

Molecular Interactions of Protein-based Materials with Ionic Liquids

SHAFAQ MUZAMMAL

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

November, 2020

ABSTRACT

Renewable polymer resources have attracted an increasing amount of attention predominantly due to two major reasons, firstly environmental concerns, and secondly the realisation that our petroleum resources are finite. Silks and keratin-based materials such as wool are rich in protein content and have an immense importance as a renewable raw material in textile industry. Processing of such complex biopolymers has been hindered by the extremely low solubility in conventional organic solvents. Room temperature ionic liquids (RTILs) have shown favourable solvation properties for these materials and open up a wide possible range for application by turning them into novel structures and composites. In this work, molecular interactions of ionic liquids (ILs) with a range of amino acids (AAs) and peptides selected as model compounds were investigated through molecular dynamic (MD) simulations. The aim was to explain the underlying molecular mechanisms observed in the macroscopic processes with biotechnological relevance, such as, the dissolution of biomaterials in ILs, application of ILs in the separation and purification of amino acid molecules and other bioprocesses involving the replacement of organic solvents with ILs.

Firstly, the effect of concentration ratio of ionic liquid (IL) to amino acid (AA) [IL:AA] at constant temperature (300K) was explored for lysine, arginine, aspartate, glutamate in pure 1-ethyl-3-methylimidazolium chloride $[C_2C_1Im][Cl]$ IL. The amino acids were constructed in their charged side chains with zwitter-ionic backbone. The strength of interactions of IL ions with AA molecules was found to decrease with the increase in concentration ratio for all amino acids. The effect of concentration ratio, temperature and water addition on solute-solvent, solute-solute and solvent-solvent interactions were investigated for lysine and aspartate in pure, and water mixtures of $[C_2C_1Im][Cl]$ and 1-ethyl-3-methylimidazolium acetate ($[C_2C_1Im][CH_3CO_2]$) ILs. The temperature increase (from 300K to 400K) was found to enhance the solute-solute interactions for both amino acids in $[C_2C_1Im][Cl]$, whereas, the effect of temperature on solute interactions was substantially less in $[C_2C_1Im][CH_3CO_2]$ IL. Water, in each case, was added in the simulation systems based on the ambient water level absorbed in the two ILs. Presence of water also caused a slight increase in the inter-amino acid interactions, and this impact was more prominent at a higher temperature of 400K.

Aromatic amino acids, histidine, tryptophan, phenyl-alanine and tyrosine were investigated in pure $[C_2C_1Im][CH_3CO_2]$ IL at constant temperature. These amino acids were constructed with the zwitter-ionic backbone and the side chains in their neutral forms. The heterocyclic aromatics, histidine and tryptophan had a strong interaction with the IL cations above the plane, along with the strong interactions with IL anion in areas where hydrogen bonding might be expected. The non-aromatic, non-polar amino acid alanine was studied in the two above mentioned ILs and the temperature variation and addition of water in these systems were explored. Strong inter-amino acid interactions were identified at a lower temperature in comparison to polar amino acids and they increased at a higher temperature. Presence of ambient water levels were shown to break the solute-solute interactions. Acetate anions showed better solvation properties than chloride anions for all amino acids studied and the simulation results were verified against experimental findings. This consistency with experiment also validated the use of General Amber forcefield (GAFF) for studying small organic molecules in ionic liquids.

Two alanine based peptides, penta-alanine and deca-alanine, in addition to a peptide segment (ACE(leucine-aspargine-glycine-lysine-threonine-leucine-lysine-glycine)NME) from protein G in its non-standard form, were analysed for peptide-IL interactions in pure $[C_2C_1Im][CH_3CO_2]$ IL and water. Secondary structures for all peptides studied were compared in two solvent media. It was observed that the organisational structuring of ionic liquid ions around the peptide residues was consistent with the single amino acid simulation results, however, the apparent strengths of interaction were found to be reduced and were affected by the position of the residue in the peptide. The current study provides a suitable framework for further development of macroscopic processes such as dissolution processes in ionic liquids.

ACKNOWLEDGEMENTS

In the name of Allah, the most beneficent and the most merciful

First and foremost, I would like to express my sincere appreciation and gratitude to my supervisor, Dr. Anna Croft, for her constant encouragement, support and for her valuable suggestions. I would also like to thank my second supervisor Dr. Christof Jaeger for his direction and support. I am grateful to Prof. Dr. Peter Licence for the knowledge he has shared.

I gratefully acknowledge the financial support provided by the Islamic development (IDB) and the University of Nottingham. I would also like to highlight and acknowledge the support from the University of Engineering and Technology, Lahore for providing me with the opportunity to conduct the doctoral research.

I am grateful to my colleagues who provided support during my PhD work through their discussions and suggestions.

I would love to thank my entire family, my sisters, aunts and all of my friends for their love, endless support and encouragement.

Lastly, and most importantly, I am sincerely grateful to my parents Muzammal Ali and Naila Muzammal who have encouraged and supported me throughout my life and my husband Kashif Javed and son Shaheer Kashif who have encouraged me the most through their love and encouragement.

DECLARATIONS

No portion of work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or institute of learning.

Shafaq Muzammal

TABLE OF CONTENTS

ABSTRACT II
ACKNOWLEDGEMENTS IV
DECLARATIONS V
TABLE OF CONTENTS VI
LIST OF FIGURES XIII
LIST OF TABLES XVIII
LIST OF ABBREVIATIONS XIX
IONIC LIQUID ABBREVIATIONS XX
1 1
Objectives and thesis Outline 1
2 4
Literature review 4
2.1 Introduction 4
2.2 Proteins 5
2.2.1 Silk 6
2.2.2 Keratin 7
2.3 Ionic liquids 10
2.3.1 Properties of Ionic liquids 10
2.3.2 Ionic liquids as solvents for natural polymers and their applications 11
2.4 Ionic liquids as solvents for keratin 18
2.5 Proteins in ionic liquids 24
2.6 Molecular dynamic simulation studies of ILs and proteins 28
3 33
Methodology 33
3.1 Molecular Dynamics Simulations and Force Field Selection 33
3.2 QM Calculations for Amino acid and Ionic Liquid Structures 35
3.3 Library building, parameter and topology files 36
3.3.1 Library building for amino acids and IL ions 36
3.3.2 Simulation boxes for amino acids and ILs simulations 37
3.4 Molecular Dynamics simulation run 38
3.4.1 Energy minimisation 39
3.4.2 Initial MD simulation, heating up the system 39
3.4.3 Second MD simulation, equilibrating the system 40
3.4.3 Third MD simulation, production stage 41
3.5 Peptide structures and MD simulation 42
4 45
Analyses methods 45
Simulation analysis 45
4.1 Radial distribution functions 45
4.2 Spatial distribution functions 46
4.3 Mean-square displacement and diffusion coefficients 46

4.4 Domain Analysis 48
4.5 Secondary Structure analysis of peptides: 48
4.6 Dihedral analysis - Ramachandran plots 49
5 51
Interactions and organisation of positively charged amino acids in imidazolium ionic liquids 51
5.1 Introduction 51
5.2 Outline of simulation conditions 53
5.2.1 QM Structures for Protonated Amino acids and Ionic Liquid 54
5.3 Results and discussion 55
5.3.1 Ionic liquid microstructure and the impact of solutes 56
5.3.2 Interactions of the ionic liquid [C_2C_1Im][Cl] with arginine and lysine 61
5.3.3 Effect of temperature on amino acid and IL interactions: 67
5.3.4 Effect of addition of water on lysine and [C_2C_1Im][Cl] interactions 70
5.4 Interactions and organisation of positively charged lysine in [C_2C_1Im][CH_3CO_2] Ionic liquid 78
5.4 Results and Discussion 78
5.4.1 Effect of [C_2C_1Im][CH_3CO_2] as an amino acid solvent 78
5.4.2 Effect of temperature on lysine in [C_2C_1Im][CH_3CO_2] 80
5.4.3 Effect of water on lysine and [C_2C_1Im][CH_3CO_2] interactions 81
5.5 Conclusions 85
6 88
Interactions and organisation of negatively-charged amino acids in imidazolium ionic liquids 88
6.1 Introduction 88
6.2 Outline of simulation systems 89
6.2.1 QM Calculations for Amino acid and Ionic Liquid Structures and antechamber 90
6.3 Results and discussion 91
6.3.1 Influence of solutes on ionic liquid microstructure 92
6.3.2 Effect of Concentration ratio of aspartate and glutamate to [C_2C_1Im][Cl]on microstructural organisation 95
6.3.2.1 Interactions and organisation of negatively-charged amino acids aspartate and glutamate in [C_2C_1Im][Cl] 95
6.3.3 Effect of temperature on Aspartate-IL interactions 98
6.3.4 Effect of addition of water on aspartate and [C_2C_1Im][Cl] interactions 102
6.4 Interactions and organisation of negatively charged amino acid aspartate in [C_2C_1Im][CH_3CO_2] ionic liquid 109
6.4.1 Effect of concentration ratio on the interaction of aspartate with [C_2C_1Im][CH_3CO_2] 109
6.4.2 Effect of temperature on Aspartate and [C_2C_1Im][CH_3CO_2] IL interactions 111
6.4.3 Effect of addition of water on aspartate and [C_2C_1Im][CH_3CO_2] interactions: 114
6.5 Conclusions 119
7 121
Interactions and organisation of aromatic amino acids and alanine in [C_2C_1Im][CH_3CO_2] .. 121
7.1 Introduction 121
7.2 Materials and Methods: 123
7.2.1 Outline of simulation conditions 123

7.2.2 QM Calculations for Amino acid and Ionic Liquid Structures and antechamber 124
7.3 Results and discussions 125
7.3.1 Effect of temperature on interactions of concentrated alanine solution in pure and IL/water mixture of [C_2C_1Im][Cl] 125
7.3.2 Effect of addition of water on interactions of alanine with IL/water [C_2C_1Im][Cl] in concentrated systems 128
7.3.3 Effect of temperature on interactions of alanine with [C_2C_1Im][CH_3CO_2] at high concentration 133
7.3.4 Effect of water addition on interactions of concentrated alanine with [C_2C_1Im][CH_3CO_2] 135
7.3.5 Interactions of [C_2C_1Im][CH_3CO_2] IL with aromatic amino acids at 300K 139
7.4 Conclusions 143
8 145
Structure and interactions of peptides in water and ionic liquids 145
8.1 Introduction 145
8.2 Materials and methods 147
8.2.1 Outline of simulation conditions 147
8.3 Results and discussions 148
8.3.1 Structure and interactions of Alanine (5-mer) peptide in pure [C_2C_1Im][CH_3CO_2] IL and water at 300K 149
8.3.2 Structure and interactions of Alanine (10-mer) peptide in pure [C_2C_1Im][CH_3CO_2] IL and water at 300K 153
8.3.3 Structure and interactions of protein G peptide in pure [C_2C_1Im][CH_3CO_2] IL and water at 300K 156
8.4 Conclusions 160
9 161
General Discussion, conclusions and suggestions for future work 161
9.1 Summary and conclusions 161
9.2 Recommendations for future work 164
References 166
APPENDICES 183
Appendix A. 1: Karl Fischer measurement 183
Appendix A. 2: Calculation of number of water molecules 183
Appendix A. 3: Input Files for Lysine Molecule QM calculations 183
Appendix A.3. 1: Geometry Optimisation 183
Appendix A.3. 2: Resp Charge Derivation 186
Appendix A. 4: Gaussian optimisation output files for IL ions and amino acids 186
Appendix A. 4.1: Gaussian frequency analysis (shortened) output file for lysine 191
Appendix A. 5: Library, mol2 and frcmod files for lysine molecule 191
Appendix A.5. 1: Lysine mol2 file 191
Appendix A.5. 2:Library file 192
Appendix A.5. 3: frcmod file for lysine 195
Appendix A. 6: Input files for Packmol, co-ordinates and MD Simulation 196
Appendix A.6. 1: Packmol input file for building [C_2C_1Im][Cl]:[LYS] box for high concentration (5:1) system 196
Appendix A.6. 2: Input file for co-ordinate files (commands.in) 196
Appendix A.6. 3: Input file for first energy minimisation stage 197
Appendix A.6. 4: Input file for second energy minimisation stage 197
Appendix A.6. 5: Input file for first and second MD simulation run 197

Appendix A.6. 6: Input file for third MD simulation production run 198
Appendix A. 7: Chem draw structures for AAs and IL ions with scaled charges 199
Appendix A. 8: Graphs from heating and equilibration MD stages for density, temperature and energy for all Amino acid simulations 202
Appendix A.8. 1: Density graphs from equilibration MD stage for all Amino acid- IL systems and IL/water systems 203
Appendix A.8. 2: Box dimensions for lysine and aspartate systems in pure IL and IL/water mixtures of $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ ILs 205
Appendix A.8. 3: Density graphs from equilibration MD stage for all Peptides (LNGKTLKG, ALA-5 and ALA-10) in $[C_2C_1Im][CH_3CO_2]$ IL and water systems 207
Appendix A.8. 4: Temperature graphs from heating and equilibration stages of MD simulation for all Amino acid- IL and IL/water systems 207
Appendix A.8. 5: Temperature graphs from heating and equilibration stages of MD simulation for (LNGKTLKG, ALA-5 and ALA-10) in $[C_2C_1Im][CH_3CO_2]$ IL and water systems 211
Appendix A.8. 6: Energy graphs from heating and equilibration stages of MD simulations for all AA-IL and IL/water systems 212
Appendix A.8. 7: Energy graphs from heating and equilibration stages of MD simulations for all peptides (LNGKTLKG, ALA-5 and ALA-10) in $[C_2C_1Im][CH_3CO_2]$ IL and water systems 217
Appendix A. 9: Cation-Anion interactions for positively and negatively charged amino acids in pure ILs 218
Appendix A.9. 1: Site-site radial distribution function of $[C_2C_1Im]$ cation and chloride anion around LYS, ARG, ASP and GLU at 300K and for LYS and ASP at 400K low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ 218
Appendix A.9. 2: Centre-of-mass radial distribution function around $[C_2C_1Im]^+$ cation atomic sites and O1 atom of $[CH_3CO_2]^-$ anion for LYS in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K 219
Appendix A. 10: Effect of concentration ratio of IL:AA on solute-solvent interactions for lysine and arginine systems at 300K in $[C_2C_1Im][Cl]$ ionic liquid 220
Appendix A.10. 1: Centre-of-mass radial distribution function of $[C_2C_1Im]$ cation around carboxylate group of lysine and arginine molecule for Low concentration *vs* asp high concentration systems in $[C_2C_1Im]$ [Cl] at 300 K 220
Appendix A.10. 2: Site-site radial distribution function of $[C_2C_1Im]$ cation around backbone Nitrogen atom of lysine and arginine for low *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K 220
Appendix A.10. 3: Site-site radial distribution function of $[C_2C_1Im]$ cation around alpha-hydrogen and Nitrogen atom of lysine and arginine for Low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K 221
Appendix A.10. 4: Spatial distribution function of $[C_2C_1Im]^+$ cation around backbone nitrogen and carboxyl group of arginine molecule for low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K 222
Appendix A.10. 5: Site-site radial distribution function of chloride anion around carboxylate group of lysine and arginine for low *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K 222
Appendix A. 11: Effect of Temperature on Interactions for Lysine amino acid in $[C_2C_1Im][Cl]$ Ionic liquid at 300K *vs* 400K 223
Appendix A.11. 1: Site-site radial distribution function of cation atomic sites around carbonyl group on lysine molecule for low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K *vs* 400K 223
Appendix A.11. 2: Site-site radial distribution function of chloride anion around backbone nitrogen atom on lysine molecule for low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K vs 400K 223
Appendix A.11. 3: Overall Centre-of-mass radial distribution function of solute-IL, IL-IL and solute-solute interactions in pure $[C_2C_1Im][Cl]$ at 300 K *vs* 400K 224
Appendix A. 12: Effect on interactions for lysine high concentration systems at 300K *vs* 400K in $[C_2C_1Im][Cl]$ IL/ water mixture 224
Appendix A.12. 1 Overall Centre-of-mass radial distribution function of solute-IL, IL-IL and solute-solute interactions in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]/H_2O$ mixture 224

Appendix A.12. 2: Spatial distribution function for $[C_2C_1Im][Cl]$:LYS high concentration system for solute-solvent interactions in pure and $[C_2C_1Im][Cl]/H_2O$ mixture 225
Appendix A.12. 3: Spatial distribution function for cation-anion and cation-cation interactions for lysine high concentration systems in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]/H_2O$ mixture 225
Appendix A.12. 4: Snapshots of lysine high concentration systems in pure $[C_2C_1Im][Cl]$ IL and $[C_2C_1Im][Cl]$-water mixture, after 100 ns simulation..... 226
Appendix A. 13: Diffusion coefficients, MSD and log-log plots for lysine in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ at 400K 227
Appendix A.13.1: MSD curves and log-log plots for lysine in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ at 400K..... 227
Appendix A.13.2: Self-Diffusion coefficients values calculated from 100ns simulation for $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixture..... 228
Appendix A. 14: Effect of concentration ratio of IL:AA on interactions for lysine and systems at 300K in $[C_2C_1Im][CH_3CO_2]$ ionic liquid..... 228
Appendix A.14. 1: Site-site radial distribution function of atomic sites on lysine molecule around $[C_2C_1Im]^+$ cation and acetate anion in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K 228
Appendix A.14. 2: Overall solute-solute and solute-solvent interactions for lysine systems at 300K determined by MD simulation $[C_2C_1Im][CH_3CO_2]$ IL at 300 K 229
Appendix A.14. 3: Centre-of-mass radial distribution function of atomic sites on lysine molecule around $[C_2C_1Im]^+$ cation in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K 230
Appendix A. 15: Effect of Temperature on Interactions for Lysine amino acid in $[C_2C_1Im][CH_3CO_2]$ Ionic liquid at 300K *vs* 400K..... 231
Appendix A.15. 1: Centre-of-mass radial distribution of solute-solute, solute-solvent and solvent-solvent interactions for LYS in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K 231
Appendix A.15. 2: Site-site radial distribution around lysine atomic sites in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K 231
Appendix A.15. 3: Spatial distribution functions comparing cation-anion, cation-cation interactions for lysine in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K..... 232
Appendix A. 16: Effect of concentration and temperature on interactions for lysine systems at 300K *vs* 400K in $[C_2C_1Im][CH_3CO_2]$ IL/H_2O mixture..... 233
Appendix A.16. 1: Site-site radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: lysine around alpha-hydrogen atom in pure IL and IL/water-mix 233
Appendix A.16. 2: Centre-of-mass radial distribution of solute-solute, solute-solvent and solvent-solvent interactions for lysine in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K in pure IL and IL/water-mix 234
Appendix A.16. 3: Spatial distribution function for $[C_2C_1Im][CH_3CO_2]$:lys high concentration system for solute-solvent interactions in pure and $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture..... 234
Appendix A.16. 4: Spatial distribution function for cation-anion and cation-cation interactions in high concentration systems of lysine in pure and $[C_2C_1Im][CH_3CO_2]$/water mixture 235
Appendix A.16. 5: Self-diffusion coefficients values from entire 100ns simulation for pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K and 400K..... 235
Appendix A. 17: Comparison of lysine systems in $[C_2C_1Im][Cl]$ *vs* $[C_2C_1Im][CH_3CO_2]$ 236
Appendix A.17. 1:Spatial distribution functions for comparison between pure and water mixtures of $[C_2C_1Im][Cl]$and $[C_2C_1Im][CH_3CO_2]$at 300K and 400K 236
Appendix A. 18: Input files for aspartate..... 236
Appendix A.18. 1: Geometry optimisation..... 236
Appendix A.18. 2: RESP charge derivation 238
Appendix A.18.3: Distribution of $[C_2C_1Im][Cl]$ IL ions around central aspartate molecule at 300K . 238
Appendix A. 19: Effect of concentration ratio of IL:AA on interactions for aspartate and glutamate systems at 300K in $[C_2C_1Im][Cl]$ ionic liquid 239
Appendix A.19.1: Centre-of mass radial distribution functions for $[C_2C_1Im][Cl]$: [AA] for overall solute-solute and solute-solvent interactions for low *vs* high concentration systems at 300K..... 239
Appendix A.19. 2: Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$ for overall cation-anion and cation-cation interactions for aspartate and glutamate at 300K. 239

Appendix A. 20: Effect of Temperature on Interactions for aspartate amino acid in $[C_2C_1Im][Cl]$ Ionic liquid at 300K *vs* 400K 240

Appendix A.20. 1: Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: [ASP] for overall solute-solute and solute-solvent interactions for aspartate at 300K *vs* 400K 240

Appendix A. 21: Effect of temperature on interactions for aspartate high concentration systems at 300K *vs* 400K in pure and IL/water mixture of $[C_2C_1Im][Cl]$ 240

Appendix A.21. 1: Comparison for spatial distribution functions for aspartate interactions around central aspartate molecule for pure and $[C_2C_1Im][Cl]$ IL/water mixture 240

Appendix A.21. 2: Comparison for centre-of-mass radial distribution functions for interactions around central aspartate molecule for pure and $[C_2C_1Im][Cl]$ IL/water mixture..... 241

Appendix A.21. 3: Mean square displacement (MSD) curves and log-log plots for aspartate in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ at 400K 241

Appendix A.21. 4: Self-Diffusion coefficients values calculated from 100ns simulation for Aspartate in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixture 242

Appendix A. 22: Effect of concentration ratio of IL:AA on interactions for aspartate systems at 300K in $[C_2C_1Im][CH_3CO_2]$ ionic liquid..... 243

Appendix A.22. 1: Centre-of-mass radial distribution function of atomic sites on aspartate molecule around $[C_2C_1Im]^+$ cation (H1 atom) in $[C_2C_1Im][CH_3CO_2]$ at 300 K 243

Appendix A.22.2: Overall mass-centred distribution of cation, anion and aspartate molecule around the central aspartate molecule..... 244

Appendix A.22. 3: Centre-of-mass radial distribution functions for overall cation-anion and cation-cation interactions in pure $[C_2C_1Im][Cl]$ for aspartate systems at 300K..... 244

Appendix A.22. 4: Centre-of-mass radial distribution function of $[CH_3CO_2]^-$ anion (O1 atom) around $[C_2C_1Im]^+$ cation atomic sites for aspartate systems in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K..... 245

Appendix A.22. 5: Spatial distribution function for anion-anion and anion-cation interactions in pure $[C_2C_1Im][CH_3CO_2]$ for high concentration aspartate systems..... 245

Appendix A. 23: Effect of Temperature on Interactions for Aspartate amino acid in $[C_2C_1Im][CH_3CO_2]$ Ionic liquid at 300K *vs* 400K..... 246

Appendix A.23. 1: Centre-of-mass radial distribution functions for overall solute-solute and solute-solvent interactions for aspartate in $[C_2C_1Im][CH_3CO_2]$ at 300K *vs* 400K..... 246

Appendix A.23. 2: Centre-of-mass radial distribution function around $[C_2C_1Im]^+$ cation atomic sites and $[CH_3CO_2]^-$ anion for aspartate systems in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K *vs* 400K..... 246

Appendix A. 24: Effect of temperature on interactions for aspartate high concentration systems at 300K *vs* 400K in pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$ IL/ water mixture 247

Appendix A.24. 1: Comparison for centre-of-mass radial distribution functions for interactions around central aspartate molecule for pure and $[C_2C_1Im][CH_3CO_2]$ IL/water mixture..... 247

Appendix A.24. 2: Comparison for centre-of-mass radial distribution functions for interactions around central aspartate molecule for pure and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ IL 247

Appendix A.24. 3: Comparison for centre-of-mass radial distribution functions for interactions around $[C_2C_1Im]$ cation for pure and $[C_2C_1Im][CH_3CO_2]$ IL/water mixture..... 248

Appendix A.24. 4: Snapshots of aspartate high concentration systems in pure $[C_2C_1Im][CH_3CO_2]$ IL and IL/water mix respectively at 300K and 400K..... 248

Appendix A.24. 5: Self-Diffusion coefficients values calculated from 100ns simulation for aspartate in pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$ /H_2O mixture at 300K and 400K..... 249

Appendix A. 25: Effect of temperature on interactions for alanine amino acid systems at 300K *vs* 400K in pure and IL/water mixture of $[C_2C_1Im][Cl]$ 249

Appendix A.25. 1: Centre-of-mass radial distribution for overall interactions between solute and solvent molecules 249

Appendix A.25. 2: Centre-of-mass radial distribution for interaction between atomic sites on $[C_2C_1Im]$ cation and chloride anions 250

Appendix A.25. 3: Comparison of solute-solvent, solute-solute and solvent-solvent interactions for alanine systems in pure $[C_2C_1Im][Cl]$ IL and $[C_2C_1Im][Cl]$/water mixture 250

Appendix A.25. 4: Comparison of solute-solvent, solute-solute and solvent-solvent interactions for pure $[C_2C_1Im][Cl]$ IL and $[C_2C_1Im][Cl]$/water mixture..... 251

Appendix A.25. 5: Centre-of-mass radial distribution for interaction between atomic sites on $[C_2C_1Im]^+$ cation and chloride anions in pure *vs* IL/water mixture of $[C_2C_1Im][Cl]$ IL 251
Appendix A.25. 6: Comparison of centre-of-mass radial distribution for overall interactions between solute and solvent molecules for alanine systems in pure and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ IL 252
Appendix A.25. 7: Comparison of distribution of acetate anions and $[C_2C_1Im]^+$ cations around alanine atomic sites for pure IL and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ as a function of temperature ... 252
Appendix A.25. 8: Snapshots for alanine systems in pure IL and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ at 300K *vs* 400K..... 253
Appendix A.25. 9: log-log plots for MSD curves for the calculation of diffusive regime for alanine systems in pure $[C_2C_1Im][Cl]$, $[C_2C_1Im][Cl]/H_2O$ mixture, pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 400K..... 253
Appendix A.25. 10: Effect of temperature on the cation and anion interactions with histidine molecule 254
Appendix A.25. 11: Cation-anion and cation-cation interactions for aromatic amino acids systems in $[C_2C_1Im][CH_3CO_2]$ IL 254
Appendix A.25. 12: MSD curves for aromatics systems in pure $[C_2C_1Im][CH_3CO_2]$ IL at 300K..... 255
Appendix A. 26: Peptide simulations in water and $[C_2C_1Im][CH_3CO_2]$ ionic liquid..... 256
Appendix A.26. 1: Input file for simulation of peptide sequence from Protein G in water 256
Appendix A.26. 2: Ramachandran plots comparison for penta alanine peptide in $[C_2C_1Im][CH_3CO_2]$ IL *vs* Water 257
Appendix A.26. 3: Ramachandran plots comparison for deca alanine peptide in $[C_2C_1Im][CH_3CO_2]$ IL *vs* water 258
Appendix A.26. 4: Radial distribution functions for cation and anion distribution around backbone atoms of residues in protein G sequence 260
Appendix A.26. 5: Ramachandran plots and secondary structures comparison for peptide segment of protein G in $[C_2C_1Im][CH_3CO_2]$ IL *vs* water..... 260
Appendix A.26. 6: Ramachandran plots comparison for peptide segment of protein G in $[C_2C_1Im][CH_3CO_2]$ IL *vs* water 261

LIST OF FIGURES

Figure 2. 1: Schematic representation of SF structure. Insets show the fibril overall structure and the fine β-sheet antiparallel alignment of SF polypeptide chains. Adapted from[24] 7
Figure 2. 2: Schematic diagram of different components of keratin; (A) polypeptide single unit, (B) amino acid sequence, (C) α-helix, and (D) β-Sheets conformations of keratin protein. Taken from [37] 8
Figure 2.3: Inter- and intra-molecular bonding in keratin.[15] 8
Figure 2.4: Structures of (A) lysine, (B) aspartate, (C) arginine and (D) glutamate indicating important bonding sites. 9
Figure 2. 5: Cellulose polymer chain structure[80] 12
Figure 2. 6: Schematic illustration of the structure of a wool fibre 21
Figure 2. 7: POM images of a wool fibre in $[C_4C_1Im][CH_3CO_2]$ at 120 °C: (a) original wool fibre, for (b) 5 s, (c) 50 s, (d) 100 s, and (e) 180 s 22
Figure 2.8: Structures of the peptides protected with BOC (tert-butoxycarbonyl) group and ILs used in the study[128] 26
Figure 2. 9: Chemical structure of applied ILs and peptide Ala-Xaa-Pro-Phe, Xaa corresponds to Ala, Asp, Glu, Gly, Lys and Phe.[134] 27

Figure 3.1: A two-dimensional array of boxes representing the movement of molecules, leaving the box and being replaced by their images from other boxes from example, molecule 1 replaced by molecule from box and the movement is replicated all boxes.[183] 39
Figure 3. 2: Flow chart of the procedure followed for the amino acid/IL/IL-H_2O simulations 42

Figure 5. 1: Optimised structures of (A) Lysine(+), (B) Arginine(+), (C) $[C_2C_1Im]^+$ cation, (D) $[CH_3CO_2]$ anion and (E) water molecule with atom numbering generated by antechamber.[181] Atoms encircled yellow are in each molecule show the atoms that are used as reference to create respective SDFs around these molecules. Backbone atoms are labelled in lysine and arginine cations. Atoms encircled black are the ones that are used to discuss results in RDFs analysis 55

Figure 5.3. 1: Chem draw structures of (A) positively charged lysine molecule with zwitter-ion backbone, (B) positively charged arginine molecule with zwitter-ion backbone, (C) $[C_2C_1Im]^+$ cation and (D) $[CH_3CO_2]^-$ anion. Atoms encircled red in the molecules are those used as reference to create respective SDFs around the molecule. Backbone atoms are labelled in lysine and arginine. Atoms encircled green are the ones that are used to discuss results in RDFs analysis. 57
Figure 5.3. 2: Spatial distribution functions for cation-cation and cation-anion interactions around central $[C_2C_1Im]^+$ cation for (A,C) lysine low concentration *vs* high concentration system respectively and (B,D) arginine low concentration *vs* high concentration system respectively, in $[C_2C_1Im][Cl]$ IL at 300K. Central $[C_2C_1Im]$ structures are presented in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}. The central molecule that is used to show the distribution around is generated by TRAVIS[186] while constructing the SDF 58
Figure 5.3.3. Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$ for overall cation-anion and cation-cation interactions for lysine (solid lines) and arginine (dashed lines) for (A) low concentration and (B) high concentration systems at 300K determined by MD simulation 60
Figure 5.3. 4: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for lysine (A) low concentration system *vs* (B) high concentration system in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3} 60
Figure 5.3.5. Spatial distribution functions around the central (A,B) lysine low concentration system *vs* high concentration system and (C,D) arginine low concentration *vs* high concentration system in $[C_2C_1Im][Cl]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3} 62
Figure 5.3.6: Site-site radial distribution functions for $[C_2C_1Im][Cl]$:[lys] at 300K for H1, C1, C6 atomic sites on $[C_2C_1Im]^+$ cation *vs* (A) O1 Oxygen atom on lysine molecule and (B) O1 Oxygen atom on arginine molecule [Low concentration (200:1) solid lines] and [high concentration (5:1) dashed lines] determined by MD simulation 63
Figure 5.3.7: Comparison of radial distribution functions for $[C_2C_1Im][Cl]$: [AA] at 300K for chloride anion around backbone nitrogen on lysine and arginine at (A) low (200:1) *vs* (B) high (5:1) concentration systems and around ammonium nitrogen on lysine and arginine at low (C) low (200:1) *vs* (D) high concentration (5:1) determined by MD simulation 65

Figure 5.3.8. Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: [AA] for overall solute-solute and solute-solvent interactions for lysine (solid lines) and arginine (dashed lines) for (A) Low concentration *vs* (B) high concentration systems at 300K determined by MD simulation 66
Figure 5.3.9: Spatial distribution functions around the central (A) lysine low concentration system at 300K (B) lysine low concentration system at 400K (C) lysine high concentration system at 300K and (D) lysine high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3} 68
Figure 5.3.10. Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: lysine low concentration ((200:1 (left A,C)) and high concentration ((5:1) (right B,D)) at 300K (solid lines) *vs* 400K (dashed lines). (A,B) show chloride anion distribution around backbone nitrogen, carboxyl group, terminal nitrogen and alpha-hydrogen *vs* (C,D) show H1 $[C_2C_1Im]^+$ cation distribution around backbone nitrogen, carboxyl group and O1 determined by MD simulation 69
Figure 5.3.11. Spatial distribution functions for cation-cation and cation-anion interactions around the central $[C_2C_1Im]^+$ cation for (A) lysine low concentration system at 300K, (B) lysine low concentration system at 400K, (C) lysine high concentration system at 300K and (D) lysine high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3} 70
Figure 5.3.12. Comparison of centre-of-mass radial distribution functions around lysine atomic sites in pure IL and IL/water mixture of $[C_2C_1Im][Cl]$. High concentration (5:1) systems with water (dashed lines) and without water (solid lines). Chloride anion around lysine atomic sites at (A) 300K *vs* (B) 400K. and $[C_2C_1Im]^+$ cation around lysine atomic sites at (C) 300K *vs* (D) 400K determined by MD simulation. 72
Figure 5.3. 13: Domain analysis across complete simulations of (A) lysine in $[C_2C_1Im][Cl]$ and (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K(blue) and at 400K (red) 74

Figure 5.4. 1: Spatial distribution functions around the central lysine molecule for (A) low concentration system *vs* (B) high concentration system, in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}. 79
Figure 5.4.2. Spatial distribution functions around the central lysine molecule for (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K and (D) high concentration system at 400K in $[C_2C_1Im][CH_3CO_2]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3} 81
Figure 5.4.3. Snapshots of lysine high concentration systems in $[C_2C_1Im][CH_3CO_2]$ IL after 100 ns simulation for (A,C) 300K for pure and IL/WAT mix respectively and (B,D) 400K for pure and IL/WAT mix respectively 82
Figure 5.4.4. Domain analysis for (A) lysine in $[C_2C_1Im][CH_3CO_2]$ and (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K(blue) and at 400K (red) 82

Figure 6.3.1:Spatial distribution functions for cation-cation and cation-anion interactions around the central $[C_2C_1Im]^+$ cation for aspartate (A) low concentration *vs* (B) high concentration system and for glutamate (C) low concentration *vs* (D) high-concentration system in $[C_2C_1Im][Cl]$ IL at 300K in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3} 93
Figure 6.3. 2: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for aspartate in (A) low concentration system *vs* (B) high concentration system in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3} 94
Figure 6.3. 3: Spatial distribution functions around the central (A,B) aspartate low and high concentration systems respectively and (C,D) glutamate low and high concentration systems respectively in $[C_2C_1Im][Cl]$ IL at 300K in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= aspartate at iso-value=10 nm^{-3} 96
Figure 6.3. 4: Site-site radial distribution functions for $[C_2C_1Im][Cl]$:[AA] at 300K for acidic H1 on $[C_2C_1Im]^+$ cation around N1 backbone nitrogen atom, COO^- carboxyl group on □-carbon and side-chain carboxyl group for (A) low concentration (200:1) *vs* (B) high concentration (5:1) system determined for aspartate (solid lines) and glutamate (dashed lines)) by MD simulation 97
Figure 6.3. 5: Comparison of site-site radial distribution functions for $[C_2C_1Im][Cl]$:AA for (A) low concentration (200:1) *vs* (B) high concentration (5:1) asp (solid lines) *vs* glu (dashed lines). (A) and (B) both show chloride anion distribution around backbone nitrogen and carboxyl groups determined by MD simulation 98
Figure 6.3. 6: Spatial distribution functions around the central aspartate molecule for (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= aspartate at iso-value=10 nm^{-3} 99

Figure 6.3. 7: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for aspartate in pure $[C_2C_1Im][Cl]$ (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3} 100
Figure 6.3. 8: Comparison radial distribution functions for $[C_2C_1Im][Cl]$:aspartate low concentration ((200:1 (left A,B)) *vs* high concentration ((5:1) (right C,D)) at 300K (solid lines) *vs* 400K (dashed lines). (A,C) show the distribution of chloride anions and (B,D) show the distribution of H1 $[C_2C_1Im]^+$ cation around backbone nitrogen, carboxyl groups determined by MD simulation..... 101
Figure 6.3. 9: Centre-of-mass radial distribution functions for aspartate in pure $[C_2C_1Im][Cl]$ IL for high concentration systems. RDFs show overall solute-solute and solute-solvent interactions for aspartate at 300K (solid lines) and 400K (dashed lines), determined by MD simulation 102
Figure 6.3. 10: Spatial distribution functions around the central aspartate molecule in (A) pure $[C_2C_1Im][Cl]$ at 300K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][Cl]$ at 400K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 400K. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= aspartate at iso-value=10 nm^{-3}. 103
Figure 6.3. 11: Centre-of-mass radial distribution function for high concentration systems of aspartate in (A) pure $[C_2C_1Im][Cl]$ *vs* (B) IL/water mixture at 300K (solid lines) *vs* 400K (dashed lines) determined by MD simulation 104
Figure 6.3. 12: Snapshots of aspartate high concentration systems in $[C_2C_1Im][Cl]$ IL after 100 ns simulation for (A,C) at 300K for pure and IL/water mix respectively and (B, D) at 400K for pure and IL/water mix respectively. 105
Figure 6.3. 13: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: aspartate high concentration (5:1) systems with water (dashed lines) and without water (solid lines) at 300K (A,C) *vs* 400K (B,D). (A,B) backbone nitrogen, carboxyl group and alpha-hydrogen on aspartate *vs* Chloride anion and (B,D) backbone nitrogen and carboxyl groups on aspartate *vs* H1 $[C_2C_1Im]^+$ cation determined by MD simulation 105
Figure 6.3. 14: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for high concentration systems of aspartate in (A) pure $[C_2C_1Im][Cl]$ at 300K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][Cl]$ at 400K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 400K in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}. 106
Figure 6.3. 15: Domain analysis for aspartate in (A) pure $[C_2C_1Im][Cl]$ and (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300 K (blue) and at 400 K (red)..... 107

Figure 6.4. 1: Spatial distribution functions around the central aspartate molecule for (A) low *vs* (B) high concentration system, in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$ and green = aspartate at iso-value=10 nm^{-3}.
Figure 6.4. 2: Centre-of-mass radial distribution function of $[C_2C_1Im][CH_3CO_2]$:aspartate for acetate anion O1 atom around (A) COO^- groups, backbone N1 and α-proton H4 on aspartate molecule and H1 atom $[C_2C_1Im]^+$ distribution around (B) COO^- groups and backbone N. High concentration systems (dashed lines) *vs* low concentration (solid lines) at 300K
Figure 6.4. 3: Centre-of-mass radial distribution function for high concentration systems of aspartate in pure $[C_2C_1Im][Cl]$ (dashed lines) and pure $[C_2C_1Im][CH_3CO_2]$ (solid lines). RDFs show distribution of α-hydrogen atom on aspartate around anions (acetate anion O1 atom and chloride) and O1 atom on aspartate molecule at 300K
Figure 6.4.4. Spatial distribution functions around the central aspartate molecule for (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][CH_3CO_2]$. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$ and green= aspartate at iso-value=10 nm^{-3}
Figure 6.4.5. Spatial distribution functions around the central $[C_2C_1Im]^+$ cation (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][CH_3CO_2]$ IL. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.
Figure 6.4.6. Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:aspartate low concentration ((200:1 (left A,B)) and high concentration ((5:1) (right C,D)) at 300K (solid lines) *vs* 400K (dashed lines). (A,C) show distribution of O1 atom on acetate anion around backbone nitrogen, carboxyl groups and α-hydrogen and (B,D) backbone nitrogen, carboxyl group and O1 *vs* (B) H1 $[C_2C_1Im]^+$ cation determined by MD simulation
Figure 6.4.7. Spatial distribution functions around the central aspartate molecule in (A) pure $[C_2C_1Im][CH_3CO_2]$ at 300K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][CH_3CO_2]$ at 400K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 400K. Blue = $[C_2C_1Im]^+$, red = $[CH_3CO_2]^-$and green = aspartate at iso-value=10 nm^{-3}.

Figure 6.4. 8: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for high concentration aspartate systems in (A) pure $[C_2C_1Im][CH_3CO_2]$ at 300K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][CH_3CO_2]$ at 400K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 400K in a view to keep acidic proton H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}
Figure 6.4. 9: Domain analysis for (A) aspartate in $[C_2C_1Im][CH_3CO_2]$ and (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K(blue) and at 400K (red)

Figure 7. 1: Chemdraw structures of neutral (A) tryptophan (B) phenyl-alanine(C) alanine, (D) histidine and (E) tyrosine with zwitter-ionic backbones and atom numberings generated by antechamber. Atoms encircled red in the molecules are those used as reference to create respective SDFs around the molecule. Backbone atoms are labelled in all amino acids. Atoms encircled green are the ones that are used to discuss results in RDFs analysis. 122
Figure 7. 2: Optimised structures of aromatic amino acids (A) Histidine (B) Phenylalanine, (C) Alanine (D) Tryptophan and (E) Tyrosine with atom numbering generated by antechamber.[180] Atoms encircled yellow are in each molecule show the atoms that are used as reference to create respective SDFs around these molecules. Backbone atoms are labelled in lysine and arginine. Atoms encircled black are the ones that are used to discuss results in RDFs analysis. 125
Figure 7. 3: Spatial distribution functions around the central (A) alanine molecule at 300K, (B) alanine at 400K, (C) $[C_2C_1Im]^+$ cation at 300K and (D) $[C_2C_1Im]^+$ cation at 400K in pure $[C_2C_1Im][Cl]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= alanine at iso-value=10 nm^{-3} 126
Figure 7. 4: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$:[alanine] at 300K for H1, C1, C6 atomic sites on $[C_2C_1Im]$ cation *vs* (A) backbone ammonium (B) backbone carboxylate, (low concentration (200:1) solid lines and (high concentration (5:1) dashed lines determined by MD simulation..... 127
Figure 7. 5: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$:[alanine] 5:1 for (A) chloride distribution around backbone nitrogen, carboxyl group and alpha-hydrogen (B) H1 $[C_2C_1Im]^+$ distribution around backbone nitrogen, carboxyl groups and O1 cation determined by MD simulation 127
Figure 7. 6: Spatial distribution functions around the central alanine molecule in (A) pure $[C_2C_1Im][Cl$ at 300K, (B) $[C_2C_1Im][Cl$/water mixture at 300K, (C) pure $[C_2C_1Im][Cl$ at 400K, (D) $[C_2C_1Im][Cl$/water mixture at 400K. Blue = $[C_2C_1Im]^+$, red = Cl^-, and green = alanine at iso-value=10 nm^{-3}. 128
Figure 7. 7: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$:[alanine] high concentration system for (A) chloride distribution around backbone nitrogen, carboxyl group and □-hydrogen and (B) H1 $[C_2C_1Im]^+$ distribution around backbone nitrogen, carboxyl groups and O1 determined by MD simulation. Solid lines show pure IL and dashed lines show IL/water mixtures of $[C_2C_1Im][Cl]$. 129
Figure 7. 8: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation at (a) 300k for pure IL (b) at 300k for IL/water mixture, (c) 400k for pure il and (d) 400k for IL/water mixture in a view to keep ring hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3} 130
Figure 7. 9: Snapshots for Alanine systems in pure $[C_2C_1Im][Cl]$ IL (A,C) at 300K *vs* 400K and in $[C_2C_1Im][Cl]$/water mixture (B,D) at 300K *vs* 400K respectively from 100 ns MD simulations. The positioning of the alanine molecules is shown at the right-hand side of each panel in the absence of water and IL ions to see the cluster formation more clearly..... 131
Figure 7. 10: Domain analysis over time for (A) lysine in $[C_2C_1Im][Cl]$ and (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K(blue) and at 400K (red)..... 131
Figure 7. 11: Spatial distribution functions around the central (A) alanine at 300K, (B) alanine at 400K, (C) $[C_2C_1Im]^+$ cation at 300K and (D) $[C_2C_1Im]^+$ cation at 400K for $[C_2C_1Im][Cl]$ IL. Blue= $[C_2C_1Im]^+$, red = Cl^-, and green = alanine at iso-value=10 nm^{-3}..... 133
Figure 7. 12: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: [alanine] at 300K for H1, H7, H8, N1, C1, C4, C6 atomic sites on $[C_2C_1Im]$ cation *vs* (A) backbone ammonium (B) backbone carboxylate, (low concentration (200:1) solid lines and (high concentration (5:1) dashed lines determined by MD simulation 134
Figure 7. 13: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:[alanine] 5:1 for (A) chloride distribution around backbone nitrogen, carboxyl group and alpha-hydrogen (B) H1 $[C_2C_1Im]^+$ distribution around backbone nitrogen, carboxyl groups and O1 cation determined by MD simulation..... 134
Figure 7. 14: Spatial distribution functions around the central alanine molecule in (A) pure IL at 300K, (B) IL/water mixture at 300K, (C) pure IL at 400K , (D) IL/water mixture at 400K. Blue= $[C_2C_1Im]^+$, red= Cl–, and green= alanine at iso-value=10 nm^{-3}. 136

Figure 7.15. Spatial distribution functions around the central (A) histidine, (B) tryptophan (C) phenylalanine and (D) tyrosine at 300K [C_2C_1Im][CH_3CO_2] IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$ and green= amino acid at iso-value=10 nm^{-3}. 139
Figure 7. 16: Comparison of centre-of-mass radial distribution functions for [C_2C_1Im][CH_3CO_2]: [AA] (5:1) at 300K for O1 atom on acetate anion *vs* backbone nitrogen, carboxyl group and alpha-hydrogen on (A) Histidine (B) Tryptophan, (C) Phenylalanine and (D) tyrosine determined by MD simulation 140
Figure 7. 17: Comparison of centre-of-mass radial distribution functions for [C_2C_1Im][CH_3CO_2]: [AA] (5:1) at 300K for H1 atom on imidazolium cation *vs* backbone nitrogen (dashed lines) and *vs* carboxyl group (solid lines) on aromatic amino acids determined by MD simulation 141

Figure 8. 1: Three-dimensional representation of peptide LNGKTLKG in (A) octahedron water box and (B) cubic [C_2C_1Im][CH_3CO_2] IL box. Image taken from 1000ns simulation of deca-alanine peptide in two solvent media. 148
Figure 8.2. Starting structures of peptides penta-alanine, deca-alanine and G-protein fragment LNGKTLKG. 148
Figure 8.3. Centre-of-mass radial distribution functions for (A) acetate anion oxygen atoms around backbone nitrogen atoms and (B) acidic proton on imidazolium cation around oxygen atoms on alanine residues in alanine-5 peptide dissolved in pure [C_2C_1Im][CH_3CO_2]. 149
Figure 8.4. Spatial distribution functions around the central (A) alanine peptide and (B) [C_2C_1Im] cation in peptide-[C_2C_1Im][CH_3CO_2] solution at 300K. Blue = $[C_2C_1Im]^+$ and red = $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}. 150
Figure 8.5. Ramachandran plots for penta-alanine or $(ALA)_5$ peptide in (A) water and (B) pure [C_2C_1Im][CH_3CO_2]. 152
Figure 8.6. Secondary structure analysis of penta-alanine in (A) water and (B) pure [C_2C_1Im][CH_3CO_2].. 153
Figure 8.7. Centre-of-mass radial distribution functions for (A) acetate anion oxygen atoms around backbone nitrogen atoms and (B) acidic proton on imidazolium cation around oxygen atoms on alanine residues in $(ALA)_{10}$ peptide in pure [C_2C_1Im][CH_3CO_2]..... 153
Figure 8.8. Spatial distribution functions around the central (A) $(ALA)_{10}$ peptide and (B) [C_2C_1Im] cation in pure [C_2C_1Im][CH_3CO_2] IL at 300K. Blue = $[C_2C_1Im]^+$ and red = $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.. 154
Figure 8.9. Ramachandran plots for penta ALA or $(ALA)_5$ peptide in (A) water and (B) pure [C_2C_1Im][CH_3CO_2]. 155
Figure 8.10. Secondary structure analysis of deca-alanine in (A) water and (B) pure [C_2C_1Im][CH_3CO_2]. 155
Figure 8.11. Centre-of-mass radial distribution functions for (A) acetate anion oxygen atoms around on two lysine residues' backbone nitrogen atoms and (B) acidic proton on imidazolium cation around two lysine residues' oxygen atoms on G-protein fragment LNGKTLKG in pure [C_2C_1Im][CH_3CO_2] IL..... 157
Figure 8.12. Spatial distribution functions around the central (A) G-protein fragment LNGKTLKG and (B) [C_2C_1Im] cation in pure [C_2C_1Im][CH_3CO_2] at 300K. Blue = $[C_2C_1Im]^+$ and red = $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}. 157
Figure 8.13. Centre-of-mass radial distribution functions for tail carbon on imidazolium cation around tail carbon atoms of each residue on G-protein fragment LNGKTLKG in pure [C_2C_1Im][CH_3CO_2] IL... 158
Figure 8.14. Ramachandran plots for protein G peptide sequence in (A) water and (B) pure [C_2C_1Im][CH_3CO_2]. 158
Figure 8.15. Secondary structure analysis of protein G peptide sequence in (A) water and (B) pure [C_2C_1Im][CH_3CO_2]. 159

LIST OF TABLES

Table 2. 1: Solubility of wool keratin fibres in ionic liquids 19
Table 2. 2: Dissolution of wool in ILs and time taken to dissolve the cuticle.[118] 22
Table 2. 3: Ionic Liquids used as solvents or co-solvents for the protein-based materials 23

Table 5. 1: Lysine and Arginine simulation conditions 54
Table 5. 2: Domain analysis for $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixtures 75
Table 5. 3: Self diffusion coefficients calculated from MD simulations for lysine systems in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixtures 77

Table 5.4. 1: Domain analysis for $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$/water mixtures 83
Table 5.4. 2: Experimental Solubility values for L-Lysine and L-Aspartic acid in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ ILs.[242] 84
Table 5.4. 3: Self-diffusion coefficients calculated from MD simulations for lysine systems 84

Table 6. 1: List of Aspartate and Glutamate simulations 90

Table 6.3. 1: Domain analysis for aspartate in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixtures. 107
Table 6.3. 2: Self diffusion coefficients calculated from MD simulations for aspartate systems 108

Table 6.4. 1: Experimental Solubility values for L-Aspartic acid and L-Glutamic acid in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$[242] 116
Table 6.4. 2: Domain analysis for $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$/water mixtures 117
Table 6.4. 3: Self diffusion coefficients calculated from MD simulations for aspartate systems 118

Table 7. 1: Simulation conditions for Alanine and aromatic amino acids 124
Table 7. 2: Domain analysis for alanine in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixtures 132
Table 7. 3: Domain analysis for $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$/water mixtures 137
Table 7. 4: Self-diffusion coefficients calculated from MD simulations for alanine systems in pure and IL/water mixtures of $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ 138
Table 7. 5: Domain analysis for aromatics systems in pure $[C_2C_1Im][CH_3CO_2]$ IL 142
Table 7. 6: Self diffusion coefficients calculated from MD simulations for aromatic amino acids in pure and IL/water mixtures of $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ IL 143

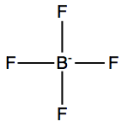
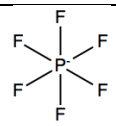
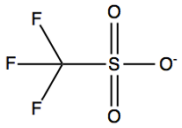
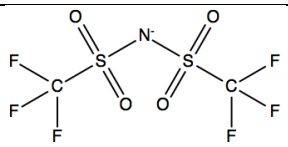
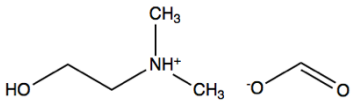
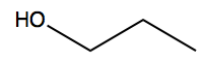
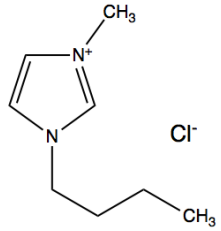
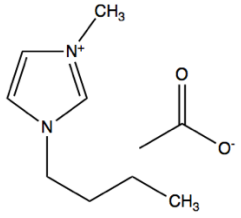
Table 8. 1: List of peptides simulations carried out in water or in $[C_2C_1Im][CH_3CO_2]$ IL 147

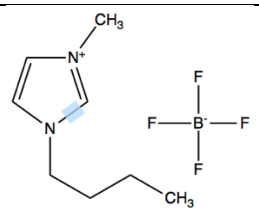
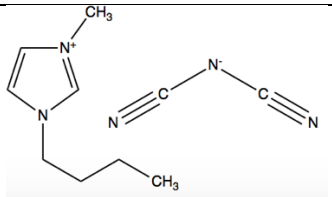
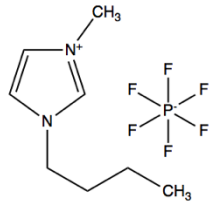
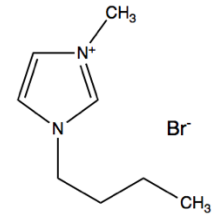
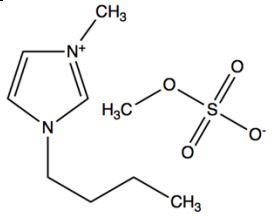
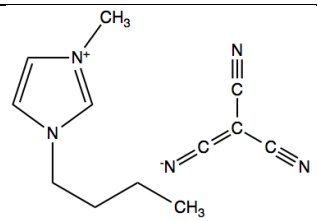
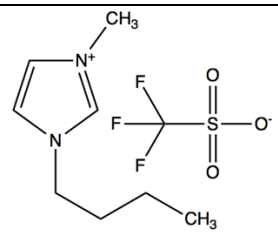
Table A. 1: Karl Fischer measurements for water in Ionic liquids taken from Steven Maxwell Hogg's work 183
Table A. 2: Calculation for number of water molecules from Karl Fischer measurements 183

