31G* level of theory.

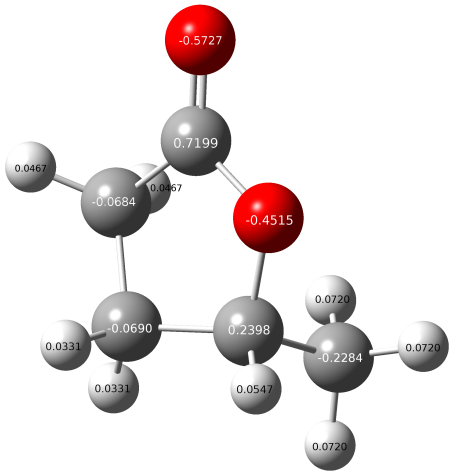


Figure 12.3 Partial charges of γ-valerolactone as calculated by the R.E.D. Server at the HF/6-31G* level of theory.

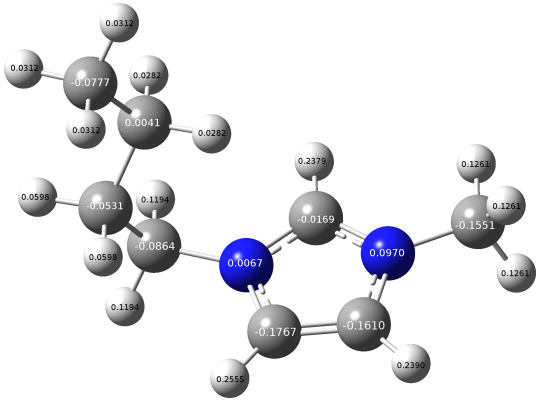


Figure 12.4: Partial charges of 1-butyl-3-methylimidazolium as calculated by the R.E.D. Server at the HF/6-31G* level of theory.

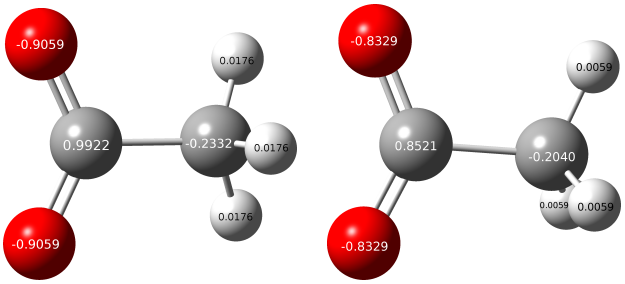


Figure 12.5: (The partial charges of acetate as calculated by the R.E.D. Server at the HF/6-31+G* (left) and HF/6-31G* (right) levels of theory.

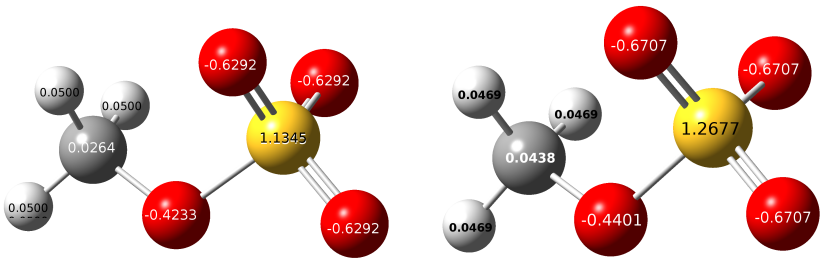


Figure 12.6: The partial charges of methyl sulfate as calculated by the R.E.D. Server at the HF/6-31G* (left) and HF/6-31+G* (right) levels of theory.

Table 12.1: Comparison of the bond lengths of 1-butyl-3-methylimidazolium from this work to the experimentally known bond lengths of the cation from 1-butyl-3-methylimidazolium chloride

Bond	Bond length (this work)	Bond length Experimental
C5-C6	1.37235	1.343
C4-N1	1.34241	1.332
C4-N2	1.34405	1.337
C3-N1	1.47780	1.485
C8-N2	1.46959	1.470
C5-N1	1.37520	1.369
C6-N2	1.37530	1.393

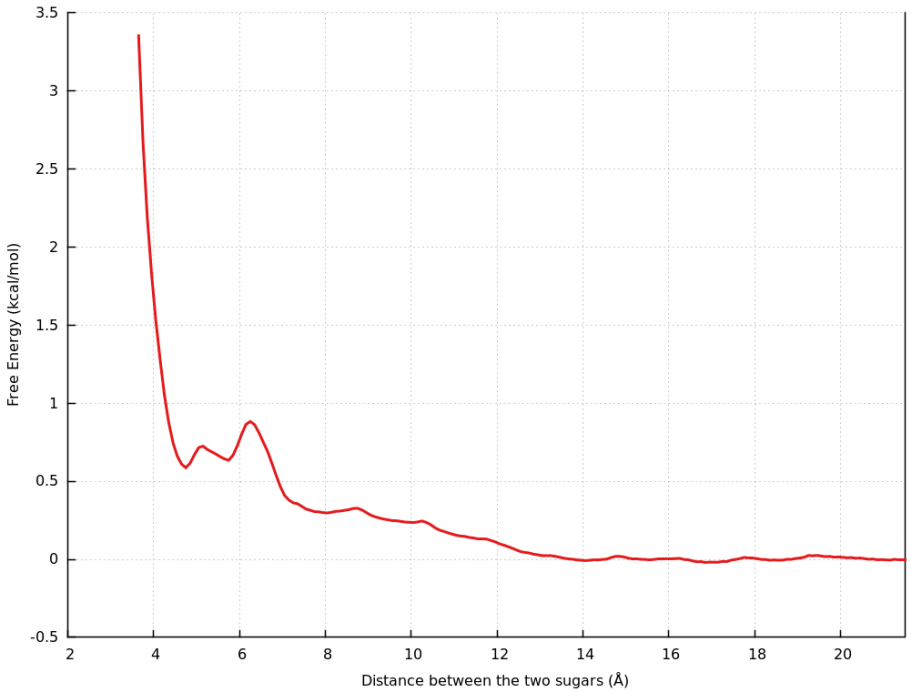


Figure 12.7: The PMF curve of separating two $(GlcNAc)_2Me_2$ monomers in pure $[C_4C_1im][CH_3COO]$ at 353.15 K in which the production phase was run for 150 ns.

Even after 150 ns, the energy barrier remains, suggesting that it is important to the separation profile.