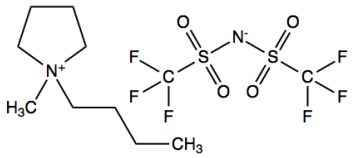
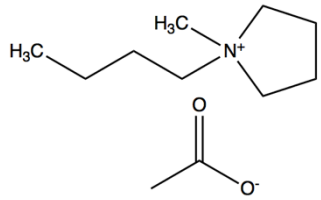
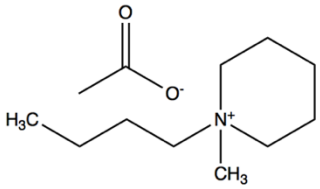
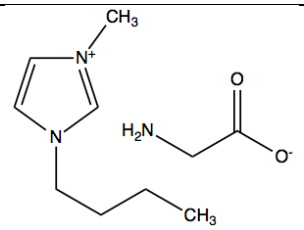
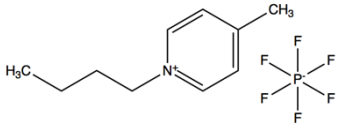
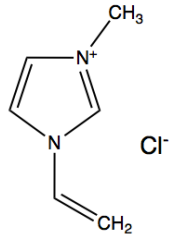
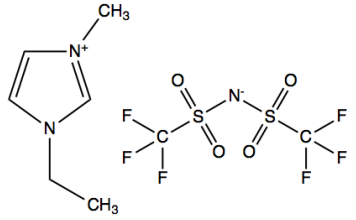
LIST OF ABBREVIATIONS

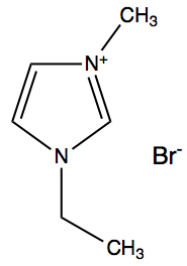
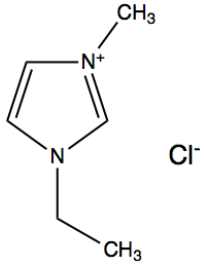
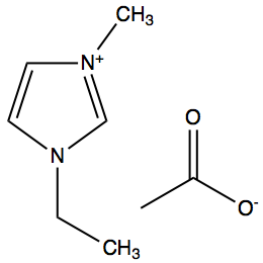
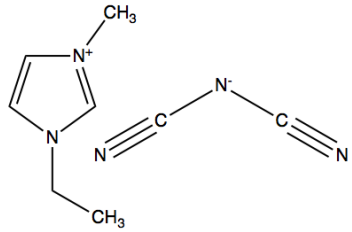
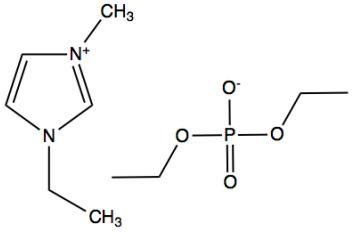
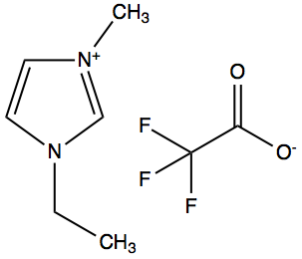
Alanine	ala
Alpha	α
Amino Acids	AAs
Amino acid-based Ionic Liquids	AAILs
Aspartate	asp
Assisted Model Building with Energy Refinement	AMBER
Beta	β
Candida Antarctica lipase B	CAL-B
Circular Dichroism	CD
Graphics Processing Units	GPUs
General Amber Force Field	GAFF
Glycine	gly
Glutamate	glu
GROningen MAchine for Chemical Simulation	GROMACS
Ionic Liquids	ILs
Infrared	IR
Lysine	lys
Molecular Dynamics	MD
Mean square displacement	MSD
Nuclear Magnetic Resonance	NMR
Polarizing optical microscope	POM
Quantum Mechanics	QM
Restrained Electro-Static Potential fit	RESP
Root mean square deviation	RMSD
Simulated Annealing with NMR-Derived Energy Restraints	SANDER
Self-Diffusion coefficient	D
Serine	Ser

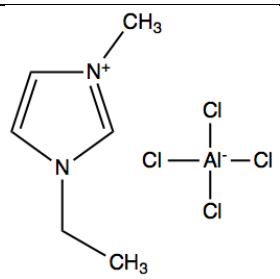
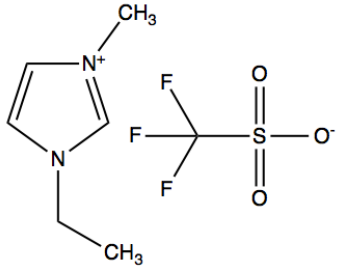
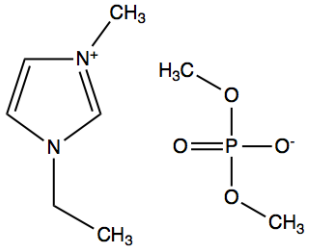
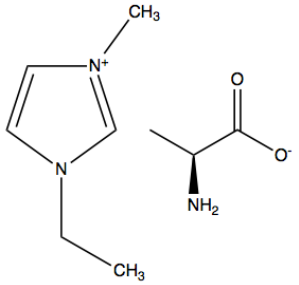
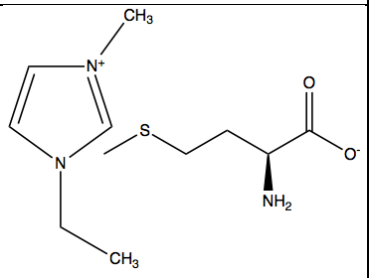
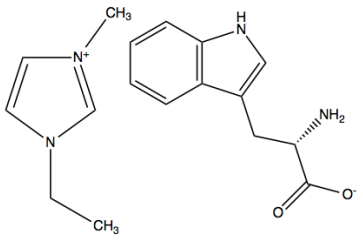
IONIC LIQUID ABBREVIATIONS

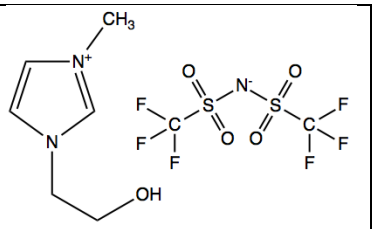
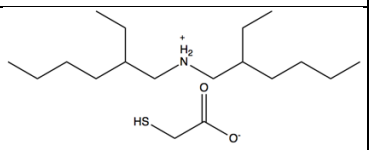
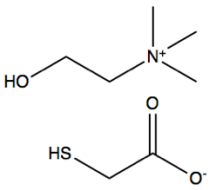
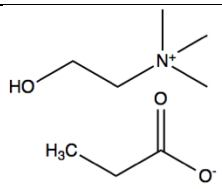
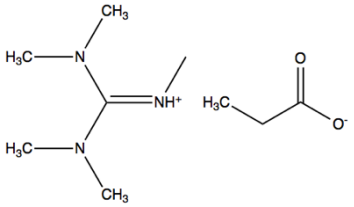
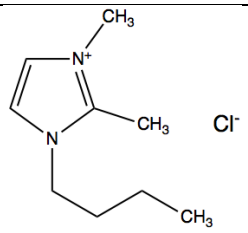
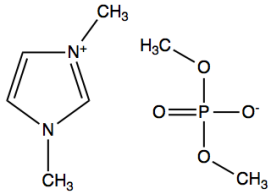
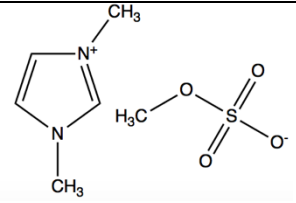
Name	Symbol/Abbreviation	Structure
Chloride	Cl^-	
Bromide	Br^-	
Iodide	I^-	
Acetate	OAc^-	
Tetrafluoroborate	BF^-_4	
Hexafluorophosphate	PF^-_6	
Trifluoromethanesulfonate	$CF_3SO^-_3$	
bis(trifluoromethanesulfone) imide	$CF_3(SO_2)_2N^-$ [NTf_2]	
Di-methyl-ethanol-ammonium formate	[DMEA][HCOO]	
1-propanol		
1-butyl-3-methylimidazolium chloride	[BMIM][Cl] [C_4C_1Im][Cl]	
1-butyl-3-methylimidazolium acetate	[BMIM][OAc] [C_4C_1Im][OAc] [C_4C_1Im][CH_3CO_2]	

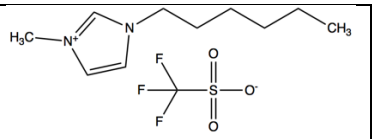
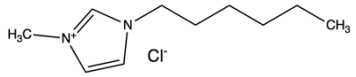
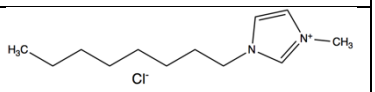
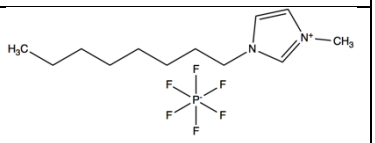
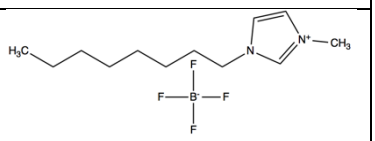
1-butyl-3-methylimidazolium tetrafluoroborate	[BMIM][BF_4] [C_4C_1Im][BF_4]	
1-butyl-3-methylimidazolium dicyanamide	[BMIM][dca] [C_4C_1Im][dca]	
1-butyl-3-methylimidazolium hexafluorophosphate	[BMIM][PF_6] [C_4C_1Im][PF_6]	
1-butyl-3-methylimidazolium bromide	[BMIM][Br] [C_4C_1Im][Br]	
1-butyl-3-methylimidazolium methyl sulfate	[BMIM][C_1SO_4] [C_4C_1Im][C_1SO_4] [C_4C_1Im][$MeSO_4$]	
1-butyl-3-methylimidazolium tricyanomethane	[BMIM][$C(CN)_3$] [C_4C_1Im][$C(CN)_3$]	
1-butyl-3-methylimidazolium trifluoromethanesulfonate	[BMIM][CF_3SO_3] [C_4C_1Im][CF_3SO_3]	

1-butyl-1-methylpyrrolidinium bis(tri-fluoromethane-sulfonyl)-imide	[BMpyrr][NTf_2] [C_4C_1pyrr][NTf_2]	
1-butyl-1-methylpyrrolidinium acetate	[BMpyrr][OAc] [C_4C_1pyrr][OAc]	
1-butyl-1-methylpiperidinium acetate	[BMpip][OAc] [C_4C_1pip][OAc]	
1-butyl-3-methylimidazolium glycine	[BMIM][Gly] [C_4C_1Im][Gly]	
1-butyl-4-methylpyridinium hexafluorophosphate	[C_4C_1py][PF_6]	
1-allyl-3-methylimidazolium Chloride	[AMIM][Cl]	
1-ethyl-3-methyl-imidizolium bis(trifluoromethyl)sulfonyli mide	[EMIM][NTf_2] [C_2C_1Im][NTf_2]	

1-ethyl-3-methyl-imidizolium bromide	[EMIM][Br] $[C_2C_1Im][Br]$	
1-ethyl-3-methyl-imidizolium chloride	[EMIM][Cl] $[C_2C_1Im][Cl]$	
1-ethyl-3-methyl-imidizolium acetate	[EMIM][OAc] $[C_2C_1Im][OAc]$ $[C_2C_1Im][CH_3CO_2]$	
1-ethyl-3-methyl-imidizolium dicyanamide	[EMIM][dca] $[C_2C_1Im][dca]$	
1-ethyl-3-methylimidazolium diethyl phosphate	$[EMIM][Et_2PO_4]$ $[C_2C_1Im][(C_2H_5)_2PO_4]$	
1-ethyl-3-methylimidazolium trifluoro acetate	$[EMIM][CF_3CO_2]$ $[C_2C_1Im][CF_3CO_2]$	

1-ethyl-3-methylimidazolium tetrachloro aluminate	[EMIM][$AlCl_4$] [C_2C_1Im][$AlCl_4$]	
1-ethyl-3-methylimidazolium trifluoro methane sulfonate	[EMIM][CF_3SO_3] [C_2C_1Im][CF_3SO_3]	
1-ethyl-3-methyl-imidizolium dimethyl Phosphate	[EMIM][$(CH_3)_2PO_4$] [C_2C_1Im][$(CH_3)_2PO_4$]	
1-ethyl-3-methyl-imidizolium alaninate	[EMIM][Ala] [C_2C_1Im][Ala]	
1-ethyl-3-methyl-imidizolium methioninate	[EMIM][Met] [C_2C_1Im][Met]	
1-ethyl-3-methyl-imidizolium tryptophanate	[EMIM][Trp] [C_2C_1Im][Trp]	

1-hydroxyethyl-3-methylimidazolium bis(trifluoro-methane-sulfonyl)imide	[HOEMIM][NTf_2]	
[bis(2-ethylhexyl)ammonium][thioglycolate]		
Choline Thioglycolate	[Choline][TG]	
Choline Propionate	[Choline][Pn]	
1,1,3,3-tetramethylguanidinium propionate	[TMG][Pn]	
1-butyl-2,3-dimethylimidazolium chloride	[DMBIM][Cl]	
1,3-dimethylimidazolium dimethyl phosphate	[MMIM]Me_2PO_4] [C_1C_1Im][$(CH_3)_2PO_4$] [C_1C_1Im][DMP]	
1,3-dimethylimidazolium methyl sulfate	[MMIM]$MeSO_4$] [C_1C_1Im][CH_3SO_4] [C_1C_1Im][C_1SO_4]	

1-hexyl-3-methylimidazolium trifluoro methane sulfonate	$[C_6C_1Im][CF_3SO_3]$	
1-hexyl-3-methylimidazolium chloride	$[C_6C_1Im][Cl]$	
1-octyl-3-methylimidazolium chloride	$[C_8C_1Im][Cl]$	
1-octyl-3-methylimidazolium hexafluorophosphate	$[C_8C_1Im][PF_6]$	
1-octyl-3-methylimidazolium tetrafluoroborate	$[C_8C_1Im][BF_4]$	

This Thesis is dedicated to my beloved parents, my sisters, my husband and my son

1

Objectives and thesis Outline

Keratin is the most abundant structural fibrous protein of hair, skins, bristles, horns, hooves, bird feathers, wool, nails and claws. [1, 2] Millions of tonnes of keratinous wastes are generated annually globally, especially in the wool textile industry and in poultry slaughter houses.[3] The three dimensional structural stability of wool keratin is attributed to inter-molecular disulfide bonds and to inter and intra-molecular interactions between polar and non-polar amino acids that renders it difficult to solubilise in conventional organic solvents.[4] To facilitate the processing of these materials into high-value, biocompatible products, ILs have been identified as potential solvents. Many ionic liquids are able to dissolve various polymers and protein-based materials. The ionic liquids 1-ethyl-3-methylimidazolium chloride ($[C_2C_1Im][Cl]$) and 1-ethyl-3-methylimidazolium acetate ($[C_2C_1Im][CH_3CO_2]$) are among the best solvents for cellulose and silk dissolution,[5, 6] and so were selected for the current project.

The primary goal of the study was to elucidate the molecular mechanism underlying the dissolution processes of protein-based materials in ILs through molecular dynamics simulation studies. For this purpose, various amino acid model compounds were identified primarily based on their abundance in wool keratin fibres. The first objective was to analyse the molecular interactions arising between different amino acids and IL ions, in pure ILs and in IL/water mixtures. This was achieved by investigating the origin and pattern of organisation of solvent ions around the solute molecules by exploring two sets of conditions of concentration and temperature. The second objective was to match the results of these simulations, through various analysis methods, with the experimental solubility values of the studied amino acids in two ILs. Solubility values were made available to the project. The relation of MD simulation results with the experimental solubility data was developed. This would enable the validation of the

force-field employed for running these simulations and also validate the parameters for the non-standard amino acid model compounds derived from them. Peptide segments were identified in the literature,[7] and simulations were run in water and pure $[C_2C_1Im][CH_3CO_2]$ IL, to compare the role of interactions between peptide and amino acids. The final objective therefore, was to develop the organisational and structural comparison between peptides and amino acids, in addition to comparison of the secondary structures of peptides in both water and IL.

The outline of thesis is as follows. Chapter 2 covers a general introduction about various protein-based materials, for example wool, silk and other keratin- based materials, their structural properties and applications. Ionic liquids are discussed, and their general properties are outlined. The applications of ILs as solvents for natural polymers such as cellulose and lignin, and particularly for keratin-based materials are reviewed. In addition to this, the use of ILs for proteins are discussed and the roles of IL cations and anions for stabilisation/destabilisation of protein structures are highlighted. Furthermore, this chapter provides a detailed review of experimental and computational studies conducted for proteins and ILs.

In Chapter 3, the overall methodology used for running the molecular dynamics simulations is discussed. QM calculations for the preparation and optimisation for all amino acids and IL ion structures used in the simulation are explained along with the discussion on parameters and their derivation. Peptide structures construction and forcefield selection for different sets of simulations are outlined. The procedure for building simulation boxes using packmol[8] and the steps followed for running MD simulations using AMBER[9] are explained.

In Chapter 4, the methods of analysis used for the interpretation of results of all the simulations are discussed along with the theoretical background. The methods of analysis include, radial and spatial distribution functions that give a detailed overview about local structure in amino acid and IL solutions; domain analysis, which provides information about whether amino acid solute molecules are aggregating in the given system; and finally self-diffusion coefficients are discussed. Peptide simulations analysis methods including Ramachandran plots and secondary structure analysis, are

also explained. The software tools used to run these analyses and the visual depiction of results are also outlined.

Chapters 5 and 6 cover the detail on interactions of positively charged amino acids and negatively charged amino acids in two ILs $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$. Each chapter is divided into two subchapters to separately deal with the two ILs in their neat form as well as their mixtures with water. The concentration ratio and temperature effects on the organisational or the structural behaviour of the respective solute and the solvent molecules are discussed through radial and spatial distribution functions. Attention is focused on understanding the role of interactions established between different atomic sites of amino acid molecules and ionic liquid ions. Domain count analysis is done to relate the data with the experimental solubility values. Self-diffusion coefficients are measured through mean square displacement curves to explain the dynamics of the systems.

Chapter 7 discusses amino acids with neutral side chains including alanine and the aromatic amino acids. Alanine is investigated on the similar lines of positively and negatively charged amino acids in two ILs in their pure and in water mixture forms, by exploring the effect of temperature so that a comparison is developed between charged and neutral solutes. Based on the better solubility of amino acids in $[C_2C_1Im][CH_3CO_2]$ IL aromatic amino acid were investigated for this ionic liquid. The solubilities of neutral solutes and their organisational preferences for each ionic liquid ions were analysed through the similar analyses run in chapter five and six.

In Chapter 8, peptide segments penta-alanine, deca-alanine and a peptide segment from protein G were simulated in water and $[C_2C_1Im][CH_3CO_2]$ IL to analyse the organisational similarity in comparison with single amino acids.

Lastly, Chapter 9 summarises the main discussion and conclusions obtained through these studies. Recommendations for future work are also included in the end.

2

Literature review

2.1 Introduction

Polymers derived from renewable resources have gained an increasing amount of attention from both the scientific and industrial communities over the last two decades, that has led to the utilisation and exploitation of natural polymers as a source of energy, fuels and raw materials for high value and specialty products. There have been predominantly two major reasons for this focussed attention, firstly environmental concerns, and secondly the realisation that our petroleum resources are finite.[10,11]

Natural polymers are produced by living organisms such as plants, animals and microorganisms and these biopolymers can be classified into three main categories such as polysaccharides, proteins and nucleic acids depending on their main components.[12] Protein-based materials, such as silks and wool have an evident history of their use as raw materials for the textile industry. These materials have unique physical properties and have an advantage of being renewable, biodegradable and readily biosourced.[13] Many structural proteins such as silks, keratins, elastins and collagens have been studied and their applications in biomedical research are being explored. Their biocompatibility and properties such as mechanical strength, elasticity, porosity and high surface area-to-volume ratio have made these materials an attractive choice to use them in scaffolds, sutures, wound healing, ligament replacement, and drug delivery technology.[14] Owing to these properties and applications, protein based materials are important for developing a circular economy and reducing the environmental impacts. But with the exception of high-quality wool, that is used to produce garments and rugs, several low grades and trimmings obtained from slaughterhouses cannot be used in the wool industry and end up as a waste stream. Also, feathers waste, a by-product of poultry industry, poses threat to the environment and economy when most of it is

disposed of by incineration or buried in landfills.[15] This not only causes the waste of keratin resources, but also pollutes the environment.[16] Therefore, it is valuable to regenerate and convert biodegradable wool keratin into high-value-added commercial biomaterials from both the environmental and economic standpoint. The development of energy efficient, cost effective and environmentally friendly dissolution processes for keratin are the foundation for further reusing and recycling waste wool keratin.[17] However, it is extremely difficult to dissolve wool keratin and many natural polymers in conventional solvents due to the presence of strong inter and intra-molecular hydrogen bonds that are responsible for the tight packing of the secondary structures in the polypeptide chains and this greatly limits the applications of natural polymers in many fields.

2.2 Proteins

Proteins, also known as polypeptides, are large complex molecules that are inherently composed of twenty different naturally-occurring amino acids, where residues are linked covalently by peptide bonds (–CO–NH–).[18] They are essential macromolecules in biological systems. The sequence of residues in the polypeptide is known as the primary structure that forms the backbone of the biomolecule. Standard proteins can be categorized as polar, non-polar, aromatic, anionic and cationic in nature. [19] Moreover, due to the presence of weak basic and acidic character of the amino and carboxylic groups respectively, these can exist in the corresponding ionised forms, leading to the zwitterionic form of AAs.[11] Proteins have different levels of structural organization called primary, secondary, tertiary and quaternary structures and the underlying amino acid (AA) sequences of proteins play a vital role in several biological processes. The proteins need a distinct 3D structure within the surroundings to perform specific functions. The properties of proteins are dependent on the nature and the arrangement of the AA side chain along the protein backbone[20, 21] and being polymeric materials, the range of materials properties that are accessible is substantial. The different amino acid functionalities lend these protein-based materials to further post-processing, which can extend their property scope.

Proteins can be classified according to their shape as globular or fibrous proteins. Globular proteins often exhibit mixed secondary structures with specific domains of α-

helix, β-sheet and random coil. Fibrous proteins are characterized by repetitive structural motifs that impart a significant influence on their macroscopic functional properties.[22] Keratin is the major structural fibrous protein providing outer covering for hair, wool, feathers, nails, and horns of mammals, reptiles, and birds.[23] Silk filament is a similar fibrous protein consisting of a double strand of fibroin, which is held together by a gummy substance called sericin.

2.2.1 Silk

Silk is a natural fibre produced by silkworms or spiders, and other insects. Silk fibres are composed of two monofilaments and consist of two different proteinaceous substances: silk fibroin (SF) and sericin. Fibroin is a protein that forms the core of silkworm silks and these fibroin-based fibres are coated and strengthened together by glue-like protein called sericin to produce a cocoon that provides a protective environment for the silkworm to undergo its transformation into a moth.[13, 24] Figure 2.1 shows the schematic diagram of SF. Bombyx mori (*B. mori*) silkworm fibres have been used in the textile industry for thousands of years. Owing to the mechanical strength, softness, biodegradability and biocompatibility, silk fibres have found varied range of applications in biomedicine and biotechnology, for example in tissue regeneration, biosensors, drug release, cell adhesion and growth, and also as the potential starting material for the preparation of medical devices.[25]

The amino acid composition of SF from B. mori consists primarily of glycine, alanine, and serine amino acids in the molar ratio of 3:2:1 which forms typical $-(-gly-ala-)_n-$ repeating motif of the hexapeptide GAGAGX (where X = S, Y, A) broken up periodically by a (GAAS) motif. [26, 27] These three residues consisting of short side chains permit close packing of crystals through the stacking of hydrogen-bonded β-sheets,[28] that contribute greatly to the rigidity and tensile strength of SF. The hydrophobic domains in the silk structure are formed by the highly repetitive blocks in heavy chain of SF, of hexapeptide sequence and repeats of Gly-Ala/Ser/Tyr dipeptides. The amorphous domains are characterized by the presence of amino acids with bulkier side chains and allow the crystalline domains to orient under strain and contribute to the elasticity of silk fibres. [29, 30]

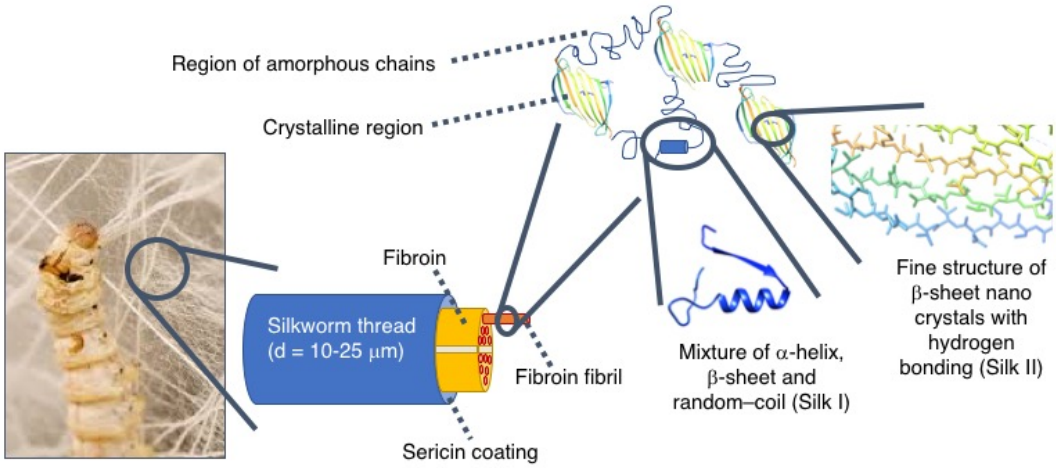


Figure 2. 1: Schematic representation of SF structure. Insets show the fibril overall structure and the fine β-sheet antiparallel alignment of SF polypeptide chains. Adapted from[24]

2.2.2 Keratin

Keratin is an insoluble structural protein that makes up the bulk of the epidermis, including hair, horns, and fingernails. The presence of cell binding motifs on keratin, as well as its ability to self-assemble, make it an ideal natural polymer to be used in the creation of biomaterials for tissue regeneration[31]. They can serve as raw material in the production of eco-friendly biocompatible materials that can further be employed for the production in textile industry, building materials, films, microcapsules, sponges, bioplastics, edible materials, compostable packaging and the components of composites.[23, 31-33]

The keratin protein can be characterised as a soft or hard keratin depending on its amino acid sequence and the sulfur content. Hard keratins, like wool, have a high sulfur content, and therefore more covalent (disulfide) crosslinks whereas soft keratins can be found in the outer layer of the skin and have lower sulfur content and hence fewer covalent (disulfide) cross-links.[34] The structure and composition of keratin fibres is quite complex and varies depending on its source such as wool, human hair or chicken feathers. The α-helix or β-sheet chains are tightly packed into a coiled polypeptide chain (Figure 2.2) and keratin fibres exhibit a stable three-dimensional structure owing to the presence of a range of disulfide bridges, hydrogen bonds, hydrophobic forces and electrostatic forces[35] (Figure 2.3). This structural stability not only is responsible for its desirable mechanical properties but also makes it a challenge to find suitable solvents for its dissolution.[36]

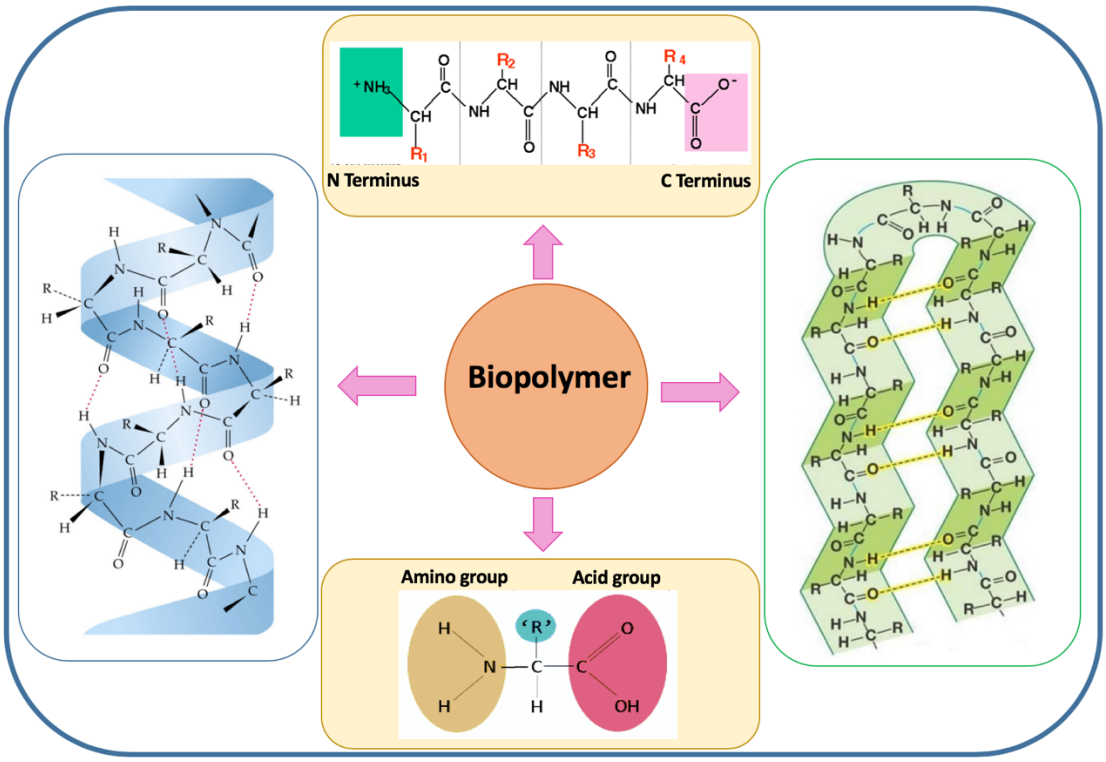


Figure 2. 2: Schematic diagram of different components of keratin; (A) polypeptide single unit, (B) amino acid sequence, (C) α-helix, and (D) β-Sheets conformations of keratin protein. Taken from [37]

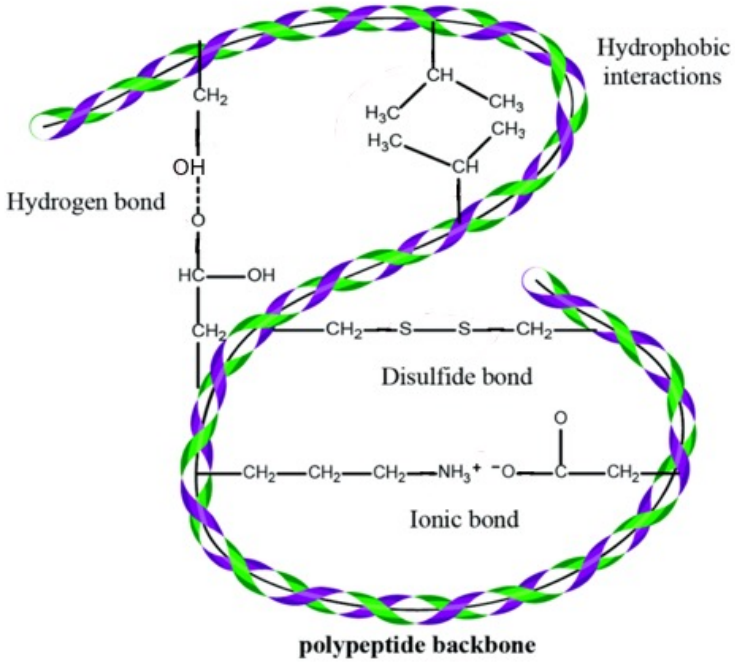


Figure 2.3: Inter- and intra-molecular bonding in keratin.[15]

In addition to the high content of cysteine (11-17%) in wool keratin, there are several AAs that dominate the keratin structure and thus contribute to the high stability of the

protein. Lysine, arginine, aspartate and glutamate are highly concentrated within the keratin structure when compared to the other AAs present in the protein sequence.[38] The neutral structures for these amino acids are shown in the figure 2.4, indicating important hydrogen-bonding sites.

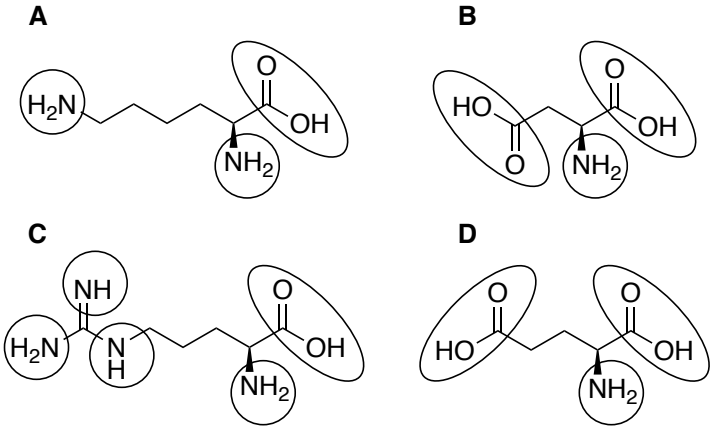


Figure 2.4: Structures of (A) lysine, (B) aspartate, (C) arginine and (D) glutamate indicating important bonding sites.

All these inter- and intra-peptide chain interactions are responsible for the extremely high stability of wool keratin fibres that render them extremely difficult to be dissolved by regular solvents. The major methods, such as reduction, oxidation, microwave irradiation, alkali extraction, steam explosion, sulfitolysis for keratin dissolution or extraction have been reviewed by Shavandi *et al*.[39] The solvents or mixtures of solvents used in these methods play a key role in breaking covalent or non-covalent bonds, but they are toxic and are non-renewable. These reagents lead to excessive degradation of keratin and yields a lower molecular weight of regenerated keratin, which decreases its value.[17] In addition to toxicity and lack of renewability, the other disadvantages associated with these solvent regimes have been highlighted; such as volatility, corrosiveness, environmental impact and difficulty of recycling.[40-42] There has been a clear demand to develop sustainable dissolution process for keratin without degrading the protein structure. ILs have been investigated for this purpose and have shown promising results.[16, 43, 44,45] Small model compounds or peptides that can mimic the structure of keratin segments, along with exploring the prominent amino acids present in the keratin structure, and can form a basic framework to help in understanding and development of the keratin dissolution process.

2.3 Ionic liquids

ILs are termed as molten salts with melting points below 100 °C. ILs are purely ionic, salt-like materials that are liquid at unusually low temperatures. In particular, salts that are liquids at room temperature are called room-temperature ionic liquids (RTILs). RTILs, also known as organic liquid, molten, or fused salts, are a class of typically molecular ionic solvents with low melting points.[46, 47] RTILs are constituted of organic cations such as imidazolium, pyridinium, pyrrolidinium or ammonium derivatives. These cations can be combined with either organic or inorganic anions such as Cl^-, Br^-, I^-, $[BF_4]^-$, $[PF_6]^-$, $[CF_3SO_3]^-$, $[(CF_3SO_2)_2N]^-$, $[CH_3COO]^-$, $[N(CN)_2]^-$ etc. The constituting ions can be varied independently to provide task-specific materials for defined applications and therefore they are often termed as "Designer Solvents".[48] Ionic liquids have an elaborate network of often asymmetric ions that are bonded by electrostatic charges and hydrogen bonding.

2.3.1 Properties of Ionic liquids

ILs are characterised by high densities and viscosities, are thermally and chemically stable and have low vapour pressure that distinguish them from their molecular solvent counterparts.[49] The strong electrostatic interactions of the constituent ions results in a number of useful properties, some of which underpin their "green" reputation. This includes minimal harmful vapor when handling, low flammability, and the lack of vaporization, that provides an opportunity to recycle these liquids across many cycles and make them industrially attractive solvents to be used in various types of synthesis processes.[13, 50, 51] The properties of ionic liquids result from a complex interplay of intermolecular interactions between ions, their geometry and charge distribution that extends over larger distances in comparison to molecular solvents.[52, 53] All possible universal (Coulomb, van der Waals, dipole-dipole) and specific (*e.g.* H-bonding) intermolecular interactions can define the behaviour of ILs. Long-range order in simple salts is governed by Coulomb's law, whereas electrostatic forces in ILs are reduced due to the presence of van der Waals interactions.[53]

2.3.2 Ionic liquids as solvents for natural polymers and their applications

The interest in establishing alternative solvents for organic synthesis to provide safe and environmentally clean processes has increased dramatically over the last decade.[54] There, a broad range of uses and applications of ILs including separation and extraction,[45, 55, 56] polymerisation,[57] organic synthesis and catalysis,[58, 59] inorganic synthesis,[60] and as binders to prepare biosensors.[46] Particularly, ILs possess the ability to dissolve polar, nonpolar, organic, inorganic and polymeric compounds and have been identified as a promising alternative to organic solvents for various applications, including those involving biological macromolecules, e.g., enzymes[48, 61] and proteins where ILs have shown to solubilise them and also lead to structural stabilisation.[62]

Due to their favourable solvation properties, ILs find applications in dissolution of many polar substrates, including carbohydrates[63], proteins.[53, 54, 58, 64, 65] and other natural polymers, such as cellulose[66-68], starch and lignin.[69, 70] These polymers are not only used individually but also in blends and composites, that further improve the properties and functionalities of these materials and extend their application fields. The unique properties of some of these natural polymers open up perspectives for the development of innovative biomaterials.[71] By using ILs to dissolve these materials in 5-25 wt% solution, unachievable with standard solvents, methodologies for creating films,[72] composite mixes,[73] and 3-D printing[74] are being explored and have particular current relevance in the manufacture of medical implants.

As a case study for the use of ionic liquids, lignocellulose stands out. Lignocellulose is a composite material synthesised by plant cells and mainly consists of polymeric carbohydrates (cellulose and hemicelluloses) and the aromatic polymerlignin.[75] The composition of lignocellulose depends on the species, the plant tissue and the growth conditions. Lignin is a complex, amorphous biopolymer that holds cellulose and hemicellulose together,[76] and unlike proteins or other natural polymers such as cellulose, it lacks clearly defined secondary and tertiary order structure.[77] The growing interest in developing new lignin-based products is driven by the fact that lignin is a renewable, low-cost bioresource and environmentally benign with unique functionalities, that it can be directly transformed to aromatic specialty and fine chemicals that can function as a basic building block chemical—performing similarly

to petroleum-based analogues (benzene, xylene, toluene).[78] Lignin can also be employed in composite materials to improve the performance of other polymers. Several advantageous properties can be realized (antioxidant, anti-bacterial, anti-UV, etc.) by incorporating small amounts of lignin with other polymers.[79] Cellulose is a linear polymer of cellobiose that consists of two glucose sugar units that are linked by glucosidic linkages (C-O-C) at the C1 and C4 positions. (Figure 2.5) The size of its chains depends on the degree of polymerization (DP), which is the number of repeating units of glucose. Multiple hydrogen-bonding between cellulose molecules results in the formation of highly ordered crystalline regions; furthermore, these are mixed or covered by lignin. It is almost impossible to solubilize cellulose with water and common organic solvents. The high chemical and mechanical stability of cellulose and its hydrophobic nature is due to the strong glycosidic bonds and the huge degree of both intra- and intermolecular hydrogen-bonding. They not only act as interconnections between the cellobiose units but are also responsible for the hydrophilic nature of the biopolymer.[80, 81]

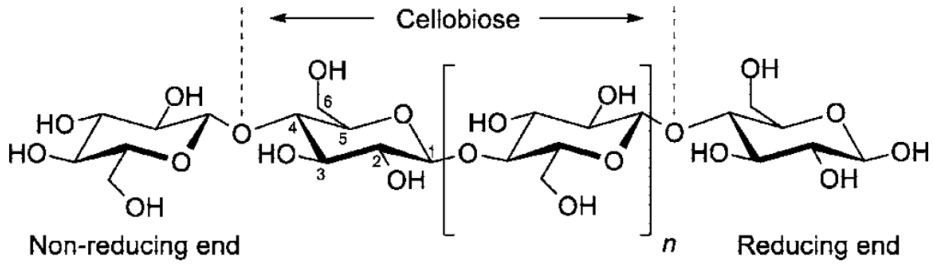


Figure 2. 5: Cellulose polymer chain structure[80]

Lignin dissolution in ILs have been the subject of many studies, considering either the whole biomass or commercially available Kraft lignin. Yunqiao Pu *et al.*,[77] investigated the applicability of aprotic ILs for lignin dissolution. Dissolution experiments were carried out employing lignin isolated from pine kraft pulp. In this study, the dissolution of a softwood lignin in different ILs including, $[C_1C_1Im][MeSO_4]$, $[C_6C_1Im][CF_3SO_3]$, $[C_4C_1Im][BF_4]$), $[C_4C_1Py][PF_6]$, and $[C_4C_1Im]^+$ cations with a range of anions ($[C_1SO_4]^-$, Cl^-, Br^-, and $[PF_6]^-$) was examined. It was observed that $[C_1C_1Im][C_1SO_4]$ and $[C_4C_1Im][C_1SO_4]$ were able to dissolve lignin at room temperature((~74 g/L) and (~62 g/L) respectively. Rapid dissolution along with increased solubility of lignin was noticed at higher temperature of 50°C resulting in highly viscous solutions (i.e., 344 g/L in $[C_1C_1Im][MeSO_4]$ and 312 g/L in $[C_4C_1Im][MeSO_4]$. Lignin could not be

dissolved in $[C_6C_1Im][CF_3SO_3]$ under ambient conditions but comparable amounts of 275 g/L were observed to be dissolved at 70°C. For the $[C_4C_1Im]^+$containing ionic liquids, the order of lignin solubility for varying anions was found to in order ($[MeSO_4]^-$ > Cl^-, Br^->>>$[PF_6]^-$), indicating that the solubility of lignin was principally influenced by the nature of the anions.

Lee *et al*.,[82] demonstrated a set of ILs that possess very high solubility for lignin with low solubility for cellulose and other constituents of wood flour. They were able to selectively extract lignin from lignocellulose and simultaneously yield a highly biodegradable cellulose fraction. The selective extraction of lignin from lignocellulosic biomass can therefore provide a route towards increased total utilization of lignocellulosic biomass. $[C_4C_1Im][CF_3SO_3]$ and $[C_1C_1Im][C_1SO_4]$ could preferentially dissolve lignin from lignocellulose while having low solubility for wood-flour. $[C_4C_1Im][PF_6]$ and $[C_4C_1Im][BF_4]$ ILs did not favour lignin solubility. An interesting compromise between high lignin solubility and low wood flour solubility was achieved when $[C_2C_1Im][OAc]$ IL was used as a pretreatment solvent for wood flour. The cellulose in the pretreated wood flour becomes far less crystalline without undergoing solubilization. When 40% of the lignin was removed, the cellulose crystallinity index dropped below 45, resulting in >90% of the cellulose in wood flour to be hydrolyzed by Trichodermaviride cellulase. $[C_2C_1Im][OAc]$ was easily reused without significant loss in the yield of extracted lignin and subsequent cellulose hydrolysed after four cycles. And the resulting solution of highly concentrated unmodified lignin thus obtained should be of value in preparing lignin-based products.

Fu *et al,*[83] reported $[C_2C_1Im][OAc]$ IL as an effective solvent for lignin extraction from triticale straw, flax shives and wheat straw. 52.7 wt % of original lignin in triticale straw was extracted using $[C_2C_1Im][OAc]$ IL at 150°C temperature after 90 min exposure time. They also reported that $[C_2C_1Im][OAc]$ was better at delignification of straw as compared to $[C_4C_1Im][Cl]$ IL. Moniruzzaman and Ono in 2012, also demonstrated the use of $[C_2C_1Im][OAc]$ IL as a co-solvent as well as a pre-treatment reagent for delignification of wood chips. Higher delignification yield of lignin(50%) was achieved with pretreatment of wood chips as compared to 20% without the pretreamtment.[84] The applicability of $[C_4C_1Im][HSO_4]$ IL for the pretreament of lignocellulose was also examined by Brandt *et al*[85]. They found that the IL liquors containing methyl sulfate,

hydrogen sulfate and methane sulfonate anions were the most effective for lignin/cellulose fractionation and cellulose digestibility.

$[C_4C_1Im][Cl]$ and dicyanamide anion based ILs also remained the focus of research for dissolving cellulose and carbohydrates, owing to their strong ability to disrupt hydrogen bonds under relevant conditions. Forsyth *et al*.,[86] first used dicyanamide anion based ionic liquids as solvents for the dissolution of carbohydrates in 2002. The ILs $[C_4C_1Im][dca]$ and $[C_2C_1Im][dca]$ were shown to dissolve glucose in these liquids to greater than 10 wt% at room temperature. The solubility of disaccharides and trisaccharides (e.g. sucrose and raffinose respectively) was found to be less than glucose, although the solubility of all saccharides increased with rising temperature. Liu *et al.* also reported that ILs based on the dicyanamide anion were highly effective aprotic solvents that can dissolve carbohydrates such as glucose, starch and even cellulose in large amounts.[87]

Swatloski *et al*.,[88] investigated ILs containing $[C_4C_1Im]^+$ cation with a range of anions from small, hydrogen-bond acceptors Cl^- to large, noncoordinating anions $[PF_6]^-$ including Br^-, $[SCN]^-$, and $[BF_4]^-$ for cellulose dissolution. In addition, for chloride salts, alkyl-substituents on imidazolium cation were varied from butyl through octyl to investigate the effect of chain length on the dissolution capacity. Dissolution experiments were carried out using cellulose-dissolving pulps (from cellulose acetate, lyocell, and rayon production lines), fibrous cellulose (Aldrich), and Whatman cellulose filter papers. The untreated cellulose samples were added to the ILs in glass vials and heated without agitation on a heating plate or in a domestic microwave oven. The group reported that $[C_4C_1Im][Cl]$ had the highest solubility for cellulose to yield increasingly viscous solutions., and prepared an up to 10 wt% of cellulose/IL solution by heating in an oil bath at 100 °C and 25 wt% using a microwave-assisted heating. The longer-chain substituted ILs ($[C_6C_1Im][Cl]$ and $[C_8C_1Im][Cl]$) appear to be less efficient at dissolving cellulose. Chloride ions are non-hydrated in these ILs and $[C_4C_1Im][Cl]$ has a chloride concentration of approximately 20 mol % that is almost 3 times higher than compared to dimethylacetamide [(DMAC)/LiCl] solvents. They postulated that this high chloride concentration and activity in $[C_4C_1Im][Cl]$ makes it highly effective in breaking the extensive hydrogen-bonding network present in cellulose fibres thus

allowing for much quicker dissolution times, and dissolving higher concentrations of cellulose than the other Cl^- based ILs or the traditional solvents.

The lignocellulosic biomass, being a recalcitrant material poses a major challenge for its sustainable and cost-effective utilization. An efficient pre-treatment is decisive for processes based on lignocellulose. Some studies relied on the solubility of the individual wood polymers in ionic liquids, while others compared the amount of each polymer(lignin, hemicellulose or cellulose) that could be dissolved when the wood is treated with an ionic liquid.[89] Zavrel *et al*,[90] conducted a study where ILs were screened for their ability to dissolve cellulose and lignocellulose. All 21 ILs tested were dried thoroughly in a drying oven to reduce the water content to approximately 0.7%. Cellulose and wood chips were used as substrates without further purification. The dissolution profiles indicated $[C_2C_1Im][OAc]$ IL to be the most efficient solvent for cellulose. For the dissolution of hardwood and softwood, [AMIM][Cl] turned out to be the most suitable IL. This was attributed to not only the ability of chloride ions to disrupt the hydrogen bonding, but also the pi-pi interactions that establish between the of π-electrons on $[AMIM]^+$ cation (within its imidazolium ring, and on its sidechain) and cellulose. They suggested that the aromatic structure of cations could help stabilize the H-bonds between the cations and cellulose hydroxyl groups, which needed further validation by experiments or simulation. The role of anions and cations in the dissolution of lignin and cellulose have been highlighted in various studies.[88, 91, 92]

The anion has a strong impact on the dissolution capability and selectivity of the ionic liquid with lignocellulosic systems. H-bond accepting ability (β value) of the anions was closely linked to the solubility of cellulose and the β value obtained in the solvatochromatic study might be a good indicator of the ability to dissolve cellulose.[93] Sun *et al.,*[6] were able to dissolve both the hard and the softwood after milling in $[C_2C_1Im][OAc]$ IL and showed that this IL was better at dissolution than $[C_4C_1Im][Cl]$ under same conditions. Vo *et al*.,[94] tested the dissolution and regeneration of poplar wood using $[C_4C_1Im][Cl]$ and $[C_4C_1Im][OAc]$ ILs and observed that at 130°C 96 wt% of poplar was dissolved in $[C_4C_1Im][OAc]$, while $[C_4C_1Im][Cl]$ could dissolve 25.3wt% under the same reaction conditions. The acetate anion $[OAc]^-$ could efficiently break hydrogen bonding due to its high basicity, and therefore allowed better wood dissolution compared to chloride ion $[Cl]^-$. The most efficient anions reported for

cellulose dissolution reported so far, are $[Cl]^-$, $[OAc]^-$, $[HCOO]^-$ and $[((C_2H_5)O)_2PO_2]^-$ while non-coordinating anions such as $[BF_4]^-$ and $[PF_6]^-$ cannot dissolve cellulose. ILs with imidazolium, pyridinium, ammonium, and phosphonium cations are often better cellulose solvents and all have unsaturated aromatic structures in their cations.[95] Moreover, several aprotic co-solvents, such as dimethyl sulfoxide (DMSO) and N, N-dimethylformamide (DMF), can enhance the dissolution of cellulose in ILs.[96, 97]

Endo *et al*.,[98] investigated the effect of cellulose concentration on the interactions between cellulose and $[C_2C_1Im][OAc]$ IL by wide-angle X-ray scattering (WAXS) and ^{13}C solid-state NMR spectroscopy. They reported that the increase in concentration of cellulose past more than 30 mol% showed a transition in the structure of cellulose, from cellulose chain alignment to ordered layers where the $[C_2C_1Im]^+$ cations and $[OAc]^-$ anions intercalated. They also proposed an anion–bridging structure transformation at high concentrations (>30 mol%) where anions would interact with multiple OH groups resulting in a bridging state.

Alongside experimental studies, computational methods such as COSMO-RS, quantum chemistry calculation and molecular dynamics simulation, have proven to be quite helpful in revealing the structural details at the atomic level. And also provide an understanding of the dissolution mechanism of lignocellulosic materials in ILs. Li *et al*.,[95] examined the influence of the unsaturated heterocyclic ring structure of cations on cellulose dissolution in four kinds of ILs composed of unsaturated ($[C_4C_1Im]$, $[C_4Py]$) and saturated ($[C_4C_1pyrr]$ 1-butyl-4-methylpyrrolidinium, and $[C_4C_1pip]$ 1-butyl-1-methylpiperidinium) cations with the acetate anion through molecular dynamics simulation and DFT calculations. In their results, they showed that unsaturated ILs could effectively separate the cellulose bundles structure whereas, the cellulose structure remained unchanged in saturated ILs. Interaction energies were calculated for all the systems and they reported that unsaturated cations had a larger interaction energy with cellulose and the delocalisation of π-electrons on the five- or six-membered ring makes the cations have stronger contact with the glucose ring of cellulose. The saturated cations, however, introduced a steric hindrance for the anions to interact weakly with the hydroxyl groups in cellulose, that resulted in a much fewer H-bonds between anions and cellulose. The dynamic properties of ILs were also shown to be impacted by the nature of cations. The larger volume of cations with the saturated

heterocyclic ring, resulted in a slow transfer of both cations and anions, hence, have lower diffusion coefficient values as compared to unsaturated cations. Higher viscosity and slower diffusion rates reduced the contact between ions and cellulose and ultimately resulted in lower solubility of cellulose in these ILs.

Molecular dynamics simulations of β-glucose in 1,3-dimethylimidazolium chloride were performed by Youngs *et al*.[99] in order to explain the solvation environment of carbohydrate. Solvent−solute interactions were analysed by simulating isolated monomers in solution. They reported that the predominant solvation interaction is through the chloride anions that formed hydrogen bonds with the hydroxyl groups of the sugar. The average chloride coordination number was found to be four, and weak hydrogen bonding interactions occurring through the acidic hydrogen at the C(2) position on the imidazolium ring were also observed. This suggested that the cation could be tailored to play an important role in the dissolution process. MD simulations were run for small bundles of crystalline cellulose in ILs by Ismail *et al.*[100] Cellulose bundles with six strands of eight to ten glucose units were simulated in $[C_4C_1Im][Cl]$, $[C_2C_1Im][OAc]$ and $[C_1C_1Im][DMP]$ ILs. A common mechanism was observed among all ILs, that the anions bound strongly to hydroxyl groups on the exterior of the individual cellulose strands and formed negatively charged moieties. Cations associated with these anions then inserted themselves between the cellulose strands that further weakened the inter-chain cellulose bonding. The breakup of cellulose strands however was found to be much more rapid in $[C_2C_1Im][OAc]$ IL than compared to the other ILs investigated.

Zhao *et al*. performed molecular dynamics simulations to systematically study the influence of different types of cations[101] and anions[102] on cellulose dissolution in the ILs. The Cl^- anion based ILs were used with 1-alkyl-3- methyl imidazolium cations $[C_nC_1Im]^+$ (n=2,3,4,6,8,10). The comparison of the radial distribution functions and the interaction energies of imidazolium based ILs showed that the interaction of the ILs with cellulose decreases with the increase in the alkyl chain length. They studied the effect of different anionic structures through MD simulations,[102] in $[C_2C_1Im]^+$ cation based ILs and found a decreased order of interaction energy as $[Cl]^- > [Ac]^- > [(CH_3O)_2PO_2]^- > [SCN]^- > [PF6]^-$ suggesting that Cl^- and O atom of $[CH_3COO]^-$ and

$[(CH_3O)_2PO_2]^-$ are better HB acceptors than N atom of $[SCN]^-$ and F atom of $[PF6]^-$ because of higher electron density.

ILs have demonstrated huge potential for the dissolution and processing of lignocellulosic materials. They offer possibilities for the utilisation of these materials for the production of bio-fuels, blends and composites and other chemicals. The fundamental mechanisms have been extensively studied with experimental and computational approaches where roles of individual IL species, effects various processing conditions and the effects of solvents and co-solvents were explained. However, there are some challenges that still need attention. ILs recyclability and toxicity are one of the issues that hinders the industrial application of these processes. Many side reactions, such as acetylation and thermal degradation of ILs at high cellulose processing temperatures have been reported. Key aspects of dissolution are integrated with the hydrogen bond disruption caused by the anionic component and mediated by the cation. Simulations have shown to be an excellent approach to obtain a more molecular view of these interactions.

2.4 Ionic liquids as solvents for keratin

ILs for the dissolution of natural protein fibres have been extensively studied,[41, 43, 69, 85, 88, 103, 104] and represents a new avenue for bio-renewables processing, which can include the development of new materials, and the selective extraction of protein for tissue engineering. [105-107]

In 2004, Phillips *et al*.,[5] demonstrated that bombyx mori silk fibroin could be dissolved by $[C_4C_1Im]^+$ and $[C_2C_1Im]^+$ based ILs. A range of 1-methyl-3-alkylimidazolium ILs have been successfully utilised for the dissolution of silks, particularly when they are coupled with chloride or carboxylate anions,[72, 108, 109] and have been summarised in Table 2.3. Small methyl and ethyl alkyl groups on these cations result in better dissolution of silk, which is in line with the dissolution of other proteins in ILs. However, this trend for smaller side chains, does not extend to the protic $[C_1Im][Cl]$ IL, in which silk was reported to be insoluble.[110]

Given the success in lignocellulosic dissolution, $[C_4C_1Im][Cl]$ was naturally examined as a potential solvent for the dissolution of keratin based materials and Xie *et al*.,

investigated the relationship of solubility for wool keratin fibres and the structure of ionic liquids. $[C_4C_1Im]^+$ cation with different anions such as Cl^-, Br^-, $[BF_4]^-$ and $[PF_6]^-$ and [AMIM][Cl] were tested and compared.[104] Prior to solubility experiments, the commercially available wool fibres were washed with acetone/ethanol mixture and subsequently cleaned with distilled water and dried at 100°C for 12 h. ILs were prepared according to literature[111] and dried in vacuum oven at 100°C. The results from the study (Table 2.1) show that the chloride anion containing ILs had better solubility for wool keratin fibres than ILs incorporating bromide anion or non-coordinating anions including $[BF_4^-]$ and $[PF_6^-]$. Temperature also had an impact on the dissolution capacity of ionic liquids, as the increase in temperature from 100°C to 130°C increased the solubility of wool fibres from 4 to 11 wt% (Table 2.1).

Table 2. 1: Solubility of wool keratin fibres in ionic liquids

IL	T °C	Time hr	Solubility wt%
$[C_4C_1Im][Cl]$	100	10	4
	130	10	11
$[C_4C_1Im]Br]$	130	10	2
[AMIM][Cl]	130	10	8
$[C_4C_1Im][BF_4]$	130	24	Insoluble
$[C_4C_1Im][PF_6]$	130	24	Insoluble

Idris *et al*.,[41] investigated the solubility of barbs of cleaned, commercially-obtained turkey feathers in $[C_4C_1Im][Cl]$ and [AMIM][Cl] and thioglycolate anion ([choline][TG] and [bis-(2-ethylhexyl)ammonium][TG]) based ILs. The dissolution process was carried out at 130°C for 10h following similar conditions used by Xie *et al*.[104] The limiting solubility in each IL except [bis-(2-ethylhexyl)ammonium][thioglycolate]) was found to be around 50 wt% (lowest for [choline][TGA] was around 45%) of keratin in ILs. High solubility for chloride-based IL is presumably related to the ability of chloride ion to disrupt hydrogen bonding in the feather keratin, similar to the understanding that has emerged on the role of Cl^- in cellulose dissolution in ionic liquids.[66] There was no significant increase in solubility observed in [choline][TG] as compared to $[C_4C_1Im]Cl$ and [AMIM]Cl taking in view that thioglycolate solutions, have been used as reducing agents to break the disulfide linkages in the dissolution of keratin[112], which suggested that disulfide cleavage is not a significant factor in limiting

the solubility. No dissolution was observed for turkey feather in [bis-(2-ethylhexyl) ammonium][TG] and this was attributed to either the longer chain of the cation that makes it difficult to penetrate the keratin structure or to the viscosity that is much higher compared to [choline][TG], thus decreasing the solvating properties of the ionic liquid.

The same group reported that [DMEA][HCOO] was the best solvent when a series of distillable N,N-dimethylethanol-ammonium cation based protic ionic liquids (PILs) were investigated as potential distillable and recyclable solvents for turkey feathers.[43] Protic ionic liquids (PILs) are a subclass of ILs formed on proton transfer from a Brønsted acid to a Brønsted base.[113] Several PILs have been previously reported as distillable for recycling after use in the extraction of lignin and cellulose from biomass.[114, 115] PILs provide several advantages over aprotic ionic liquids (APILs), including that they can be relatively cheaper, easier to prepare and that in some cases they can be distilled from a reaction mixture. [DMEA][HCOO] IL was able to dissolve cleaned, dried barbs of turkey feathers at 100°C. Keratin was regenerated by using methanol and IL was recovered with 99% purity by further distillation.

In another study, Wang and Cao,[116] extracted keratin from chicken feathers using a hydrophobic [HOEMIM][NTf_2] IL. Prior to dissolution, chicken feathers were washed with water, filtered and dried and milled into small pieces. The water content of equilibrated IL was 4.89%(w/v) and was dried under vacuum to reduce the water content to 0.453%(w/v). The solubility of IL in water was 7.941%(w/v), that indicated that [HOEMIm][NTf_2] was slightly soluble in water and the maximum recovery of it could be up to 96%. The extraction and dissolution process were carried out in the presence of $NaHSO_3$ to reduce the disulfide bonds. The yield of keratin increased slowly with the increase in the mass ratio of feathers to ionic liquid from 1:40 to 1:45 and declined with a further increase in mass ratio. This effect was considered to be related to the ability of IL [HOEMIM][NTf_2] to produce electrostatic and hydrogen-bond interactions between itself and feather keratin.[117] It also indicated that part of peptide bond of feather keratin was disrupted when excessive amounts of ionic liquid were used. From these results, the mass ratio of feathers to ionic liquid 1:40 was adopted in the subsequent experiments. The extracted keratin was analysed by infrared spectrometry (IR) and it was shown that the regenerated keratin has similar IR bands to

the raw feather keratin that implied that the keratin structure was retained after dissolution in IL.

Chen, Vongsanga *et al*.,[118] studied the mechanism for the dissolution of wool and hair in $[C_4C_1Im]OAc$ IL. IL was dried and the water content was measured to be 0.8-2%. Merino wool (19.5 µm) and silk cocoons (bombyx mori) were commercially obtained Silk cocoons were degummed in water at 120°C for 1h and dried at 55 °C overnight. Prior to experimentation, virgin black hair was washed with detergent, rinsed in water, and dried at 55°C. The *in-situ* dissolution of a single wool fibre in the IL was observed using polarized optical microscopy. A wool fibre consists of an inner component called cortex that makes up 90% of the mass of a fibre, is comprised of cortical cells, which are bound together with each other and with the cuticle cells by a cell membrane complex. Cuticle is the outer layer that further comprises three layers, epicuticle, exocuticle and endocuticle as shown in the figure (Figure 2.6).

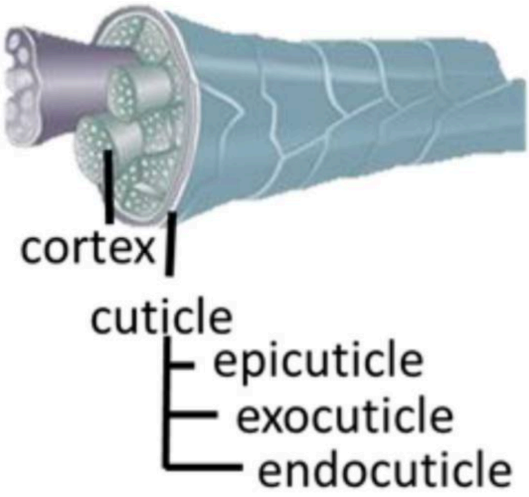


Figure 2. 6: Schematic illustration of the structure of a wool fibre

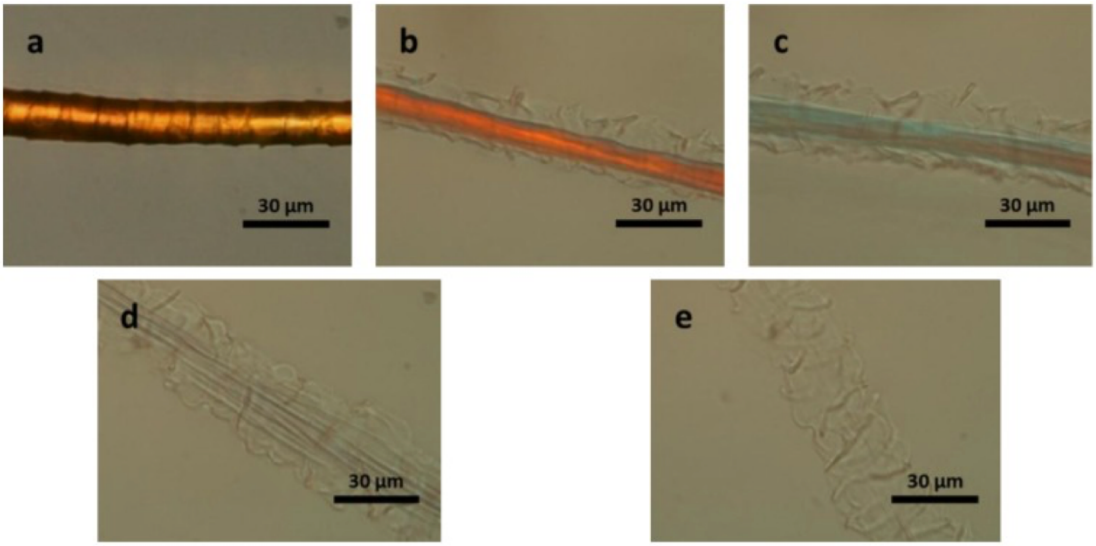


Figure 2. 7: POM images of a wool fibre in $[C_4C_1Im][CH_3CO_2]$ at 120 °C: (a) original wool fibre, for (b) 5 s, (c) 50 s, (d) 100 s, and (e) 180 s.

Polarizing optical microscope (POM) images showed the sequence of dissolution process (Figure 2.7) started with the swelling of cuticle, followed by the swelling of cortex. The cortex then becomes transparent suggesting the dissolution of the fibre occurs first in cortex, completely disappears (3 min) leaving the cuticle behind (1 h). (Figure 2.7 (e)). The extended time required to dissolve the cuticle was considered to be linked to the high content of disulfide bonds within the cuticle. In comparison, it was shown that silk dissolves progressively from the surface, from the outside to inside.

Table 2. 2: Dissolution of wool in ILs and time taken to dissolve the cuticle.[118]

IL	T (°C)	T (min)	Cuticle swollen	Cortex dissolved
$[C_4C_1Im][OAc]$	120	3	Yes	Yes
[Choline][TGA]	120	10	Yes	Yes
[Choline][Pn]	120	45	Yes	Yes
$[C_4C_1Im][Cl]$	120	90	Yes	Yes
[TMG][Pn]	100	390	Yes	No

Other studies on the dissolution of wool and hair keratin or silk in ILs have been grouped together below in the Table 2.3.

Table 2. 3: Ionic Liquids used as solvents or co-solvents for the protein-based materials.

Properties / Ionic Liquid	Protein Source	Solubility wt%	Dissolution	References
[DMEA][HCOO]	Turkey feathers	150 mg/g	100 °C, 7 h	[43]
[HOEMIM][NTf_2]	Chicken feathers	21.50	80 °C, 4 h	[116]
[C_4C_1Im][Cl]	Merino Wool		99 °C, 18 h	[65]
	Wool	4	100 °C, 10 h	[104]
	Wool	11	130 °C, 10 h	[104]
	Wool	15	130 °C, 8.91h	[119]
	Human hair	3	110 °C, 8 h	[69]
		6	120 °C , 8 h	[69]
		1	130 °C , 8 h	[69]
	Wool	Cortex soluble	120 °C	[118]
	Silk fibroin	13.2	100 °C	[5]
[C_4C_1Im][OAc]	Wool	Cortex soluble	120 °C	[118]
	Human hair	Soluble	120 °C	[118]
	Silk fibroin	Soluble	120 °C	[118]
[C_4C_1Im][Br]	Wool	2	130 °C, 10 h	[104]
	Human hair	10	130 °C, 8 h	[69]
	Silk fibroin	0.7	100 °C	[5]
[C_4C_1Im][BF_4]	Wool	Insoluble	130 °C, 24 h	[119]
	Human Hair	Insoluble	130 °C ,8 h	[69]
	Silk fibroin	Insoluble	100 °C	[5]
[C_4C_1Im][PF_6]	Wool	Insoluble	130 °C, 24 h	[119]
[C_4C_1Im][I]	Silk fibroin	0.2	100 °C	[5]
[AMIM][Cl]	Wool	8	130 °C, 10 h	[104]
	Wool	21	130 °C, 10.6 h	[119]
	Human Hair	19	130 °C	[118]
	Silk fibroin	1-15	90 °C, 1.5 h	[108]
[C_2C_1Im][Cl]	Silk fibroin	23.3	100 °C	[5]
[C_2C_1Im][BF_4]	Silk fibroin	Insoluble	100 °C	[5]
[C_2C_1Im][$AlCl_4$]	Silk fibroin	Insoluble	100 °C	[5]
[DMBIM][Cl]	Silk fibroin	8.3	100 °C	[5]
[Choline]TGA	Wool	Cortex soluble	120 °C	[118]
[Choline][Pn]	Wool	Cortex soluble	120 °C	[118]
[TMG][Pn]	Wool	Insoluble	100 °C	[118]

2.5 Proteins in ionic liquids

Tanford and Kirkwood in 1957,[120] explained the interaction of charges with solvent, and showed how the burial of charges upon protein folding was an important factor in stability.[120] The influence of their model, with further elaborations, can still be seen in much subsequent work on protein electrostatics and implicit solvent models. Another dominant factor is the hydrophobic effect that is responsible for the thermodynamic stability of folded conformation of many globular proteins. Hydrophobic amino acid side chains could stabilise proteins and play a role in determining their structure.[121] The opinion that hydrophobicity is the major contributor to protein stability is generally accepted, although the current studies of solvation recognise the importance of other types of solvent-protein interactions, including polar, charged, ionic and hydrogen bonding interactions.[122]

The wider variety of complex interactions that ILs can provide than standard organic solvents, means that this is an important area to understand with respect to proteins. In 2011, Huang *et al.*[123] stated that "an investigation of smaller peptide systems is desirable for understanding the effect of ILs on well-defined secondary structures in order to better understand and predict the composite effect felt by more complex proteins", and performed a study on the folding behaviour of helical (AKA_2, mini protein Trp-cage) and β-hairpin (Trpzip4) peptides in neat $[C_4C_1pyrr][NTf_2]$ IL using circular dichroism (CD) spectroscopy. Thermal transitions of small peptides were recorded and compared with aqueous data, in order to understand the effect of IL on the secondary structures. In contrast to aqueous data, the far-UV CD spectra of AKA_2 and Trp-cage in $[C_4C_1pyrr][NTf_2]$ IL indicated minimal helical structure at low temperatures, and the peptides were unable to fold. A heat-induced folding behaviour was observed, and the proteins were able to maintain a high degree of helical content at high temperatures (96°C). Moreover, AKA_2 in IL was shown to retain its helical structure when it was cooled after a melting experiment. This observation suggested that initial low temperature folding of the peptides is kinetically inhibited by IL reorientation dynamics, but upon cooling, the peptides are able to access the thermodynamically preferred conformation and related it to a strong preference for intra-peptide H-bonds over peptide-IL H-bonds. In contrast to helical peptides, β-hairpin (Trpzip4) structure was destabilised by $[C_4C_1py][NTf_2]$ IL at high temperatures

due to the possible interactions with the peptides, where the IL hydrophobic alkyl chains could interact with the four tryptophans of Trpzip4 and disrupted the Trp-Trp stacking interaction, which is essential to the thermodynamic stability of the Trpzip fold.

Some studies have indicated that ILs can be even used for refolding proteins.[123] The conformation of a globular protein is essential for its enzymatic function. A number of enzymes of widely diverging types have been observed to maintain or even improve their catalytic activity in IL solutions, along with the enhancement of structural stability. [124, 125] A great number of experiments have indicated that, given suitable anions (such as, BF_4^-, PF_6^-, and Tf_2N^-), ILs can stabilise the native structures of proteins. Iborra *et al*., found an over-stabilisation of *Candida Antarctica lipase* B (CAL-B) at 50°C in several IL aqueous solutions containing an imidazolium cation and $[BF_4]^-$, $[PF_6]^-$, or $[NTf_2]^-$ anion.[126] Using differential scanning calorimetry (DSC) and fluorescence spectroscopy experiments, the same research group also reported that the protein α-chymotrypsin has both enhanced melting temperature and heat capacity in $[C_2C_1Im][NTf_2]$ solution, compared to that in either water or conventional organic solvent (1-propanol).[127]

Seitkalieva *et al*.,[128] conducted a comparative study of l-alanine and l-valine-containing peptides using an NMR spectroscopy to investigate the effects of neat ILs, namely, $[C_4C_1Im][Cl]$, $[C_2C_1Im][Cl]$, $[C_6C_1Im][Cl]$, $[C_4C_1Im][BF_4]$, $[C_4C_1Im][PF_6]$, and $[C_4Pyr][BF_4]$. The study revealed a fascinating ability of ILs to distinguish slight variations in the structure of peptides, for example, di-valine and di-alanine, though structurally similar were able to be separated out in IL/organic solvent mixture where alanine showed greater affinity towards IL.[128] The study identified a suitable simple model to investigate peptide-ionic liquid interactions but couldn't explain clearly the molecular origins of the behaviour of the studied peptide systems.

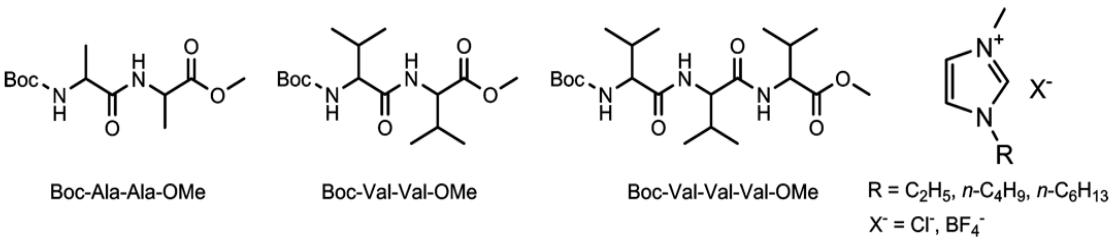


Figure 2.8: Structures of the peptides protected with BOC (tert-butoxycarbonyl) group and ILs used in the study[128]

Studies that have been previously reported mostly focussed on the effect of ILs on the stabilisation or destabilisation of the native structure of proteins, however, little attention is given to actually explain the interactions of ILs with the proteins. Many questions arise as to whether each IL ion plays their roles in stabilising or destabilising the protein structure. Also, the roles of various interactions such as electrostatic and hydrophobic interactions need to be explained and how the balance between various interactions at play affect or contribute to the destabilisation.[129] The IL cation and anion present different effects on structure and stability. The overall stability of the protein structure is dominantly dependent on the IL anion, but the denaturing effect of the anions can be partially overcompensated by pairing them with a suitable cation. Cation effects cannot be disregarded, as suggested also by other studies using imidazolium-based ILs.[130] More recent studies reveal that under the given conditions, the overall effect that a particular IL can exert on a particular solute is dependent on the balance of solvation energetics among all partners, that is, solute, IL cation and anion, and eventually the water molecules. The balance between directional, short-range hydrogen bonds, non-directional, long-range electrostatic interactions, and hydrophobic interactions were considered as key aspects to understand protein stabilisation by Shim *et al.*[131]

Various analytical techniques have been used for the structural studies. Solution NMR spectroscopy is a suitable method for structural studies on a molecular level, however serious problems are often encountered in IL/solute systems. These problems include low spectral resolution due to the high viscosity of ILs,[132] the necessity for using a deuterated NMR solvent, the uncertain effect of the kind and concentration of added deuterated solvent on the IL chemical shifts, the correct choice of the chemical shift reference and the large signal intensities differences between the ILs and solutes.

High-resolution magic angle spinning (HR-MAS) NMR spectroscopy has been demonstrated as a powerful method to solve these problems. Hence, line-broadening effects caused by dipolar coupling, chemical shift anisotropy, and magnetic susceptibility differences are reduced by a fast sample spinning along the magic angle. In spite of the unique advantages of the HR-MAS technology, only a few examples of such investigations of solutes dissolved in ILs have been presented in the literature. Regarding peptides or small biomolecules as solutes, little has been published so far.[133]

Richardt, Mrestani-Klaus *et al.,*[134] have reported a systematic study on the impact of ILs on the structure of peptides via HR-MAS NMR spectroscopy. The investigation was based on three distinct aqueous *N*,*N*-dialkyl-imidazolium-derived ILs ($[C_1C_1Im][(CH_3)_2PO_4]$, $[C_2C_1Im][(C_2H_5)_2PO_4]$ and $[C_2C_1Im][CF_3CO_2]$) and six tetrapeptides of the general structure Ala-*Xaa*-Pro-Phe (*Xaa*: Ala, Asp, Glu, Gly, Lys and Phe)(Figure 2.9).

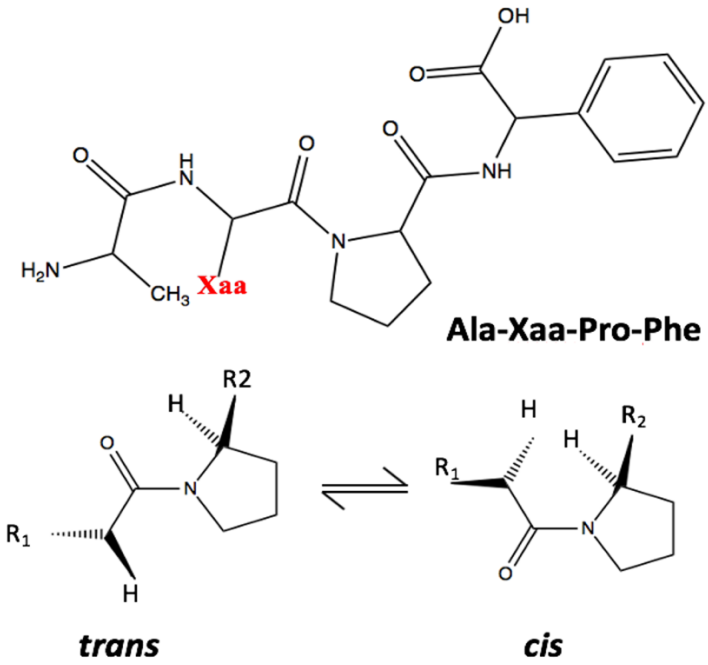


Figure 2. 9: Chemical structure of applied ILs and peptide Ala-Xaa-Pro-Phe, Xaa corresponds to Ala, Asp, Glu, Gly, Lys and Phe.[134]

Analyses of the peptides' chemical ^1H shift changes revealed the existence of direct interactions between the ILs and the peptide solutes. The interactions ranged from hydrogen bonds over hydrophobic contacts to ring stacking effects and were shown to be dependent on the individual peptide sequence and, most importantly, the nature of

the IL itself. Hydrogen bonds between the peptides' protons and IL's anion could be found for $[C_2C_1Im][CF_3CO_2]$ whereas for $[C_2C_1Im][(C_2H_5)PO_4]$ and $[C_1C_1Im][(CH_3)_2PO_4]$ ILs, a broader spectrum of interaction suggested that additional hydrophobic and ring stacking interactions were mainly mediated by the cation of the ILs.[134] This HR-MAS NMR study laid the foundation for the investigation of IL-peptide interaction patterns and their effects on inducing structural/conformational changes of peptide molecules.[134] This is important because its is expected that similar interactions may not only hold for flexible shorter peptides, but also for more elongated structured peptides and even for proteins. Thus, detailed insights into the interaction behaviour of ILs with peptidic structures could help to explain or even rationalise effects already found, for example, in protein folding or enzyme catalysis.

In order to get a good representation and idea of specific molecular interactions, computational approaches can be very helpful.[135] The field of molecular simulation and especially molecular dynamics has achieved significant contributions both to the field of simulating IL-solute systems and to the field of bio-molecular simulations.[136] MD simulation methods have proved to be a useful tool for the study of the molecular interactions in biochemical systems, including aqueous solutions and aqueous saline solutions of amino acids, peptides, proteins, lipids, and hydrophobic solutes, [61, 137-145] and have been playing an important role in the understanding of the fundamental chemistry of ILs and of their mixtures with water, enzymes, and hydrocarbons.[146]

2.6 Molecular dynamic simulation studies of ILs and proteins

In recent years, significant algorithmic progress has been made to enhance the simulation and analysis of molecular models. Among these developments are software packages,[147-149] the widespread utilisation of GPUs in existing software packages (*e.g.* AMBER[9], GROMACS[150], DL-POLY[151]) allowing access to more rapid and larger simulations, and automated procedures to derive force-field parameters to access a wider range of materials.

The structure and dynamics of a model protein (the serine protease cutinase from *Fusarium solani pisi*) was investigated through MD simulations, in two kinds of imidazolium-based ILs $[C_4C_1Im][PF_6]$ and $[C_4C_1Im][NO_3]$ by Micaelo and Soares. .

They calculated the RMSD values for enzyme structure in both ILs as a function of increasing hydration in the simulated systems, and, compared it with the reference X-ray structure of enzyme. The enzyme was shown to maintain its structure in $[C_4C_1Im][PF_6]$ IL and converged to an average native like structure at an optimal water content of 5-10% (w/w). $[C_4C_1Im][NO_3]$ IL destabilised the enzyme structure and a strong interaction of the $[NO_3]^-$ anion with the enzyme hydrogen donor groups in the α-helix and loop regions, was considered to be one of the main mechanisms of enzyme structure destabilization by this IL. Both the ILs differ with respect to anions and therefore, the correct choice of anion could help in the rational design of ILs for the biomolecules solvation or enzyme structure stabilisation.[152]

Solvation free energies of methylated nucleobases were calculated in pure and hydrated $[C_2C_1Im][OAc]$ IL, and pure water using classical MD simulations. It was observed that strong hydrogen bonds were formed between the acetate anions of the IL and the secondary amine hydrogen atoms of the nucleobases and the strength of these interactions decreases with the increase in water concentration as the both the anions and water compete to form hydrogen bonds with the amine hydrogen atoms. These hydrogen bonds contributed majorly to the Coulombic contribution to the solvation free energy. The solubility of nucleobases was greater in pure or hydrated IL than in pure water, evident from the higher values of solvation free energies. The stacking of cations above and below the nucleobase ring provided excess solvation free energy in the form of van der Waals interaction.[153]

Extraction of an aromatic amino acid tryptophan was studied through MD simulations in aqueous ILs $[C_4C_1Im][PF_6]$, $[C_8C_1Im][PF_6]$, and $[C_8C_1Im][BF_4]$ and the partition coefficient values were calculated.[56] The solvation energies were found to be higher in $[C_4C_1Im][PF_6]$ as compared to $[C_8C_1Im][PF_6]$ IL. The hydrophobic octyl groups limited diffusion of water into the IL phase, thereby reducing favourable electrostatic interactions of cationic Trp with water dipoles compared to $[C_4C_1Im][PF_6]$, which in turn reduced Trp extraction. The solvation of Trp was substantially improved in $[C_4C_1Im][BF_4]$ IL as compared to $[C_4C_1Im][PF_6]$. This is in line with the experimental findings[154] that due to the hydrophilic nature of BF_4^- anions, a higher concentration of water molecules in the ILs phase was observed, that intensified the electrostatic interactions and thereby enhanced the solvation of cationic Trp.

Smiatek J. *et al*.,[155] studied the stability of a small β-hairpin peptide under the influence of an aqueous $[C_2C_1Im][OAc]$ IL solution via MD simulation and observed a denaturation effect on the peptide structure which was validated by a significant broadening of the end-to-end distance between the centre-of-mass of the first and the last residue. The electrostatic Coulomb and the Lennard-Jones interaction energies between ions ($[C_2C_1Im]^+$ and OAc^-) and the folded and the unfolded peptide conformations were calculated. It was shown OAc^- anions interacted dominantly via electrostatic interactions with the peptide and Lennard-Jones interactions were less pronounced in contrast to those observed between $[C_2C_1Im]^+$ and the peptide. The most pronounced difference for un-folded conformation as compared to the folded conformation, was given by the strong increase of the net electrostatic energy between acetate anions and the peptide. The stabilisation study of two proteins (ubiquitin and zinc finger motif) was done in $[C_2C_1Im][CF_3SO_3]$ IL solution by Steinhauser *et al*.[156] They also attributed the protein stabilisation to the interactions between IL ions and protein, and the accumulation of IL cations and the binding of IL anions to specific spots (residues) on protein surface was reported. The stability of lipase *CAL*-B enzyme in eight ILs with varying cations and anions, was studied at high temperatures through MD simulations.[157] Two different factors were identified that caused the destabilisation of protein structure in these ILs. Anions variation dominated the overall influence on the instability of protein surface through Coulomb interactions. NO_3^- anions, with its compact structure and strong charge polarisation was found to be most effective in unravelling the helix structure. The cations with the long alkyl chains were reported to enhance the destabilisation effect through protein core-IL Lennard-Jones interactions.

Paluch *et al*.,[158] studied the solvation thermodynamics of amino acid analogues in water and Tf_2N^- anion-based imidazolium ILs with varying alkyl chain lengths. Solvation of hydrophobic amino acid analogues were dependant on the nonpolar domains on the cation, with a negligible influence from the anions. For the hydrophilic analogues, the solvation was dominated by the interactions between the amino acid analogues and the anion's polar domains, where the cation's influence was irrelevant. Tomé *et al.*[159] performed MD simulations to explain the interactions between five amino acids (glycine, alanine, valine, isoleucine, and glutamic acid) in aqueous solutions of $[C_4C_1Im][NTf_2]$ IL. Non-polar groups of hydrophobic amino acids (Ile and Val)

preferably interacted with non-polar regions of the cation while such interactions were not observed for the hydrophilic amino acids (Gly and Ala). Glutamic acid, that has a polar functional group showed interactions with both the ions of IL and water. This was considered to be due to the less hydrophobic character of the amino acid.

Amino acid-based ILs have also been investigated in regard to proteins and natural polymers, in view of their benign and biodegradable nature. In 2005 Fukumoto *et al.*[160] developed AAILs, where the $[C_2C_1Im]^+$ cation was paired with the carboxylate anion form of the 20 natural amino acids. Mantz *et al*.,[110] demonstrated that $[C_2C_1Im][Gly]$ was an effective solvent for silk dissolution. AAILs provide some of the highest dissolution capacities for silk, showing promise in the area for more biocompatible dissolution solvents.[13] Choline based AAILs have been reported to slightly decrease the thermal stability of collagen fibres when the dielectric and rheological behaviour was studied for serinate, threoninate, lysinate, and phenylalaninate anions.[161] The results from CD spectroscopy and FTIR indicated a change in the helical structure of collagen, shown by comparing it with spectra of native collagen. They postulated that this was due to the change in hydrogen bonding environment around collagen structure and anions were playing the major role as, with the same cation, phenylalaninate anion showed the most pronounced change in the CD spectrum relative to other anions. The deformation of collagen structure increased with the increase in concentration of choline based AAILs. Chevrot *et al.*,[162] investigated the thermal stability and real-time denaturation of a model mini-protein in pure water, $[C_2C_1Im][Ala]$ (5 mol% in water), $[C_2C_1Im][Met]$ (5 mol% in water), and $[C_2C_1Im][Trp]$ (5 mol% in water). They found that model mini-protein remained stable at 30–40 K higher temperatures in aqueous AAILs than in water.

From above discussion it can be concluded that the interactions between amino acids and ILs in aqueous solution are essentially dependent on the nature and properties of the amino acid side chain, namely its size, polarity, and charge, and can be understood in terms of a balance between direct or indirect (water-mediated) interactions established between amino acids and the ionic liquids.

Several applications have been reported that showed the effective use of ILs as mentioned above but apart from the beneficial effects found for the use of suitable,

biocompatible ILs in relation to proteins, little is known about the molecular basis of the underlying mechanism involving IL–solute interactions.[54] The interactions between ILs and proteins are complex and with the current literature it suggests that proteins stabilization/destabilization depends on primarily on the nature and selectivity between ILs and proteins. The nature of ILs ions play vital role in affecting the protein biological and conformational structure, either by establishing direct interactions with the protein molecules or indirectly by the modification of the microenvironment in, and around, the protein solute, due to their different physicochemical properties. Anions seem to be playing dominant role but the contributing effects from cations cannot be ignored.

A deeper understanding of the interactions of IL cations and anions with specific side chains and the peptide backbone, its dependence on the IL environment, and the effect of absence or presence of water is still missing. A detailed analysis of these interactions would support the rational selection and design of suitable ILs for particular applications concerning the various polymer dissolution processes, generation of a peptide sequence or its conformation, and the optimisation of purification and characterisation protocols. A comprehensive study is still required that could explain and clearly identify the type, strength and origin of these interactions that lead to the favourable effects observed, for example, with enzymes or other proteins and ILs as solvents or co-solvents.

3

Methodology

3.1 Molecular Dynamics Simulations and Force Field Selection

In a classical framework, the principal computational methods are molecular dynamics (MD) and Monte Carlo statistical mechanics (MC). These approaches provide a convenient path to predict the thermo-physical properties, structure and dynamics of ILs and ILs mixtures and also are proven to be helpful in understanding the molecular details of various systems including enzyme activation/deactivation processes and protein-ligand interactions.[163-165] The outcome of the simulations is primarily controlled by the expressions for the total energy, which are collectively referred to as the force field. The accuracy of the predictions relies on the accuracy of the force field that is used to describe the intra- and intermolecular, bonded and non-bonded interactions of the solvent molecules. Most classical force fields have a similar form consisting of two sets of terms, bonded and non-bonded as shown in Eq 3.1.

$$E_{pair} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\emptyset - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

Eq 3.1

Several force fields such as OPLS[166], CHARMM[167], AMBER[9], and the general AMBER force field[168] (GAFF) have been optimized to reproduce experimental data for small organic molecules such as ethane, propane, butane, isobutane, cyclohexane, methanol and ethanol etc. These force fields have been used as templates to generate force fields for ILs, because of the similarity in structure between ILs and some of the molecules that are used for the parametrisation of these forcefields (FFs).[169] However, most of these FFs have been developed for only a small subset of ILs. GAFF has been used in IL simulations,[170,164,171,172] and contains all of the needed parameters for

calculation of the bonded and Lennard-Jones terms. GAFF was designed to improve the applicability of AMBER FFs by introducing 35 basic atom types defined on the basis of the element type, hybridization, aromaticity, and chemical environment that improved its transferability across small organic molecules such as modified amino acids that otherwise are well known for their use to study proteins and nucleic acids. The use of GAFF facilitates the detailed study of organic molecules such as those present in ILs in conjunction with studies of biomolecules.[168]

In a classical force field, molecules are simply defined as a set of atoms that are held together by simple elastic (harmonic) forces and the forcefield replaces the true potential with a simplified model valid in the region being simulated. Bond stretching is often represented with a simple harmonic function that controls the length of covalent bonds shown in Eq 3.1, and a cosine function is used to evaluate the dihedral term. Coulombic and 12-6 Lennard-Jones terms are used to calculate the electrostatic and van der Waals non-bonded interactions, respectively, considering all pairs of atoms (i and j) in different molecules or in the same molecule. The overall sum of both the bonded and non-bonded interaction energies represent the total potential energy of the system. Parameters include the force constants K_r, K_θ, and V_n, the equilibrium structural parameters r_{eq} and θ_{eq}, the partial atomic charges q_i and q_j, the Lennard-Jones well depth, ε, and radii, A_{ij} and B_{ij}; where n is multiplicity; and λ is a phase angle for torsional angle parameters.

K.G Sprenger *et al.*[169] performed a comprehensive evaluation of GAFF in order to determine how well the standard molecular simulation force fields perform for a new class of IL solvents for which the parameters were not optimized. Results showed that, for a wide range of ionic liquids, GAFF, with no tuned parameters, could reproduce a variety of thermodynamic and transport properties with similar accuracy to the other IL-specific, force fields. For small organic molecules, such as all kinds of amino acids, and the IL systems investigated in this study, all the force constants and equilibrium and Lennard-Jones parameters therefore have been taken from GAFF[168] that has recently been shown to be a highly effective choice for IL solvents,[169] with ff99SB used for the peptides in these simulations. The ff99SB force field was chosen because it has been employed previously for the alanine based peptides and showed a good agreement between the experiments and the simulations.[173] The electrostatic point charges have to

be determined in addition to all available parameters in GAFF. ILs characteristic properties are a result of strong electrostatic interactions that drive molecular-scale fluctuations, accurately modelling the electrostatic landscape of each cation and anion is therefore of primary importance.

3.2 QM Calculations for Amino acid and Ionic Liquid Structures

The structures of all the amino acids and IL ions were built using Webmo[174], a web-based interface to computational chemistry packages. This has been utilised for the generation of the structures of all amino acid model compounds chosen for the study namely, lysine, arginine, aspartate, glutamate, alanine, histidine, phenylalanine, tryptophan, tyrosine and IL ions $[C_2C_1Im]^+$ cation and Cl^- and $[CH_3CO_2]^-$ anions. The AA structures were made keeping the zwitter-ionic backbone in each case. Lysine and arginine were constructed in their cationic forms (NH_3^+ group in the side chain), aspartate and glutamate in their negatively charged state (COO^- group in the side chain), aromatic amino acids and the alanine were prepared with neutral side chains. All the structures were optimised by performing individual quantum mechanics calculations involving geometry optimisation and energy calculations using the electronic structure program Gaussian 03/09.[175, 176] Structure optimisation is important to find stable minimum-energy structures in order to run smooth simulations. The B3LYP (6-31G*) theoretical method was used for geometry optimisation followed by the frequency calculation to confirm the minima and obtain the lowest energy structures. For energy calculations and derivation of ESP (electrostatic potentials); the Hartree–Fock (HF) (6-31G*) level of theory was employed in accordance with standard literature procedures for force field derivation.[169] The charges produced are fit to the electrostatic potential at points selected according to the Merz-Singh-Kollman (MK) scheme[177, 178] using the population keyword Pop=MK IOp(6/50=1) in Gaussian 09.[176] The restrained electrostatic potential (RESP) method[179] is then used to fit to the atomic charges for IL ions and AAs. The Gaussian 09 program was used on the UoN HPC to generate the optimised structures and RESP charges data files to be used by simulation software. The input files for geometry optimisation and resp charge derivation for lysine are shown as example in appendix A.3. The Gaussian output files for geometry optimisation for all AAs and IL ions are shown in appendices A.4.

3.3 Library building, parameter and topology files

3.3.1 Library building for amino acids and IL ions

In this work, molecular dynamics simulations are carried out using the AMBER 14 simulation package. The data files obtained from Gaussian calculations were further used by antechamber[180] to generate the respective library, topology and coordinate files for the AA and ionic liquid ions. Antechamber is the main program in the Antechamber module of the AMBER package that can not only perform molecular format transformations, but also reads in and writes out several commonly used molecular formats, including prep (AMBER residue topology file), pdb and mol2. It assigns atomic charges, atom types and bond types. By default, bond types are always assigned when antechamber reads in a molecule and the forcefield GAFF was selected using -at flag. The mol2 and pdb files for each molecule were also generated during this step. The files generated for lysine are shown in appendix A.5. The mol2 files for IL ions and AAs were modified where the derived RESP charges were scaled by multiplying each by a factor of 0.8 because it has been shown that non-integer charges better replicate the dynamic properties of ILs and that the reduced partial charges can mimic the average charge screening caused by the polarisation as well as the charge transfer effects, these methods are very similar to some of those employed by Maginn and co-workers.[164,181.] The structures with their scaled atomic charges are drawn using chemdraw and can be found in appendix A.7.

The library files were saved by loading the modified mol2 files in LEaP (graphical editor xleap). Once the atom typing and bond type assignment was done, the generated library file was then checked manually and compared with the gaff.dat file to see if the typing of atoms was correct and that all the atoms are in their required form *e.g*. sp3 carbon is typed correctly according to forcefields. Parmchk was used to produce the missing angle or bond parameters in the form of *.frcmod files. The example files for lysine structure are shown in appendix A.6.

3.3.2 Simulation boxes for amino acids and ILs simulations

Cubic boxes were made with side dimensions of 45 Å were constructed by using the PDB structures of the molecules using packmol.[8] Packmol is a software package used for building the initial configurations for running molecular dynamic simulations. An example input file for lysine and $[C_2C_1Im][Cl]$ IL system is shown in appendix A.6.1. 200 ion pairs of ILs ions were used for AA simulations and the number of AA molecules were determined by the concentration ratio (IL:AA) used in the system. For low concentration systems, one AA molecule was added to the simulation box (IL:AA, 200:1) and for high concentration systems, the concentration ratio (IL:AA) of 5:1(200:40) was used to match the experimental concentration ratio that were made available to the project. So, 40 molecules of AA and 200 ILs ion pairs were added for high concentration systems along with the counter IL ions to balance the charge in order to ensure neutrality. In addition to concentration, effect of temperature variation was also explored, so simulations were run for 300K and 400K temperatures, the list of simulations are shown in the respective chapters for all amino acids. Aqueous ionic liquids play a major role in the stability of enzymes and in other proteins. Water is also present naturally, even in pure ionic liquid applications, so it was important to see if these ambient levels of water can affect the organisation of IL ions around amino acids. Hence, simulations were run for pure ILs and in IL/water mixture. The number of water molecules to be added were determined from the grams of water absorbed from atmosphere using the Karl Fischer method. Experimental Karl Fischer measurements of the concentration of water in the selected ILs were carried out by Steven Maxwell-Hogg and appear in appendix A.1, Table A.1. These concentrations were used to determine the appropriate number of water molecules to be included in the IL/water simulations, resulting in 26 water molecules being added for $[C_2C_1Im][Cl]$ and two water molecules for $[C_2C_1Im][CH_3CO_2]$ in the "high concentration" systems, keeping other concentrations constant.

PDB structure for water molecule was generated in packmol and was used as such in the AA/IL-water mixtures. TIP4P-EW water model was applied for water and anion parameters that work well in combination with ff14SB force field that was employed alongside GAFF. TIP4P-EW is the newer version of the TIP4P water model and contains the improved parameters with the inclusion of both Coulomb and Lennard-

Jones long-range interactions[182] and has shown better bulk water properties than TIP3P which is still in common use for peptides and proteins. The amino acids were still defined by their respective libraries (created using RESP charges) and GAFF forcefield as non-standard residues and ff14SB force field was only used to derive parameters for water molecules and ions because it works in combination with the water model. To prevent the AAs from becoming standard residues, atom type and residue name section of ff14SB forcefield was manually modified/turned off. The PDB structure for each simulation box were saved as box.pdb files.

The PDB structure of the box along with the library file produced from LEaP and frcmod file from Parmchk2 were used to build parameter and coordinate files (*.prmtop and *.inpcrd) by loading these files into LEaP program of AMBER 14 package. The Amber simulation package outlines the set of procedures to be followed. The sets of commands were written as commands.in file and an example file is shown in appendix A.6.2 for lysine and [C_2C_1Im][Cl] system. The parameter and topology files were then saved for the box as box.prmtop and box.inpcrd fies.

3.4 Molecular Dynamics simulation run

All boxes, with different parameters of concentration and temperature, were simulated with periodic boundary conditions (PBCs). PBCs allow a more realistic model of a solution to be developed by preventing the outer solvent molecules from boiling off into space, and to allow a relatively small number of solvent molecules to reproduce the properties of the bulk, as compared to simple boxes of solute in a solvent, which result in some solvent molecules being within the bulk of the solvent, with a large number at the edge of the solvent and the surrounding vacuum. PBCs have particles simulated in an enclosed box that is then replicated in all three dimensions to give a periodic array, a two-dimensional representation of which is shown in Figure 3.1. Particles that leave one side of the box re-enter from the opposite side as their image. Hence, the total number of particles in the central box remains constant and molecules that would otherwise be exposed to vacuum are instead integrated with other solvent molecules to more closely mimic a bulk solution.

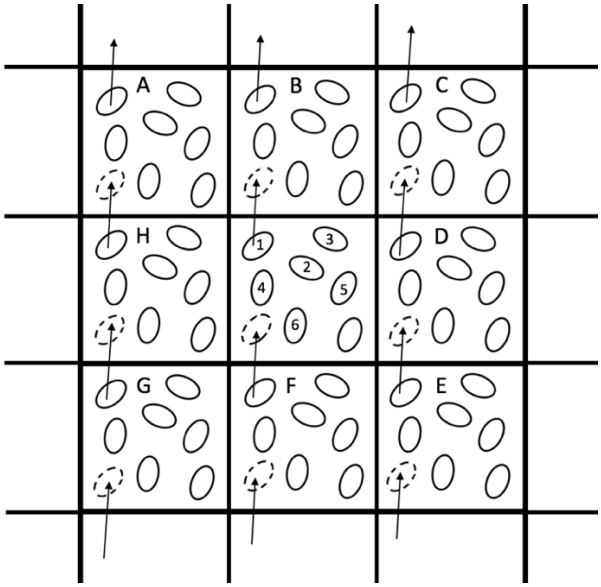


Figure 3.1: A two-dimensional array of boxes representing the movement of molecules, leaving the box and being replaced by their images from other boxes from example, molecule 1 replaced by molecule from box and the movement is replicated all boxes.[183]

3.4.1 Energy minimisation

The simulation procedure for each box includes two energy minimisation steps. Sander, the Amber module which carries out energy minimization, molecular dynamics, and NMR refinements, was used for performing the minimisation. Minimisation is done to lower the system's energies prior to running MD simulations to ensure the stability. The first minimisation stage was run for 20,000 cycles (maxcyc = 20000) using steepest descent for the first 2,500 cycles and conjugate gradient for the second 17,500 cycles (ncyc = 2500). Also, the minimisation was run at a constant volume (ntb = 1) with no pressure scaling. The weight of 5.0 kcal/mol−Å^2 for the positional restraints (ntr=1) was applied in the first minimisation stage on the solute using the "restraint mask" and "restraint_wt" keywords e.g. on lysine to allow the mixing of the solvent molecules so that the system is stably moved towards lower energy levels. The second minimisation stage was run without any restraints with the same parameters.

3.4.2 Initial MD simulation, heating up the system

The output file from the second minimisation stage was used as the starting point for this stage of the MD simulation (ntx = 1). The heating stage of MD was also run under constant volume (ntb = 1) NVT ensemble, using a cut-off distance of 10 Å without any restraints (ntr = 0). The temperature of the system was taken from 0-300 K or 0-400 K

using Langevin dynamics (ntt = 3) at a collision frequency of 1 ps-1 (gamma_ln = 1). The simulation was run for 50,000 steps (nstlim = 50,000) at 0.001 ps time step for a total of 50 ps. The random number generator (ig = -1) was used during this stage. The same initial coordinate files for the box of each system (box.prmtop and box.inpcrd) were used for running simulations at 300 K or 400K. Each simulation followed the same procedure separately as detailed in section 3.4.

3.4.3 Second MD simulation, equilibrating the system

The restart file obtained at the end of heating stage was used as the starting structure for this stage of the MD simulation. The equilibration was done at constant temperature of 300 K or 400K (depending on the simulation temperature) for 5 nanoseconds (25,00,000 steps of 0.002 picoseconds). The simulation was run under constant pressure (NPT ensemble) periodic boundaries (ntb = 2) using isotropic position scaling (ntp = 1). The output files were written every 100 steps. The remaining criteria were the same as those of heating stage. The random number generator was also applied during the equilibration phase. The properties such as temperature, density and energy (Total, Kinetic and Potential) were extracted from the respective output files and plotted using gnuplot to ensure that equilibration has been successfully achieved before starting the third MD simulation stage.

For the density equilibration graphs output file from second MD stage and for temperature and energy equilibration graphs, output files from both the heating and equilibration stages were used. There were no reference densities for our systems, so equilibration stage was continued till a constant density was achieved. The final density of the systems remains close to the density of water or ILs they are simulated in. For example, the density of pure $[C_2C_1Im][Cl]$ is reported to be 1.1238 g/cm^3 (7.6651 mol/dm^3) at 333 K[184]. The densities for our systems containing $[C_2C_1Im][Cl]$ IL fall in the range from 0.96-1.1 g/cm^3. The graphical representation of these properties for all the systems are shown in appendix A.8. The final box sizes for lysine and aspartate systems are included in appendix A.8.2.

3.4.3 Third MD simulation, production stage

For the production stage, the restart file from equilibrating stage was used to start the production stage of MD. The simulation was run for 100 ns (5,000,00,00 steps of 0.002 ps). In this stage, constant pressure periodic boundaries were applied with isotropic pressure scaling (ntb = 2, ntp = 1). Pressure was maintained using Berendsen barostat (default P=1 bar). The run was also performed at constant temperature (300 K or 400K) coupling algorithm with collision frequency of 5 ps^{-1} (gamma_ln = 5). The temperature was maintained using Langevin thermostat. The SHAKE algorithm was applied to restrain bonds involving hydrogen atoms (ntc = 2, ntf = 2). Particle Mesh Ewald (PME) summations were used to calculate long-range electrostatic interactions. Cut-off value of 10 Å was used for the production phase as well. The use of PME (by default) means we are calculating all of the 'infinite' electrostatics and so are not actually truncating the electrostatics. The VDW interactions are still needed though which means we cannot make the cut-off too small. For production calculations, the ideal range is between 8 and 10 angstroms and hence the value 10 was selected. The output files were written after every 1000 steps. During the heating phase, using constant pressure will result in an unstable system as the system is cold at the start. However, constant pressure must be used during the equilibration phase to allow for the density of the system, and solvent molecules in particular, to equilibrate. The three stages of MD simulation were applied using the CUDA version of commands for AMBER 14 that involves the use of GPUs rather than CPUs to run AMBER calculations.

The cpptraj[185] program in Amber was used for the purpose of processing coordinate trajectories and data files. Self-diffusion coefficients were measured with this tool as well (outlined in chapter 4). The analysis of the trajectories was performed with TRAVIS[186](outlined in Chapter 4). Travis is a free interactive program package for analysing and visualising Monte Carlo and molecular dynamics trajectories. It is a powerful tool that contains various important analyses techniques in one program and therefore provides a convenient way to analyse the simulations by directly providing the trajectories and coordinates information obtained from MD simulations. The volumetric data for the spatial distribution functions as well as the trajectories were visualised with VMD.[187]

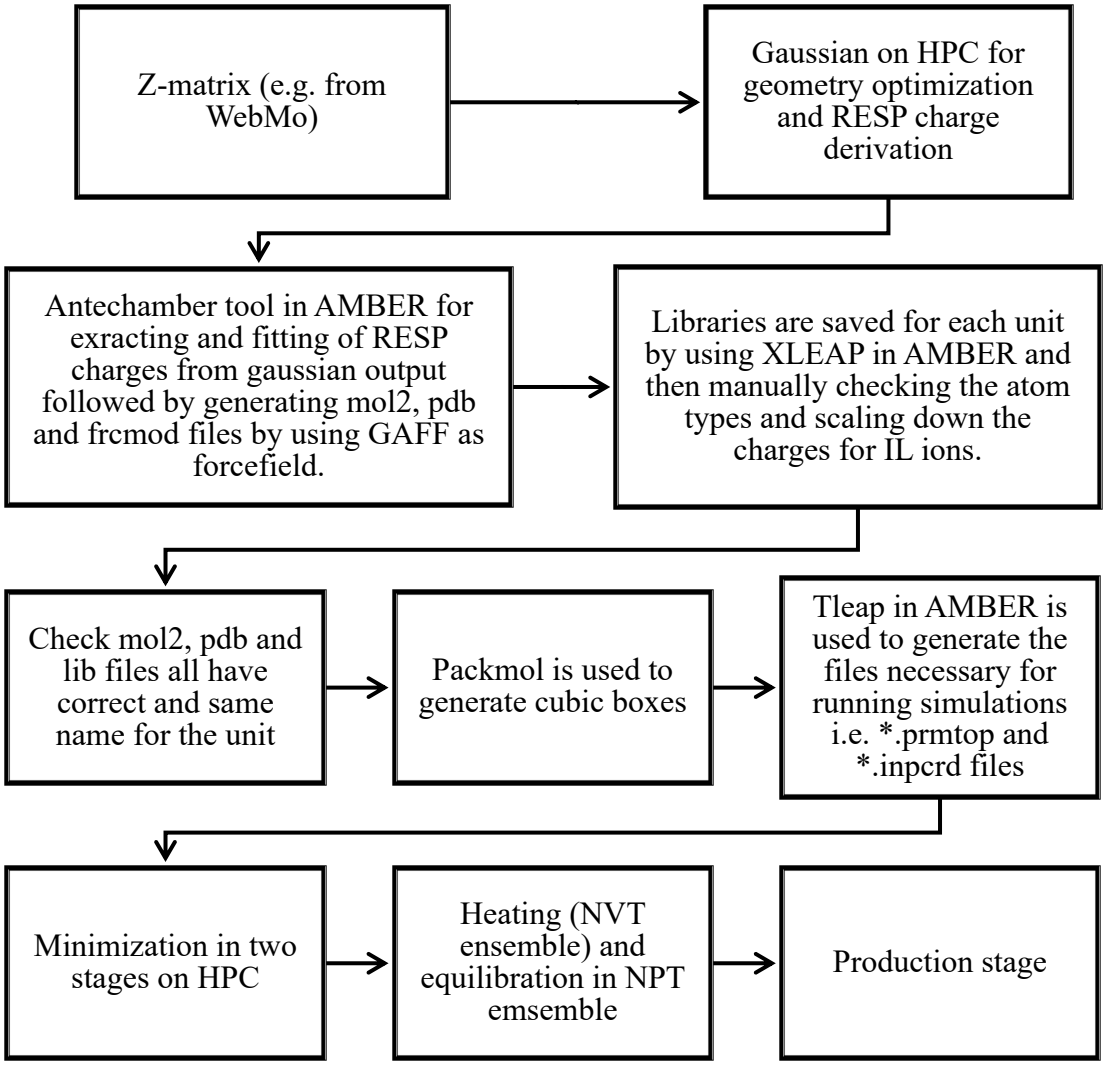


Figure 3. 2: Flow chart of the procedure followed for the amino acid/IL/IL-H_2O simulations

3.5 Peptide structures and MD simulation

Three peptides were examined: LNGKTLKG (ACE(Leu-Asn-Gly-Lys-Thr-Leu-Lys-Gly)NME), the 5-mer AAAAA (ACE-(Ala-Ala-Ala-Ala-Ala)NME) referred to as penta-alanine and the 10-mer AAAAAAAAAA (ACE-(Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala)NME) referred to as deca-alanine. Replicate simulations utilising ff99SB forcefield parameters were carried out to confirm similar folding dynamics in water to that observed in literature, and subsequent simulations were carried out using $[C_2C_1Im][CH_3CO_2]$ IL. All peptides were capped at the C-terminus with an acetyl group (ACE) and at the N-terminus with an N-methyl group (NME) to ensure the charge

neutrality of the chain. All three peptides were in their non-standard extended forms to observe, especially in case of peptide sequence from protein G, whether the peptide was able to fold back to its native conformation of a hairpin structure.

The peptide structures were created using xleap by loading the parameters from the ff99SB force field with standard amino parameters applied to the peptide structure. ff9SB has been employed previously for the alanine based peptides and was able to show a good agreement between the experiments and the simulations.[173] The pdb and the library file for peptide structure were saved and were later also used for peptide simulations in $[C_2C_1Im][CH_3CO_2]$ IL. If the overall charge for the peptide was non-neutral (as for the case of LNGKTLKG, where it is +2), an appropriate number of balancing chloride ions were added to neutralise the box, regardless of whether the simulations were run in water or in the $[C_2C_1Im][CH_3CO_2]$ IL. For water simulations, a truncated water octahedron box of 10 Å thickness was constructed around peptide solute using the solvateoct command in Leap. The parameters for water and ions were taken from the JC TIP3P water model. TIP3P water model is widely applied for protein structures till date, however more recent models are developed but each model has its own advantages.[188] TIP4P-ew was applied for small amino acid molecules but is computationally expensive, hence TIP3P water model was selected for the more complex peptide systems. The pdb files for the box were then saved as box.pdb and coordinate files were generated. The examples of the sets of commands for box creation and example files are shown in appendix A.26.1. The structure of peptide in water was visualised and ensured that solute would not be hindered along the axis of a cube. Small peptides are considered, and so a cube could have been generated using a fixed number of water molecules, but the truncated octahedron is considered to be a more efficient shape to be used for simulating proteins in water.

For all peptide simulations in ILs, cubic boxes were generated using packmol,[8] following the similar format of input files that are shown for lysine in appendix A.6.1. The box size selected for peptides simulations in $[C_2C_1Im][CH_3CO_2]$ IL was 50 x 50 x 50 and was chosen to be slightly bigger than the amino acid simulation boxes. In addition to parameters used for peptide simulations in water, GAFF[168] force field was used for IL ions for peptide simulations in $[C_2C_1Im][CH_3CO_2]$ IL. The scaled and optimised parameters already prepared for AA simulations in ILs were used here.

Peptide simulations in both water and IL were run at 300K. Similar parameters were used for minimisation, heating and equilibration stages of MD simulation as were used for AA simulations. The simulation procedure included two energy minimisation steps, followed by a 50 ps NVT (number, volume, temperature) ensemble with restraint weight on the peptide of 5 kCal but with hydrogen atoms not fixed. An NPT (number, pressure, temperature) ensemble heating process was subsequently carried out until a constant density was achieved. A temperature of 300 K was fixed using the Langevin thermostat. The density, temperature and energy profiles were constructed using Gnu plot and confirmed that they had equilibrated before running the final production run to ensure the stability of these simulations. The production run was carried out for 1000 ns in stages, each stage being 100 ns long. The trajectories were written every 2 ps. Finally, for the analysis, all trajectories were combined into a single trajectory of 1000 ns.

4

Analyses methods

Simulation analysis

4.1 Radial distribution functions

Radial distribution functions (RDFs) are calculated to obtain information about the local structure in a fluid[189, 190] and particularly in our study were calculated to analyse various atomic sites for all the possible interactions arising between the biomolecules and ionic liquid ions. The RDFs were calculated from the MD simulations for a series of amino acids with different polarities and have been analysed in a similar fashion to Tomes *et al*.[159]

RDF calculation allows a detailed analysis of the ordering of a system to be conducted. It shows how the density of surrounding matter varies as a function of the distance from a particular point, as compared to the uniform density, *i.e.* the density that would be obtained if all the observed particles would be uniformly distributed in the simulation cell. Thus, g(r) is a local measure of how close the observed distribution is to a uniform one. RDFs were calculated using TRAVIS.[186] Equation 4.1 shows the in-built function used by TRAVIS to compute RDFs, where r_i and r_j denote the position vectors of the i-th and the j-th particle. The function indicates if there is local structure in the fluid and a uniform distribution of particles would have g(r) = 1 for all r. Accumulation of particles, relative to the uniform distribution, shows up as g(r) > 1 and depletion as g(r) < 1.

$$g_{ab}(r) = \frac{V}{N_a . N_b} \sum_{i=1}^{Na} \sum_{j=i+1}^{Nb} \langle \delta(r - |\vec{r}_i\ (t) - \vec{r}_j\ (t)| \rangle_t \qquad \text{Eq 4.1}$$

4.2 Spatial distribution functions

Spatial distribution functions (SDFs) are analysed for each system and describe the spatial distribution of any types of atoms/molecules around a central atom/molecule. These provide an extension of radial distribution functions to the three dimensions of space.[191] SDFs give the three dimensional density distribution of ILs around the biomolecules and are directly related to the corresponding RDFs. The SDFs depict the probability of finding a particle at a certain position in space around a fixed reference system of other particles. These functions were calculated using TRAVIS. A reference molecule was selected, and for setting up a reference frame, three atoms, for example for amino acids the Nα-Cα-C(CO) were selected. Every instance of the reference molecule in every time step is transformed into this unique orientation, which enables averaging of probabilities in the space around the reference molecule.

4.3 Mean-square displacement and diffusion coefficients

The properties such as mean-square displacement (MSD), diffusion coefficient, electrical conductivity, and viscosity of ILs play a critical role in many applications of ILs such as in electrochemical devices,[192] lubricants for heat transfer,[193] and most importantly as industrial solvents.[194] An estimation of the transport properties of ions in ionic liquids is important for designing new ionic liquids for electrolytes and for selecting an IL for a specific application. The viscosity and conductivity of ionic liquids depend on the translational motion of the ions. The motion of the ions can be simulated by molecular dynamics calculations.[195, 196] To look at the aggregation and potential gelation behaviour of amino acid-ionic liquid solutions, self-diffusion coefficients were calculated for "high concentration" systems at 300K and 400K. Keaveney *et al.*,[197] calculated the self-diffusion coefficients through NMR studies and demonstrated that these can be valuable in understanding various solute-solvent and solvent-solvent interactions in ionic liquid systems. So, in order to observe in detail, the change of the mobility of the ions as a function of the temperature and water concentration, self-diffusion coefficients were calculated using Einstein's relation (Eq 4.2).

$$D = \frac{1}{6}\lim_{t\to\infty}\frac{d}{dt}\left\langle \left|\vec{r}_i(t) - \vec{r}_i(0)\right|^2 \right\rangle \qquad \text{Eq 4.2}$$

Mean-square displacements (MSDs) were calculated using diffusion command in cpptraj[185] software package in AmberTools 15/16 . The graphs of MSD *vs* time were then plotted using grace or excel and the self-diffusion coefficient were calculated by linear fitting of MSD graphs and extracting their slopes.[198] In newer AMBER 18 package the self-diffusion values are printed along with the slopes by running the diffusion command.

The GAFF force field that we have used for the derivation of parameters in our study has been shown to be able to accurately capture the trends of decreasing cation/anion self-diffusivity with increasing cation chain length of [RMIM][NTf_2] class of ILs, anion self-diffusivity tracking that of the cation and also the trend that the self-diffusivity of the [Tf_2N] anion was highest when paired with the fastest cation, [EMIM], and was lowest when paired with the slowest cation, [OMIM],[169] suggesting that cations and anions diffuse through the liquid in pairs or clusters, rather than on their own; as previously noted by others.[196] However, Sprenger *et al*.,[169] also observed that from a quantitative perspective, deviations in cation/anion self-diffusivity from experiment ranged from 49 to 73% and that, there is about a 25% increase in the mean absolute error for the GAFF cation and anion data sets as compared to the OPLS-AA/AMBER framework that has been specifically modified to reproduce densities and self-diffusion coefficients of ions of [RMIM][Tf_2N] ILs. Therefore, it was suggested that this method should be used as a way to obtain a qualitative trend of self-diffusivity.[169]

Following standard methodology, as also reported by Cadena and co-workers,[196] to assess the true diffusive behaviour of the ions, MSDs are displayed in log-log plots, together with their slopes $\beta(t)$ as a function of time,

$$\beta(t) = \frac{d \log MSD(t)}{d \log t} \qquad \text{Eq 4.3}$$

At very long times, a real diffusive regime is achieved where β approaches 1 and the mean-square displacement of ions grow linearly with time. At very short times, a value of β equal to two is expected, corresponding to a free, ideally ballistic, motion of the ions. This analysis is crucial in determining whether the system has achieved a true diffusive regime, or the dynamics of the system are still in the sub-diffusive regime. This is particularly important in the case of slow, viscous liquids such as RTILs.[198] They also noted that the calculated co-efficients were at least a factor lower than the self-

diffusivity measured from NMR experiments, which is consistent with subdiffusive motion. So, longer simulations and higher temperatures could give $\beta(t)$ values close to one. The "apparent" self-diffusivity measured from the first 500 ps of simulation were closer to the experimental NMR values. Also, the AMBER package's cpptraj tool give accurate values at the start of the simulation and the noise increases as the simulation time increase. Keeping in view, all the shortcomings of this analysis, the method is used for the qualitative estimation of the systems' dynamic behaviours, and for the explanation of the interactions occurring among the solute and solvent molecules.

4.4 Domain Analysis

One of the characteristic properties of ionic liquids is their nanostructure, that is, the ions can segregate into polar and non-polar domains. Such behaviour has been reported in the literature.[53, 196] To analyse this behaviour, a new tool has been developed in TRAVIS to explain this segregation, both qualitatively and quantitatively, and is referred as domain analysis. Brehm *et al.*,[52] showed the dissection of the fluorinated ILs into polar, fluorous, and alkyl subsets. The domains are built up with the aid of radical Voronoi tessellation. In this study, domain analysis was utilised to understand the segregation or solvation of amino acids in the particular ionic liquids by extracting the domain count values, where a value close to 1 was considered to be indicative of amino acid aggregation and a value close to 40, that is the maximum number of amino acids in the system, would indicate complete solvation.

4.5 Secondary Structure analysis of peptides:

Secondary structure analysis for the whole 1000 ns trajectory containing IL-solvated peptides was performed using the "secstruct" command in cpptraj[185]. It calculates the secondary structure following the DSSP ("Dictionary of Secondary Structure of Proteins") method by Kabsch and Sander[199]. The secondary structure (sheet and helix) assignment to the given protein/peptide structure is done based solely on backbone-backbone hydrogen bonds according to this method. The output data having secondary structure assignment per residue is then plotted as a 2-D representation of secondary structure *vs* time or secondary structure *vs* frame number.

4.6 Dihedral analysis - Ramachandran plots

The two torsion angles in the polypeptide chain, N-Cα (called Phi, φ) and Cα-C ((CO) called Psi, ψ) describe the rotations of the polypeptide chain around the backbone atoms and are considered to be most important structural parameters that control protein folding. The Ramachandran plot[200] provides an easy way to view the distribution of dihedral angles in a protein structure and provides an overview of energetically allowed and excluded or forbidden regions that show which rotations of the polypeptide are not allowed due to steric hindrance. The Ramachandran plot of a particular protein may also serve as an important indicator of the quality of its three-dimensional structures. So, if Ramachandran angles for a particular protein can be predicted, it would help in predicting its fold. Ramachandran plots were generated by performing the dihedral analysis with the cpptraj[185] tool in Amber and then plotting the subsequent phi and psi angles using gnuplot. The plots were then analysed to explain the different permissible regions in the plots for respective amino acids in both water and $[C_2C_1Im][CH_3CO_2]$ ionic liquid. A typical Ramachandran plot for alanine and alanine like proteins is shown in the figure 4.1.

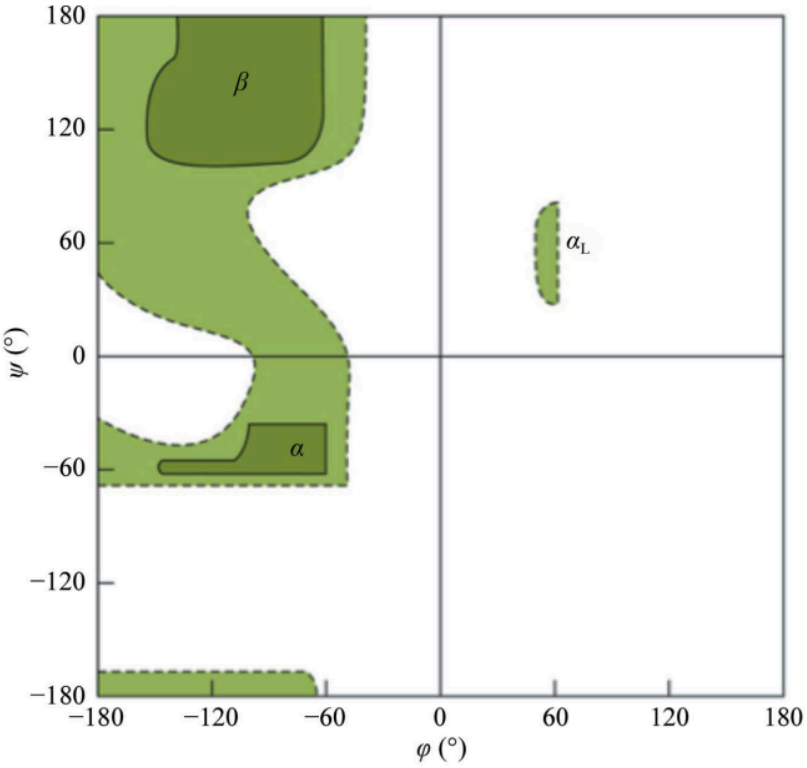


Figure 4. 1: Ramachandran plot for Ala and Ala-like residues, showing the allowed regions (continuous lines) and the partially allowed regions (dotted lines).[201]

The dark green and light green regions represent the favoured and allowed regions respectively. The dark green regions correspond to conformations where there are no steric clashes in the model peptide. These favoured regions include the dihedral angles typical of the alpha-helical and beta-sheet conformations. The white areas correspond to conformations where atoms in the protein come closer than the sum of their van der Waals radii. These regions are sterically forbidden for all amino acids with side chains (the exception is glycine which has no side chain).[202]

5

Interactions and organisation of positively charged amino acids in imidazolium ionic liquids

5.1 Introduction

Lysine and arginine are both positively-charged, polar and basic amino acids, that differ in respect to the functional group attached at the end of alkyl chain. These amino acids were selected for study on the basis that they are highly concentrated in the keratin/wool structure.[203,204] These amino acids are known to provide stability to the protein structure, are mostly exposed to the protein surfaces and have the ability to effectively interact by forming ionic interactions and hydrogen bonds in the proteins as well as by interacting with water molecules. Geometric structure of both these amino acid is quite different and the guanidinium group in arginine allows interactions in three possible directions through its three asymmetrical nitrogen atoms whereas lysine is able to interact through one functional group.[205, 206] Both amino acids play an important role in biological environments and their have peptides been identified as the promising drug delivery carriers owing to the readiness with which they can enter the cells.[207]

Lysine and arginine both were made with the zwitter-ionic backbone, as mentioned in Chapter 3. Depending on the pH of the solution, amino acids can be present in different equilibrium forms due to the ionization/protonation of their characteristic functional groups, −COOH and $-NH_2$. The majority of the biomolecules possesses neutral side chains and, therefore, have no net charge, but Arg, Lys, Asp, and Glu in their zwitterionic forms comprise acidic/basic groups in their side chains and are ionized/protonated at physiological pH. As a result, they acquire positive/negative net charges, depending on their acidic/basic character.

Protein surface charge has been found to be significant in enzyme stabilisation and activity in ILs. The interactions especially involving enzymes and ILs have been shown to be influenced by the nature of the amino acids that makeup the enzyme surface. The studies of cellulases in ILs have demonstrated that if the surface charge is positive, symmetrical, small, negatively-charged ions (Cl^-) are attracted to the surface and disrupt the secondary structure substantially, with the coordinating strength of the anion to the surface depending on its H-bond acceptor ability.[13] This disruption does not automatically lead to a major disturbance of the active centre, unless inhibition by the IL ions occurs.

The hydrophilic anions seem to prefer bonding with the positively charged surface as they try to establish hydrogen bonds to lysine, arginine, or histidine (if positively charged) and show strong Coulomb interaction with these amino acids due to their high charge density.[125] Experimental data (experiments performed by Steven Maxwell) that was made available to the project also revealed the difference in the solubility of these amino acids in various ILs as also shown in Table 5.4.2. Both lysine and arginine showed 10% (w/w) solubility in $[C_2C_1Im][Cl]$ than compared to $[C_2C_1Im][CH_3CO_2]$ where they have 90% and 70% solubility respectively. These preferential solubility values make it intriguing to further explore these differences using MD simulations.

In view of the importance of these nitrogen-containing amino acids, lysine and arginine were chosen to provide specific insight on how different types of positively-charged amino acid sidechains might display variations in the strength and pattern of interactions with different ionic liquids. The ILs $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ were chosen based on their known application in the dissolution of biomolecules,[5,208] which is facilitated by the anionic part being a strong hydrogen-bond acceptor. Protonated lysine was simulated in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ ionic liquids and also with the mixture of IL/H_2O to observe how the interactions strengths are affected as water is added in small amounts.

5.2 Outline of simulation conditions

Three different experimentally-relevant conditions were explored, namely concentration, temperature and whether the mixture was either pure or ambient water-containing. (Table 5.1). Temperature can have an impact on both structuring and interactions, and was explored for protonated lysine at 300 and 400 K, representing room-temperature and heated solutions, respectively. Protonated lysine was simulated in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ ILs at both 300K and 400K. The IL to amino acid concentration ratios of IL:AA 200:1 and IL:AA 5:1 were chosen to match low concentration systems and experimental data made available to the project. The experimental melting points (m.p.) values for $[C_2C_1Im][Cl]$ IL is ~78°C and $[C_2C_1Im][CH_3CO_2]$ IL is around 30°C. The simulation temperatures selected here for studying IL solutions are to observe the trends over two temperature values. The addition of solute molecules will cause depression in m.p. of AA/IL solutions. The m.p. values calculated computationally are highly dependent on the selected forcefield so a difference of 20-40K between the experimental m.p. and the simulation temperature are generally considered to be well within range *e.g.* for $[C_4C_1Im][Cl]$, the melting point has been computed to be 320 K,[209] which was considered to be well within the experimental value of 337–339 K.

Table 5.1 highlights two classes of mixture with respect to concentration as also mentioned in Chapter 3: “low”, which contains charge-scaled 201 ion pairs of IL (either $[C_2C_1Im][Cl]$ or $[C_2C_1Im][CH_3CO_2]$) and one charge-scaled molecule of protonated amino acid, and represents a dilute solution with which to examine the specific impact of ionic liquid on the amino acid ion; the other, more concentrated mixture contains 200 ion pairs of scaled $[C_2C_1Im][Cl]$ and 40 charge-scaled molecules of protonated amino acid in order to replicate the experimental concentrations (termed “High”). The overall charge of the more concentrated simulation box is balanced by adding counter ions of the IL so that, for example for the high concentration system, if the total positive charge is (200 x 0.8 = 160) and negative charge is ((200 x 0.8) + (40 x -0.8)) = -200, so 40 counter ions of reduced-charge IL anion are added to balance the charge. The simulation systems that were run for each of the protonated lysine and arginine amino acids are listed in the table 5.1.

Table 5. 1: Lysine and Arginine simulation conditions

IL	Amino acid	Number of IL ions and AA molecules	Number of water molecules added	Concentration ratio IL:AA and Temperature			
$[C_2C_1Im][Cl]$	Lysine(+)	For (IL:AA)= (200:1)	n/a	200:1 (300K)	200:1 (400K)	5:1 (300K)	5:1 (400K)
	Arginine (+)	200 IL ion pairs one AA molecule	n/a	200:1 (300K)		5:1 (300K)	
$[C_2C_1Im][CH_3CO_2]$	Lysine(+)	For (IL:AA)=(5:1)	n/a	200:1 (300K)	200:1 (400K)	5:1 (300K)	5:1 (400K)
$[C_2C_1Im][Cl]$-WAT	Lysine(+)	200 IL ion pairs	26			5:1 (300K)	5:1 (400K)
$[C_2C_1Im][CH_3CO_2]$ -WAT	Lysine(+)	40 AA molecules	2			5:1 (300K)	5:1 (400K)

5.2.1 QM Structures for Protonated Amino acids and Ionic Liquid

Protonated Lysine and Arginine, and IL ion structures were prepared as outlined in Chapter 3, using the WebMo web-based interface. The MD simulations were run for all boxes according to the procedure outlined in chapter 3. The optimised lowest energy structures obtained from QM calculations are shown in figure 5.1. The atoms are labelled using numbering generated by antechamber.

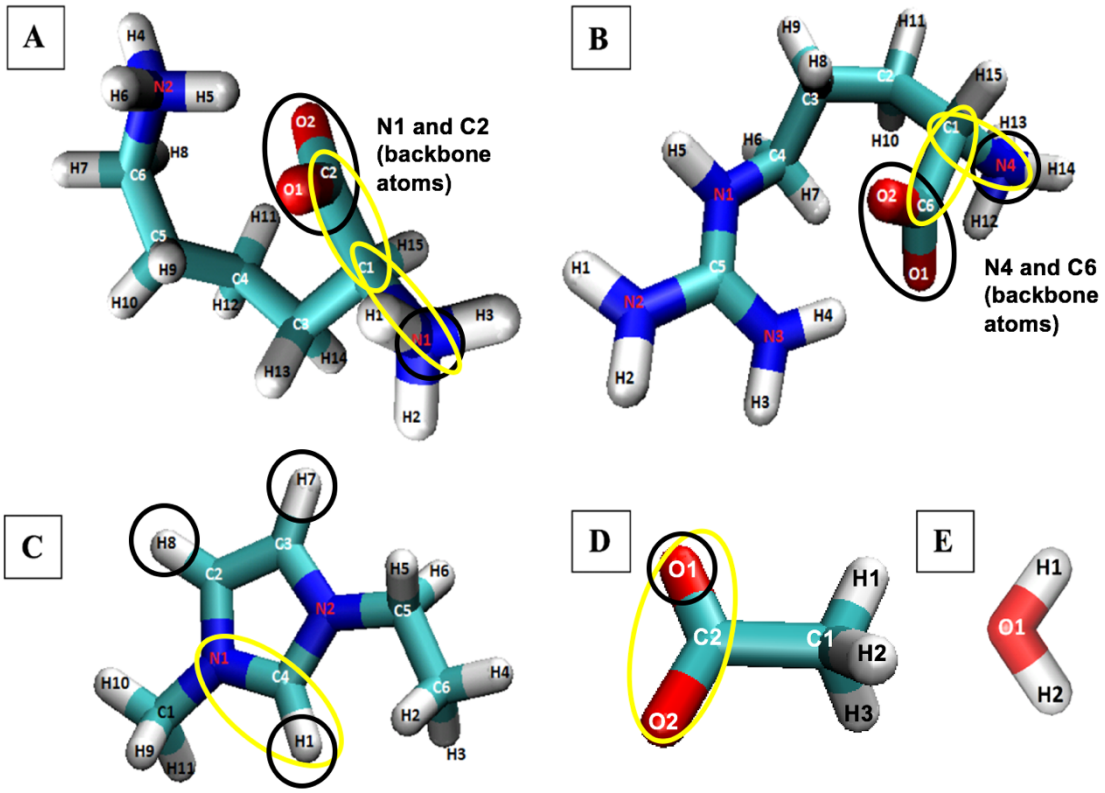


Figure 5. 1: Optimised structures of (A) Lysine(+), (B) Arginine(+), (C) [C_2C_1Im] cation, (D) [CH_3CO_2] anion and (E) water molecule with atom numbering generated by antechamber.[180] Atoms encircled yellow are in each molecule show the atoms that are used as reference to create respective SDFs around these molecules. Backbone atoms are labelled in lysine and arginine cations. Atoms encircled black are the ones that are used to discuss results in RDFs analysis.

5.3 Results and discussion

To understand molecular interactions at the microscopic level, molecular dynamics simulations provide a number of mechanisms to interrogate these interactions. By far the most visual and easy to comprehend of these analysis is that of probability functions – the probability that either one molecule or atom is a certain distance and/or orientation at any particular time from another. Because these functions are a measure of probability across a simulation, this data reflects the relative strengths of interactions and can provide rational chemical insight.

The simplest of these, radial distribution functions (RDFs), can describe the relative distribution probability of a particle or a molecule as a function of distance from the centre of another particle. Centre of mass radial distribution functions are useful for whole molecules and can provide visualisations to show the interactions among cations,

anions, amino acids and water molecules. Overall, mass-centred molecule interactions, along with the site-site radial distribution functions, can be analysed to explain the origin and strength of the interactions established between different atomic sites. Here we have used them specifically to understand the interactions of IL (cation and anion), AA, and water and they can also be used to visualise interactions *within* flexible cations, anions and amino acid molecules, where relevant.

Spatial distribution functions, which provide a three-dimensional representation of radial distribution functions, give a better understanding of the organisation within a solution – a highly important feature of ionic-liquid solutions. This organisation is also reflected in self-diffusion coefficients, which are measured from the mean square displacement (MSD) of molecules throughout the solution. In this chapter, simulations were carried out and analysed using the above approaches for non-aqueous ionic liquids and are compared to those with water to understand the influence of addition of water to proteinaceous-IL mixtures.

5.3.1 Ionic liquid microstructure and the impact of solutes

ILs are well recognised for possessing a nano-structuring at the microscopic scale.[21] Coulombic interactions among ions tend to dominate the structure of ionic liquids combined with the rotational and/or torsional fluxionality of ions, which helps in achieving the efficient packing of ions. These two interactions give rise to the underlying liquid structure of ILs, comprising of alternating ions shells in which each cation is surrounded by anions and *vice versa* and this structural relation extends out to a second and even third shell.[210] In addition, hydrogen bonding interactions also play a vital role in explaining the structure of ionic liquids and are dependent on the specific features and nature of ions forming the ionic liquid.[211]In the case of ionic liquids possessing aromatic cations, cation-cation interactions are also possible arising from π-π and cation-π interactions.

To confirm the structuring in the two ionic liquids of interest, $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$, and the impact of solute (lysine and arginine) additions, simulations using the GAFF forcefield and containing 200 IL ion pairs and either one (low concentration) or forty (high concentration) amino acids were carried out.

Specifically, lysine molecules having a positively charged side chain with zwitterionic backbone or arginine molecules having a positively charged side chain with zwitterionic backbone were chosen as representative solutes to investigate the organisation of ionic liquids with these positively charged amino acids.

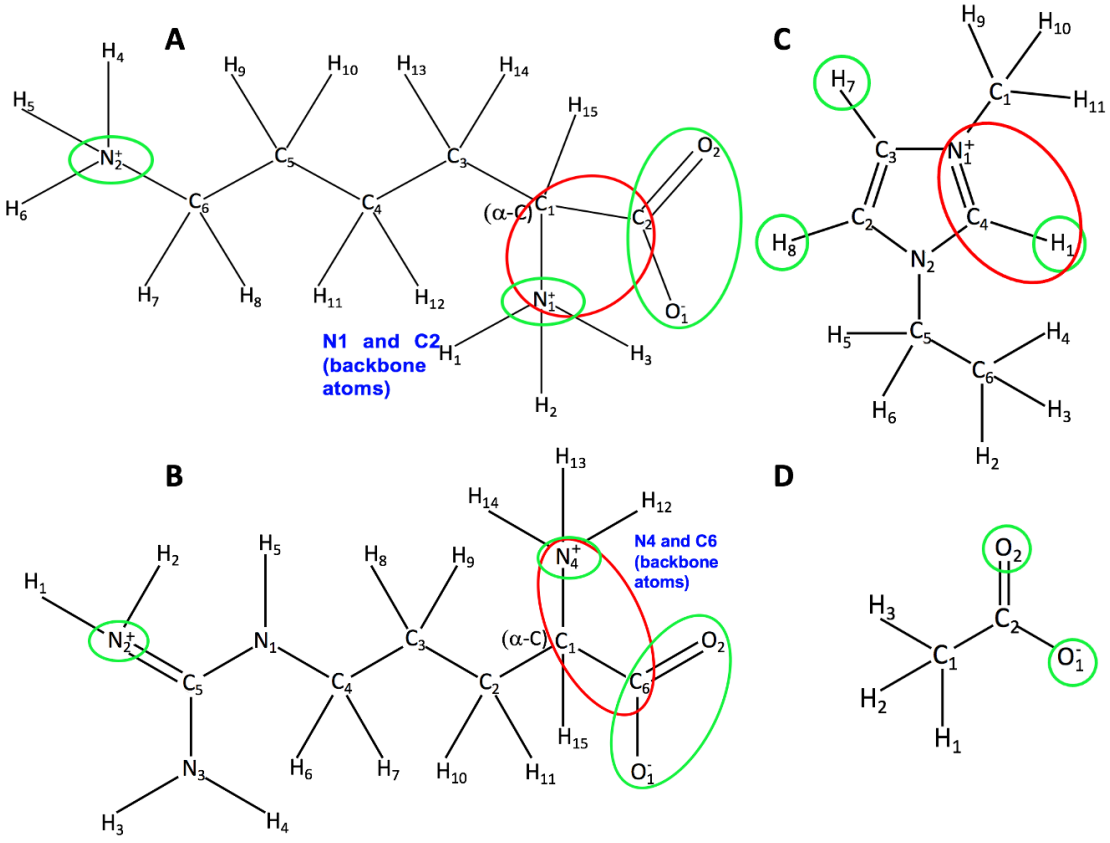


Figure 5.3. 1: Chem draw structures of (A) positively charged lysine molecule with zwitter-ion backbone, (B) positively charged arginine molecule with zwitter-ion backbone, (C) $[C_2C_1Im]^+$ cation and (D) $[CH_3CO_2]^-$ anion. Atoms encircled red in the molecules are those used as reference to create respective SDFs around the molecule. Backbone atoms are labelled in lysine and arginine. Atoms encircled green are the ones that are used to discuss results in RDFs analysis.

Figure 5.3.2 illustrates the changes in the IL microstructure in $[C_2C_1Im][Cl]$ when different amino acids are added, alongside the effect of increasing the concentration of these amino acids. The SDFs were produced using TRAVIS[186] by taking the N1, C4, H1 atoms of $[C_2C_1Im]^+$ cation and, in case of amino acids lysine and arginine, their backbone and alpha atoms, as the centre of the SDF as depicted in Figure 5.3.1. Visualisation was achieved with VMD.[187] Blue has been used to represent the cation organisation, with red for chloride anions.

For all systems, there are specific features relating to the general interactions of the solvent anions and cations that remain consistent in MD simulations. Cation-cation

interactions (blue) are found above and below the plane of the imidazole ring, consistent with π-π and cation-π interactions. Anion-imidazole interactions are strongest around the ring annulus, primarily around the H1, H7 and H8 atoms of the ring. This is strongest for the H1 proton owing to its more acidic nature, as has been reported previously for similar pure ionic liquid systems through theoretical studies.[189, 210, 212, 213]

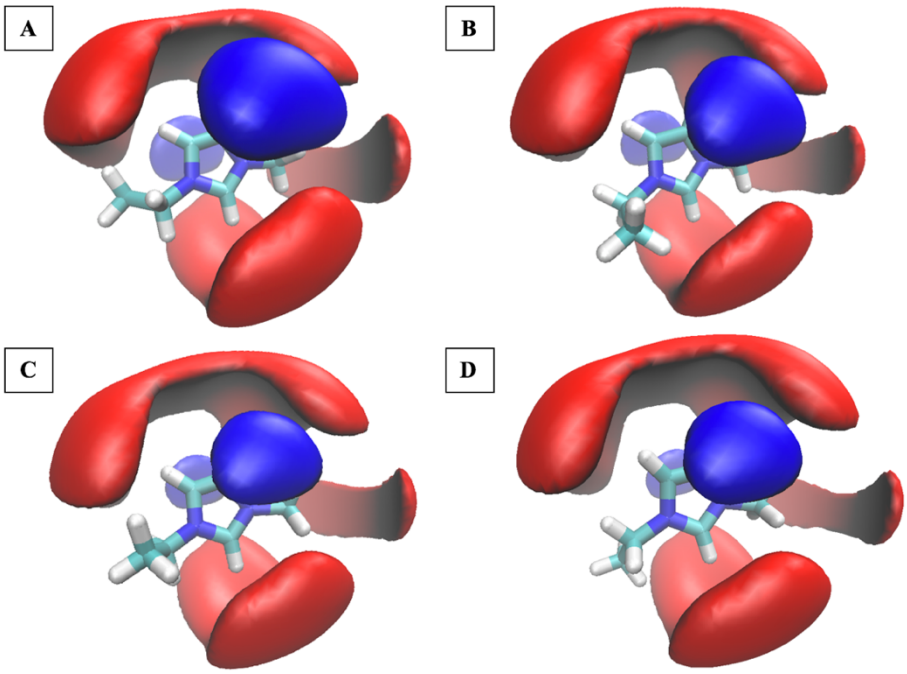


Figure 5.3. 2: Spatial distribution functions for cation-cation and cation-anion interactions around central $[C_2C_1Im]^+$ cation for (A,C) lysine low concentration *vs* high concentration system respectively and (B,D) arginine low concentration *vs* high concentration system respectively, in $[C_2C_1Im][Cl]$ IL at 300K. Central $[C_2C_1Im]$ structures are presented in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}. The central molecule that is used to show the distribution around is generated by TRAVIS[186] while constructing the SDF.

Anion distribution is not symmetric about the central axis of the imidazolium ring owing to the anisotropy that exists between the regions associated to the H7 and H8 carbons because of the presence of the ethyl side chain. The access of anions over the H7 position can be hindered by the free rotation of the ethyl group and the density is slightly tilted toward the H8 hydrogen of the cation near methyl group. The anion density associated with the H1 position is also offset more towards the methyl group of the imidazolium cation for the same reason.[210] On increasing the concentration of solute, the IL-IL begin to differ for two basic amino acids, there is a more significant reduction in cation-anion interaction than cation-cation interactions for lysine, whereas both IL-IL interactions slightly increase for arginine, suggesting that anions bond more tightly and are involved in alternative interactions with the solute. The general patterns

of interaction seen are consistent with those already reported extensively in the literature.[214]

To probe the extent of the change in ordering in a more quantitative fashion, RDF plots were used. The atom numberings shown in Figure 5.1 or 5.3.1 are used for the radial distribution functions (RDFs) under discussion. In Figure 5.3.3, the impact at low concentrations of the solute on IL-IL interactions is not different for either solute or can be regarded as having little to no impact. At higher concentrations, slight differences occur, with lower interactions, consistent with the ions interacting with other species (in this case amino acids) and thus having less average interaction with one another, although the effect is minor. The value of g(r) show that cation-cation and cation-anion both decrease for lysine at high concentration whereas for arginine both show an increase. Lysine has g(r) value for the first peak around 2.3 for cation-anion interactions that decreases slightly to 2.1 at a higher concentration with an even slight decrease of g(r) value for cation-cation interactions from 1.27 to 1.25 is observed. However, for arginine solute, there is a slight increase in IL-IL interactions with the g(r) value going from 1.3 to 1.4 for cation-cation and from 1.3 to 1.35 for cation-anion interactions. The RDFs shown in Appendix A.9.1 suggest that the highest peak for the chloride anion around cation atomic sites occurs around H1 atom of [C_2C_1Im] cation. This shows that the first solvation shell of chloride ion is present mostly around aromatic hydrogens of the imidazolium ring. The strength of interaction is highest for the acidic proton H1 that is capable of forming hydrogen bonds with Cl^- ion and also with the carboxyl oxygens of all amino acids as shown in their respective RDFs. The overall mass centred RDFs in figure 5.3.3 show that the electrostatic interaction between cation and anion is dominant and is not much affected by the nature of amino acid present in the system.

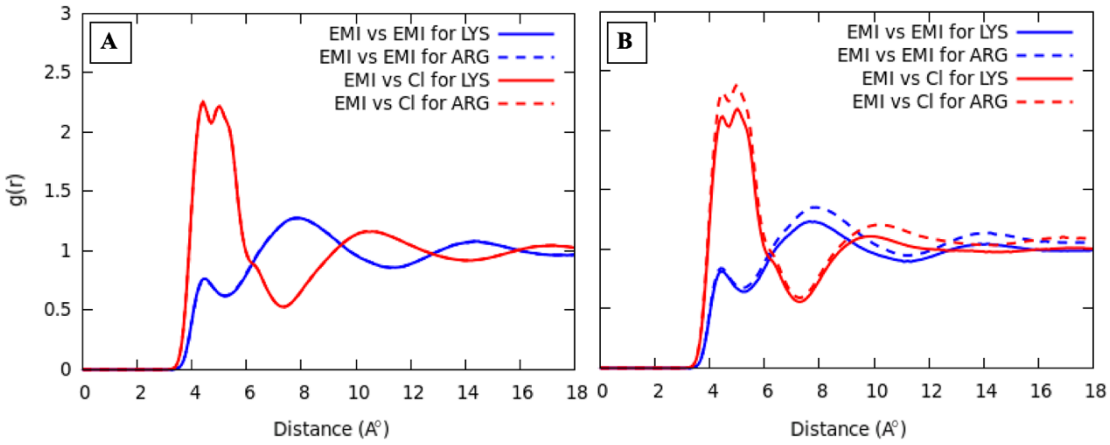


Figure 5.3.3. Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$ for overall cation-anion and cation-cation interactions for lysine (solid lines) and arginine (dashed lines) for (A) low concentration and (B) high concentration systems at 300K determined by MD simulation

For $[C_2C_1Im][CH_3CO_2]$ the distributions show a less comprehensive interaction with the acetate anion (mass-centred), and a much lower interaction of the cation at the same isovalue (Figure 5.3.4). Bowron *et al.*[210] examined the structure and dynamics of $[C_2C_1Im][OAc]$ and showed a similar distribution of anions about the central cation.

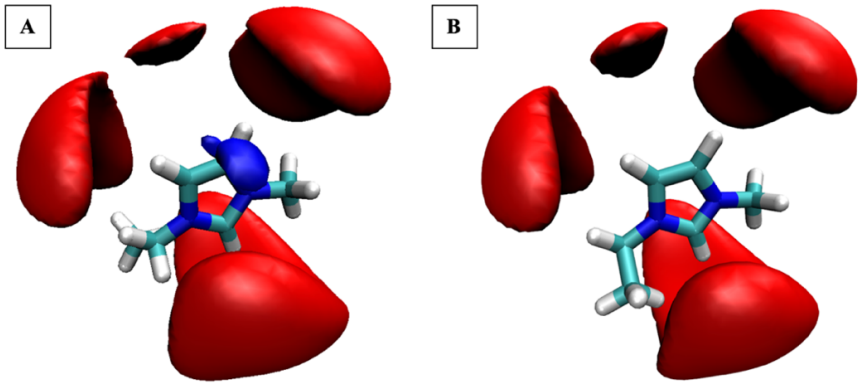


Figure 5.3. 4: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for lysine (A) low concentration system *vs* (B) high concentration system in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.

Radial distribution functions between cation and acetate anion atomic sites are shown in appendix A.9.2 for low concentration *vs* high concentration systems at 300K. They show that the closest peaks for acetate anion O1 atom occurs for ring hydrogens H1, H7 and H8, the highest one being for H1 atom and they all slightly decrease as the concentration is increased. Comparing it with SDFs shown here confirms that acetate anions mainly coordinate to the cation within the ring plane and adjacent cations are almost exclusively located on top and below the $[C_2C_1Im]^+$ ring. This is in agreement with other recent studies for various pure and mixture of ILs.[189, 190, 215-217]

The trend of decrease in interactions with concentration is similar to the lysine system in $[C_2C_1Im][Cl]$ although the organization within the chloride IL is stronger, with more concentrated regions of cation and anion around central cation as compared to $[C_2C_1Im][CH_3CO_2]$ IL.

5.3.2 Interactions of the ionic liquid $[C_2C_1Im][Cl]$ with arginine and lysine

Structuring around solutes can have a profound impact on their ability to dissolve these components, participate in reactions,[146] and interact with other species in solution, for example participation in the formation of structurally important salt bridges within proteins. As such it is important to characterise these IL-amino acid interactions further. The SDFs constructed for the mass-centred molecules are shown in Figure 5.3.5 and indicate that the central lysine molecule is surrounded by IL cation and anion and no lysine-lysine interactions are observed at the chosen iso-value of 10 nm^{-3}. For these charged lysine molecules, the high probability regions for anions are around the backbone ammonium and high probability regions for the cation are either as a second shell near these areas or directly interacting with the lysine carboxylate oxygens. A similar trend was observed by Youngs *et al.*[208] for glucose dissolution in $[C_2C_1Im][CH_3CO_2]$. The anions were found to be predominant in the regions around glucose and are directed by hydrogen bonding to the hydroxyl groups with cations majorly occupying spaces around the anions because of the electrostatic attractions. In addition, some cations had direct interactions with ring oxygens.[208]

For arginine, there is a highly-pronounced set of interactions of the anion with the guanidinium side group that also influence the extent of the cationic interactions. This interaction can be rationalised due to the more diffuse nature of the guanidinium group affording a greater surface-area for ion-ion interactions. This suggests both that the side chain can influence the ionic interactions at the backbone, even when relatively distant, and that changes to the nature of the anion are likely to have a greater influence on guanidinium-based molecules than for ammonium-based species.

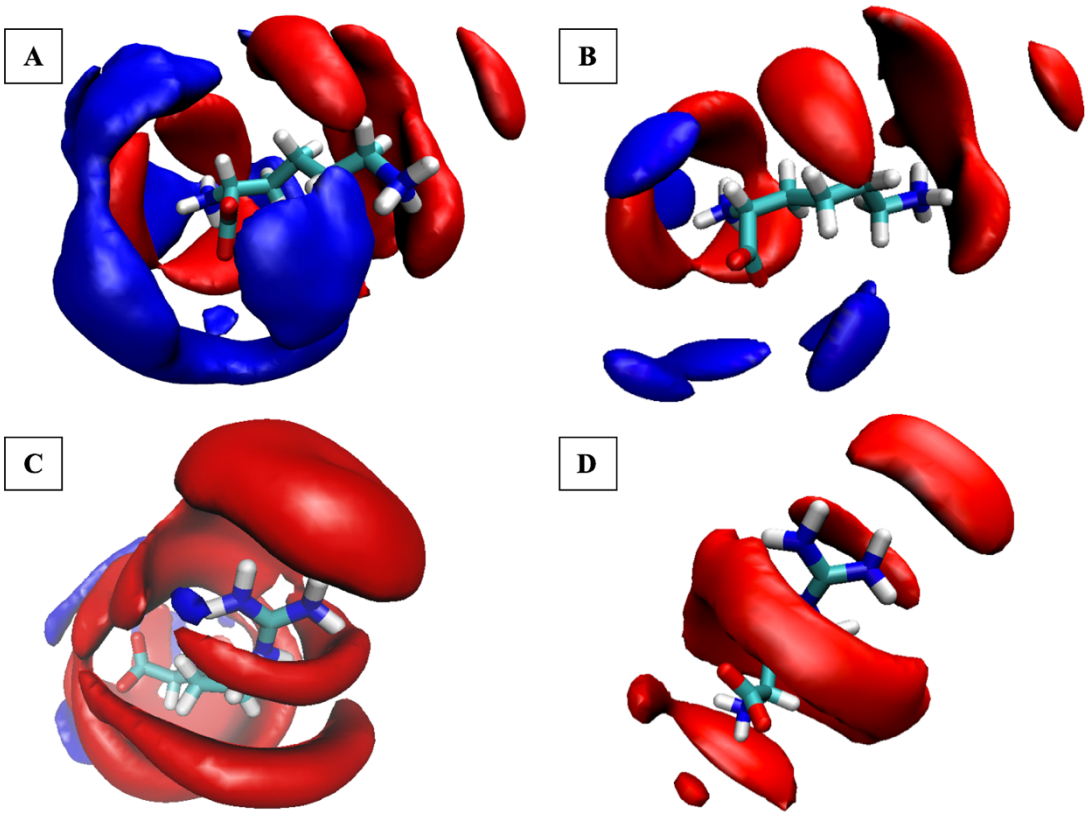


Figure 5.3.5. Spatial distribution functions around the central (A,B) lysine low concentration system *vs* high concentration system and (C,D) arginine low concentration *vs* high concentration system in $[C_2C_1Im][Cl]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}

As the concentration ratio is increased, both anionic and cationic interactions are affected with a more substantial relative change for the cations, seen by the increased loss of cationic density from the SDF, relative to that for the anions. These SDFs illustrate clearly that the first solvation shell around the backbone ammonium is dominated by strong, electrostatic chloride-amino acid interactions. This might suggest a sustained tighter interaction between the anions and the amino acid ions that is harder to disrupt, consistent with accounts of the importance of anion choice above cation choice in directing reactivity.[218] The comparison for the H1 atomic site on the $[C_2C_1Im]$ cation RDFs shows that the imidazolium H1 is closer to carboxylate group on lysine molecule, compared to chloride anions to ammonium ion on lysine group and is consistent with the SDFs in appendix for these specific interactions (figure A.10.4). This suggests that under conditions of changing concentration, the cation has the potential to influence the lysine-IL interaction, alongside the chloride, by having an entropy-driven impact that can be tuned by amino-acid concentration, with the impact of the anion being more concentration independent. Further studies, including

experimental studies, would need to probe this concept further, and to examine the underlying structuring that may underpin these variations in ion-ion interaction.

The important atomic sites from solutes that are expected to play key role in interacting with the surrounding IL ions are the functional groups attached to the alpha-carbon atom in each amino acid system, along with the side chain functionality. For analysing the distribution and structuring of [C_2C_1Im] cations around the central amino acid molecule, various atomic sites from these cations are selected i.e. H1, the acidic proton, H7 and H8, N1 and the tail and methyl carbon (C1 and C6, respectively, shown in appendix A.10.2 & A.10.3). RDFs are constructed by keeping the lysine and arginine molecules as reference molecules.

Figure 5.3.6 shows the distribution of various atomic sites of the IL cation as a function of distance from oxygen atoms in the reference lysine and arginine molecules and shows the effect of increasing the concentration of the two amino acids as well. Appendix A.10.1 shows the [C_2C_1Im] cations structuring comparison around carboxylate groups for two amino acids, and is pretty much similar to figure 5.3.6, but the strengths are more for oxygen atom as for carboxylate group, the whole group was chosen as the central molecule.

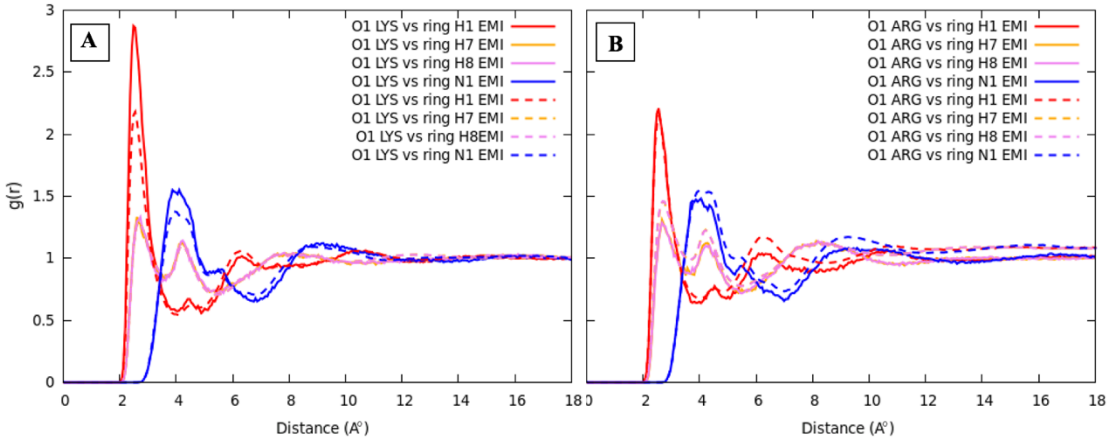


Figure 5.3.6: Site-site radial distribution functions for [C_2C_1Im][Cl]:[lys] at 300K for H1, C1, C6 atomic sites on [C_2C_1Im]$^+$ cation *vs* (A) O1 Oxygen atom on lysine molecule and (B) O1 Oxygen atom on arginine molecule [Low concentration (200:1) solid lines] and [high concentration (5:1) dashed lines] determined by MD simulation

Figure 5.3.6 (A) and (B) show the organisation around both the mass-centred O1 oxygen atom of the carboxyl group of lysine and arginine, respectively. The closest peaks for distribution around O1 at 2 Å occur for ring protons on [C_2C_1Im] cation but

the intensity is not the same for all the ring protons, suggesting some directionality of interaction, with the peaks sharper and more intense for H1 proton. For example, the low concentration system of lysine has a g(r) value of 2.8 compared to 1.4 for the ring protons (H7 and H8), suggesting the acidic proton H1 interacts more strongly with the lysine oxygen atom. The peaks for H7 and H8 are broader peaks suggesting that the interaction with the lysine O1 is less likely and thus weaker. This contrasts with the second peaks for both H1 and the ring hydrogens, which occur around 4.2 Å. The co-ordination between acidic hydrogen and O1 atom on arginine molecule is significant but is less intense than compared to lysine and is not affected with increasing concentration. At higher concentrations, the interactions of both H1 and N1 appear to be reduced for lysine, with decreased probability at the same distance. This could also indicate a decrease in the directionality and/or rigidity of organisation at these higher concentrations, which could be a preliminary indicator of a less organised system. This could also be a concentration effect that molecules on average has less interaction with each other and manifests itself as a weaker co-ordination. This contrasts with the structuring of ring hydrogens around arginine that show slightly increased interactions at higher concentrations but remains unchanged around acidic proton, but the overall effect observed in SDFs is in line with the overall RDFs. (Figure 5.3.8)

Figure A.10.2 in appendix reveals that the comparison of the interactions of the two positively charged amino acids lysine and arginine, for both the examined concentrations, the organisation pattern is similar especially around the oxygen atoms on carboxylate groups, the strength of interaction however is higher for single amino acid (low concentration) systems but the pattern differs around the backbone nitrogen. The ammonium ions in the side chains of both amino acids are not interactive with any atomic sites of cation as expected for them being the same charge species (appendix figure A.10.3). However, although these amino acid regions are structurally similar, it is clear that the side chain does impact on these interactions in a measurable way, and indicates that backbone-IL interactions may show localised variations within peptides and proteins. This would be consistent with longer-range impacts of organisation of the IL ions that may be important for protein solubility and especially functional changes. The latter could play a role in indirectly attenuating enzyme reactions.

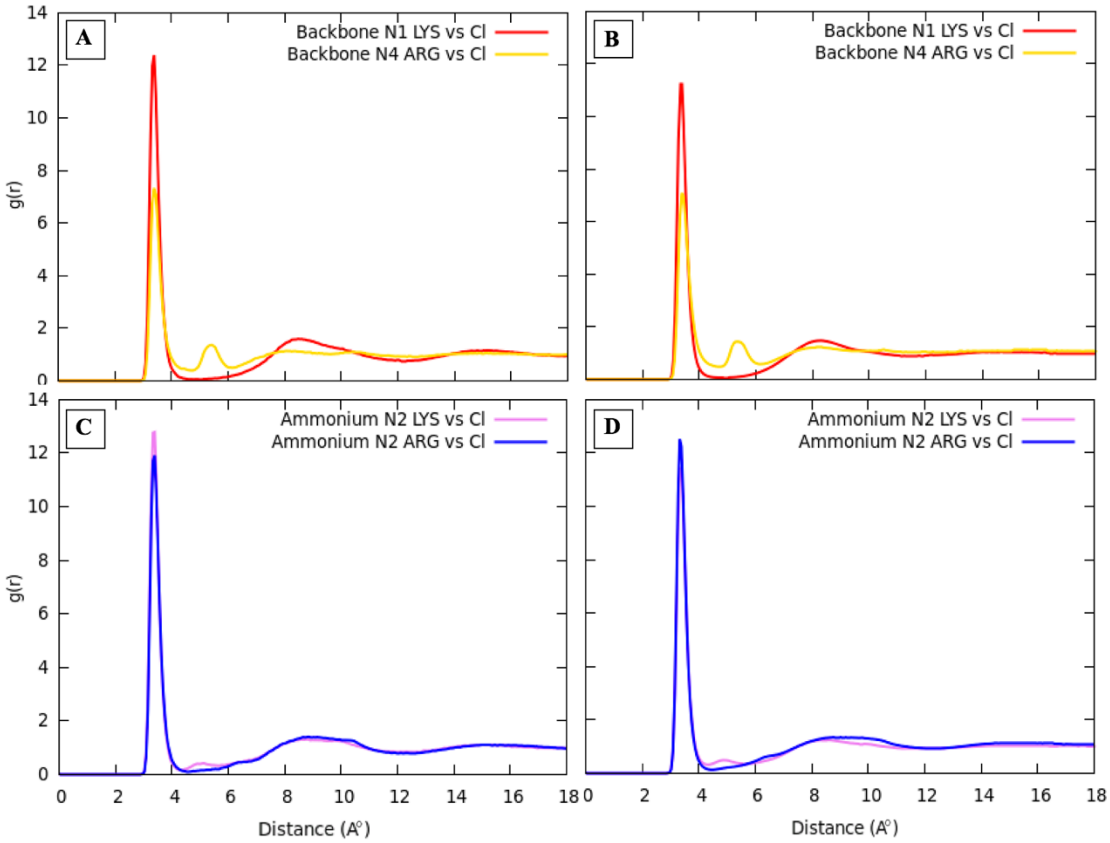


Figure 5.3.7: Comparison of radial distribution functions for $[C_2C_1Im][Cl]$: [AA] at 300K for chloride anion around backbone nitrogen on lysine and arginine at (A) low (200:1) *vs* (B) high (5:1) concentration systems and around ammonium nitrogen on lysine and arginine at low (C) low (200:1) *vs* (D) high concentration (5:1) determined by MD simulation

Anions have been shown to play an important part in interacting with protein-based biomaterials, thereby leading to their dissolution. The anions are thought to initiate this dissolution process through hydrogen-bonding interactions, with the more bulky cations moving in through the space created by anions and further aiding dissolution through hydrophobic interactions.[172] Chloride is considered a small and effective anion in this regard. The distribution around the alpha-hydrogen atom on both lysine and arginine molecules, labelled as H15, shows the closest peaks for chloride anion occur around this region at 2.5 Å, which is consistent with this being a closer distance than for the corresponding cations.(appendix Figure A.10.5) The first maxima and origin of interaction for lysine and arginine molecule are the same and peak heights are comparable, however the second peak is more pronounced for lysine molecule and hence indicates a more ordered structuring, reflected by the presence of two solvation shells. The probability of finding an anion in this position is however decreased by the increase in the concentration ratio. Figure 5.3.7 shows the distribution of chloride ion around the central backbone nitrogen atom and the nitrogen atom N2 atom for each

polar amino acid at low and high concentration at 300K. Both backbone nitrogen atoms (N1 for lysine and N4 for arginine) also show strong association with Cl^- ion with the first solvation shell at a distance of 3.5 Å. The peak is quite sharp, showing that the electrostatic interactions between the amino acid and chloride ion are quite high in this region and are stronger than those formed with the alpha-hydrogen atom. The interaction is somewhat higher for lysine than for arginine. The trend and ordering of the Cl anion for high concentration system is similar to that observed for the dilute system, with the slightly decreasing probabilities for the high-concentration system. The decreasing pattern observed here is expected as on the average the same number of anions are interacting with the higher number of amino acids.

By contrasting the pronounced and sharp organisation seen for the IL anion (Figure 5.3.7 and appendix Figure A.10.5) with the more diffuse organisation of cation (appendix Figure A.10.2), the role of anionic interactions with the backbone amino nitrogen is highlighted.

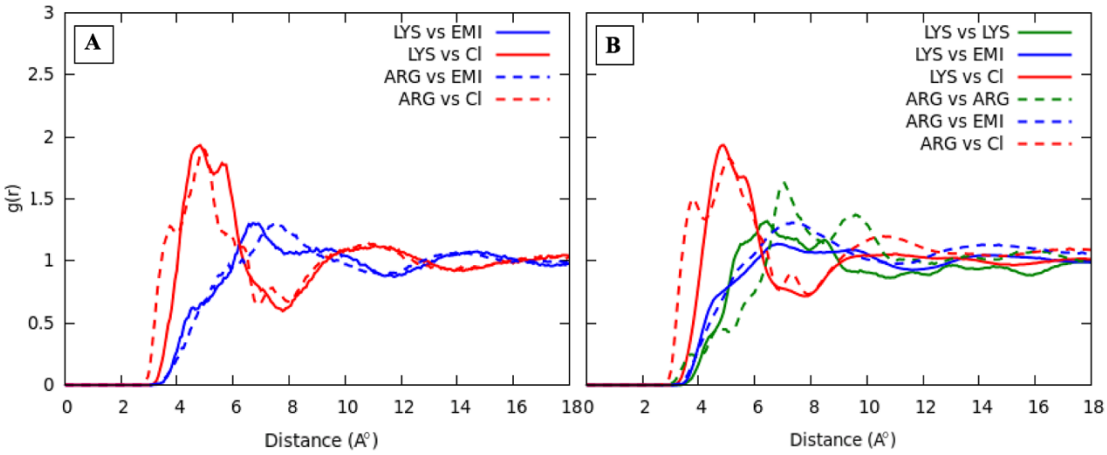


Figure 5.3.8. Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: [AA] for overall solute-solute and solute-solvent interactions for lysine (solid lines) and arginine (dashed lines) for (A) Low concentration *vs* (B) high concentration systems at 300K determined by MD simulation

The overall RDFs between mass-centred amino acids and ionic liquid ions are shown in figure 5.3.8. Arginine in both concentrations (high and low) exhibits closer interactions with the chloride and, on average, a greater probability of finding a chloride close than lysine.

Chloride ions dominate the shell around the arginine molecule much more. The side chain attached to the backbone carbon contains the positively charged guanidinium ion, which potentially results in a greater probability of chloride ions around the arginine

molecule due to the spread-out nature of the charge and this substantial interaction is supported by the broad peak of the chloride ion in the RDFs. Lysine and arginine differ in respect to the groups attached to alkyl chain i.e. ammonium cation for lysine and guanidinium cation for arginine and so the difference in the accumulation of chloride ions around them can be explained on the basis of difference of the strength of hydrogen bonds formed by these groups. The similar behaviour was observed by Fromm *et al* [219] where they showed that hydrogen bonds formed by guanidine group are more close to ideal 180° and therefore stronger than those formed with the ammonium side-chain of lysine. This suggests also that the anion effect is likely to be more significant for arginine than for lysine when comparing ionic liquid solvation.

5.3.3 Effect of temperature on amino acid and IL interactions:

Density, viscosity, surface tension and heat capacity are a few of the thermo-physical properties of imidazolium-based ionic liquids that been shown previously to be sensitive to temperature, and are quantities that can be obtained experimentally and computed theoretically for ionic liquids.[198, 220-222] Work by the collaborating group of Licence has shown that temperature also significantly impacts on the state of ionic liquid-amino acid interactions. In an experimental study, Maxwell-Hogg was able to show that a selection of charged amino acids were able to induce gel-like properties when dissolved at temperatures of around 100 °C in ionic liquids at high concentrations (approximately 1:5 molar ratio of amino acid to ionic liquid) and subsequently cooled. These gel-like materials were thermo-reversible, becoming liquid again once heated. Thus, to observe the effect of temperature on the structuring of ionic liquid and lysine mixtures, in line with the experimental data, simulations were run at 400K for lysine in $[C_2C_1Im][Cl]$ $[C_2C_1Im][CH_3CO_2]$ for two molar concentration ratios of IL:AA; 200:1 and 5:1. The same initial coordinate files were taken for running simulations at both 300K and 400K as explained in Chapter 3. The choice of this higher 400K temperature ensures a lower viscosity and, therefore, a more liquid-like state and a better sampling. The thermal stability also lies within this temperature range.[216]

A summary of the overall impact of structuring of the ionic liquid around the amino acid lysine is shown in Figure 5.3.9. The most substantial changes afforded by either an increase in concentration, or an increase in temperature are in the density of the

cationic interactions, which become noticeably weaker, being imperceptible at the high-concentration 400 K condition at an isovalue of 10 nm^{-3}. This suggests that these are weaker interactions, more able to be broken, than the corresponding anionic interactions with the ammonium groups. This could have impact on selective reactivity and interactions, as at higher temperature the carboxylate group of the amino acid is relatively more exposed, and entropic gain from breaking the IL-solute interactions will be less at this position.

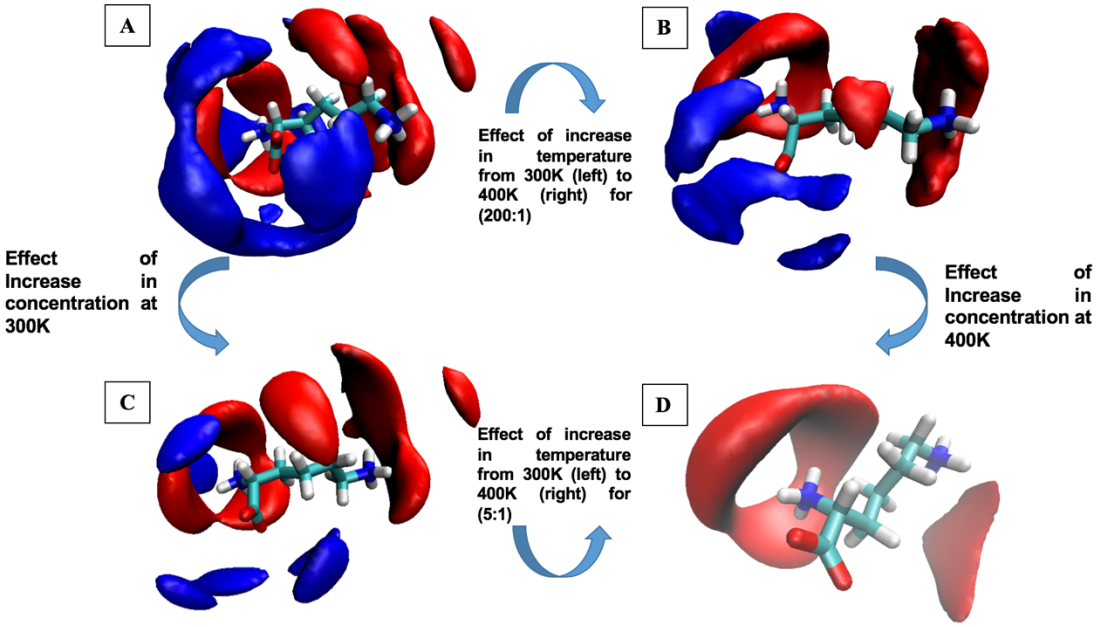


Figure 5.3.9: Spatial distribution functions around the central (A) lysine low concentration system at 300K (B) lysine low concentration system at 400K (C) lysine high concentration system at 300K and (D) lysine high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}

Figure 5.3.10 shows the detail of interactions of lysine atomic sites with chloride anions and the H1 acidic proton on $[C_2C_1Im]$ cation as the temperature increases from 300K to 400K for both low concentration and high concentration systems. The plotted data shows that the interactions with chloride anion are slightly decreased around the alpha-hydrogen and the side-chain N2 nitrogen atom (shown in appendix A.9.2), but a pronounced effect is seen in the magnitude of the interaction around the backbone nitrogen with a change in g(r) value from 11 to 9 on going from low to high temperature. There is also a significant change in going from low to high concentration, which implies that the competitive interactions between the lysine molecules also plays its role to further reduce the anion-lysine interactions as the temperature is increased. This conclusion is further supported centre-of mass RDFs drawn between lysine and IL ions (Appendix figure A.11.3), which shows that the lysine-lysine interaction has slightly

strengthened at higher temperature condition, although this is not visible in the SDF at the iso-value selected.

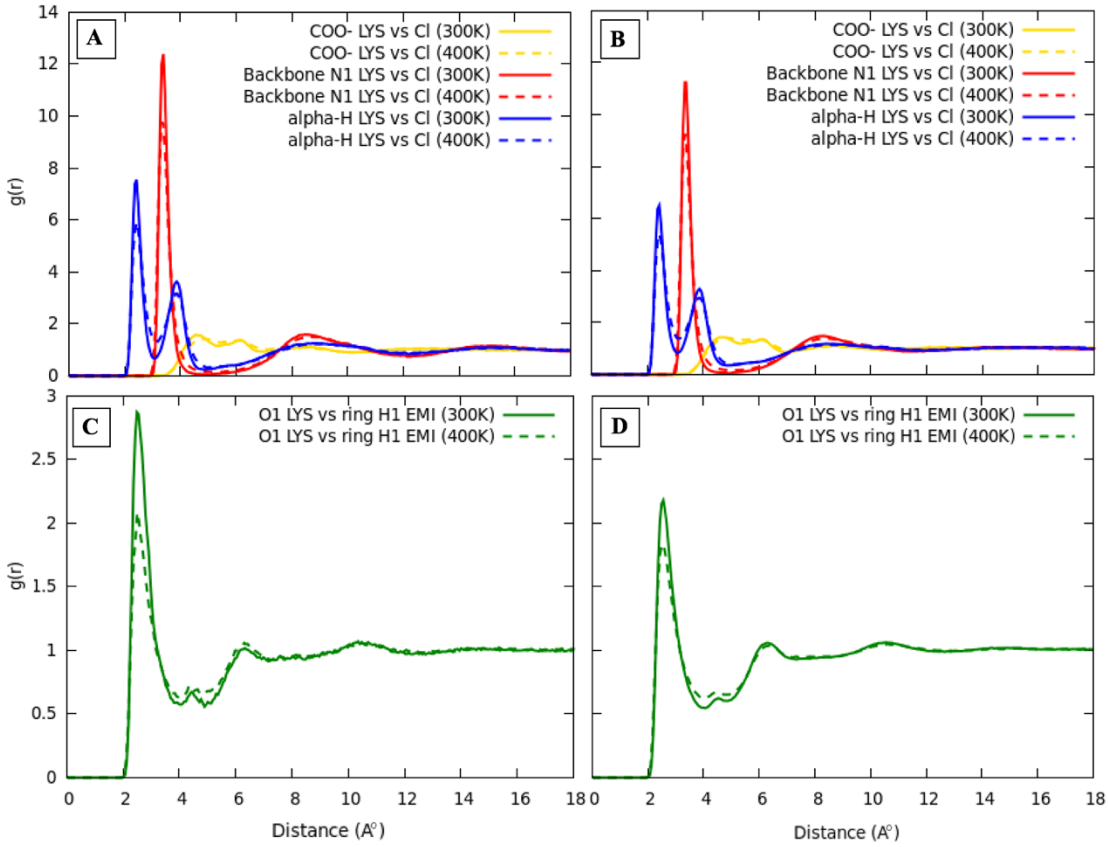


Figure 5.3.10. Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: lysine low concentration ((200:1 (left A,C)) and high concentration ((5:1) (right B,D)) at 300K (solid lines) *vs* 400K (dashed lines). (A,B) show chloride anion distribution around backbone nitrogen, carboxyl group, terminal nitrogen and alpha-hydrogen *vs* (C,D) show H1 $[C_2C_1Im]^+$ cation distribution around backbone nitrogen, carboxyl group and O1 determined by MD simulation

The temperature change in lysine solutions in $[C_2C_1Im][Cl]$ has a clear impact. At 300K, both the positively charged amino acids are quite dispersed and there is no substantial aggregation seen (appendix figure A.11.3). At the higher temperature, the chlorine-lysine interaction substantially drops, more prominent in above figure 5.3.10 than in the overall RDFs, however, the lysine-lysine interactions are slightly enhanced.

The SDFs shown in Figure 5.3.11 compare the organisation in the solvent as a function of change in the concentration ratios and temperature for lysine. Temperature affects both the cation-anion and cation-cation interactions, with the most substantial impact on the cation-cation interactions. In case of the low concentration lysine system, cation concentration around cation rings is significantly reduced, but can be seen at higher

temperature whereas the anion concentration around the methyl group completely disappears showing that the temperature causes a decrease in cation-anion interactions. For higher concentration system, anions disappear from the region around methyl group and the cation-cation interaction is lost as the temperature is raised. This emphasises the combined impact of temperature and solute concentration in determining the organisation in the ionic liquid itself and thus the potential impact of solutes and impurities on wider ionic liquid properties, and highlights the complexity of the equilibria in these systems.

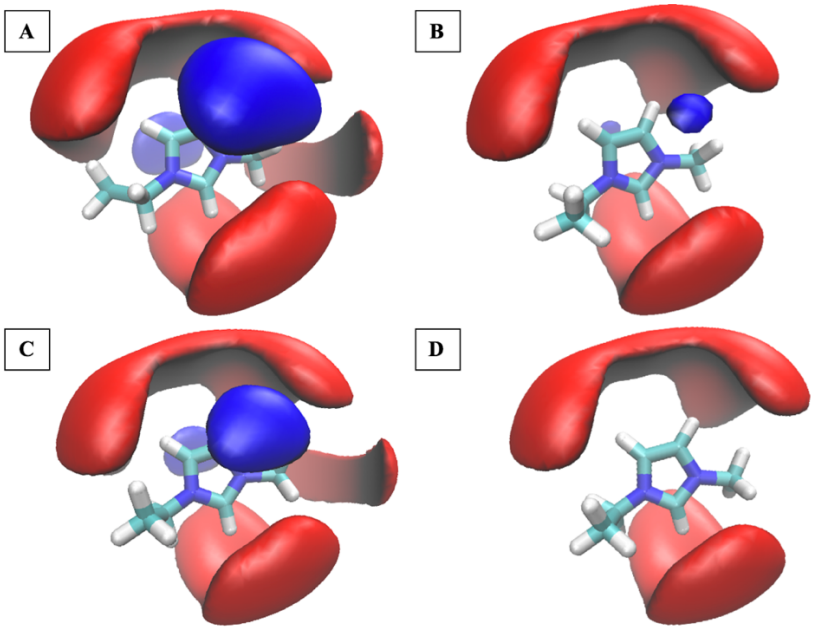


Figure 5.3.11. Spatial distribution functions for cation-cation and cation-anion interactions around the central $[C_2C_1Im]^+$ cation for (A) lysine low concentration system at 300K, (B) lysine low concentration system at 400K, (C) lysine high concentration system at 300K and (D) lysine high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}

5.3.4 Effect of addition of water on lysine and $[C_2C_1Im][Cl]$ interactions

ILs have been widely investigated in a large range of processes in the chemical industry, including separations processes, reactions and biomass processing owing to their unique properties.[223] Most ionic liquids (ILs) are either water soluble or have a non-negligible miscibility with water and it is known that IL phase equilibria and their thermophysical properties are significantly influenced by the presence of water. A significant body work has built up, addressing the thermophysical and thermodynamic properties of water-saturated ILs, in particular for densities and viscosities[224]. The ionic

nature of ILs means that many are highly hygroscopic and, therefore, any large-scale application involving ILs must take this issue into account. Most IL thermophysical properties are sensitive to the presence of water and other impurities and Seddon and co-workers show that the viscosity is dramatically decreased in the presence of water,[222] with IL viscosity one of the properties most affected by the water content.[225] Rodríguez and Brennecke[226] studied the density and viscosity of aqueous solutions of imidazolium-based ionic liquids with ethyl-sulfate [$EtSO_4$], trifluoroacetate [CF_3CO_2]$^-$ and tri-fluoro-methane-sulfonate [CF_3SO_3]$^-$. Both properties were found to decrease with an increase in either temperature or in mole fraction of water. In most studies, it was found that the water content has a strong effect on the viscosity of the ILs, while the effect is less significant on their densities.

Viscosity is one of the most important physical properties when considering the scale-up and application of IL-based processes. Low viscosity is usually desired for solvent applications in order to increase the mass transfer rates and to minimize pumping costs.[80] The viscosity of an ionic liquid influences the solubility of cellulose-containing natural products. Undesirable high viscosity impedes the dissolution of biomass composites. So, in view of this, Karl Fischer measurements (Appendix A.1) were carried out for the samples of ILs used for experiments and, based on these measurements, the number of water molecules to be added in the computational system was calculated as shown in Appendix A.2 and also mentioned in chapter 3. On this basis, 26 water molecules were added to the high-concentration system containing IL:AA 5:1. The amount of water reflects the amount absorbed from atmosphere and is not added during experiments. The simulations followed the procedure laid out in methodology, and were run at two temperatures: 300K and 400K. The results were compared to the simulations run without water.

The addition of water affects the interaction strengths as shown in figure 5.3.12. The figure panels A and C show radial distributions at 300K, and B and D at 400K. The simulations with water that were run at 300K and 400K have the same starting structure. Figure 5.3.12 A shows that the distribution of chloride ion around various atomic sites of lysine molecule slightly decrease as water is added to the system at both temperatures, the g(r) value for backbone nitrogen reduces from 11.4 to 11.2 at 300K and the curves also fractionally shift to the right as temperature is raised, indicating that

on average these shells are expanded on addition of water. Comparing the RDFs both in the absence and presence of water in panels A and B shows that the temperature increase for lysine systems in IL/wat mixtures decreases the interactions, with the biggest impact on the ammonium interaction with chloride, where the g(r) value reduces from 11.4 to 8.5.

The IL cationic interactions are shown in panels C and D. Water addition and then running simulation with ionic liquid/water mixture at 300K and 400K show that the interactions for the cationic H1 atom around lysine O1 oxygen atom only decrease as a function of temperature and water addition has little effect. Temperature has a prominent effect on the interaction strength, both in the presence or absence of water, whereas in the presence of water, the interactions still suffer a decrease, but water has a much smaller impact. Overall, ion-ion interactions (Appendix figure A.12.1) with lysine appear to slightly reduce at the higher temperature in the water solution, however lysine-lysine interactions are strengthened as water is added.

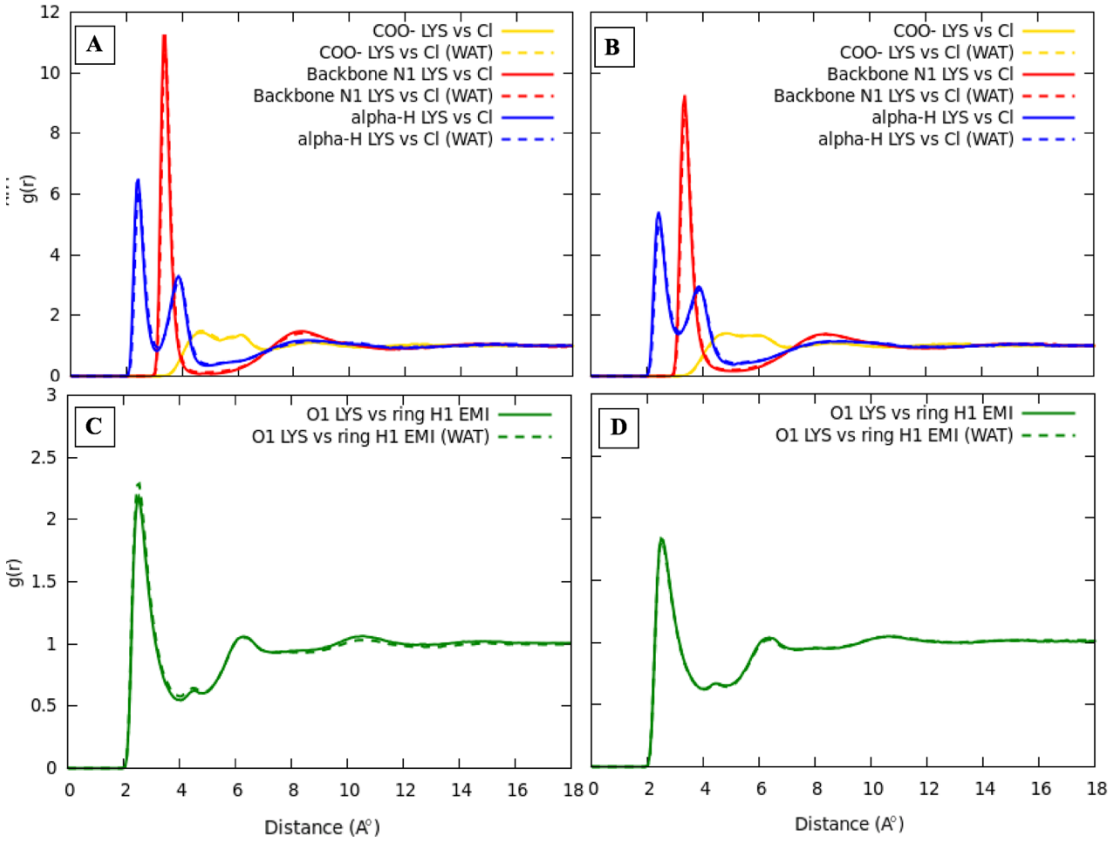


Figure 5.3.12. Comparison of centre-of-mass radial distribution functions around lysine atomic sites in pure IL and IL/water mixture of $[C_2C_1Im][Cl]$. High concentration (5:1) systems with water (dashed lines) and without water (solid lines). Chloride anion around lysine atomic sites at (A) 300K *vs* (B) 400K. and $[C_2C_1Im]^+$ cation around lysine atomic sites at (C) 300K *vs* (D) 400K determined by MD simulation.

Nano-structuring is characteristic of ILs with $C_n \geq 4$ and can be considered through segregation of the polar and non-polar phases.[53] Wang and Voth[227] in their article on coarse-grained simulations of 1-butyl-3-methylimidazolium nitrate $[C_4C_1Im][NO_3]$, discussed that "the neutral tail groups of cations can aggregate to form spatially heterogeneous domains of the tails, while the charged head groups and anions distribute as uniformly as possible due to the strong electrostatic interactions." Lopes and Pádua[228], later studied several $[C_nC_1Im][PF_6]$ IL varying their chain lengths and also confirmed the micro-heterogeneity. They showed that the charged domains are not homogeneously distributed but there is a continuous 3-D network of ionic or polar domains that co-exist with the nonpolar domains.[228]. The range of forces present in ILs, have different impacts in different domains. Columbic interactions including hydrogen bonding are typical for the polar domains and dispersion forces are characteristic to the non-polar domains.[229, 230] The ability of ILs to interact with the different species depend on the nature of the solute and is associated with the complex nature of these forces.[231] M. Brehm *et al.,*[232] introduced the concept of domain analysis that can further provide insight into the micro-heterogeneous structure of liquids. [52]

For our systems, we have applied this analysis to check the aggregation of amino acid molecules in the solutions and not really looking at the heterogeneity of IL alkyl chains. The number of domains or domain counts in each case can indicate if the amino acids are fully dispersed or not. A series of domain analyses was run for all the systems using the TRAVIS[186] software, by considering all the atoms present in the system as the base population and then separating the domains into solute (amino acid) and solvent consisting of cations and anions. Here the maximum domain count for the amino acid can be considered as the number of amino acid molecules present, i.e. 40 for the high concentration systems, and this maximum value indicates complete solvation of the solute, with lower values indicating increasing aggregation. The domain analysis across the simulations for lysine high concentration systems in pure $[C_2C_1Im][Cl]$ and the $[C_2C_1Im][Cl]$/water mixture is shown in the figure 5.3.13, at the two simulation temperatures 300K and 400K. Here, the most obvious apparent trend is the greater fluctuation of domain numbers in the water-based solutions, probably representing break up of aggregates through dynamic interstitial waters.

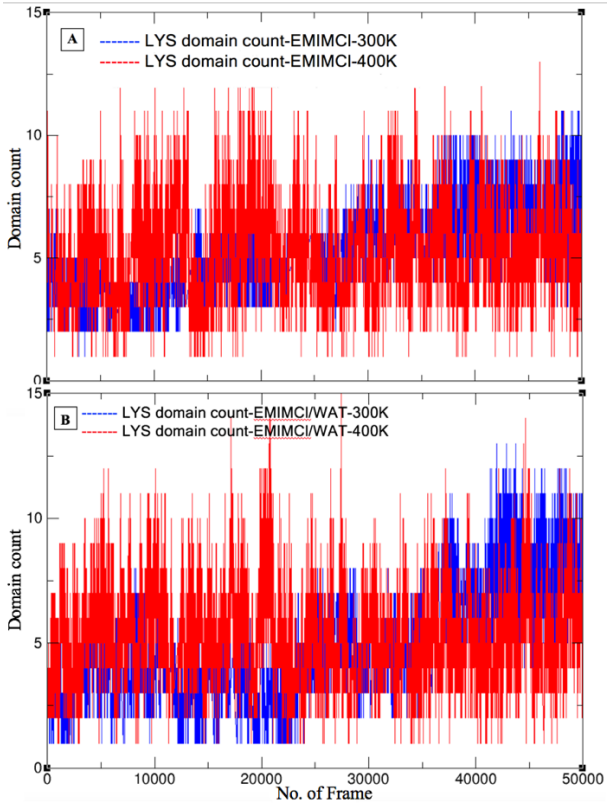


Figure 5.3. 13: Domain analysis across complete simulations of (A) lysine in $[C_2C_1Im][Cl]$ and (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K(blue) and at 400K (red)

Domain counts averaged over the complete simulation time are presented in Table 5.2. Overall, at low temperature and in pure IL, the average domain count for lysine is 4.97, that slightly increases to an average of 5.15 as the temperature is increased from 300K to 400K for pure $[C_2C_1Im][Cl]$ system. This increase suggests that the lysine molecules might be breaking away from each other as the temperature is increased and there is a minor rearrangement happening that is indicated by the slightly enhanced lysine-lysine peaks (Appendix figure A.11.3). This might be indicative of more opportunity for lysine molecules to interact with different lysine molecules at a higher temperature. Hence, they break away and have more but transient interactions. For the $[C_2C_1Im][Cl]$/water system the average domain count for lysine is 4.6 at 300K, increasing to 5.19 at 400K. The average values here don't give a clear picture about the system. An interesting note is that the water molecules are fairly evenly dispersed throughout the solution, indicating high miscibility with the IL. These values of domain counts are not significantly different. Snapshots of the solution at 100 ns simulation time (Appendix A.12.4), when the system is well equilibrated, indicate that the lysine molecules aggregate with either increasing water or increasing temperature. This aggregation is

supported through a slight increase in lysine-lysine interactions seen in the RDF (Appendix A.12.1).

Table 5. 2: Domain analysis for $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixtures

Ionic Liquid	Temperature (K)	Domain count (average)			
		lysine (40)	Cl (240)	$[C_2C_1Im]$ (200)	O1-WAT (26)
$[C_2C_1Im][Cl]$	300K	4.97	173.2	2.27	
$[C_2C_1Im][Cl]$	400K	5.15	163.8	2.11	
$[C_2C_1Im][Cl]/H_2O$	300K	4.66	170.9	2.21	24.98
$[C_2C_1Im][Cl]/H_2O$	400K	5.19	162.64	2.2	24.83

The average domain count in each simulation is closer to one than the maximum possible value of 40 in the high concentration systems. This suggests that $[C_2C_1Im][Cl]$ is poorly able to solvate the lysine molecule. A comparison with the domain analysis performed for $[C_2C_1Im][CH_3CO_2]$ as the solvent is shown later in the next sub-chapter 5.4 to highlight the different solvating abilities of different ionic liquids.

To look at the aggregation and potential gelation behaviour of lysine-IL solutions, self-diffusion coefficients were calculated. The self-diffusion coefficients calculated through NMR studies have been found to be useful in understanding various solute-solvent and solvent-solvent interactions in IL systems.[197, 233] So, in order to observe in detail, the change of the mobility of the ions as a function of the temperature and water concentration, self-diffusion coefficients were calculated using Einstein's relation as outlined in chapter 4, by linear fitting of MSD graphs and calculating their slopes.

The diffusion coefficients calculated initially were based on the whole 100ns trajectory, following Einstein's relation however, it is suggested that the values must be calculated at the end of simulation time or at very long times to ensure the dynamics of the system are in diffusive regime.[196] The initial values of diffusion coefficients are placed in Appendix A.13.2 for reference in relation to improved calculated values. The log-log-plots from the entire simulation showed that the slopes of the graph were not close to linear and hence the systems were not close to true diffusive behaviour. As explained in Chapter 4, the diffusion coefficients calculated through MD are close to the

experimental values when they are calculated from the start of simulation. These are apparent diffusion values and should be used for qualitative discussion. For accurate D values, the simulation should be run for longer times at a higher temperature or run in NVE ensemble.

For the systems under discussion, D values were calculated again from the first 50 ps of simulation as suggested by Maginn and co-workers[196]. The procedure followed for the calculation of D values is already explained in chapter 4. Using this methodology, the diffusion coefficients for pure $[C_2C_1Im][Cl]$ IL and $[C_2C_1Im][Cl]$/water systems are much higher as compared to those initially calculated. The slopes of the re-plotted log-log plots (Appendix A.13.1) calculated using eq 4.3 are still not linear but have increased, for example for lysine in pure $[C_2C_1Im][Cl]$, from 0.3 to 0.7. This indicates that the systems are close to being in a true diffusive regime. These D values however were considered as the apparent D values and for validation, the diffusion coefficients must be calculated through *e.g*. NMR diffusion methods.

The calculated diffusion coefficients for lysine in pure $[C_2C_1Im][Cl]$ IL and $[C_2C_1Im][Cl]$/water systems are shown in Table 5.3. By example, the slope is 0.1184 $Å^2\ ps^{-1}$ for lysine high concentration system in pure $[C_2C_1Im][Cl]$. To find the diffusion coefficient, this number is divided by six and multiplied by 10, to afford 1.973 E-10 m^2s^{-1}(0.1973 E-5 cm^2s^{-1}). The cpptraj[185] software package in Amber tools 2017/18 was used for the calculation of MSD values and 2D plots generated using excel (Appendix Figure A.13). Self- diffusion coefficients are temperature dependent and, as shown in the table, the D values increase as the temperature is raised for all the species involved in the systems. Water addition at both temperatures also shows the increase in D values that indicates that even the small amount of water or impurities in the system can affect the interactions and the overall mobility and viscosity of the system to make it more fluid. Generally, the lower diffusion coefficients indicate the strengthening of interactions among the molecules in the system. So, comparison of D values for lysine systems in two different ILs would indicate which IL is better at the solvation of lysine molecules. The D values for lysine in $[C_2C_1Im][CH_3CO_2]$ IL are shown in section 5.4 Table 5.4.3.

Table 5. 3: Self diffusion coefficients calculated from MD simulations for lysine systems in [C_2C_1Im][Cl] and [C_2C_1Im][Cl]/water mixtures

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s	
		T=300K	T=400K
		5:1	5:1
lysine-charged	lysine	1.97	5.14
[C_2C_1Im][Cl]	[C_2C_1Im]	2.29	10.39
	CL	1.67	6.51
lysine-charged	lysine	4.57	5.83
[C_2C_1Im][Cl]-wat	[C_2C_1Im]	5.67	11.73
	CL	4.37	7.19
	WAT	12.09	40.62

The key considerations here are that the order of magnitude is consistent and that the cation diffusion coefficients are higher than those for the anions. In all cases this is in line with the literature.[198] The higher self-diffusion coefficients of cations can be attributed to the flexibility of cation alkyl chains. Conversely, previous studies have found that cations diffuse faster than anions, even when their corresponding effective hydrodynamic radii are calculated to be larger than those of the counter-ions.[234] The D values for lysine and Cl^- ions are quite close to each other specifically at higher temperature. The cations diffusing faster than anions could also highlight a more prominent role of anion interactions with the lysine molecule as compared to the cations that is highlighted even more at higher temperature 400K. This is consistent with the observations from the SDFs.

5.4 Interactions and organisation of positively charged lysine in $[C_2C_1Im][CH_3CO_2]$ Ionic liquid

5.4 Results and Discussion

5.4.1 Effect of $[C_2C_1Im][CH_3CO_2]$ as an amino acid solvent

The structure and dynamics of the $[C_2C_1Im][CH_3CO_2]$ ionic liquid has been studied extensively either as a pure solvent or with water as co-solvent, with a specific focus on the solvation of enzymes, cellulose, and other bio-renewable materials.[208, 210, 216, 235-237] The interest in ionic liquids with acetate anions developed from the initial report of cellulose dissolution in 1-alkyl-3-methylimidazolium chloride ionic liquids[88] and the subsequent explosion of interest[237] in the application of ILs in biomass processing to access or generate polymeric materials,[36, 74] 1-ethyl-3-methylimidazolium acetate $[C_2C_1Im][CH_3CO_2]$ has been identified as perhaps the "best" cellulosic solvent[238] and, additionally, also shows significant ability to absorb CO_2[239] and has also known to be useful for enzyme[240, 241] and environmental-friendly features.

5.4.1.1 Key interactions of lysine in $[C_2C_1Im][CH_3CO_2]$

Lysine was the focus of this study, as the interactions are more varied than those with arginine. The simulations were run for 100ns for lysine in $[C_2C_1Im][CH_3CO_2]$ ionic liquid at 300K and then at 400K for two different concentration ratios (200:1 and 5:1, IL:AA) to provide a detailed comparison with the $[C_2C_1Im][Cl]$ systems. Systems were generated based on the procedure laid out in methodology section.

On initial inspection, the trends of the anion and cation dominant regions around the central lysine (Figure 5.4.1) are similar to those observed for $[C_2C_1Im][Cl]$ (Figure 5.3.5), consistent with the domination of electrostatic interactions. Here, however, cation interactions appear more dominant, with less dominant anion interactions, suggesting a weaker anion-ammonium interaction for acetate than with chloride. The SDFs are drawn (Figure 5.4.1) by taking backbone atoms from lysine molecule as shown in figure 5.3.1 and the centre-of-mass of acetate anion.

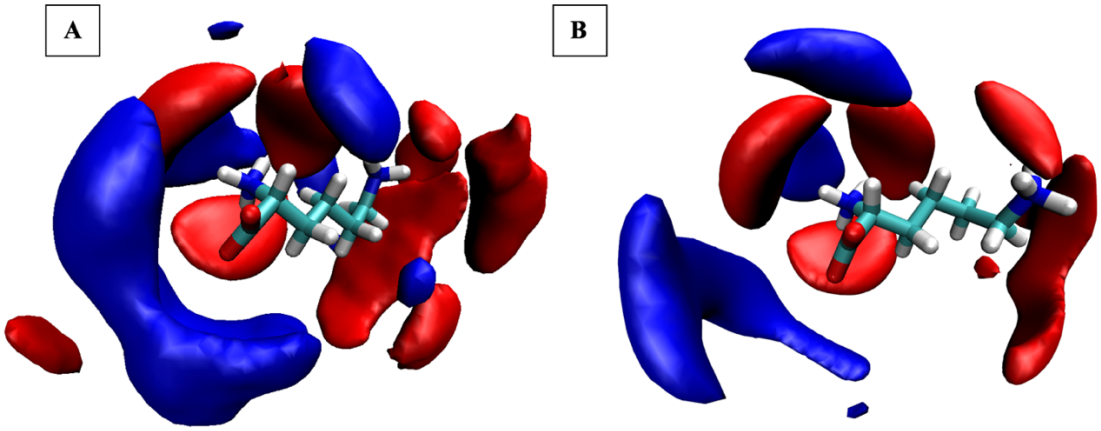


Figure 5.4. 1: Spatial distribution functions around the central lysine molecule for (A) low concentration system *vs* (B) high concentration system, in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.

Figure A.14.1 in appendix, compares the distribution of O1 atoms on acetate anions and ring proton H1 on $[C_2C_1Im]$ cation around various atomic sites of lysine for the two systems differing in concentration of lysine amino acid at 300K. The ring proton H1 is chosen for the comparison of interactions as it shows the strongest peaks for organisation around various lysine atomic sites as shown in appendix A.14 in figure A.14.3 (panels A, B, and C). The pattern of organisation and strength of interaction is not much different for the two systems. In comparison with the lysine systems in two ILs, the main difference lies in organisation on changing from chloride to acetate, with the backbone nitrogen-chloride distance around 3.5 Å whereas for O1 atom on acetate anion two peaks are observed at 3 Å and around 4.5 Å. This can be related to the large ionic radius of chloride ion relative to that of the acetate oxygen atoms. The strength of interactions with both the anions are comparable in the sense that in both cases, they decrease slightly with increasing concentration. The two peaks of acetate anions suggest more structured ordering of acetate anions and the closeness of shells can just be attributed to the bigger ionic radius of chloride anion as compared to acetate oxygen atom. This is consistent with acetate anions having a stronger ability to form hydrogen bonds, compared to chloride anions.[237] The distribution of cation is not affected by the change in nature of anion and a similar pattern can be seen for both ionic liquids. However, the concentration effect has more impact on $[C_2C_1Im][Cl]$ IL where the cation organisation around lysine atomic sites decrease with the increase in concentration. (Figure 5.3.5)

Figures in appendix A.11.1 and A.15.2 highlights the interactions with the lysine high concentration system between both the $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ solvents at 300K and 400K. Comparing acetate with chloride-containing ILs directly shows sharper and higher g(r) peaks for the cation in the acetate solution, indicating a stronger cationic association with lysine molecule when the acetate anion is present. The acetate anions and chloride anions have comparable peak heights but acetate anions form a tighter shell with lysine. Inter-amino acid interactions are not affected for lysine as the temperature is increased in pure $[C_2C_1Im][CH_3CO_2]$ IL (Appendix A.15.1), as opposed to pure $[C_2C_1Im][Cl]$ IL (figure 5.3.8). Here, in each case lysine-lysine interactions are quite dispersed and imply that $[C_2C_1Im][CH_3CO_2]$ is able to solvate lysine molecules better than $[C_2C_1Im][Cl]$.

5.4.2 Effect of temperature on lysine in $[C_2C_1Im][CH_3CO_2]$

For lysine systems in $[C_2C_1Im][Cl]$ changes in temperature showed distinct changes in the organisation. Therefore, the simulations for lysine systems in $[C_2C_1Im][CH_3CO_2]$ were run at 400K to provide a commentary on the relationship between temperature and interactions. Temperature values higher than 100°C are used for the dissolution of keratin-based materials as also mentioned in Table 2.1. The effect of temperature on both low and high concentration systems (figure 5.4.2) shows that the interactions strengths of both cation and anion with central lysine molecule are decreased as the temperature is increased, in line with the trends seen in the chloride-based IL. The most pronounced change in interaction is that of the IL cation with the lysine carboxylate that is significantly reduced at higher temperature and high concentration of lysine. In addition, the side-chain ammonium loses the bulk of its interaction with the acetate under the high-concentration, high temperature condition. This contrasts with the situation in the chloride-based IL, where some interaction is still remaining on this side group. This has potential implications for choice of solvent when considering protein-based solutes.

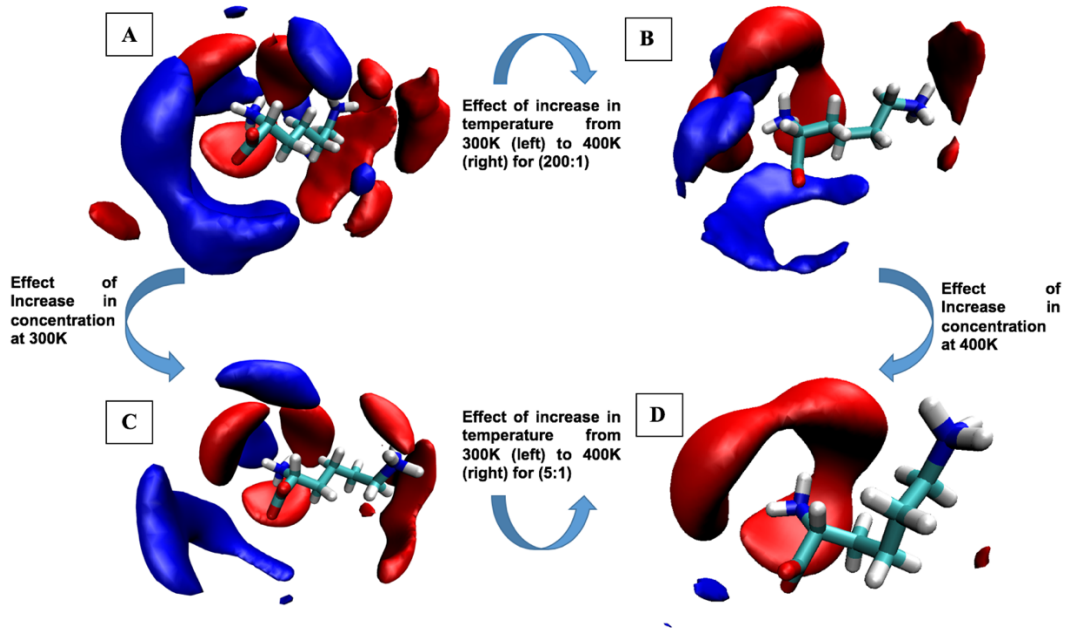


Figure 5.4.2. Spatial distribution functions around the central lysine molecule for (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K and (D) high concentration system at 400K in $[C_2C_1Im][CH_3CO_2]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}

5.4.3 Effect of water on lysine and $[C_2C_1Im][CH_3CO_2]$ interactions

Karl Fischer measurements for $[C_2C_1Im][CH_3CO_2]$ were carried out and the water content values are reported in appendix (A.1), table A.1.[242] $[C_2C_1Im][CH_3CO_2]$ is substantially less hygroscopic than the corresponding chloride, and as a result, only two water molecules were added to the $[C_2C_1Im][CH_3CO_2]$ systems. Simulations were run for 100ns each at 300K and 400K. The results were compared with the simulations run with pure IL at both the temperatures. Acetate anion interacts more strongly with water than with the cations but overall, the small quantity of water absorbed by $[C_2C_1Im][CH_3CO_2]$ does not have a significant impact on the interactions.

The snapshots extracted from 100ns simulations (Figure 5.4.3) show that lysine seems more dispersed than in chloride-based ionic liquid, with much less aggregation, and the corresponding SDFs comparing the simulations with and without water addition to IL are shown in appendix A.14.3 and A.14.4. In solution with the acetate-based IL, the lysine-lysine interaction is relatively minor and not observed in the SDFs.

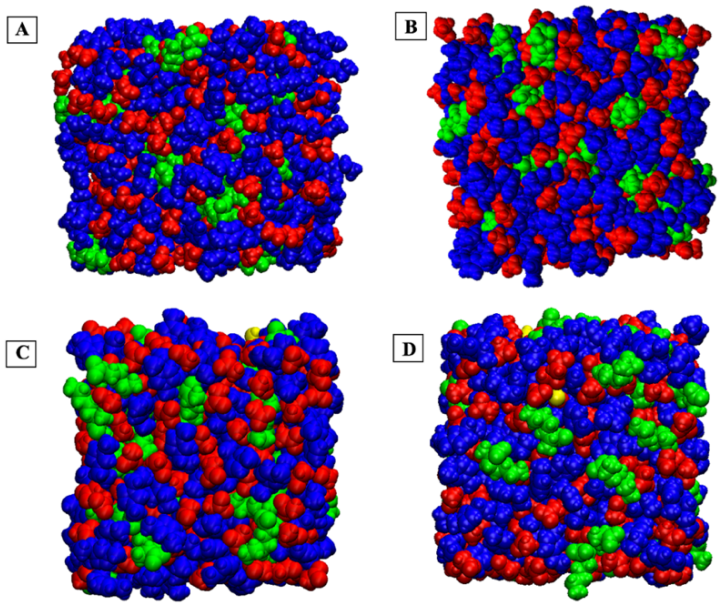


Figure 5.4.3. Snapshots of lysine high concentration systems in $[C_2C_1Im][CH_3CO_2]$ IL after 100 ns simulation for (A,C) 300K for pure and IL/WAT mix respectively and (B,D) 400K for pure and IL/WAT mix respectively

A series of domain analyses was run for all the systems and the domain analysis for the lysine high concentration systems in pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$/water mixtures is shown in the figure 5.4.4 at the two simulation temperatures of 300K and 400K.

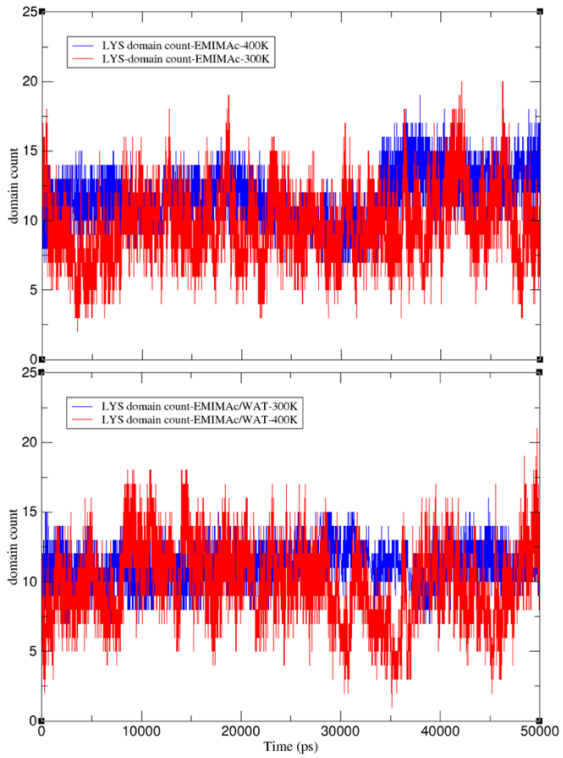


Figure 5.4.4. Domain analysis for (A) lysine in $[C_2C_1Im][CH_3CO_2]$ and (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K(blue) and at 400K (red)

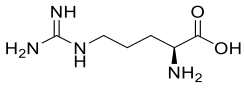
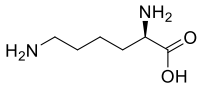
Table 5.4. 1: Domain analysis for $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$/water mixtures

Ionic Liquid	Temperature (K)	Domain count (average) Lysine (40)	O1-ACT (240)	$[C_2C_1Im]$ (200)	O1-WAT (2)
$[C_2C_1Im][CH_3CO_2]$	300K	11.58	219.4	8.12	
$[C_2C_1Im][CH_3CO_2]$	400K	9.72	217.6	7.38	
$[C_2C_1Im][CH_3CO_2]/H_2O$	300K	11.29	219.45	7.93	1.99
$[C_2C_1Im][CH_3CO_2]/H_2O$	400K	9.84	217.6	7.42	1.99

Domain counts for lysine in the pure $[C_2C_1Im][CH_3CO_2]$ system average 11.58 (maximum 40), dropping to an average of 9.72 as the temperature is increased from 300K to 400K. For the $[C_2C_1Im][CH_3CO_2]/H_2O$ systems, the domain count average is highly similar being 11.29 at 300K compared to 9.84 at 400K. This similarity is unsurprising, given the small quantity of water added. The domain count average values are still not close to 40, which would indicate complete dissolution and complete surrounding of the lysine by IL ions only, nevertheless these values are higher than those observed for $[C_2C_1Im][Cl]$. This indicates again that $[C_2C_1Im][CH_3CO_2]$ is better at solvating the positively charged lysine molecule as compared to the ionic liquid containing the chloride ion. The other domain count values are shown in Table 5.4.1 and indicate that the waters molecules are dispersed relative to one another and that the $[C_2C_1Im]$ cations are now more dispersed in the domain analysis due to the larger size of the acetate anions.

The domain analysis for lysine in both the chloride and acetate-based ionic liquids is reflective of the experimental solubilities, which are reproduced in Table 5.4.2. In acetate-based ionic liquid, the domain counts for lysine are significantly higher than for the chloride-based ionic liquid, mirroring the much poorer solubility seen experimentally in the chloride-based IL. Here, the change in anion, including ability to hydrogen-bond with the substrate, seems to play a pivotal role in solubility. Clearly, it would be interesting the impact of different cations that may interact more poorly with the carboxylate on solubility. Here, the results presented suggest that temperature may play an overriding role, as the cationic interactions seem more greatly perturbed by temperature.

Table 5.4. 2: Experimental Solubility values for L-Lysine and L-Aspartic acid in [C_2C_1Im][Cl] and [C_2C_1Im][CH_3CO_2] ILs.[242]

Ionic liquids	L-Arginine	L-Lysine
		
	Solubility (wt%)	
[C_2C_1Im][Cl]	10	10
[C_2C_1Im][CH_3CO_2]	70	90

In order to aid to the molecular insight of the interactions, the self-diffusion coefficients were measured for all lysine systems in [C_2C_1Im][CH_3CO_2] IL and the values are shown in Table 5.4.3. The self-diffusion coefficients were recalculated for lysine systems in [C_2C_1Im][CH_3CO_2] IL as was done for [C_2C_1Im][Cl] IL and the D values were extracted from the first 50 ps of the simulation. The D values from entire 100ns simulation are shown in Appendix A.16.5. The replotted MSD curves and log-log plots are shown in Appendix A.13.1.

Table 5.4. 3: Self-diffusion coefficients calculated from MD simulations for lysine systems

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s	
		T=300K	T=400K
		5:1	5:1
lysine-charged	lysine	0.77	4.0
[C_2C_1Im][CH_3CO_2]	[C_2C_1Im]	0.96	7.51
	[CH_3CO_2]	0.87	5.32
lysine-charged	lysine	1.90	2.21
[C_2C_1Im][CH_3CO_2]-wat	[C_2C_1Im]	2.53	6.11
	[CH_3CO_2]	2.25	3.13
	water	1.60	4.40

The similar trend for D values was observed for lysine systems in [C_2C_1Im][CH_3CO_2] IL where extracting values from the start of simulation gave apparent higher self-diffusion coefficients. The slopes of re-plotted log(t) *vs* log(MSD) graph showed that the system is close to being in true diffusive regime as the β(slope) value increased to 0.65 for lysine. The self-diffusion coefficients increase as a function of temperature. Comparison of D values for each molecule in two IL systems reveal that they are lower in [C_2C_1Im][CH_3CO_2] IL than compared to [C_2C_1Im][Cl] IL. This decrease in mobility could be explained on the basis that both IL cations and anions are able to interact well

with lysine in $[C_2C_1Im][CH_3CO_2]$ IL. The lower diffusion coefficients are indicative of stronger interactions in the system.[197] Their quite similar coefficient values might also show that the moieties are moving in the form of aggregates. The presence of water decreases the diffusion coefficients values for lysine as well as IL ions at 400K, suggesting that water is able to affect the already established interactions are playing an important role here. The overall comparison of lysine systems in two ILs and also keeping in view the experimental solubility values suggest that $[C_2C_1Im][CH_3CO_2]$ IL is better at solvating the lysine molecules.

5.5 Conclusions

The interactions of positively charged amino acids, lysine and arginine were explored with $[C_2C_1Im][Cl]$ IL at varying concentration ratios of IL:AA (200:1 and 5:1, respectively) and at 300 K. To analyse the resulting interactions, molecular dynamics simulations proved effective in providing an insight into the molecular detail of the systems. Amino acids in their charged, zwitterionic forms provided physiologically-relevant precursor models for peptides made from these monomers. Spatial distribution functions, extracted using TRAVIS to visualise various organisations, were quantitatively supported with radial distribution functions that are essential to explain the strength of the organisation of various atomic sites as a function of distance. These functions reflected the relative strengths of the interactions across the whole simulation time.

The analyses revealed that at 300 K temperature, the two positively-charged amino acids had different impacts on the structuring of $[C_2C_1Im][Cl]$ solvent. The IL cation-anion strengths were found to decrease as a function of increase in the concentration ratio for lysine, however for arginine a slight increase in IL-IL interaction was observed. The anions were found to be more dominantly occupying the regions around the aromatic hydrogens, which is in line with the previous studies,[210] and cation-cation interactions were found above and below the plane of the imidazole ring, consistent with established models for π-π and cation-π interactions. Electrostatic interactions between the anion and the positively-charged amino acids were prominent, with more dense regions around the ammonium ions. Cations were able to occupy the dominant positions around the carboxylate groups.

Higher strengths of interactions were noticed for the lysine with both the cation and anion relative to arginine, and were not much affected by increasing concentration. A key observation was an increase in the aggregation of arginine molecules relative to concentration that suggest a resultant decrease in its interactions with ionic liquid anions with either little or no impact on the cationic interactions. For lysine, in contrast, a slight reduction in lysine-chloride interaction and a major reduction in lysine-cation interactions was noticed. The cations mostly surround the carboxylate region and the direction of interaction was found to be towards the acidic proton.[216] These observations suggest that, in addition to chloride, the cation also plays an important role by interacting with the oppositely charged positions on amino acids and this would have particular relevance in proteins and other polymer applications.

To analyse the role of the change in the nature of anion, and to develop an understanding of the impact that lysine can have in different ILs, $[C_2C_1Im][CH_3CO_2]$ was chosen for the study because of its known dissolution properties for biomaterials such as cellulose. The available experimental data[242] also suggested higher solubilities of lysine in $[C_2C_1Im][CH_3CO_2]$ that further highlighted the need to develop a comparison between these two ionic liquids. Two concentration ratios (IL:AA 200:1 and 5:1, respectively) and temperature (300 K and 400 K) sets were explored to provide further insight into the organisational tendency of ILs with lysine with varying parameters. The cation-anion interactions observed in $[C_2C_1Im][CH_3CO_2]$ were similar to those found in $[C_2C_1Im][Cl]$, with the higher density regions of anions around the ring protons and the trend of a decrease in interactions with concentration at 300 K. These observations are similar to the lysine system in $[C_2C_1Im][Cl]$, although the organisation within the chloride-based IL is stronger, with more concentrated regions of cation and anion around central cation relative to $[C_2C_1Im][CH_3CO_2]$. The distribution of the cation is independent of the change in nature of the anion and a similar pattern can be seen for both ILs. The concentration effect has more impact in $[C_2C_1Im][Cl]$, where the cation organisation around lysine atomic sites decreases with the increase in concentration and shows a stronger cationic association with lysine molecule when the acetate anion is present. Inter-amino acid interactions are not affected for lysine as the temperature is increased in pure $[C_2C_1Im][CH_3CO_2]$ in contrast to pure $[C_2C_1Im][Cl]$. Lysine-lysine interactions are quite dispersed and imply that $[C_2C_1Im][CH_3CO_2]$ is able to solvate

lysine molecules better than [C_2C_1Im][Cl], as indicated by the corresponding RDFs and the domain analysis.

Nano-structuring is a characteristic feature in ILs and they are found to be segregated into polar and non-polar domains and have a unique balance of different nature of interactions ranging from columbic, electrostatic and hydrogen bonding to the tail-tail binding of cations. The structuring of the ionic liquid ions around various amino acid residues in proteins have been identified to have varying stabilising or destabilising effects.[243] It is therefore imperative to understand the organisation around the smaller amino acid solutes in order to understand the effect realised in macroscopic processes, such as protein stabilisation, or destabilisation, or dissolution, of these ionic liquids and solutes.

The presence of water in ILs can greatly influence the quality of ILs and affects many important properties including viscosity, polarity, and conductivity, and can also influence the solubility of other substances in RTILs. The imidazolium-based ILs show varied hydrophilicities, however the commonly known hydrophobic ILs are actually hygroscopic and can absorb a certain amount of water from the air. The addition of water was studied by adding water based on Karl Fischer measurements and this water was found to influence the structuring through establishing more pronounced bonds with both the amino acids and anions. Water addition in both ILs caused an increase in the mobility of the ILs ions and amino acid molecules and a more pronounced effect was observed for lysine in [C_2C_1Im][Cl] IL at a higher temperature.

6

Interactions and organisation of negatively-charged amino acids in imidazolium ionic liquids

6.1 Introduction

Following on from the investigation in Chapter 5, where amino acids with positively charged side chains were examined, this chapter examines aspartate (Asp) and glutamate (Glu), which are both negatively charged amino acids at physiological pH. The zwitterionic forms of aspartate and glutamate are both similar, possessing the same carboxylate functional group on the side chain, but differ in respect to the length of side chain attached to the alpha-carbon. These two amino acids are abundant in wool and hair keratin, as identified by many studies[38, 203, 244] and with the choice of these specific amino acids, the intention was to extend the range of different amino acids that would be expected to differ in respect to their ionic interactions. The negative charge of these residues contrasts to the positive charges of lysine and arginine, and the interactions with the side chains, although not with the backbone, should thus be altered.

Modification of protein surface charge has been explored in relation to the enzyme's stability and activity in IL environments. Zhang *et al.*,[245] investigated the tolerance of family 5 structurally similar cellulases from multiple organisms, having varying surface properties to $[C_2C_1Im][CH_3CO_2]$ IL. The enzymes showed incomparable trend of tolerance to the increasing concentrations of IL and the activity for cellulase from the halophilic organism *Halorhabdus utahensis*, showed remarkable activity in ILs. This was related to the enzyme surface having high proportion of negatively charged surface residues. Recent studies have also indicated that chemical modifications that cause reduction in the ratio of positive to negative charged residues improves enzyme activity in ILs. Several model enzymes in the ILs $[C_4C_1Im][Cl]$ or 1-ethyl-3-

methylimidazolium ethyl sulfate ([C_2C_1Im][EtSO4]) in water showed increased activity retention by the reduction of this ratio specifically.[246]

Experimental investigation of solvation of aspartic acid and glutamic acid in [C_2C_1Im][Cl] and [C_2C_1Im][CH_3CO_2] ILs revealed that both the ILs are equally suitable for the dissolution of glutamic acid with a solubility value of 70% (wt/wt), wherease the IL [C_2C_1Im][CH_3CO_2] is better at solvating aspartic acid relative to [C_2C_1Im][Cl], with solubilities of 80% and 20% (wt/wt), respectivly.[242] In the same vein as the previous chapter, both ILs were chosen, and the interactions between aspartate and glutamate compared at 300 K. In addition, experimental solubilities and observation of gelation-like behaviour has been observed for these amino acids in the studied ionic liquids,[242] providing a tangible reference for the simulation observations. The structure and nano-structure of many pure ILs and their mixtures with water have been studied extensively.[52, 53, 136] This study extends this interaction to the more complex situation to where a solute in the form of an amino acid is present and looks at the role water plays in impacting the IL-AA organisation.

6.2 Outline of simulation systems

A summary of the simulations that were carried out for these two negatively charged amino acids is given in Table 6.1. Glutamate and aspartate low concentration and high concentration systems were compared at 300K in [C_2C_1Im][Cl] IL. Aspartate was selected based on its varying solubility in two ILs for further investigation and as per the positively charged amino acid simulations, three different parameters were explored, namely concentration, temperature and water content in the systems. Aspartate was simulated at both 300K and 400K in both [C_2C_1Im][Cl] and [C_2C_1Im][CH_3CO_2] ILs. The IL to amino acid concentration ratios of IL:AA 200:1 and IL:AA 5:1 were chosen to match low concentration systems and experimental data made available to the project. Simulation boxes were prepared in an identical manner as to those for lysine and arginine (Chapter 5), with the exception that these amino acids were substituted by the appropriate negatively charged amino acid. Charge balancing was achieved with the addition of IL counterions, with the exception that due to the overall negative charge of the amino acid, the scaled-charge cation $[C_2C_1Im]^+$ was used to achieve a neutral charge in the simulation box. Table 6.1 also highlight the number

of molecules and their ratios used in each system. The simulation procedure for each box included two energy minimisation steps keeping the solute fixed in first stage. The simulations were run following the procedure outlined in chapter 3. The example input files are shown in appendix for aspartate in Appendix A.18.

Table 6. 1: List of Aspartate and Glutamate simulations

IL	Amino acid	Number of IL ions and AA molecules	Number of Water molecules added	Concentration ratio IL:AA and Temperature			
$[C_2C_1Im][Cl]$	Aspartate	For (IL:AA)= (200:1)	n/a	200:1 (300K)	200:1 (400K)	5:1 (300K)	5:1 (400K)
	Glutamate	200 IL ion pairs one AA molecule	n/a	200:1 (300K)		5:1 (300K)	
$[C_2C_1Im][CH_3CO_2]$	Aspartate		n/a	200:1 (300K)	200:1 (400K)	5:1 (300K)	5:1 (400K)
$[C_2C_1Im][Cl]$/WAT	Aspartate	For (IL:AA)=(5:1)	26	5:1 (300K)	5:1 (400K)		
$[C_2C_1Im][CH_3CO_2]$/ WAT	Aspartate	200 IL ion pairs 40 AA molecules	2	5:1 (300K)	5:1 (400K)		

6.2.1 QM Calculations for Amino acid and Ionic Liquid Structures and antechamber

Aspartate and glutamate structures were prepared as detailed in Chapter 3 methodology. Both amino acids were constructed with a negatively charged side chain and a zwitterionic backbone. In contrast to the positively charged amino acids, aspartate and glutamate structure optimisation and RESP charge derivation was carried out with implicit water in order to keep the zwitterionic form intact. The PCM water solvent model in Gaussian was used with no further parameterisation. Additional preparation of parameterisation files was carried out as for the lysine and arginine (Chapter 5).

Further details, including commands for preparation and relevant input files can be found in the Appendix A.18. Atom numbering of glutamate, aspartate, [C_2C_1Im] cation, acetate anion and water molecule are shown in Fig 6.1.

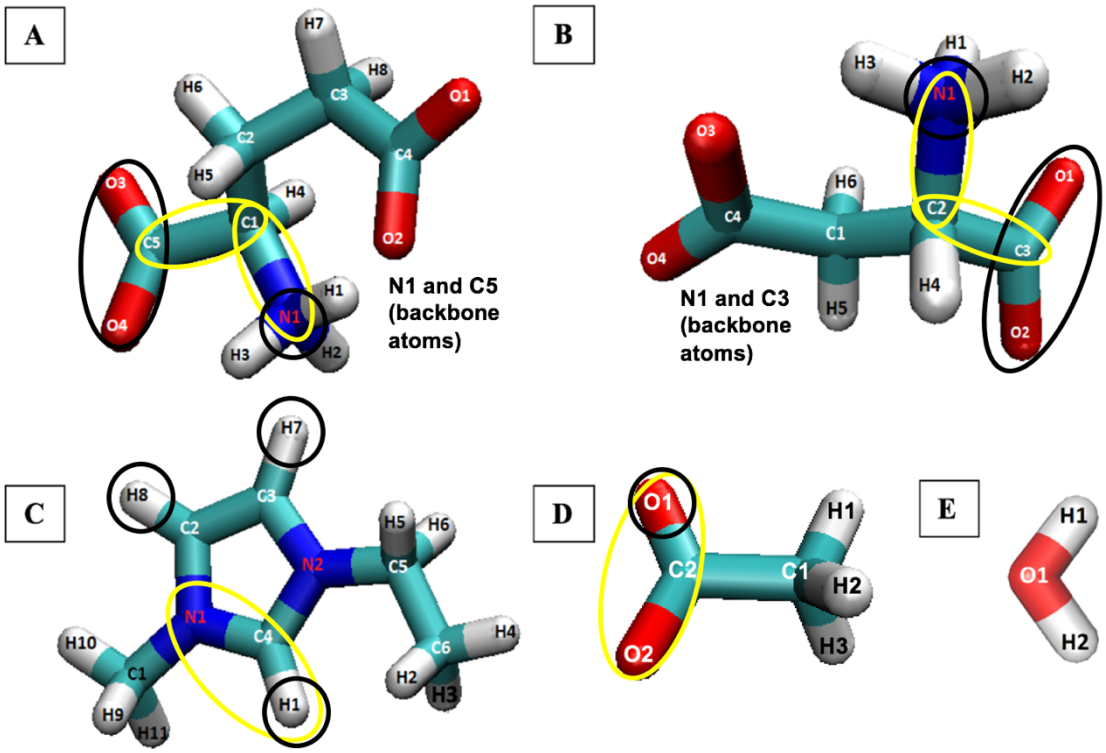


Figure 6.1: Optimised structures of (A) glutamate, (B) aspartate, (C) [C_2C_1Im] cation, (D) acetate anion and (E) water molecule with atom numbering generated by antechamber.[180] Atoms encircled yellow are in each molecule show the atoms that are used as reference to create respective SDFs around these molecules. Backbone atoms are labelled in aspartate and glutamate. Atoms encircled black are the ones that are used to discuss results in RDFs analysis.

6.3 Results and discussion

To infer the molecular origin of the interactions between aspartate and glutamate, and the pure either [C_2C_1Im][Cl] or [C_2C_1Im][CH_3CO_2] IL, and for each amino acid in IL aqueous solution, spatial distribution functions, radial distribution functions, diffusion coefficients and domain analyses were calculated from MD simulations for all the systems of these amino acids as outlined in Table 6.1. SDFs provide a three-dimensional visualisation of these RDFs that can provide a rapid qualitative assessment. All other analyses, such as domain analyses and calculations for self-diffusion coefficients, are done as described for the positively charged amino acids.

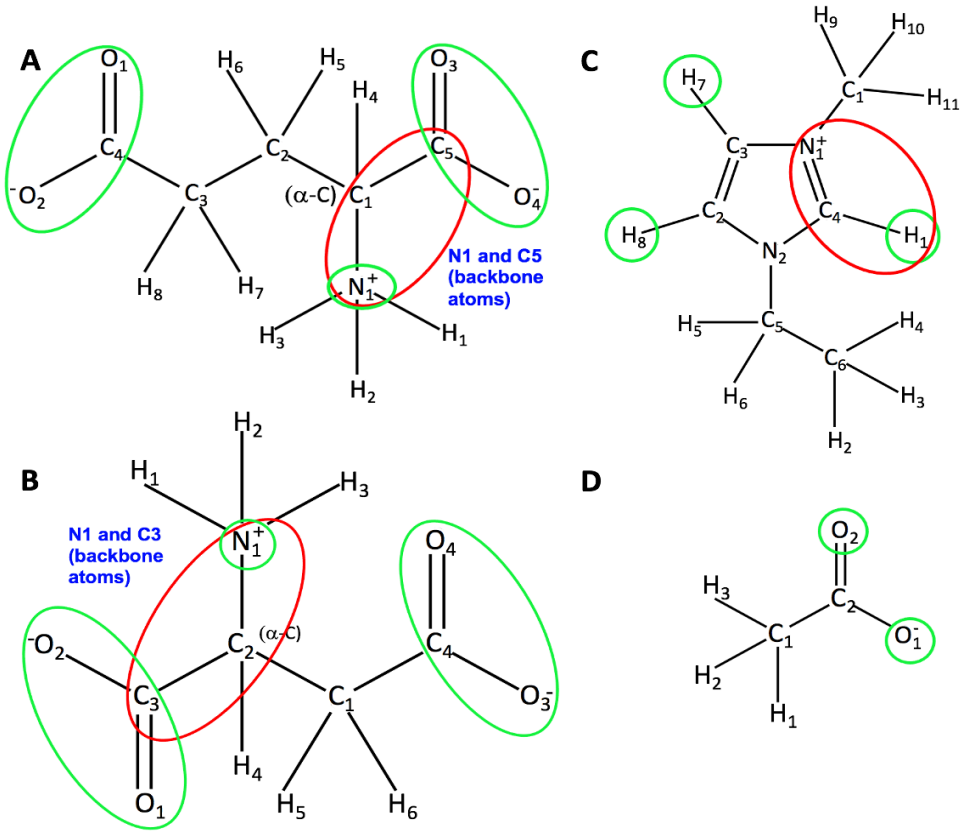


Figure 6.2: Chemdraw structures of (A) negatively charged aspartate molecule with zwitter-ion backbone, (B) negatively charged glutamate molecule with zwitter-ion backbone, (C) $[C_2C_1Im]^+$ cation and (D) $[CH_3CO_2]^-$ anion. Atoms encircled red in the molecules are those used as reference to create respective SDFs around the molecule. Backbone atoms are labelled in aspartate and glutamate. Atoms encircled green are the ones that are used to discuss results in RDFs analysis.

6.3.1 Influence of solutes on ionic liquid microstructure

Spatial distribution functions were used to visualise the impact of solute incorporation into the chloride-containing ionic liquid and demonstrate a reverse trend to that observed for the positively-charged amino acids lysine and arginine. For aspartate and glutamate, increasing the concentration ratio (IL:AA) from a 200:1 to 200:40 impacts on the magnitude of the organisation of the anion, relative to the cation, effectively weakening the IL ion-ion interaction with increasing concentration (Figure 6.3.1). This contrasts with the impact of positively-charged amino acids, where the cation-cation interaction was most substantially affected. This trend is consistent with a replacement of IL ion interaction with the corresponding charged amino acid with increasing concentration, as might be expected, in effect displacing these ions, consistent with the increased probability of such an interaction. This also suggests, however, that the predominant important interaction is with the side-chain charged residue, rather than

an influence from the zwitterionic backbone, despite these charges clearly interacting with the IL. This distribution is quite similar to what has been reported for other ionic liquids, [164, 189, 191, 210, 212, 213, 216] especially for the low concentration system where there is only one solute molecule.

The other important feature to note is that the shorter-chain aspartate appears to exert a greater impact on the depletion of cation-anion interaction, relative to glutamate. This may indicate some influence of the proximity of the charged backbone to the functionality of interest creating a stronger level of molecular organisation, potentially impacted by a cooperative shell-shell interaction that is less pronounced if the side chain is further in the bulk solution.

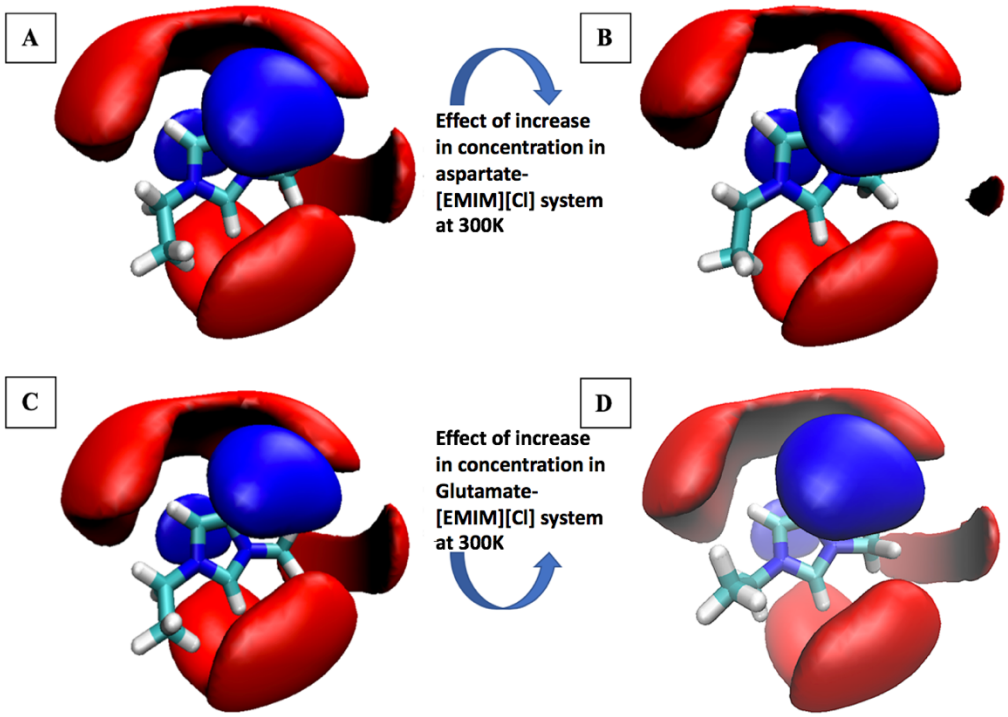


Figure 6.3.1:Spatial distribution functions for cation-cation and cation-anion interactions around the central $[C_2C_1Im]^+$ cation for aspartate (A) low concentration *vs* (B) high concentration system and for glutamate (C) low concentration *vs* (D) high-concentration system in $[C_2C_1Im][Cl]$ IL at 300K in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}

RDFs showing the distribution of chloride anions around the central cation are shown in appendix A.9.2 for aspartate and glutamate and show that the highest peak occurs with the H1 atom of $[C_2C_1Im]$ cation. The other concentrated region for chloride ions is around H7 and H8 atoms. RDFs for the mass-centred moieties are shown in appendix A.19.2. As seen in the SDFs, the first solvation shell of chloride ion is present mostly around aromatic hydrogens of the imidazolium ring, with the strength of interaction is

highest for the acidic proton H1. The electrostatic interaction between cation and anion is dominant and is not much affected as the concentration of aspartate amino acid increases in the system.

The trend in $[C_2C_1Im][CH_3CO_2]$, shown for aspartate (Figure 6.3.2), is fairly similar to that observed with $[C_2C_1Im][Cl]$, in that an increase in concentration of the aspartate results in a decrease, most substantially, in the anion interactions with the aromatic ring carbons of $[C_2C_1Im]$, and with relatively little impact on the cation-cation interactions. The acetate interactions with the $[C_2C_1Im]$ cation are weaker however, than with chloride, suggesting that these ions may be more available for interacting with solute partners. This contrasts with the cation-cation interactions, which in lysine are virtually absent in comparison, even at low concentrations.

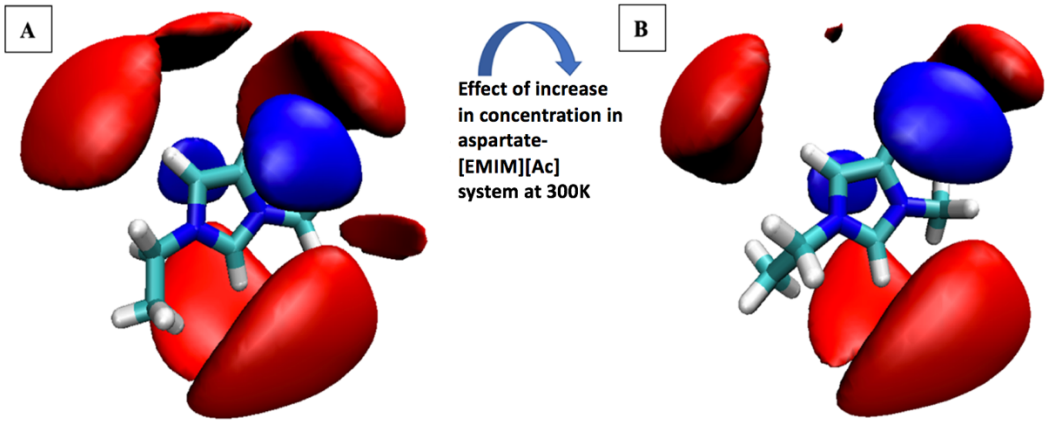


Figure 6.3. 2: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for aspartate in (A) low concentration system *vs* (B) high concentration system in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.

The SDFs for cation and anion distribution around acetate anion are shown in the appendix A.22.5, that also suggest the cations dominantly occupy the regions around the carboxylate groups. The concentration of these cations decreases as the concentration ratio of aspartate molecule is increased in the system, presumably as more aspartate anions compete for cation binding. RDFs around the different atomic sites, along with the overall mass centred cations, are shown in the appendix A.22. The partial RDFs show that the ring hydrogen atoms are able to interact with the anion, presumably through the formation of CH---O hydrogen bonds, as the probability for these atoms is the highest for H1 atoms and then for H7 and H8 atoms at almost the same distances from the central anion.

6.3.2 Effect of Concentration ratio of aspartate and glutamate to [C_2C_1Im][Cl]on microstructural organisation

6.3.2.1 Interactions and organisation of negatively-charged amino acids aspartate and glutamate in [C_2C_1Im][Cl]

The SDFs for aspartate and glutamate in [C_2C_1Im][Cl] are shown below in figure 6.3.3, with corresponding overall mass-centred radial distribution functions in appendix figure A.19.2. The RDFs and SDFs both show the dominance of [C_2C_1Im] cations around the central amino acid molecules. The anion distribution around the ammonium decreases for both aspartate and glutamate, with a larger impact on glutamate, and the anions are slightly shifted away from backbone nitrogen region. It also reinforces the picture seen with lysine and arginine, that side chain has the potential to impact on organisation of the backbone, even where the sidechain functionality/charge is identical, as in this case. The inter-amino acid interaction is evident for both aspartate (for glutamate occurs at iso-value 6 nm^{-3}) under the higher-concentration condition, and it is the increase of these interactions that may be the cause of the shell expansion around the ammonium, particularly given their placement in this vicinity. The RDFs confirm that these amino acid-amino acid interactions are more intense for the aspartate molecules. This contrasts with the lack of these non-IL interactions seen at lower temperatures for the positively-charged amino acids, and also suggests that the carboxylate side chains are effective competitors in the interaction with ammonium.

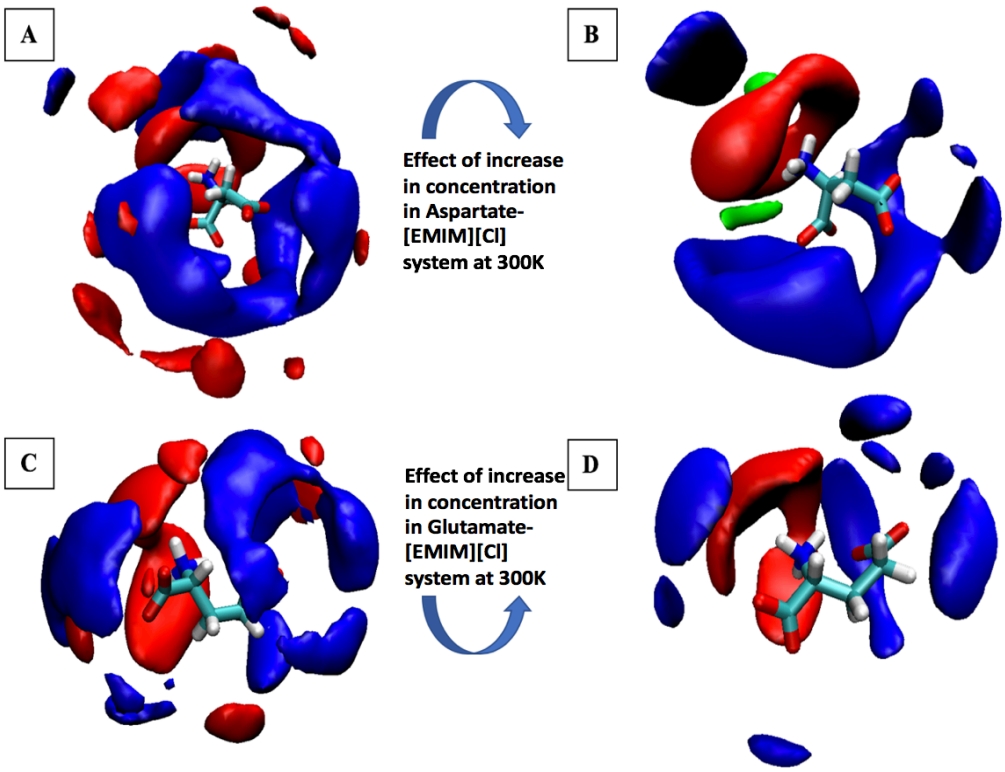


Figure 6.3. 3: Spatial distribution functions around the central (A,B) aspartate low and high concentration systems respectively and (C,D) glutamate low and high concentration systems respectively in $[C_2C_1Im][Cl]$ IL at 300K in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= aspartate at iso-value=10 nm^{-3}

Confirmation and quantification of the trends seen in the SDF was done by examining the key atomic sites identified in Figure 6.2 for RDFs for aspartate and glutamate. The atom numberings of all the molecules and ionic liquid ions are shown in figures 6.1 and 6.2. The carboxyl group, backbone nitrogen and the terminal carboxyl group were chosen for constructing the RDFs for the ring proton H1 on cation and the chloride anions. Figure 6.3.4 shows the distribution of cation atomic sites around the backbone nitrogen and carboxyl groups for the aspartate molecule. The organisation pattern in terms of distance is quite similar to that observed for lysine, as the difference lies at the terminal end of the chain, so it is expected that a similar organisation should be seen. The strength of interactions with cation is stronger as compared to the lysine molecule.

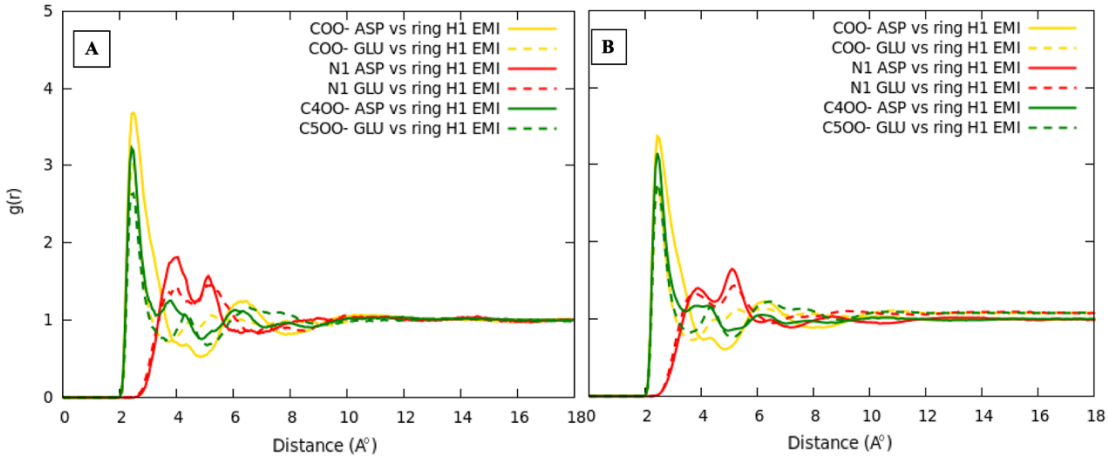


Figure 6.3. 4: Site-site radial distribution functions for $[C_2C_1Im][Cl]$:[AA] at 300K for acidic H1 on $[C_2C_1Im]^+$ cation around N1 backbone nitrogen atom, COO^- carboxyl group on α-carbon and side-chain carboxyl group for (A) low concentration (200:1) *vs* (B) high concentration (5:1) system determined for aspartate (solid lines) and glutamate (dashed lines)) by MD simulation

The first peak for ring hydrogen H1 around ammonium, is around 4 Å, indicating a second-shell interaction, while both the carboxylate groups denoted by COO^- and $C4OO^-$ (aspartate) or $C5OO^-$ (glutamate) show quite prominent interactions with the oppositely charged $[C_2C_1Im]^+$ cation, the strongest being for the backbone COO^- group. The comparison for distribution of cation atomic site H1 around the carboxylate groups, both the terminal and the backbone for aspartate and glutamate. Though the organisation pattern is similar, the peaks for aspartate are more intense and a little broader, compared to those for glutamate. For example, the g(r) for the $[C_2C_1Im]$ H1 atom with aspartate is around 3.6 decreasing to nearly 3.1, whereas for glutamate it is around 3 decreasing to 2.8, on going from low to high concentration. This confirms the density of H1 atoms is greater around the aspartate backbone carboxylate group than that of glutamate. A similar trend, although slightly lower magnitude, is noted for the side-chain carboxylates.

Aspartate and glutamate being charged and polar, prefer to be on the surface of proteins, exposed to an aqueous environment. When buried within the protein, aspartates and glutamates are frequently involved in salt-bridges, where they pair with a positively charged amino acid to create stabilising hydrogen bonds, that can be important for protein stability. Both can be quite frequently involved in protein active or binding sites and the negative charge means that they can interact with positively charged non-protein atoms, such as cations like zinc.[247] Aspartate has a shorter side-chain than glutamate that makes it slightly more rigid within protein structures. This rigidity may

contribute to the attraction of a greater density of cations, relative to glutamate, by means of fewer steric interactions due to less rotation (for example, the same effect that is seen with lower interactions of ethyl with ions than for methyl groups). A further contribution may be the reinforcement of short-range organisation possible when the side-chain is closer to other polar parts of the molecule that have attracted a density of ions. This has the potential to generate cooperative effects as a shell of anions can help to further organise the cations that are have been localised by the charged side-chain.

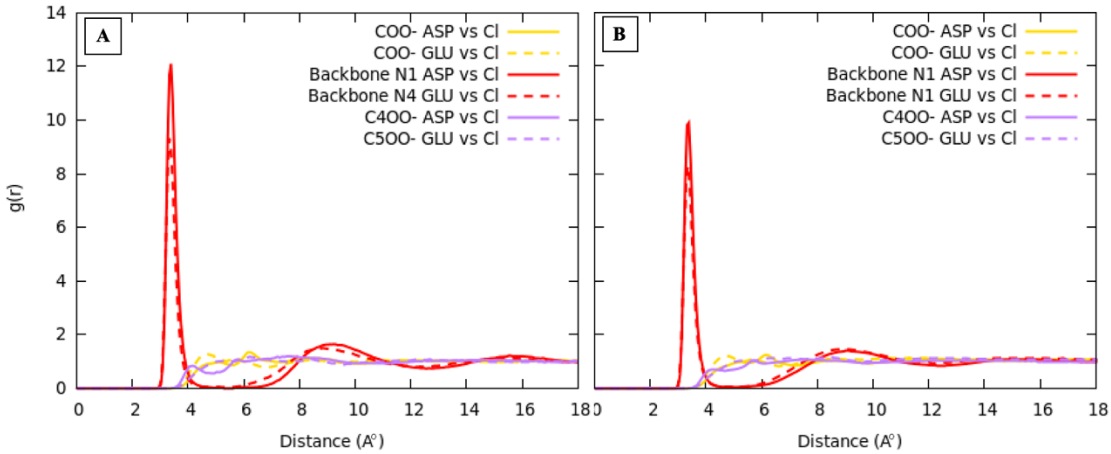


Figure 6.3. 5: Comparison of site-site radial distribution functions for $[C_2C_1Im][Cl]$:AA for (A) low concentration (200:1) *vs* (B) high concentration (5:1) asp (solid lines) *vs* glu (dashed lines). (A) and (B) both show chloride anion distribution around backbone nitrogen and carboxyl groups determined by MD simulation

The distribution of chloride anions around aspartate and glutamate atomic sites are shown in the figure 6.3.5, where the organisation for both the low concentration and high concentration systems at 300K is compared. The distribution for anions around the backbone nitrogen is quite similar in terms of peak distances, the anions form the closest shell around ammonium for aspartate molecule at a distance of 3 Å. However, the strengths for aspartate are high relative to glutamate, i.e. g(r) value of 12 relative to 9 at low concentration and these interactions decrease slightly at high concentration. A figure showing the distribution of cation based on RDFs is shown in appendix A.18.3

6.3.3 Effect of temperature on Aspartate-IL interactions

For the lysine molecule in $[C_2C_1Im][Cl]$ in chapter 5, the increase in temperature tended to increase the overall inter-amino acid interactions. To compare whether the same

trend is followed for the negatively charged amino acids, simulations at the higher temperature of 400K were run as outlined in the methodology section.

As shown by the blue regions in SDFs (Figure 6.3.6), cations dominate the region around most sites of the aspartate molecule. The SDFs also confirm that, in the high concentration condition, the aspartate molecules are able to interact with each other even at 300K, as shown by the regions marked in green, and these regions become quite enhanced as the temperature is raised to 400K. The density of anions around the backbone nitrogen is slightly affected as the concentration ratio is increased at 300K, whereas the shell thickness is greatly reduced but is not completely lost at high temperature. This contrasts with the situation for lysine, where inter-amino acid interactions were not observed at lower temperatures under the same conditions, however, the increase in interactions at higher temperature remains consistent.

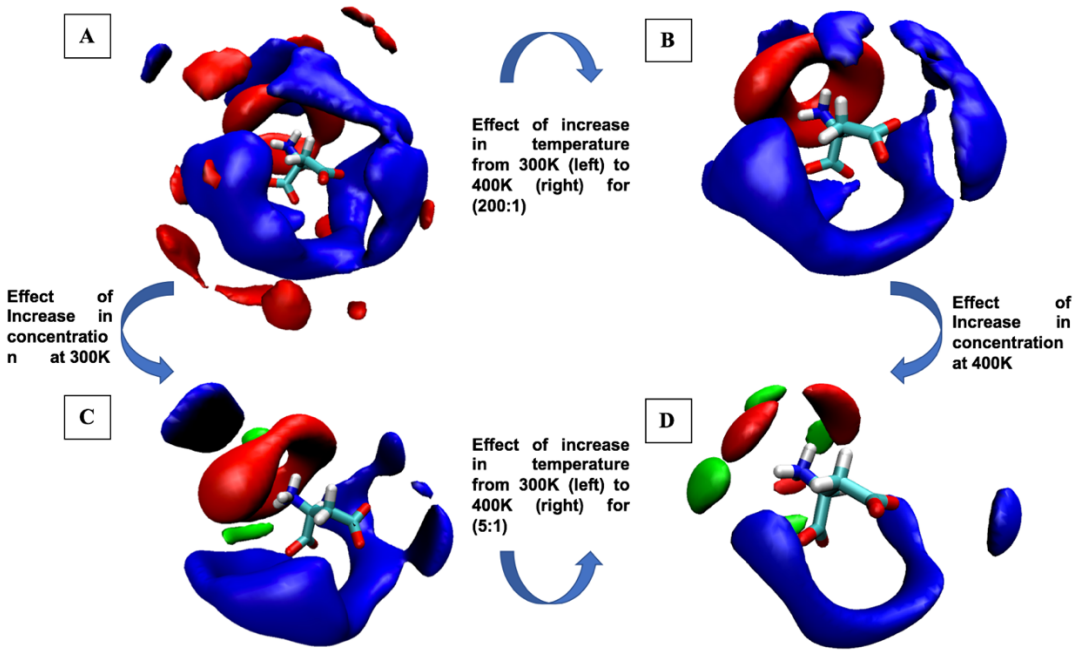


Figure 6.3. 6: Spatial distribution functions around the central aspartate molecule for (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= aspartate at iso-value=10 nm^{-3}

The effect of temperature on interactions around the $[C_2C_1Im]$ cation for the aspartate systems is shown in figure 6.3.7. The solvent-solvent interactions are affected with the temperature rise. The anions are mostly found around the aromatic regions and the density remains more or less constant, although slightly reduced, at higher temperatures, with the largest change at low concentration, where the interaction with

the [C_2C_1Im] methyl group disappears. The cation-cation organisation is reduced at higher temperature but does not completely disappear.

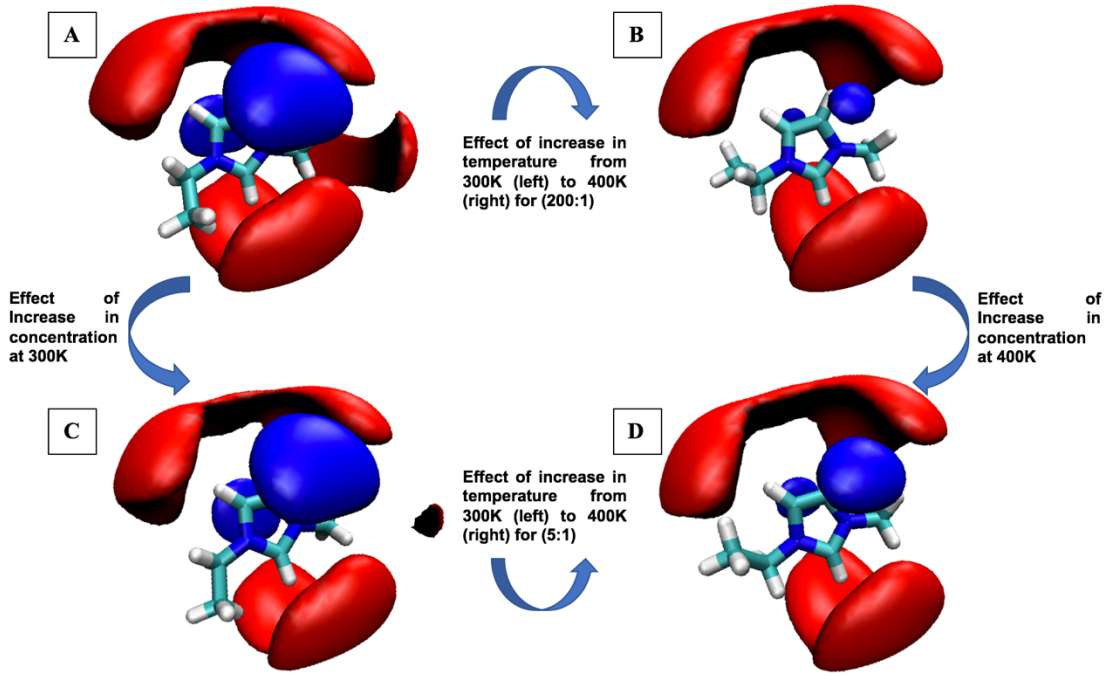


Figure 6.3. 7: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for aspartate in pure $[C_2C_1Im][Cl]$ (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][Cl]$ IL in a view the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}

Figure 6.3.8 shows the effect of temperature on interactions for both low concentration and high concentration systems of aspartate in $[C_2C_1Im][Cl]$ and compares the organisation for anion and cation atomic site around the central aspartate molecule. The solid lines in all figures shows the organisation at 300K whereas the dashed lines are for the peaks at 400K. Figure 6.3.8 panels A and B show the effect of temperature on interactions for the low concentration system and figures 6.3.8 panels C and D compare the temperature effect for high concentration systems. The organisation pattern is similar at both the temperatures, but the strength of interactions is reduced at the higher temperature, and this seems to be a non-linear effect with concentration. This is the case both for IL cation-anionic amino acid sites and, more prominently, for the chloride ion interactions with the ammonium of the aspartate where the g(r) value drops from 10 to 2 at higher temperature.

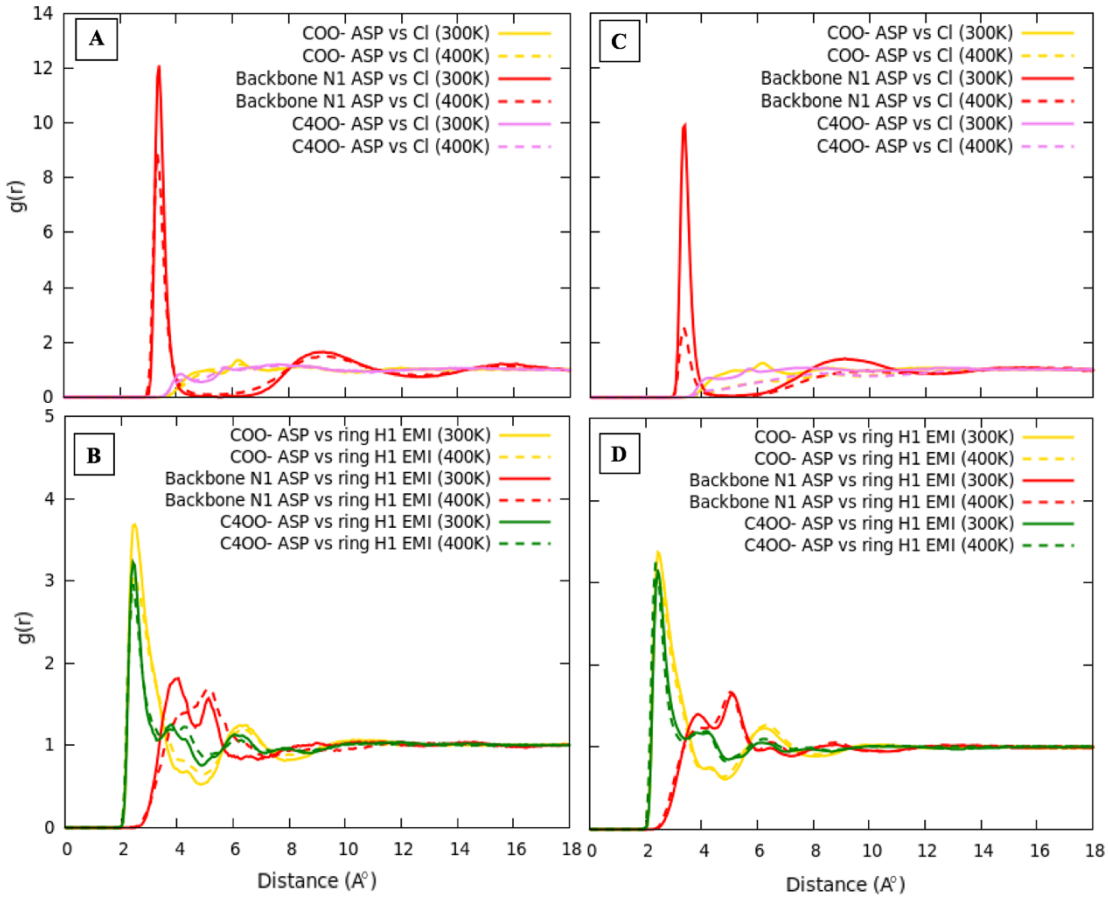


Figure 6.3. 8: Comparison radial distribution functions for $[C_2C_1Im][Cl]$:aspartate low concentration ((200:1 (left A,B)) *vs* high concentration ((5:1) (right C,D)) at 300K (solid lines) *vs* 400K (dashed lines). (A,C) show the distribution of chloride anions and (B,D) show the distribution of H1 $[C_2C_1Im]^+$ cation around backbone nitrogen, carboxyl groups determined by MD simulation

The decrease of the chloride-ammonium interaction in the high concentration systems is compensated by the significant increase in the interaction between the aspartate molecules with one another at higher temperature as shown in the figure 6.3.9, where the IL ions are compared with the organisation of aspartate molecule around the mass-centred aspartate molecule. These compensating interactions emphasise the role of temperature in mediating the complex ion-ion equilibria that are generated on amino acid dissolution.

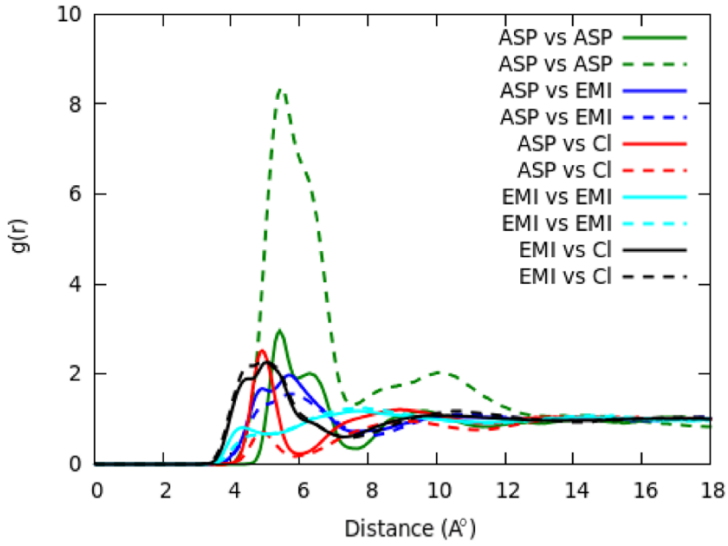


Figure 6.3. 9: Centre-of-mass radial distribution functions for aspartate in pure [C_2C_1Im][Cl] IL for high concentration systems. RDFs show overall solute-solute and solute-solvent interactions for aspartate at 300K (solid lines) and 400K (dashed lines), determined by MD simulation

6.3.4 Effect of addition of water on aspartate and [C_2C_1Im][Cl] interactions

The SDFs for these systems (Figure 6.3.10) show that the cation distribution is decreased but is still effective in solubilising the aspartate. Anion is less affective at higher temperature in IL-water mixture. The green parts show that the hydrogen bonds established between the aspartate molecules at higher temperature predominate around the backbone, and so the addition of water does not successfully break these interactions, and is further confirmed in the respective RDFs by indicating an increase in the magnitude of the aspartate-aspartate interaction as seen in figure 6.3.11. Water appears to have a more significant interaction with chloride, reducing the interaction of this anion with the aspartate molecule and is also able to impact on the cation-aspartate organisation.

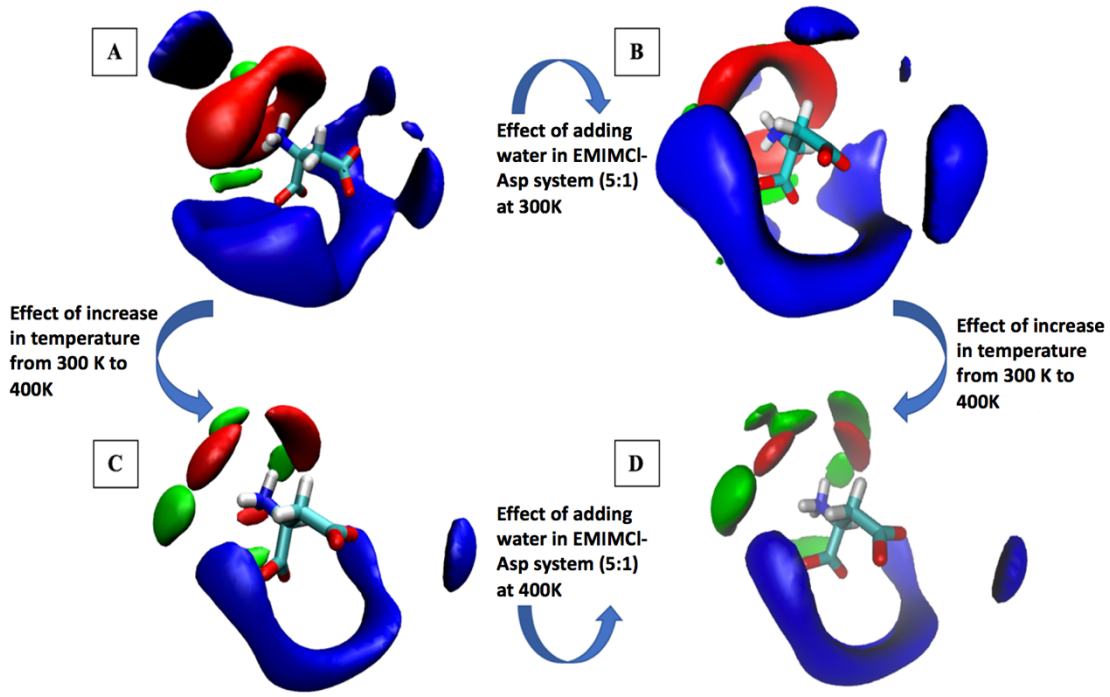


Figure 6.3. 10: Spatial distribution functions around the central aspartate molecule in (A) pure $[C_2C_1Im][Cl]$ at 300K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][Cl]$ at 400K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 400K. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= aspartate at iso-value=10 nm^{-3}.

The effect of addition of water in the simulations for aspartate molecule is further illustrated through radial distribution functions in figure 6.3.11. The presence of water drastically reduces the chloride distribution around different atomic sites on aspartate molecule is confirmed, as is evident from panels A and B at both the temperatures. Running simulations with water lowers the peak heights at 300K with even more significant effects at the higher temperature of 400K. The red lines in the graph show the interaction between aspartate and chloride anion are quite strong at lower temperature and are drastically reduced as the temperature is increased. Anions play a major role in the solubility of biomolecules in ionic liquids[248] and a drastic decrease in the aspartate-chloride anion interaction is compensated by an increase in aspartate-aspartate aggregation. Addition of water seems to enhance this affect as it can establish interactions with both the anion and aspartate. (Figure 6.3.11 & appendix A.21.2).

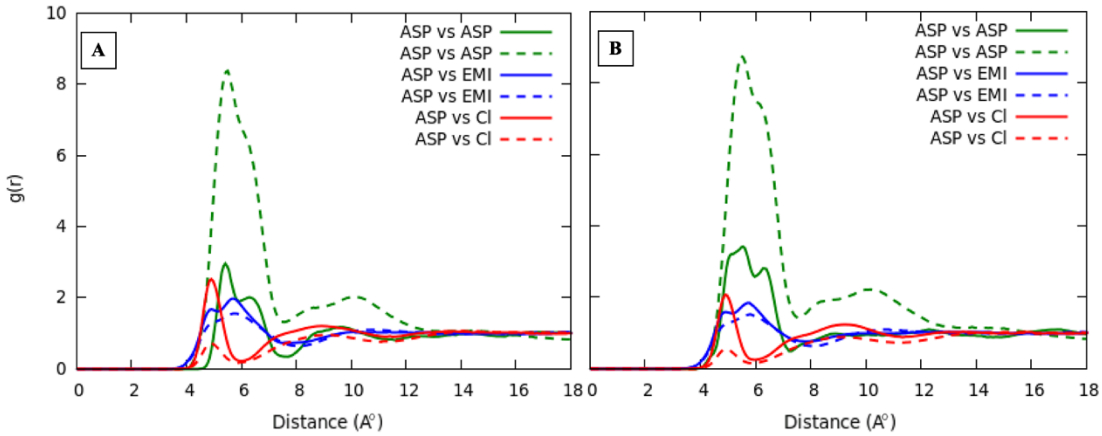


Figure 6.3. 11: Centre-of-mass radial distribution function for high concentration systems of aspartate in (A) pure [C_2C_1Im][Cl] *vs* (B) IL/water mixture at 300K (solid lines) *vs* 400K (dashed lines) determined by MD simulation

The peaks in the appendix figure A.21.2 shows the overall distribution of cations, anions and water molecules around the mass-centred aspartate molecule and show that the aspartate molecules are also able to interact with the water at the same distance as the chloride anions but the probability of finding an aspartate is quite low, relative to the probability of finding a chloride anion. Water forms two shells around the aspartate molecule and so is quite organised but these interactions are impacted more relative to chloride anion-water interactions as the temperature is raised. At both 300K and 400K, the aspartate-aspartate interaction peak intensity is not much affected, but the peaks become broader, indicating a slight rearrangement of aspartate molecules around each other. The impact of water can be further seen in the snapshots shown in figure 6.3.12, taken at end of 100 ns simulation time for each high concentration system in pure and IL/water mixture for aspartate system in [C_2C_1Im][Cl] at 300K and 400K. Here, the increased aspartate-aspartate interactions are reflected by aggregation that is visible in panel B (400 K) and most predominantly in panel D (400 K, with water).

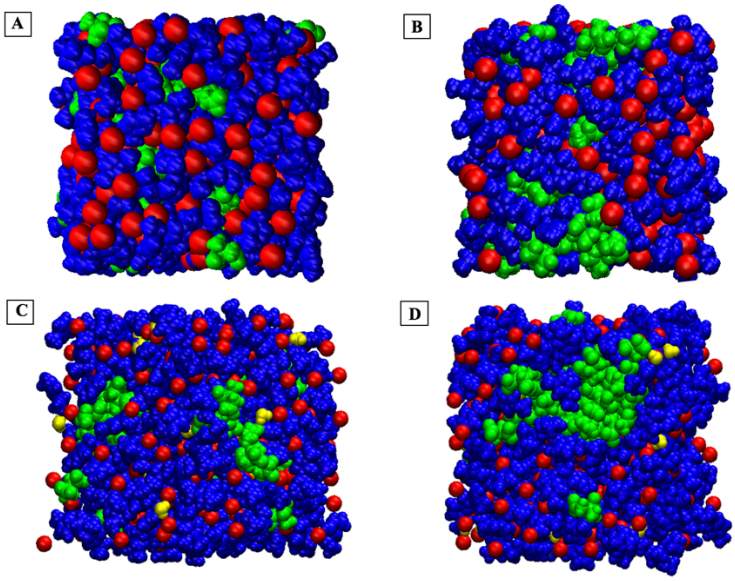


Figure 6.3. 12: Snapshots of aspartate high concentration systems in $[C_2C_1Im][Cl]$ IL after 100 ns simulation for (A,C) at 300K for pure and IL/water mix respectively and (B, D) at 400K for pure and IL/water mix respectively.

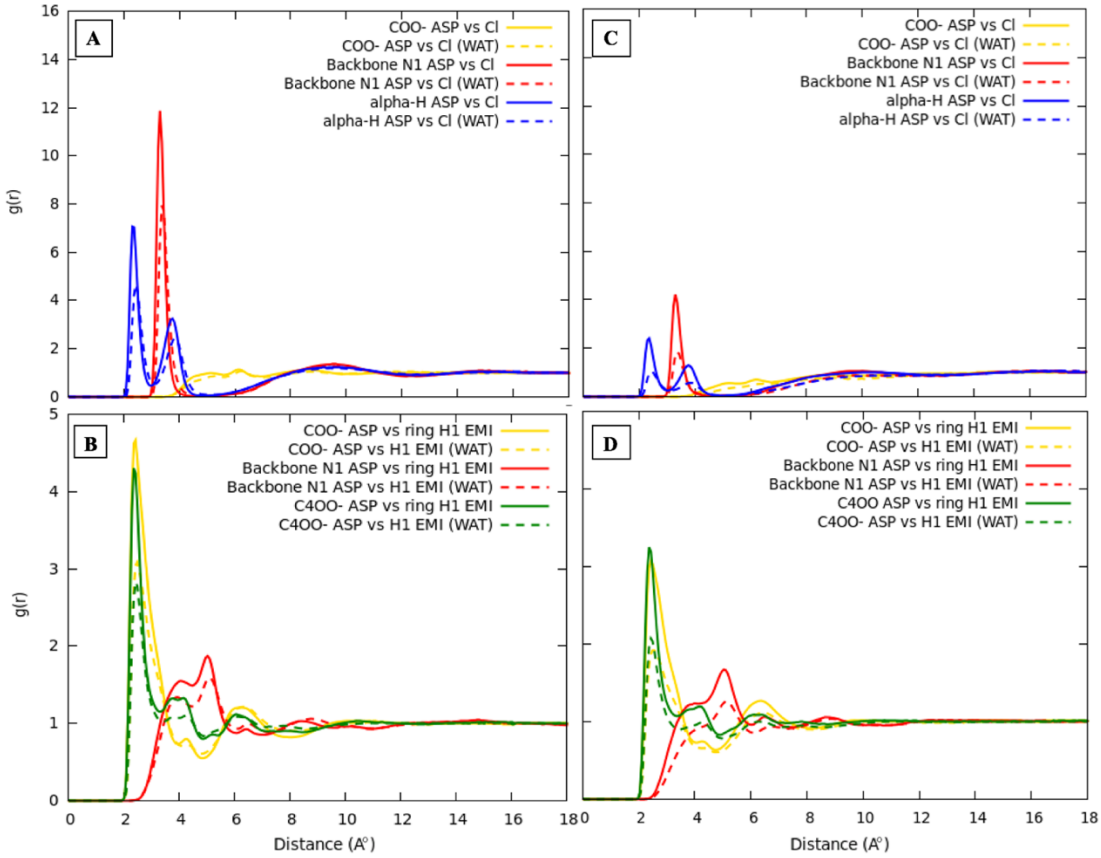


Figure 6.3. 13: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: aspartate high concentration (5:1) systems with water (dashed lines) and without water (solid lines) at 300K (A,C) *vs* 400K (B,D). (A,B) backbone nitrogen, carboxyl group and alpha-hydrogen on aspartate *vs* Chloride anion and (B,D) backbone nitrogen and carboxyl groups on aspartate *vs* H1 $[C_2C_1Im]^+$ cation determined by MD simulation

The distribution of cation H1 atoms around different atomic sites on the aspartate molecule also follow the same trend as chloride anions (figure 6.3.8) i.e., decrease in

the peak heights due to the presence of water at both temperatures but the peak heights and organisation are relatively less affected. In this instance, the cations are able to sustain the electrostatic interactions with the oppositely charged aspartate molecule. This arrangement is confirmed through SDFs in figure 6.3.10.

The impact on solvent organisation of the different conditions is shown in the figures 6.3.14. The cation-cation interactions are decreased as the water is added and is true at both temperatures.

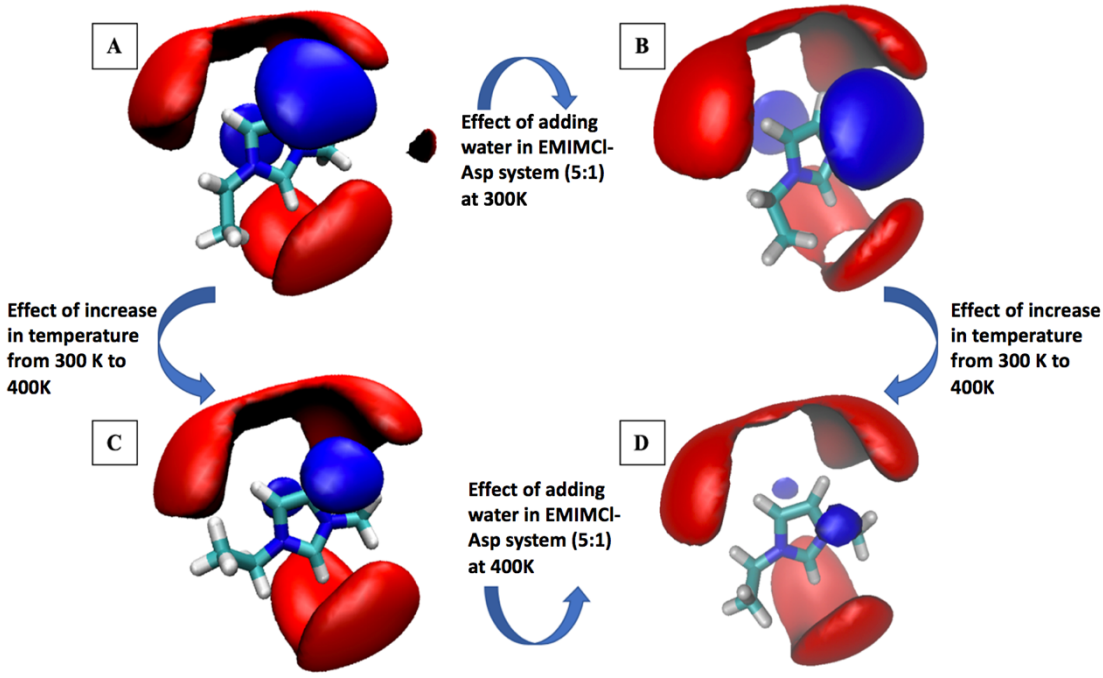


Figure 6.3. 14: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for high concentration systems of aspartate in (A) pure $[C_2C_1Im][Cl]$ at 300K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][Cl]$ at 400K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 400K in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}.

The domain counts for all the systems were measured over the entire simulation time (figure 6.3.15) and the average domain counts are shown in the Table 6.3.1. The domain count is highest for aspartate molecule for the simulation run at 300K in pure IL, which is reflected in the snapshot showing that aspartate (green) is more dispersed than for the other systems. The domain count drastically decreases to an average of 6.14 at 400K, although this is still significantly higher for the comparable average domain count for lysine, indicating a better dispersion and solubility. The addition of water also acts to lower the average domain count, with the lowest values for IL/water simulations run at 400 K of around 3, consistent with the aggregation observed in the snapshots. The water

in these simulations shows significant dispersion, having a value for the average domain count close to the maximum of 26, based on the number of added water molecules.

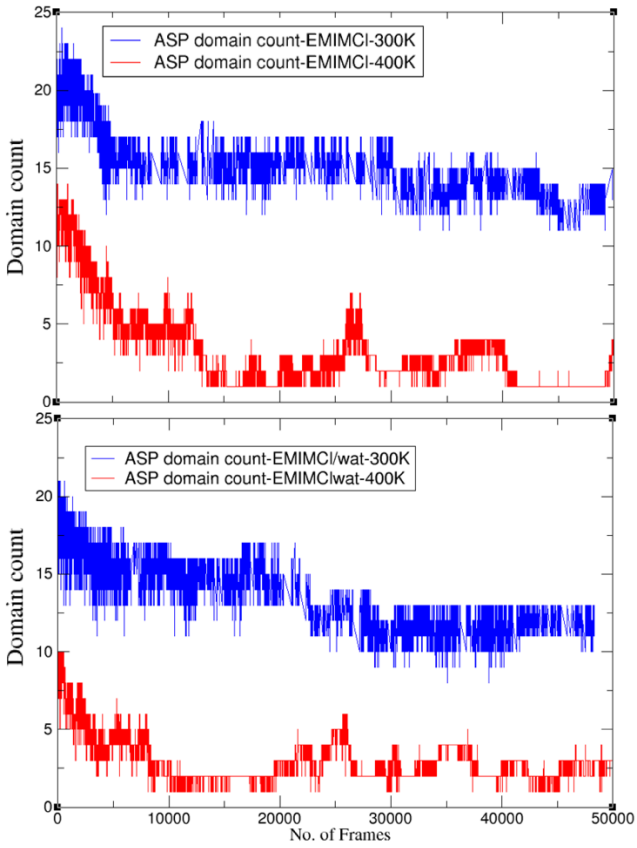


Figure 6.3. 15: Domain analysis for aspartate in (A) pure $[C_2C_1Im][Cl]$ and (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300 K (blue) and at 400 K (red).

Table 6.3. 1: Domain analysis for aspartate in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixtures

Ionic Liquid	Temperature (K)	Domain count (average)			
		aspartate (40)	Cl (200)	$[C_2C_1Im]$ (200)	O1-water (26)
$[C_2C_1Im][Cl]$	300	15.44	157.68	1.72	
$[C_2C_1Im][Cl]$	400	3.4	158.3	1.83	
$[C_2C_1Im][Cl]/H_2O$	300	13.7	160.18	1.86	25.93
$[C_2C_1Im][Cl]/H_2O$	400	3.08	156.68	2.10	24.96

Self-diffusion coefficients for all constituent species in aspartate $[C_2C_1Im][Cl]$ systems were derived by constructing MSD graphs, as shown in the appendix A.21.3, and the self-diffusion coefficient values (Table 6.3.2) were re-calculated from MSD curve slopes from the first 50 ps of simulation time, following the procedure laid out in chapter 4. The diffusion coefficients increase when the temperature is raised even for aspartate, in line with lysine systems. The individual self-diffusion coefficients also

have a similar trend, with cations diffusing faster than the chloride anions for both solute molecules, which is in line with what has been reported previously.[234] The log(MSD) *vs* log(time) graphs were plotted again and the graphs in appendix A.21, figure A.21.4 show that the system is close to being in a diffusive regime as the slope of graph is 0.6. The self-diffusion coefficients provide a qualitative comparison to explain the interaction among the species involved in the systems. The self-diffusion coefficients increase as a function of temperature. The values are calculated for $[C_2C_1Im][CH_3CO_2]$ IL in section 6.4.3.

Table 6.3. 2: Self diffusion coefficients calculated from MD simulations for aspartate systems

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s	
		T=300K	T=400K
		5:1	5:1
aspartate $[C_2C_1Im][Cl]$	aspartate	1.72	2.80
	$[C_2C_1Im]$	2.83	9.00
	CL	1.68	6.23
aspartate $[C_2C_1Im][Cl]$-wat	aspartate	1.04	3.54
	$[C_2C_1Im]$	1.93	13.14
	CL	1.15	9.57
	water	1.92	31.49

6.4 Interactions and organisation of negatively charged amino acid aspartate in $[C_2C_1Im][CH_3CO_2]$ ionic liquid

6.4.1 Effect of concentration ratio on the interaction of aspartate with $[C_2C_1Im][CH_3CO_2]$

The impact of changing concentration is reflected by the SDFs (figure 6.4.1) and the overall RDFs (figure A.22.2), which show that the anion distribution is significantly reduced with increasing concentration of aspartate. Here, there is a clear formation of shells at the isovalue of 10 nm^{-3} at 300K, indicating very strong organisation of the $[C_2C_1Im][CH_3CO_2]$ around the aspartate, where these second shell anions appear to be indirectly interacting with the aspartate molecule and are present as the resultant electrostatic force of attraction from the cations. The distribution of cations is concomitantly reduced around the carboxyl group and the backbone nitrogen at higher temperature, but not to the same degree.

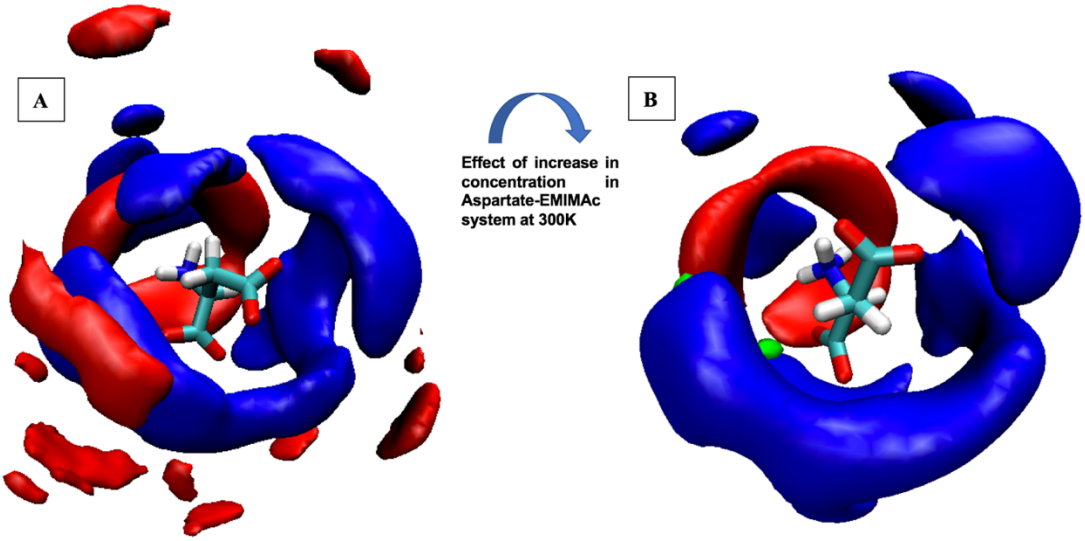


Figure 6.4. 10: Spatial distribution functions around the central aspartate molecule for (A) low *vs* (B) high concentration system, in $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$ and green = aspartate at iso-value=10 nm^{-3}.

The RDFs (appendix A.22.1) show the interaction of the two carboxyl groups and backbone nitrogen with the various atomic sites on $[C_2C_1Im]$ cation. The highest and the closest peaks occur for the H1 ring hydrogen atom. The other two ring hydrogen H7 and H8 are able to form shells at the same distance around two carboxyl groups but the probability of finding these two atoms is quite low relative to the H1 atom, and

implies some directionality to the [C_2C_1Im]-carboxylate interactions. Figure 6.4.2 compares the distribution for the acetate anion O1 atom around the backbone nitrogen, carboxyl group and alpha-hydrogen atom (panel A) and the cation H1 ring hydrogen atom interactions with the backbone nitrogen and two carboxyl groups (panel B).

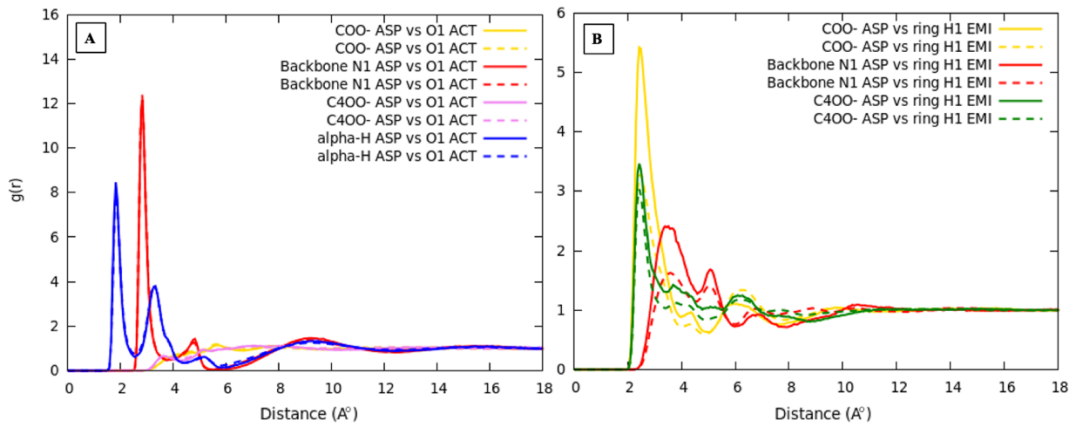


Figure 6.4. 11: Centre-of-mass radial distribution function of [C_2C_1Im][CH_3CO_2]:aspartate for acetate anion O1 atom around (A) COO^- groups, backbone N1 and α-proton H4 on aspartate molecule and H1 atom $[C_2C_1Im]^+$ distribution around (B) COO^- groups and backbone N. High concentration systems (dashed lines) *vs* low concentration (solid lines) at 300K

The increase in concentration does not affect the acetate anion distribution around the backbone nitrogen or alpha-hydrogen on aspartate molecule, similar to what has been observed in case of chloride anions with aspartate (figure 6.3.3). Nevertheless, O1 atoms on acetate anion are able to form closer and more defined shells around the aspartate molecule. The H1 ring atom on the [C_2C_1Im] cation shows a decrease in the distribution around the aspartate atomic sites as the concentration ratio is increased, with the most significant decrease around the backbone. This is also the position that shows that aspartate-aspartate interactions start to make an impact at the higher temperatures and suggests that the aspartate-ammonium interaction may generate a cascade of displacement of the ion shells. That is, as the aspartates bind more strongly to the ammonium group, the cooperative formation of localised shells is less likely to occur resulting in an overall more significant depletion of organised cations and anions. The alpha-hydrogen position gives a central backbone local to compare interactions. Anion interactions are compared for chloride and acetate anions below in figure 6.4.3 for the high concentration system at 300K. Closer shells of acetate anions as compared to chloride anions indicate that acetate anions interact more tightly and strongly with the aspartate molecules. The inter-aspartate interactions occur at the same distance for

the aspartate molecule in both ILs, consistent with the same functional-group interactions. The lower g(r) value for the chloride, along with the larger distance of the shell from the aspartate, is suggestive of better solubility of aspartate molecule in $[C_2C_1Im][CH_3CO_2]$ relative to $[C_2C_1Im][Cl]$. This is further supported by the asp-aspartate interactions showing a higher g(r) value in $[C_2C_1Im][Cl]$.

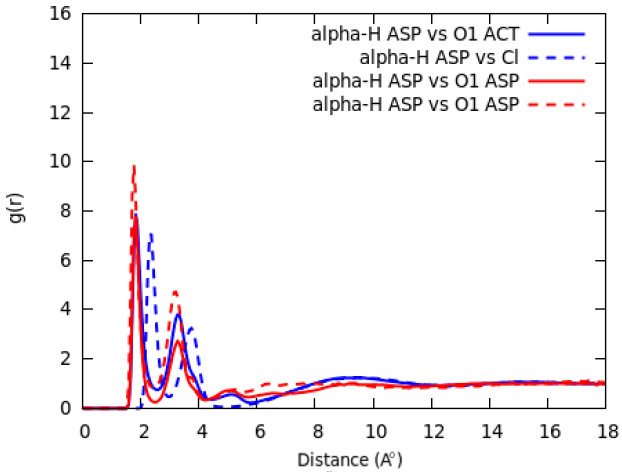


Figure 6.4. 12: Centre-of-mass radial distribution function for high concentration systems of aspartate in pure $[C_2C_1Im][Cl]$ (dashed lines) and pure $[C_2C_1Im][CH_3CO_2]$ (solid lines). RDFs show distribution of α-hydrogen atom on aspartate around anions (acetate anion O1 atom and chloride) and O1 atom on aspartate molecule at 300K

6.4.2 Effect of temperature on Aspartate and $[C_2C_1Im][CH_3CO_2]$ IL interactions

Simulations were run for aspartate molecule in $[C_2C_1Im][CH_3CO_2]$ at a higher temperature to develop an understanding about the effect that it would create on the microscopic interactions. The effect of temperature on the organisation and the solubility behaviour of lysine in two ionic liquids has already been discussed Chapter 5) and also for aspartate molecule in $[C_2C_1Im][Cl]$ (Chapter 6, section 6.3), providing a further point for comparison. The increase in temperature is reflected by a decrease in organisation, and especially a reduction in the observed second shell of anions. The impact is again more pronounced at higher concentrations of aspartate, indicative of a solute effect on the temperature-dependant organisation. Here, the effect seems more dramatic for the cation-shell.

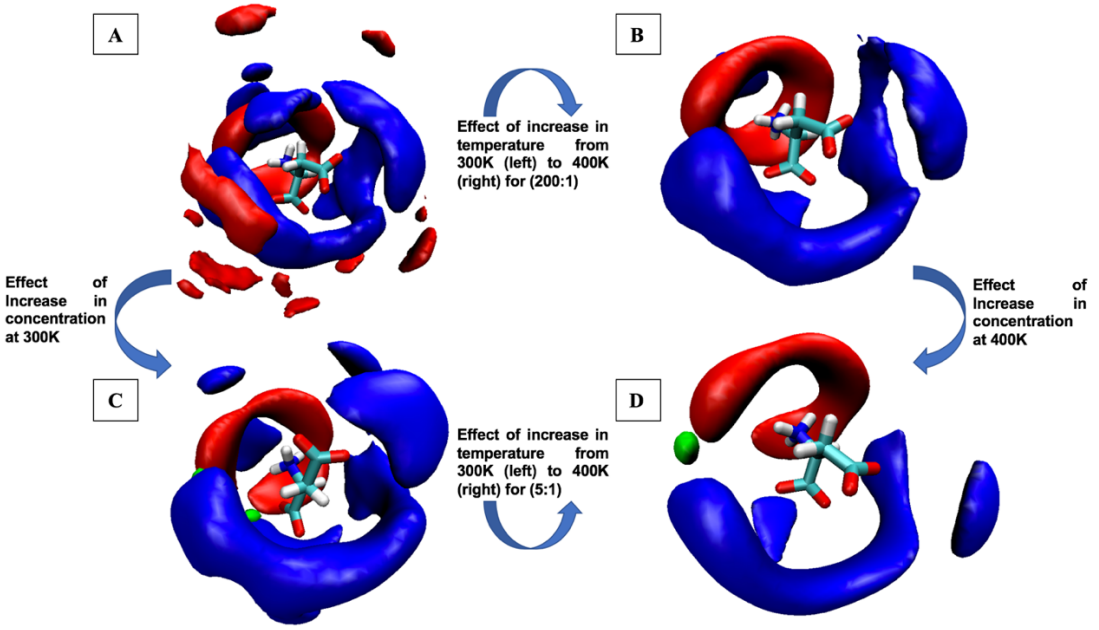


Figure 6.4.13. Spatial distribution functions around the central aspartate molecule for (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][CH_3CO_2]$. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$ and green= aspartate at iso-value=10 nm^{-3}

The impact of temperature is most prominently observed (Figure 6.4.5) in the cation-cation interactions, with the anion-cation interactions of the ionic liquid not being substantially affected.

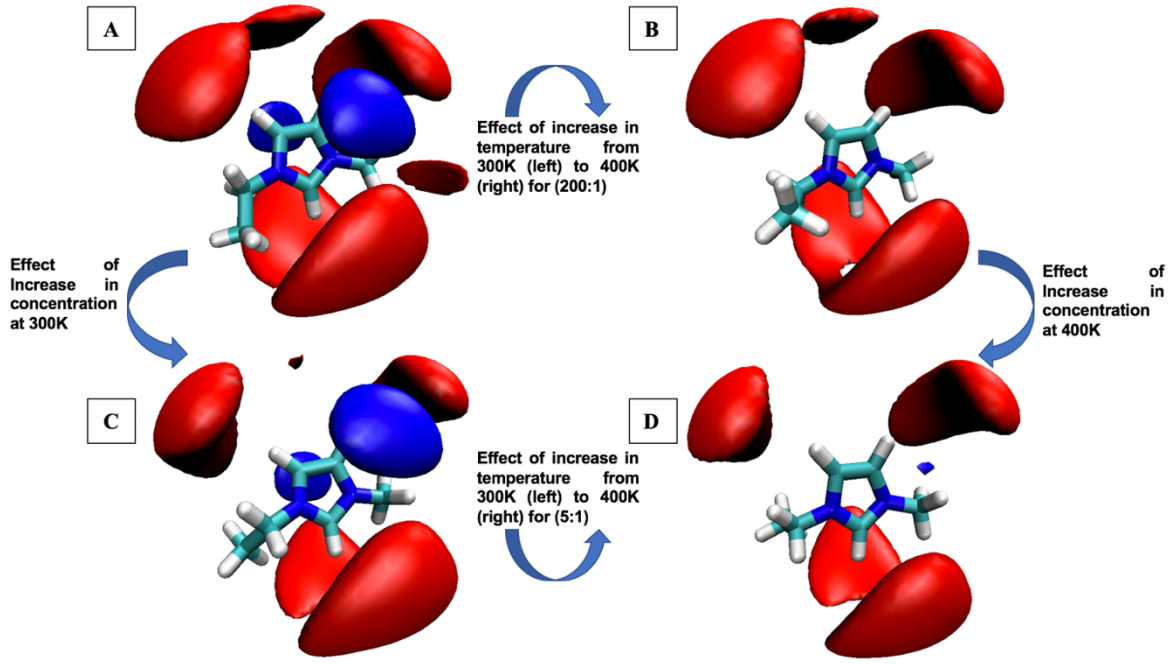


Figure 6.4.14. Spatial distribution functions around the central $[C_2C_1Im]^+$ cation (A) low concentration system at 300K (B) low concentration system at 400K (C) high concentration system at 300K (B) high concentration system at 400K in $[C_2C_1Im][CH_3CO_2]$ IL. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.

Radial distribution functions shown in the figure compare the effect of temperature on the distribution of O1 atom on acetate anion around the two carboxylate groups, backbone nitrogen and alpha-hydrogen atomic sites on aspartate molecule. The numberings used to explain the RDFs are shown in the figure 6.1 and 6.2. The organisational picture obtained in figure 6.4.6 show that the increase in temperature in each case decreases the interaction strength.

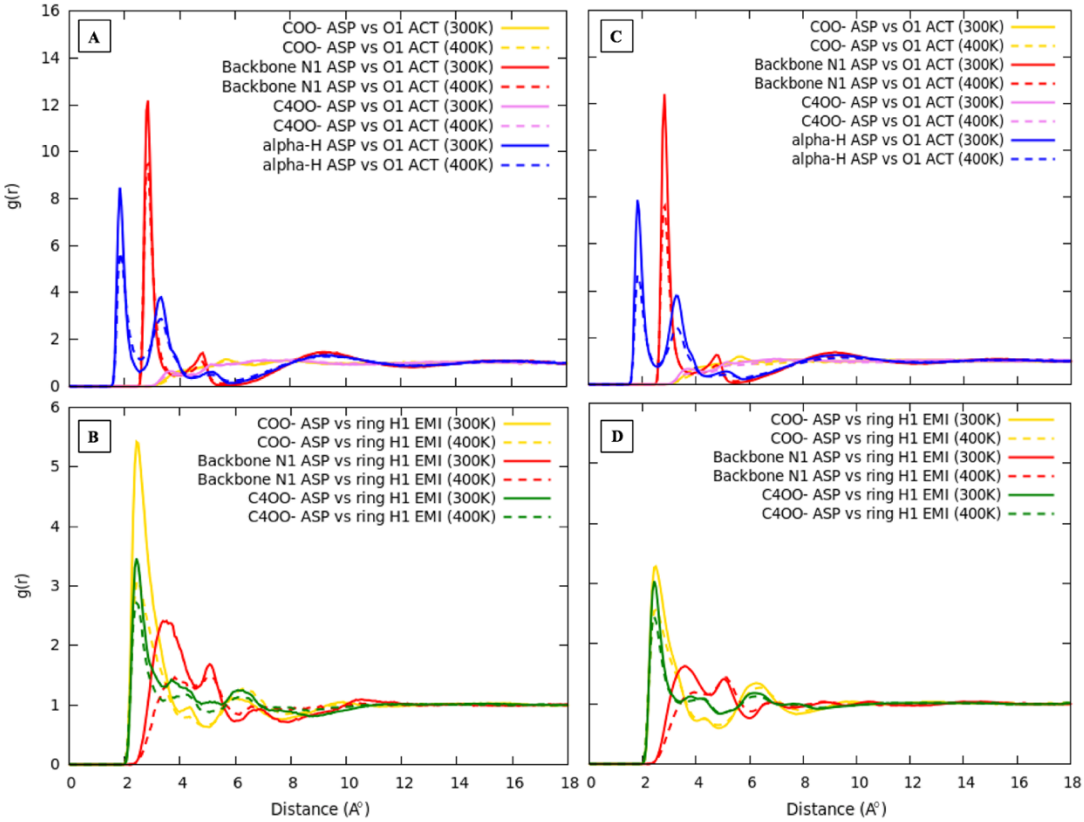


Figure 6.4.15. Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:aspartate low concentration ((200:1 (left A,B)) and high concentration ((5:1) (right C,D)) at 300K (solid lines) *vs* 400K (dashed lines). (A,C) show distribution of O1 atom on acetate anion around backbone nitrogen, carboxyl groups and α-hydrogen and (B,D) backbone nitrogen, carboxyl group and O1 *vs* (B) H1 $[C_2C_1Im]^+$ cation determined by MD simulation

This trend is similar to pattern followed for aspartate molecule in $[C_2C_1Im][Cl]$ IL systems as show in 6.3.8. The difference lies in the fact that the O1 atom interactions with aspartate molecule are quite high relative to those of the chloride anions, which suffer drastic decrease. So, acetate anions have stronger bonding with the aspartate molecules even at the higher temperature. The cations on the other hand also show a decrease around the carboxylate groups as the temperature is increased but for $[C_2C_1Im][Cl]$ the cation interactions are quite strong relative to those around the

carboxylate groups on aspartate molecules in $[C_2C_1Im][CH_3CO_2]$ IL. SDF comparisons also show that the cations are more dense at 400K around the aspartate molecule in $[C_2C_1Im][Cl]$ than in $[C_2C_1Im][CH_3CO_2]$ IL. Cation-anion overall and partial RDFs are shown in appendix A.23 that compare the effect of temperature increase on the cation-anon interactions. The overall RDFs are not much different with just a slight broadening of the cation-anion peak.

6.4.3 Effect of addition of water on aspartate and $[C_2C_1Im][CH_3CO_2]$ interactions:

Two water molecules were added to the system, in line with the Karl Fischer measurements taken experimentally for $[C_2C_1Im][CH_3CO_2]$ and simulations were run following the procedure laid out in methodology for IL:AA 5:1 at 300K and 400K. The cation-anion interactions are only slightly affected as the water is added to the system (Figure 6.4.7), consistent with the small amount of water present, and this is seen also in the overall and the partial RDFs among the solute and solvent molecules shown in appendix figures A.24.1 and A.24.2. The aspartate-aspartate interaction seems more affected at the high temperature. The small amount of water is able to break the hydrogen bonds formed between the aspartate molecules and this is evident from the SDFs with the reduction in the green regions around the central aspartate molecule. acetate anion distributions around the backbone nitrogen are dominant and are not affected much by the addition of water or the increase in temperature whereas cations are able to dominate the area around two carboxylate groups through electrostatic interactions.

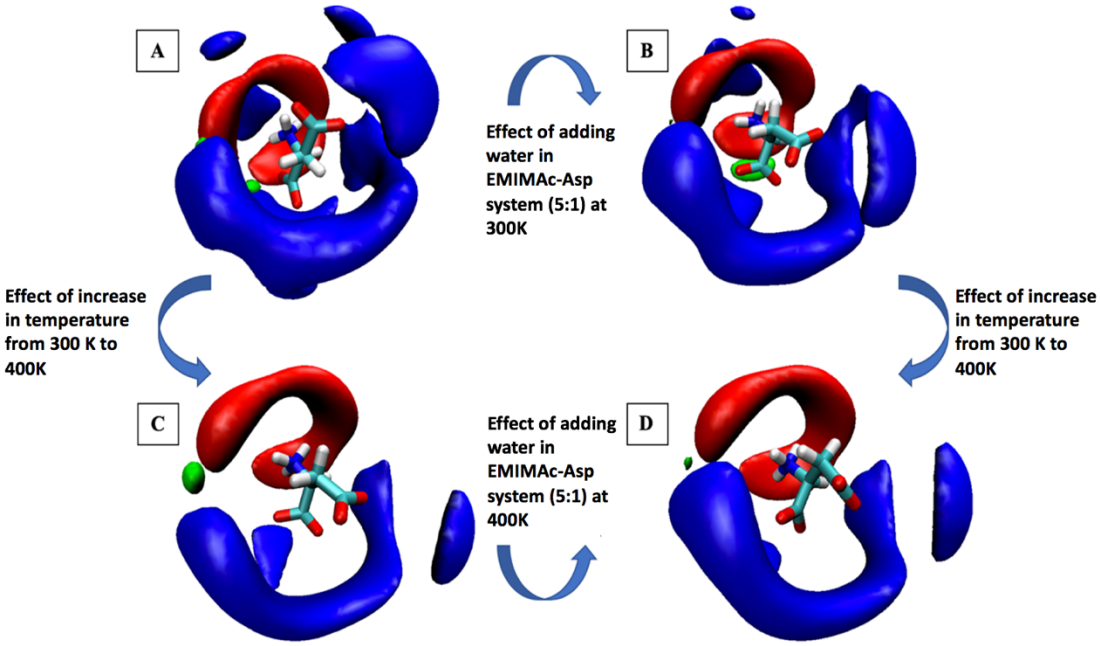


Figure 6.4.16. Spatial distribution functions around the central aspartate molecule in (A) pure $[C_2C_1Im][CH_3CO_2]$ at 300K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][CH_3CO_2]$ at 400K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 400K. Blue = $[C_2C_1Im]^+$, red = $[CH_3CO_2]^-$and green = aspartate at iso-value=10 nm^{-3}.

Appendix figure A.24.3 shows that as the temperature increases for IL/water mixtures, there is a slight reduction in the anion distribution around the central cations and a strong bonding develops between the acetate-anion oxygen atoms and hydrogen atoms on water. Cation-cation interactions are most affected and completely disappear at higher temperature for high concentration systems in both pure and IL/water mixture. Comparing with the aspartate systems in $[C_2C_1Im][Cl]$/water in figure 6.3.14 show that cation-anion and cation-cation interactions are stronger than those seen in the $[C_2C_1Im][CH_3CO_2]$/water systems shown here.

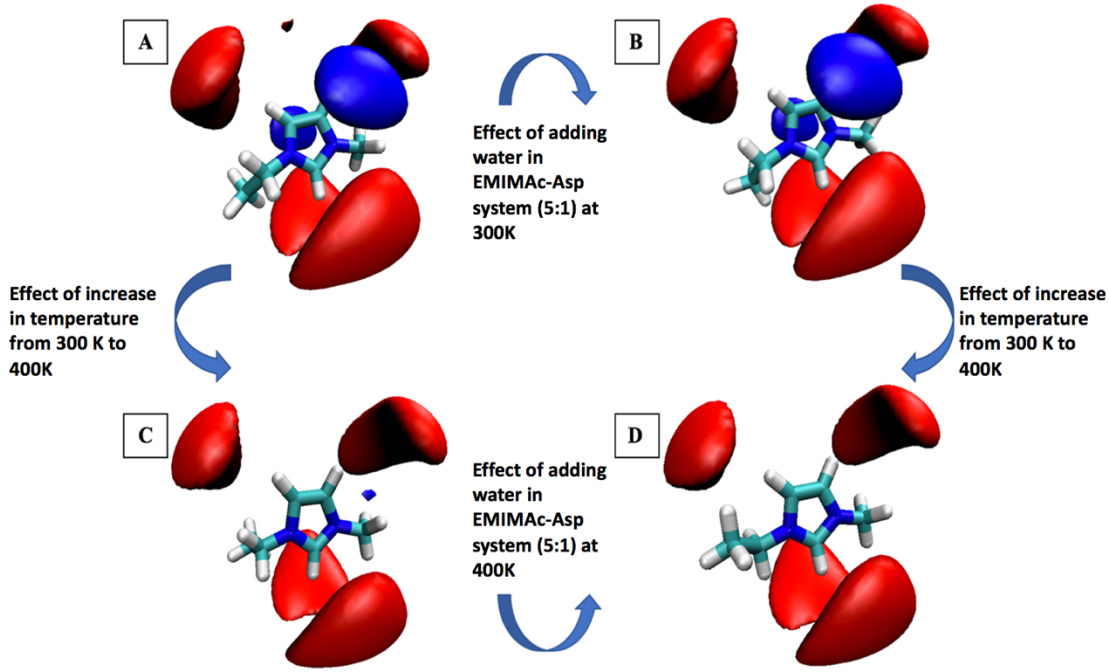
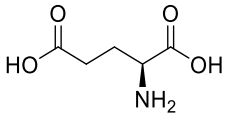
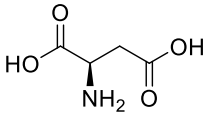


Figure 6.4. 17: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for high concentration aspartate systems in (A) pure $[C_2C_1Im][CH_3CO_2]$ at 300K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][CH_3CO_2]$ at 400K (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 400K in a view to keep acidic proton H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}

The cationic H1 atom and O1 atoms on acetate organise themselves around atomic sites on the central aspartate molecule in a similar fashion for both pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$/water mixture and there is very little difference in the absolute magnitude of these interactions, reflecting the low levels of water. Similarly, snapshots taken at the 100 ns point in the simulations at different temperatures, with and without water (Appendix A.24.4), give little indication of aggregation, as had previously been seen for either lysine solutions, or solutions in $[C_2C_1Im][Cl]$. This is consistent with the high solubility reported experimentally for aspartate in $[C_2C_1Im][CH_3CO_2]$, relative to the other systems (Table 6.4.1).

Table 6.4. 1: Experimental Solubility values for L-Aspartic acid and L-Glutamic acid in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$[242]

Ionic liquids	L-Glutamic acid	L-Aspartic acid
		
	Solubility (wt%)	
$[C_2C_1Im][Cl]$	70	20
$[C_2C_1Im][CH_3CO_2]$	70	80

Domain analysis was performed for all systems to get an idea about the solubility behaviour of aspartate for pure and water mixtures of $[C_2C_1Im][CH_3CO_2]$. The average domain counts for all the solute and solvent molecules are shown in table 6.4.2, with the progression of the domain counts across the simulations for aspartate in pure and water mixture of $[C_2C_1Im][CH_3CO_2]$ in figure 6.4.10. The average domain values confirm that the aggregation of aspartate molecules is promoted when the temperature is raised, as indicated by the decrease in the number of average domain counts of 24.22 at 300K to 14.92 at 400K. The number of domain counts at both the temperatures is still higher relative to aspartate systems in $[C_2C_1Im][Cl]$, as shown in Table 6.3.1, and this is again consistent with the better solubility of aspartate molecules in $[C_2C_1Im][CH_3CO_2]$. The lack of aggregation in $[C_2C_1Im][CH_3CO_2]$, relative to $[C_2C_1Im][Cl]$ is highlighted in Appendix figure A.24.5.

Table 6.4. 2: Domain analysis for $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$/water mixtures

Ionic Liquid	Temperature (K)	Domain count (average)			
		aspartate	$[CH_3CO_2]$	$[C_2C_1Im]$	O1-water
$[C_2C_1Im][CH_3CO_2]$	300K	24.22	189.4	4.34	
$[C_2C_1Im][CH_3CO_2]$	400K	14.92	189.27	3.8	
$[C_2C_1Im][CH_3CO_2]/H_2O$	300K	22.99	189.04	4.22	2
$[C_2C_1Im][CH_3CO_2]/H_2O$	400K	17.68	189.04	3.85	1.99

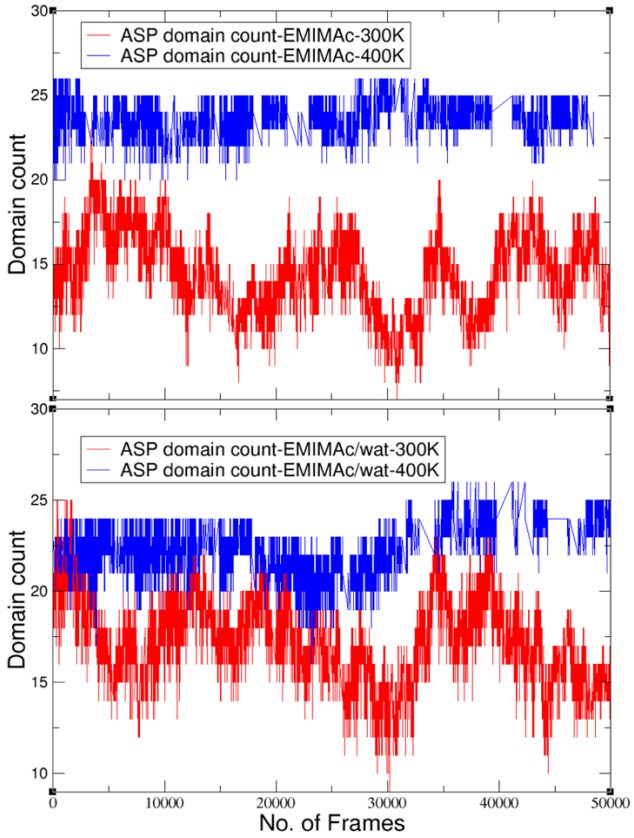


Figure 6.4. 18: Domain analysis for (A) aspartate in $[C_2C_1Im][CH_3CO_2]$ and (B) $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K(blue) and at 400K (red)

MSD and the log-log plots, shown as a function of time, are shown in Appendix A.21, Figure A.21.3 and the respective self-diffusion coefficient values are shown in table 6.4.3. These diffusion coefficient values are based on the first 50 ps of simulation.

Table 6.4. 3: Self diffusion coefficients calculated from MD simulations for aspartate systems

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s	
		T=300K	T=400K
		5:1	5:1
aspartate $[C_2C_1Im][CH_3CO_2]$	aspartate	0.81	3.28
	$[C_2C_1Im]$	1.10	6.62
	$[CH_3CO_2]$	1.10	4.52
aspartate	aspartate	1.14	1.89
$[C_2C_1Im]$ $[CH_3CO_2]$-wat	$[C_2C_1Im]$	1.91	5.84
	$[CH_3CO_2]$	2.32	3.60
	water	3.97	6.37

The self-diffusion coefficient values are indicative of fluidity of the system and inversely related to viscosity. The diffusion coefficient values increase as a function of temperature at the same concentration ratio. This trend is same as seen for aspartate in $[C_2C_1Im][Cl]$. Comparing the values of self-diffusion coefficients for aspartate and IL ions in pure IL systems show that they are lower in $[C_2C_1Im][Cl]$ than in $[C_2C_1Im][CH_3CO_2]$ IL (i.e. 2.8 and 3.28 respectively). The same is the case for IL ions, that they diffuse faster in $[C_2C_1Im][CH_3CO_2]$ IL. This would suggest a higher strength of interactions in the $[C_2C_1Im][Cl]$ system that causes slowing down of the species. However, the opposite is true for IL/water mixtures of aspartate in these two ILs which would suggest better solubility of aspartate in $[C_2C_1Im][CH_3CO_2]$ IL and hence stronger interactions. One reason could be the enhanced aspartate-aspartate interactions that are relatively weaker when acetate anion is present. This points towards the complexity of these multi-ion systems with a fine balance of relative interaction energies, and also suggest that the apparent values of self-diffusion coefficients should only be taken as a guide.

6.5 Conclusions

The impact of negatively charged amino acids on structuring in ionic liquids was investigated in this chapter. Two amino acids aspartate and glutamate both have similar nature and differ in only a single methylene with respect to the side-chain attached to the backbone. Both aspartate and glutamate were analysed at lower and high concentration ratios (IL:AA concentration of 200:1 to 5:1, respectively) in $[C_2C_1Im][Cl]$. The IL cation-anion interactions were influenced by the increase in the concentration ratio, with the dominance of cations above and below the ring plane of cation consistent with the cation-pi interactions. A similar organisational pattern was observed for lysine amino acid, but the influence of an increase in concentration was a drastic decrease in the cation-anion interactions, as opposed to cation-cation interactions. This suggested that the nature of the charge of the solute affects the structuring of ionic liquids. The other important feature was to realise that the shorter-chain aspartate tended to exert a greater impact on the depletion of this interaction, relative to glutamate, strongly suggesting that absolute charges and functionalities are not the sole determinant in directing organisation, but that this is influenced also by the structure on Å scale.

The organisational trends in $[C_2C_1Im][CH_3CO_2]$, shown for aspartate, are fairly similar to those observed with $[C_2C_1Im][Cl]$, in that an increase in concentration of the aspartate results in a decrease, most substantially, in the anion interactions with the aromatic ring carbons of $[C_2C_1Im]$, and with relatively little impact on the cation-cation interactions.

The acetate interactions with the $[C_2C_1Im]$ cation are weaker than the corresponding chloride interactions, suggesting that these acetate ions may be more available for interacting with solute partners. This contrasts with the cation-cation interactions, which in solutions with lysine are virtually absent in comparison, even at low concentrations. Negatively charged amino acids tend to aggregate at lower temperature as well, a trend missing in the lysine systems. The decrease of the chloride-ammonium interaction in the high concentration systems is compensated by the significant increase in the interaction between the aspartate molecules with one another at higher temperature. These compensating interactions emphasise the role of temperature in mediating the complex ion-ion equilibria that are generated on amino acid dissolution. The increase in temperature is reflected by a decrease in organisation, and especially a reduction in the observed second shell of anions. The impact is again more dramatic at higher concentrations of aspartate, indicative of a solute effect on the temperature-dependant organisation.

Finally, based on the above discussion and through domain analysis, it can be stated that that aspartate molecules were fairly solvated by the $[C_2C_1Im][CH_3CO_2]$ relative to the $[C_2C_1Im][Cl]$ ionic liquid, as the higher domain count values relate to more dispersion of solute in the system.

7

Interactions and organisation of aromatic amino acids and alanine in $[C_2C_1Im][CH_3CO_2]$

7.1 Introduction

Aromatic compounds are unsaturated cyclic and planar molecules that contain an aromatic ring. The presence and arrangement of π–eletrons above and below the ring plane adds to their stability. These electrons create a π–electron cloud over the ring.[249] Aromatic amino acids have unique and important properties. Phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) are generally hydrophobic, but, compared with simpler hydrophobic residues such as leucine or valine, the aromatic amino acids have additional capabilities. Both Tyr and Trp can contribute hydrogen bonds, which is an important feature. Another reason for the importance of these aromatic amino acids is their ability to get involved in cation–π interactions that are noncovalent binding interactions contributing to protein secondary structure and diverse drug-receptor interactions.[250] Histidine has the ability to exist in neutral as well as protonated forms. In the neutral and protonated forms, histidine also can interact through attractive cation–π interactions, and has properties similar to the imidazolium ionic liquids due to the imidazole side chain.

Amino acids play an important role in physiological phenomena and are critical compounds in animal and human nutrition. They are being employed as food additives, feed supplements and artificial sweeteners.[251, 252] Aromatic amino acids such as L-Phenylalanine (L-Phe), L-Tryptophan (L-Trp) and L-Tyrosine (L-Tyr), are amongst the most important for nutritional applications. These compounds can be obtained by the hydrolysis of proteins and the recovery of these amino acids, proteins and other biomolecules from natural sources and fermentation media is an essential task in

biotechnology. The demand for cost-effective and environment friendly separation and purification processes has led to a significant effort to develop clean manufacturing methods and easily scaled-up industrially relevant techniques to increase selectivity and capacity and reduce overall costs. Ionic liquids have already been shown to be useful for the separation of amino acids by extraction processes. And tryptophan, in particular has previously been chosen as a model compound to study the extraction processes using hydrophobic ionic liquids.[56, 154] Chemdraw structures of these amino acids are shown in Figure 7.1.

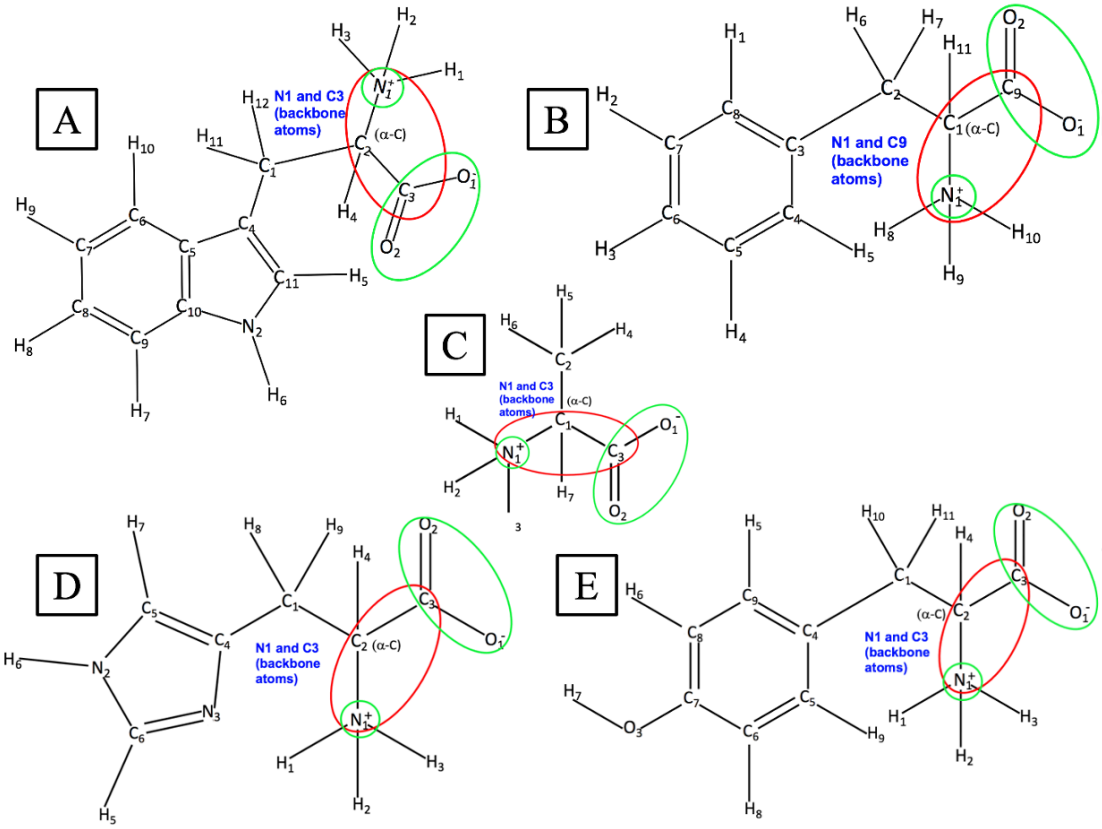


Figure 7. 1: Chemdraw structures of neutral (A) tryptophan (B) phenyl-alanine(C) alanine, (D) histidine and (E) tyrosine with zwitter-ionic backbones and atom numberings generated by antechamber. Atoms encircled red in the molecules are those used as reference to create respective SDFs around the molecule. Backbone atoms are labelled in all amino acids. Atoms encircled green are the ones that are used to discuss results in RDFs analysis.

This chapter describes the interactions of both aromatic amino acids, and the simplest, chiral amino acid alanine, with $[C_2C_1Im][CH_3CO_2]$. This ionic liquid was chosen as the focus, as it had previously shown the most promise for dissolving both positively-charged and negatively-charged zwitterionic amino acids (Chapters 5 and 6), and has been used extensively for experimental applications.[59, 74]

7.2 Materials and Methods:

7.2.1 Outline of simulation conditions

Alanine and the aromatic amino acids in their zwitterionic neutral forms were selected for comparison with the polar charged amino acids discussed in chapter 5 and 6. Among the aromatic amino acids, histidine, phenylalanine, tyrosine and tryptophan were simulated in ionic liquids to explore the behaviour of these amino acids and compare with the neutral alanine. Alanine, phenylalanine, and tryptophan are hydrophobic and non-polar, whereas histidine exists in both charged and neutral form at physiological pH and within enzymes, with a pK_a of 6 for this side chain. Alanine and aromatic amino acids were prepared in their zwitterionic neutral forms, including histidine. The overall charge was neutral for all simulations. The simulations were run at 300K for the aromatic amino acids and pure $[C_2C_1Im][CH_3CO_2]$ for a concentration of IL:AA 5:1. (Table 7.1). As seen in previous chapters, temperature can have an impact on both structuring and interactions and the impact of this was explored for histidine and alanine at 300K and 400 K for pure $[C_2C_1Im][CH_3CO_2]$, representing room-temperature and heated solutions, respectively. For alanine, other than the temperature, the simulations were also run with pure $[C_2C_1Im][Cl]$ and in $[C_2C_1Im][Cl]$ water mixture to understand the different interactions at play between the charged and neutral solutes.

Table 7.1 also highlights the high concentration systems considered for alanine and aromatic amino acids simulations, consistent with those examined in chapters 5 and 6. These systems contain 200 ion pairs of charge-scaled $[C_2C_1Im][CH_3CO_2]$ and 40 non-scaled molecules of amino acids. Experimental data was not available for these amino acids, so an effort here is made to develop relation among different amino acids in pure $[C_2C_1Im][CH_3CO_2]$ (alanine and aromatic amino acids with lysine and aspartate at 300K for IL:AA 5:1), $[C_2C_1Im][CH_3CO_2]$ IL/water mixture (alanine with lysine and aspartate) and in pure and IL/water mixture of $[C_2C_1Im][Cl]$ (alanine with lysine and aspartate). The overall charge of the more concentrated simulation box in each aromatic amino acid case is balanced as all amino acids are considered in their neutral forms.

Table 7. 1: Simulation conditions for Alanine and aromatic amino acids

IL	Amino acid	Number of IL ions and AA molecules	Number of water molecules added	Concentration ratio IL:AA and Temperature	
$[C_2C_1Im][Cl]$	Alanine	For (IL:AA)= (200:1)	n/a	5:1 (300K)	5:1 (400K)
$[C_2C_1Im][Cl]$-water	Alanine	200 IL ion pairs	26	5:1 (300K)	5:1 (400K)
$[C_2C_1Im][CH_3CO_2]$	Alanine	one AA molecule	n/a	5:1 (300K)	5:1 (400K)
	Phenyl-alanine		n/a	5:1 (300K)	
	Tyrosine	For (IL:AA)=(5:1)	n/a	5:1 (300K)	
	Tryptophan	200 IL ion pairs 40 AA molecules	n/a	5:1 (300K)	
	Histidine-neutral		n/a	5:1 (300K)	
$[C_2C_1Im][CH_3CO_2]$-water	Alanine		2	5:1 (300K)	5:1 (400K)

7.2.2 QM Calculations for Amino acid and Ionic Liquid Structures and antechamber

Alanine, histidine, phenylalanine, tyrosine, tryptophan, $[C_2C_1Im]$ cation and $[CH_3CO_2]$ anion structures were prepared as outlined in Chapter 3, using the WebMo web-based interface to extract the xyz structures. All amino acids were constructed with neutral side chains and a zwitterionic backbone. Optimised structures obtained from Gaussian are shown with atom numbering in Figure 7.2.

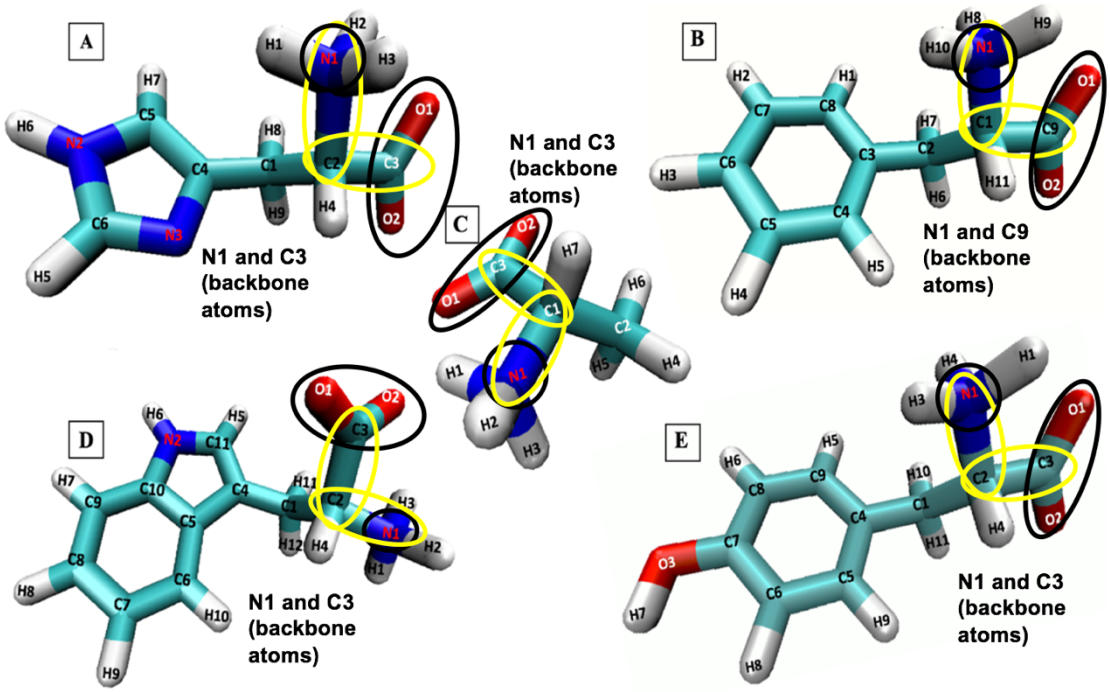


Figure 7. 2: Optimised structures of aromatic amino acids (A) Histidine (B) Phenylalanine, (C) Alanine (D) Tryptophan and (E) Tyrosine with atom numbering generated by antechamber.[180] Atoms encircled yellow are in each molecule show the atoms that are used as reference to create respective SDFs around these molecules. Backbone atoms are labelled in lysine and arginine. Atoms encircled black are the ones that are used to discuss results in RDFs analysis.

7.3 Results and discussions

7.3.1 Effect of temperature on interactions of concentrated alanine solution in pure and IL/water mixture of [C_2C_1Im][Cl]

Alanine is the simplest of chiral amino acids, and provides a good baseline for minimal ionic liquid-amino acid interactions. SDF analysis indicates that for alanine at high concentration, strong self-interaction (Figure 7.3, green) is present at low temperature, with this increasing at the cost of, particularly, cationic interaction at higher temperature. Interactions with the ionic liquid are as expected for the charge distribution, with anions interacting most strongly with the positively-charged ammonium, and cations interacting with the backbone carboxylate. At low temperature, there is also some anionic interaction with the methyl group, which is presumably a result of a second shell near the carboxylate-cation interaction. This weaker interaction is lost, along with the cationic organisation at higher temperature.

The distribution of IL anions around IL cation is similar to that previously observed in chapters 5.1 and 6.1 for lysine and aspartate systems in [C_2C_1Im][Cl] IL. The partial

RDFs in appendix figure A.25.2 also show that the interactions between chloride anions and various atomic sites on $[C_2C_1Im]$ cation are slightly reduced in line with the corresponding SDFs. This suggests that the cationic interactions are the weakest and are easily broken at higher temperature.

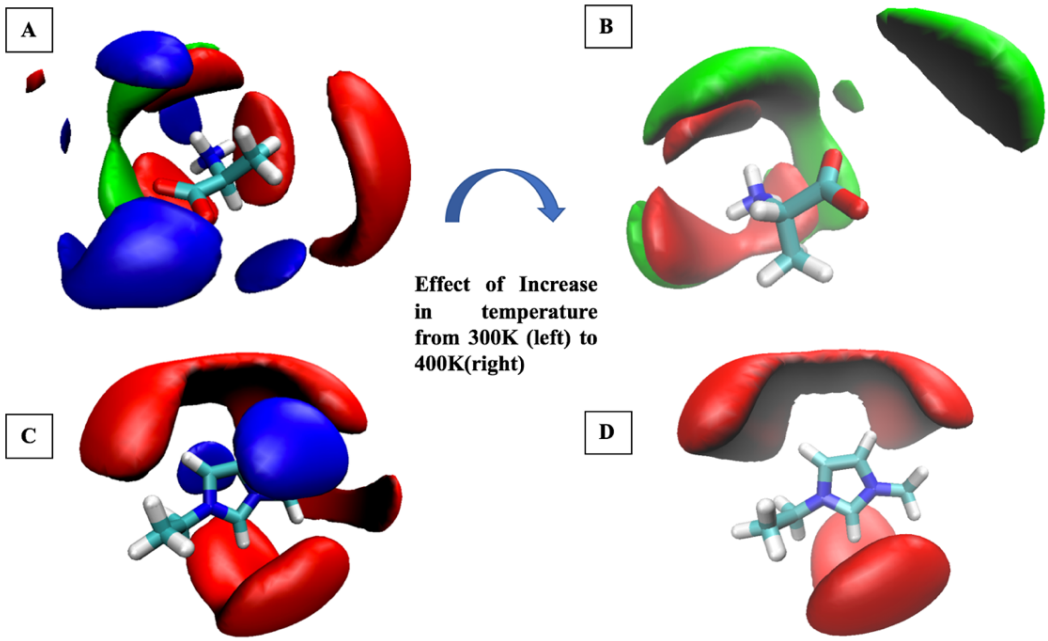


Figure 7. 3: Spatial distribution functions around the central (A) alanine molecule at 300K, (B) alanine at 400K, (C) $[C_2C_1Im]^+$ cation at 300K and (D) $[C_2C_1Im]^+$ cation at 400K in pure $[C_2C_1Im][Cl]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= Cl^- and green= alanine at iso-value=10 nm^{-3}

The radial distribution functions in figure 7.4 show the distribution of ring hydrogens, tail and methyl carbon atoms on $[C_2C_1Im]$ cation around the backbone nitrogen and carboxyl group on alanine amino acid. The cations do not show any pronounced peaks around the backbone nitrogen and the H1 atom on the cation forms the closest shell around the carboxyl group at both the temperatures, but the intensity is reduced with the rise in temperature. A similar observation can be made by looking at the figure for the carboxyl group interaction with the methyl and the tail carbon on $[C_2C_1Im]$ cation. The direction of interaction points towards the acidic hydrogen H1 on the ring carbon C4. Figure 7.4 compares the distribution of H1 atom on $[C_2C_1Im]$ cation and chloride anions around the backbone nitrogen, carboxyl group and the alpha-hydrogen atom. Chloride anions are only able to make prominent contacts with the backbone nitrogen atom and primarily with the three hydrogens as indicated by the three slightly separated red regions around the alanine ammonium ion (Figure 7.6). In each case, the strength of interaction is reduced when the temperature is increased from 300K to 400K.

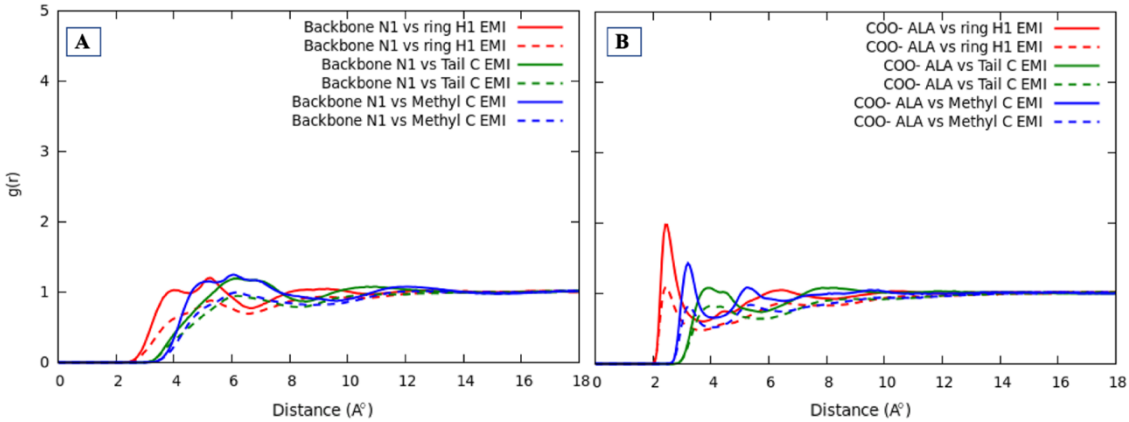


Figure 7. 4: Comparison of centre-of-mass radial distribution functions for [C_2C_1Im][Cl]:[alanine] at 300K for H1, C1, C6 atomic sites on [C_2C_1Im] cation *vs* (A) backbone ammonium (B) backbone carboxylate, (low concentration (200:1) solid lines and (high concentration (5:1) dashed lines determined by MD simulation

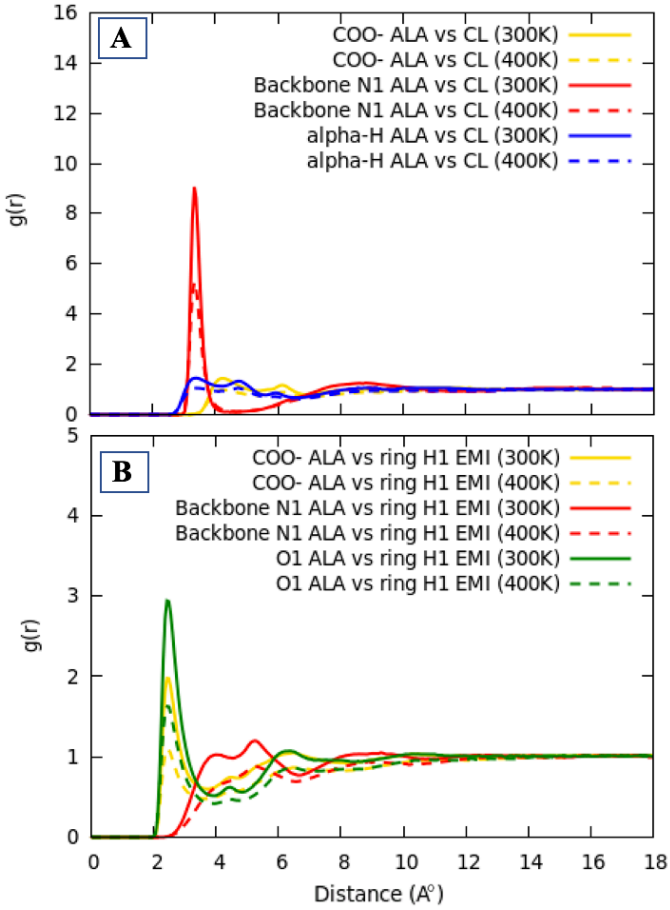


Figure 7. 5: Comparison of centre-of-mass radial distribution functions for [C_2C_1Im][Cl]:[alanine] 5:1 for (A) chloride distribution around backbone nitrogen, carboxyl group and alpha-hydrogen (B) H1 $[C_2C_1Im]^+$ distribution around backbone nitrogen, carboxyl groups and O1 cation determined by MD simulation

7.3.2 Effect of addition of water on interactions of alanine with IL/water [C_2C_1Im][Cl] in concentrated systems

As shown in previous chapters, *e.g.* in [C_2C_1Im][Cl] solutions, the addition of water reduces both the anion and cation interactions with the amino acids lysine and aspartate. Cation interactions were more affected for lysine and anion interactions were more affected for aspartate, in addition to the reduction caused by the increase in temperature. 26 water molecules were added to the alanine systems to observe the impact of addition of water in case of neutral amino acids.

At low temperatures, the change in cationic interactions seems most significant, although these are still retained, and the weak anionic density seen near to the alanine methyl group does not seem to be disrupted substantially (Figure 7.6). At higher temperatures, where there is no cationic density observed at the isovalue of 10 nm^{-3}, the most striking change is to the alanine-backbone carboxylate interaction, with a minor depletion of this density.

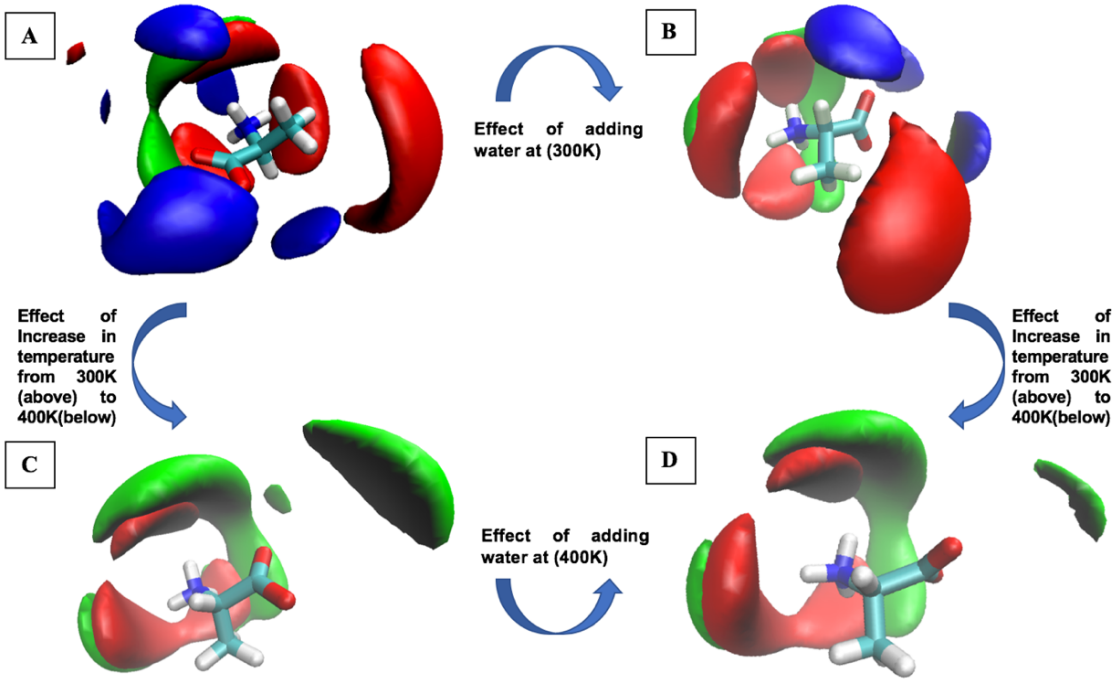


Figure 7. 6: Spatial distribution functions around the central alanine molecule in (A) pure [C_2C_1Im][Cl at 300K, (B) [C_2C_1Im][Cl/water mixture at 300K, (C) pure [C_2C_1Im][Cl at 400K, (D) [C_2C_1Im][Cl/water mixture at 400K. Blue = $[C_2C_1Im]^+$, red = Cl^-, and green = alanine at iso-value=10 nm^{-3}.

The detail of cation and anion around the alanine atomic sites is confirmed through radial distribution functions in figure 7.7 with the addition of water in each case not

affecting the organisational pattern or strength of anion or cation distributions around any of the atomic sites of alanine, but reducing slightly the inter-amino acid interactions at 400K as also shown in the appendix figure A.25.3.

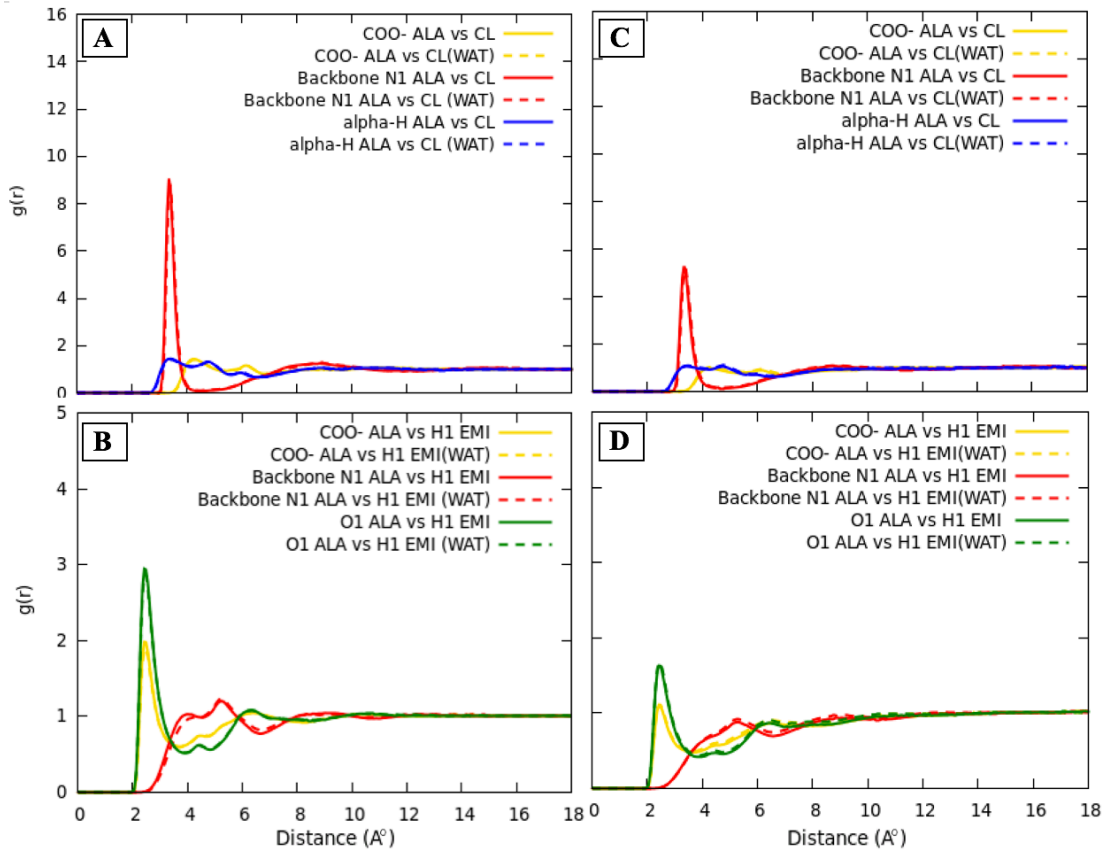


Figure 7. 7: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$:[alanine] high concentration system for (A) chloride distribution around backbone nitrogen, carboxyl group and α-hydrogen and (B) H1 $[C_2C_1Im]^+$ distribution around backbone nitrogen, carboxyl groups and O1 determined by MD simulation. Solid lines show pure IL and dashed lines show IL/water mixtures of $[C_2C_1Im][Cl]$.

The overall interactions among solute and solvent molecules are compared for pure and IL/water mixture of $[C_2C_1Im][Cl]$ at 300K and 400K in figure A.25.4 in appendix and show that the water molecules are able to interact with both the ions of the ionic liquid and alanine but the strongest and the closest peaks occur for the chloride anions. The corresponding SDFs for cation-anion and cation-cation organisation are shown in the figure 7.8. The addition of water just slightly reduces the magnitude of the density of the blue and red regions around the central cation. The distribution is affected by the increase in temperature but is not much affected by the addition of water.

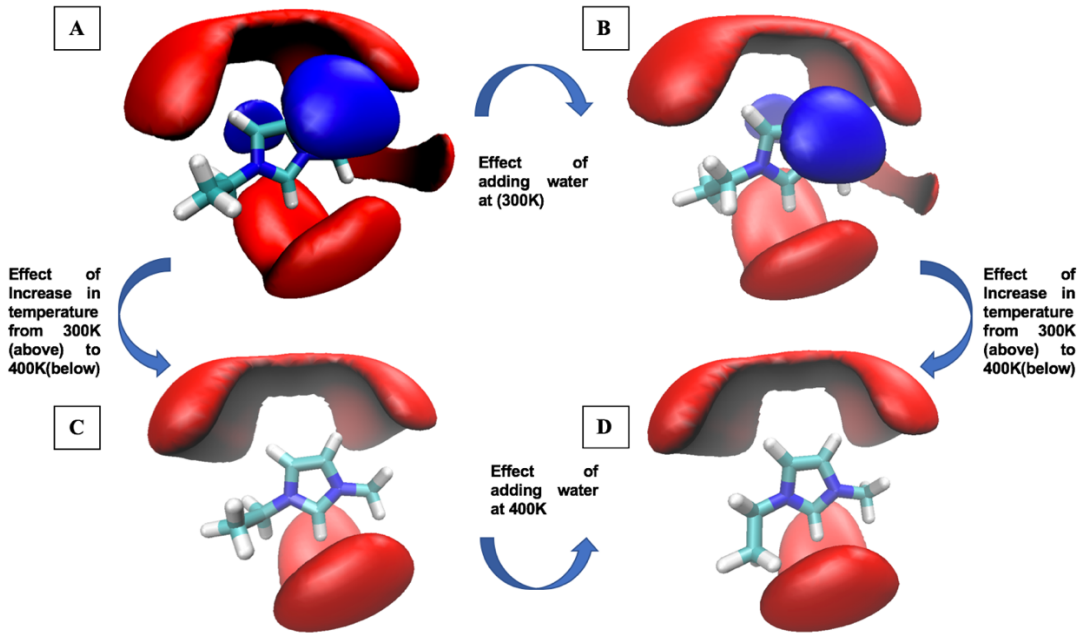


Figure 7. 8: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation at (a) 300k for pure IL (b) at 300k for IL/water mixture, (c) 400k for pure il and (d) 400k for IL/water mixture in a view to keep ring hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}

The snapshots taken from 100 ns simulations (Figure 7.9), the spatial distribution functions for solute-solvent and solvent-solvent interactions as shown in Figures 7.7 and 7.8, show that the increase in temperature for alanine in pure ionic liquid increases the inter-amino acid interactions but a broad peak in figures in appendices A.25.3 and A.25.4 for overall RDFs indicates a weak or dispersed organisation. RDFs show that the alanine-alanine interaction increases with the rise in temperature and this is reflected by the appearance of clusters in snapshot figure 7.9 (B). The addition of water at 400K reduces or breaks this interaction and alanine molecules can be seen slightly solvated in the snapshot 7.9 (D).

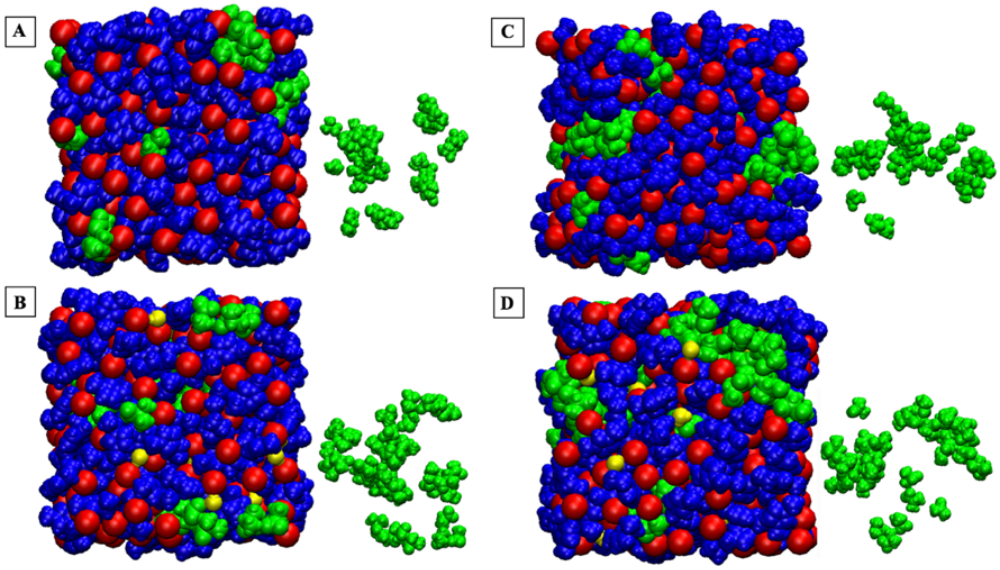


Figure 7. 9: Snapshots for Alanine systems in pure [C_2C_1Im][Cl] IL (A,C) at 300K *vs* 400K and in [C_2C_1Im][Cl]/water mixture (B,D) at 300K *vs* 400K respectively from 100 ns MD simulations. The positioning of the alanine molecules is shown at the right-hand side of each panel in the absence of water and IL ions to see the cluster formation more clearly.

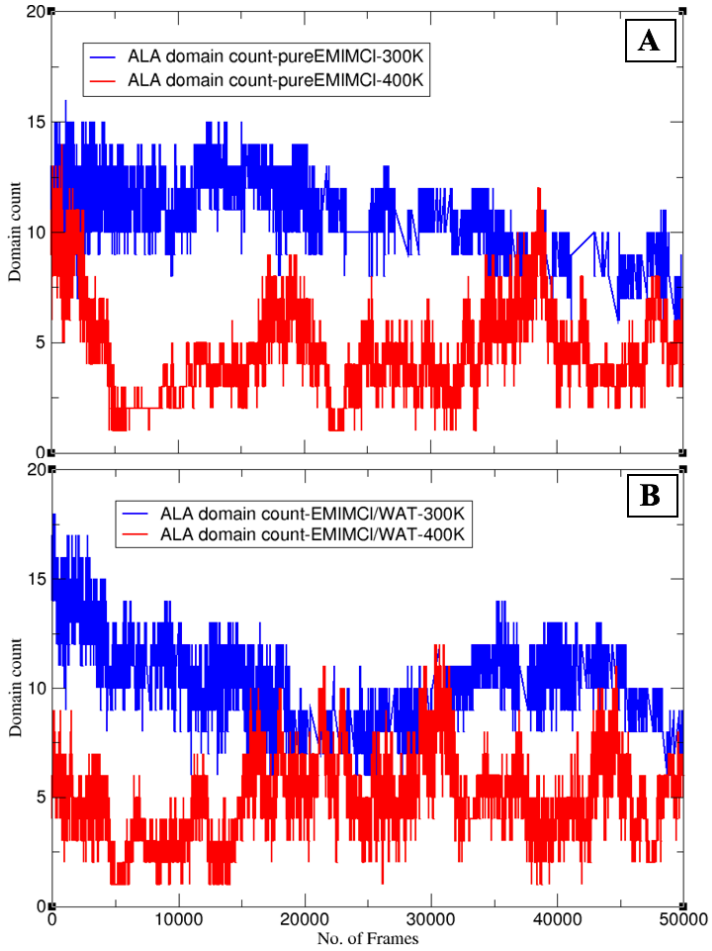


Figure 7. 10: Domain analysis over time for (A) lysine in [C_2C_1Im][Cl] and (B) [C_2C_1Im][Cl]/H_2O mixture at 300K(blue) and at 400K (red)

Domain analysis was run for all the systems to gain insight into the formation of clusters over time in the ionic liquid. The domain analysis for alanine systems in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixture is shown in the figure 7.10 compared at two temperatures 300K and 400K. The average domain counts for alanine is around 10.87, decreasing to 4.74 as the temperature is increased from 300K to 400K for the pure $[C_2C_1Im][Cl]$ system. This is consistent with the trends seen previously for the charged amino acids, and the average domain number is higher than those for the corresponding lysine systems, suggesting that alanine should be more soluble. Similarly, the average domain numbers are lower than for the corresponding aspartate systems, consistent with lower solvation of alanine than aspartate. This highlights charge of solute as a governing factor in establishing good interactions with more negatively-charged solutes favouring solubility in $[C_2C_1Im][Cl]$. This also highlights the potential role of the cation in guiding amino acid solubility.

For the $[C_2C_1Im][Cl]$/water system the domain count average is 10.51 at 300K compared to 5.01 at 400K, confirming the much smaller impact of water on the interactions with this amino acid. The average domain count values for oxygen atoms on both acetate anion and water molecules, and the ethyl chain on the $[C_2C_1Im]$ cation is shown in Table 7.2. Alanine experiences an effect similar to the charged amino acids, *i.e.* increase in clustering as the temperature is raised from 300K to 400K. Water addition slightly increases the number of domains for alanine, which is in line with the observations made from the corresponding RDFs and indicates that dispersion in the bulk solution can be maintained through interactions with water.

Table 7. 2: Domain analysis for alanine in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixtures

Ionic Liquid	Temperature (K)	Domain count (average)			
		Alanine	Cl	$[C_2C_1Im]$	O1-water
$[C_2C_1Im][Cl]$	300K	10.87	160.32	1.61	
$[C_2C_1Im][Cl]$	400K	4.74	154.9	1.79	
$[C_2C_1Im][Cl]/H_2O$	300K	10.51	157.5	1.73	25.12
$[C_2C_1Im][Cl]/H_2O$	400K	5.01	152.91	1.8	23.79

7.3.3 Effect of temperature on interactions of alanine with $[C_2C_1Im][CH_3CO_2]$ at high concentration

In comparing the solvation and microstructure around alanine on going from chloride to acetate anion (Figure 7.11), the key highlights are that the alanine has less alanine-alanine interactions (green), indicating better dispersion in the acetate-based IL than for the corresponding chloride-based IL, and that stronger cationic interactions are maintained, although reduced at the higher temperature, in contrast to the impact seen in the chloride-based IL, where these vanish completely. Ionic liquid cation-anion interactions show little temperature effect, although the already weaker cation-cation interactions in $[C_2C_1Im][CH_3CO_2]$ are further reduced as the temperature is raised.

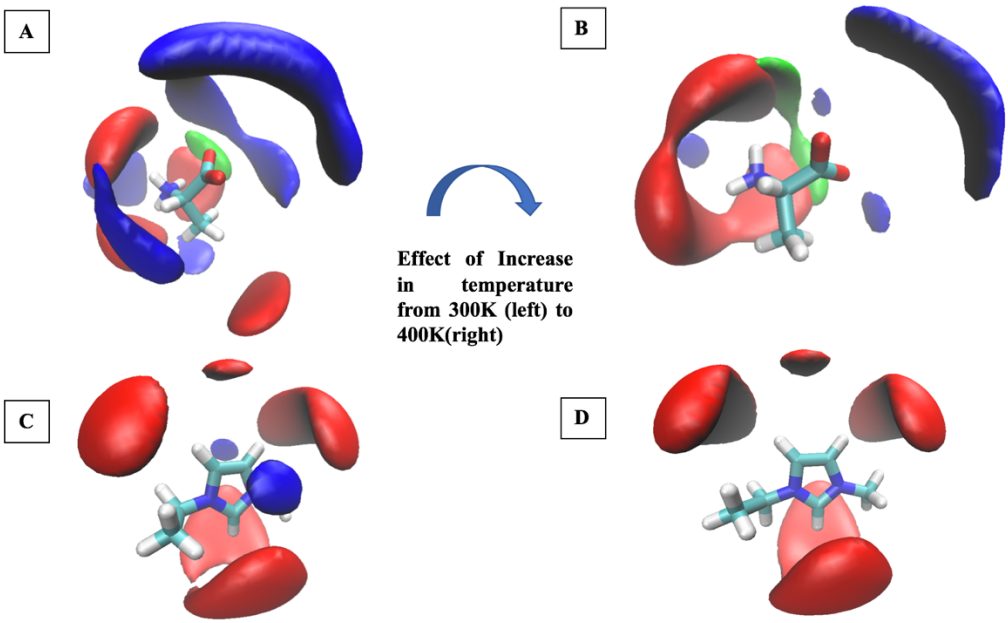


Figure 7. 11: Spatial distribution functions around the central (A) alanine at 300K, (B) alanine at 400K, (C) $[C_2C_1Im]^+$ cation at 300K and (D) $[C_2C_1Im]^+$ cation at 400K for $[C_2C_1Im][Cl]$ IL. Blue= $[C_2C_1Im]^+$, red = Cl^-, and green = alanine at iso-value=10 nm^{-3}.

Figure 7.12 confirms that the $[C_2C_1Im]$ cation forms close and defined peaks around the carboxyl group that decrease as the temperature is increased, with the highest and the closest peaks occurring for the H1 ring hydrogen on $[C_2C_1Im]$ cation. Figure 7.13 shows the acetate anions interact strongly with the backbone nitrogen and this interaction is closer and stronger than that observed for the alanine backbone nitrogen with the chloride anions in pure $[C_2C_1Im][Cl]$ (Figure 7.5). The distribution of cations is more organised around the carboxyl group of alanine and these peaks are more organised than those observed for alanine systems in $[C_2C_1Im][Cl]$.

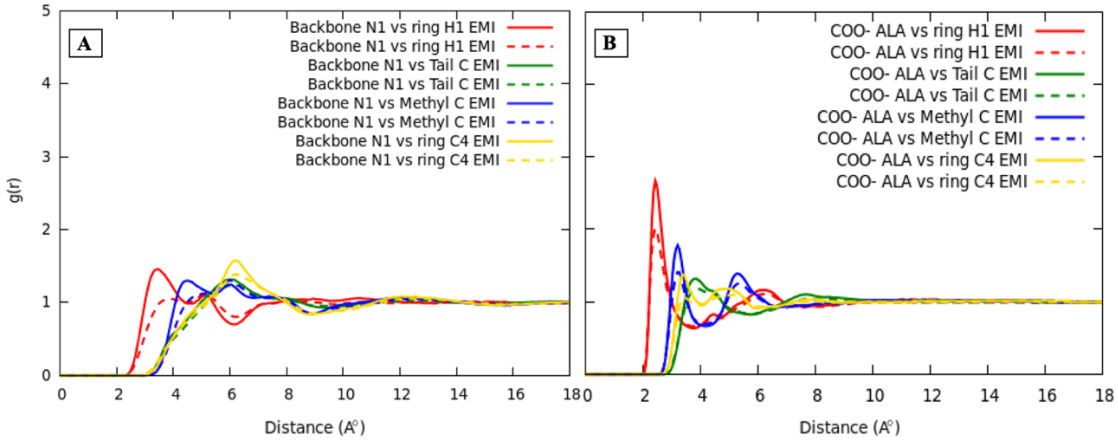


Figure 7. 12: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: [alanine] at 300K for H1, H7, H8, N1, C1, C4, C6 atomic sites on $[C_2C_1Im]$ cation *vs* (A) backbone ammonium (B) backbone carboxylate, (low concentration (200:1) solid lines and (high concentration (5:1) dashed lines determined by MD simulation

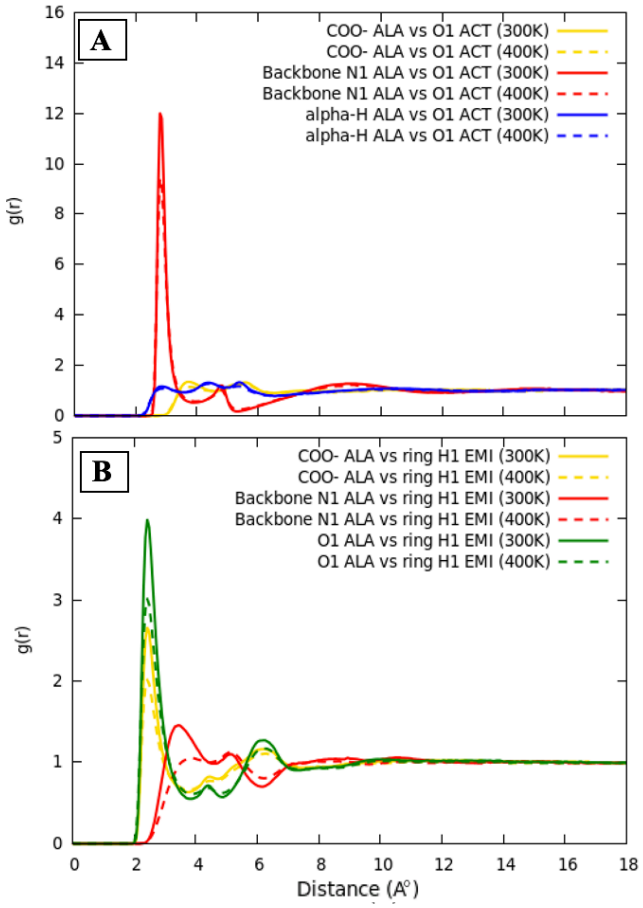


Figure 7. 13: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:[alanine] 5:1 for (A) chloride distribution around backbone nitrogen, carboxyl group and alpha-hydrogen (B) H1 $[C_2C_1Im]^+$ distribution around backbone nitrogen, carboxyl groups and O1 cation determined by MD simulation.

The inter-amino acid interactions (Appendix A.25.6) seems to highlight an increase in long-range structuring in the extended shells at higher temperatures. This might indicate the formation of a networked structure that would be interesting to interrogate experimentally.

7.3.4 Effect of water addition on interactions of concentrated alanine with [C_2C_1Im][CH_3CO_2]

The SDFs for analysing the effect of water addition for alanine systems in [C_2C_1Im][CH_3CO_2] IL are shown in figure 7.14 with corresponding RDFs in appendix figure A.25.7. As per the calculation from Karl Fischer measurements of IL water contents, only two water molecules were added to the high-concentration alanine system but are enough to affect the cation organisation around the both the alanine and the IL-cation. Both figure 7.14 and appendix figure A.25.7 (B) show that at 300K, there is a small interaction with the alpha-hydrogen/side-chain of the alanine that is broken either on water addition or when the temperature is raised. This is similar to the [C_2C_1Im][Cl] solutions where this, and the side-chain anion interactions are observed, but may be artefacts of the other organisation around the backbone.

The addition of water at 300K slightly reduces the alanine-alanine interactions. The increase in temperature increased the alanine-alanine interactions for pure [C_2C_1Im][CH_3CO_2], and this is true for IL/water mixture as well. Overall, there is good consistency with the general trends observed on changing solvent and temperature.

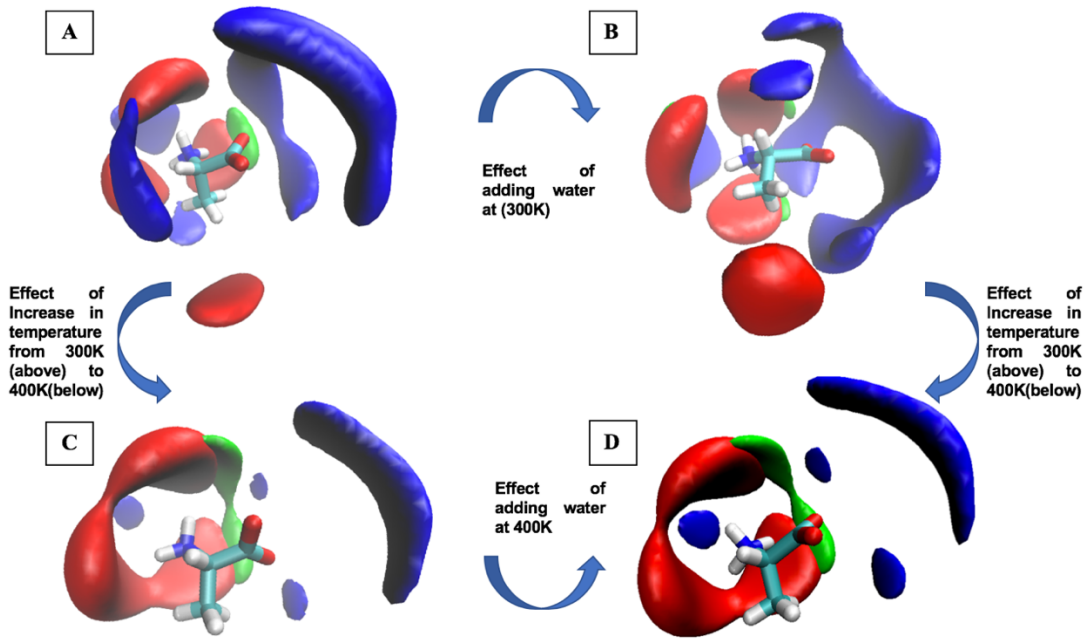


Figure 7. 14: Spatial distribution functions around the central alanine molecule in (A) pure IL at 300K, (B) IL/water mixture at 300K, (C) pure IL at 400K , (D) IL/water mixture at 400K. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$, and green= alanine at iso-value=10 nm^{-3}.

The cation-cation interaction peaks are slightly broadened but not much as the temperature is increased for the $[C_2C_1Im][CH_3CO_2]$ IL/water mixture. Cation-anion interactions are slightly reduced as observed through RDFs and SDFs. The anion interactions with water are also reduced with the increase in temperature (figure A.25.4). The snapshots taken from the end of 100ns simulations for all systems are shown in the figure in appendix A.25.8. The snapshots in case of alanine do not show a pronounced difference, in contrast to other amino acids, such as lysine in $[C_2C_1Im][Cl]$. Average domains counts for the alanine molecule, acetate oxygen atoms, the ethyl side chain for $[C_2C_1Im]$ cation, and oxygen atoms on water are shown in the table 7.3. The average domain counts for alanine in $[C_2C_1Im][CH_3CO_2]$ are higher in each case relative to the corresponding $[C_2C_1Im][Cl]$ systems reaffirming the observations made through RDFs and SDFs that $[C_2C_1Im][CH_3CO_2]$ is better at solvating alanine molecules. This trend is the same as seen for the aspartate and lysine systems. The values for alanine domain counts are highest for pure $[C_2C_1Im][CH_3CO_2]$ at 300K and the trend for average domain count values for alanine is pure $[C_2C_1Im][CH_3CO_2]$ 300K ≈ $[C_2C_1Im][CH_3CO_2]$/water 300K > pure $[C_2C_1Im][CH_3CO_2]$ 400K ≈ $[C_2C_1Im][CH_3CO_2]$/water 400K, however, these values are all fairly similar, relative to the changes seen for the other species and in other solvents.

Table 7. 3: Domain analysis for $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$/water mixtures

Ionic Liquid	Temperature (K)	Domain count (average)			
		Alanine	$[CH_3CO_2]$	$[C_2C_1Im]$	O1-water
$[C_2C_1Im][CH_3CO_2]$	300K	18.7	187.4	4.48	
	400K	15.1	186.9	4.23	
$[C_2C_1Im][CH_3CO_2]/H_2O$	300K	17.94	187.97	4.67	2.00
	400K	14.77	186.9	4.28	1.99

The self-diffusion coefficients log-log plots are shown in appendix A.25.9. Analysing the values of the self-diffusion coefficients for alanine systems in pure and IL/water mixtures of $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][Cl]$ (Table 7.4), the diffusion coefficients increase as a function of temperature, except for alanine in pure $[C_2C_1Im][Cl]$. The diffusion coefficients increase as the water is added to the system and this observation is true for both the ions and the alanine. The increase in the water content increases the self-diffusion coefficients as water acts to screen the electrostatic interactions among the ions of ILs. Generally, cation diffusion coefficients are in the same region or higher relative to anion diffusion coefficients and the difference is larger for pure ILs. This difference reduces as water content increases,[233] consistent with the coefficients of the systems that are just discussed.

Table 7. 4: Self-diffusion coefficients calculated from MD simulations for alanine systems in pure and IL/water mixtures of [C_2C_1Im][Cl] and [C_2C_1Im][CH_3CO_2]

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s	
		T=300K	T=400K
alanine [C_2C_1Im][Cl]	alanine	1.23	0.79
	[C_2C_1Im]	1.27	4.72
	Cl	1.27	3.98
alanine [C_2C_1Im][Cl]-wat	alanine	1.41	1.75
	[C_2C_1Im]	1.52	5.34
	Cl	1.50	4.47
	water	2.39	13.9
alanine [C_2C_1Im][CH_3CO_2]	alanine	0.03	1.08
	[C_2C_1Im]	0.04	2.50
	[CH_3CO_2]	0.03	1.75
alanine [C_2C_1Im][CH_3CO_2]-wat	alanine	0.20	1.22
	[C_2C_1Im]	0.20	2.87
	[CH_3CO_2]	0.19	2.10
	water	2.86	6.69

It could be speculated that acetate anions, which show better solvation of alanine molecules, will form a stronger solvation shell about alanine or other previously discussed amino acids relative to chloride anions, and thus, the self-diffusion coefficient is lower in these systems. Another possibility for slowing down of the molecules can be the inter-amino acid interactions that causes them to cluster and hence lower the diffusion coefficient. Alternatively, this decrease could arise from the amino acid molecules species residing more in the charge-enriched locales (near the charge centres on the ions) rather than in the non-polar regions, that would result in increased AA-IL ions interactions and hence an overall reduction in the self-diffusion coefficient of amino acids. By solely looking at the diffusion coefficient values, it becomes difficult to suggest the factor contributing more towards the decrease in the D values but combining these values with the observations concluded from RDFs, SDFs and domain analysis can help in understanding the dynamics of the system.[197]

7.3.5 Interactions of $[C_2C_1Im][CH_3CO_2]$ IL with aromatic amino acids at 300K

The interactions for aromatic amino acids were analysed at 300K for an IL:AA ratio of 5:1 to observe the effect on interactions as the nature of solute is changed from charged aliphatic amino acids such as lysine and aspartic acid to the neutral aromatic ones in $[C_2C_1Im][CH_3CO_2]$. Aromatic groups have been strongly established as having highly favourable interactions with ILs.[253] As such, the interaction of aromatic side chains could be expected to have important interactions. The SDFs for the four (neutral) aromatic amino acids (Figure 7.15) show that the interactions vary substantially between the different classes of aromatic.

The heterocyclic aromatics, histidine and tryptophan show a strong interaction with the cations above the plane, and additional strong anion interactions in areas where hydrogen bonding might be expected. This contrasts with the two benzene-based derivatives. Here, at the selected isovalue of 10 nm^{-3}, there is very little interaction of the ring seen with the ionic liquid ions, with the predominant interactions established with the charged zwitterionic components, and, in the case of tyrosine, with the hydrogen bonding phenolic OH.

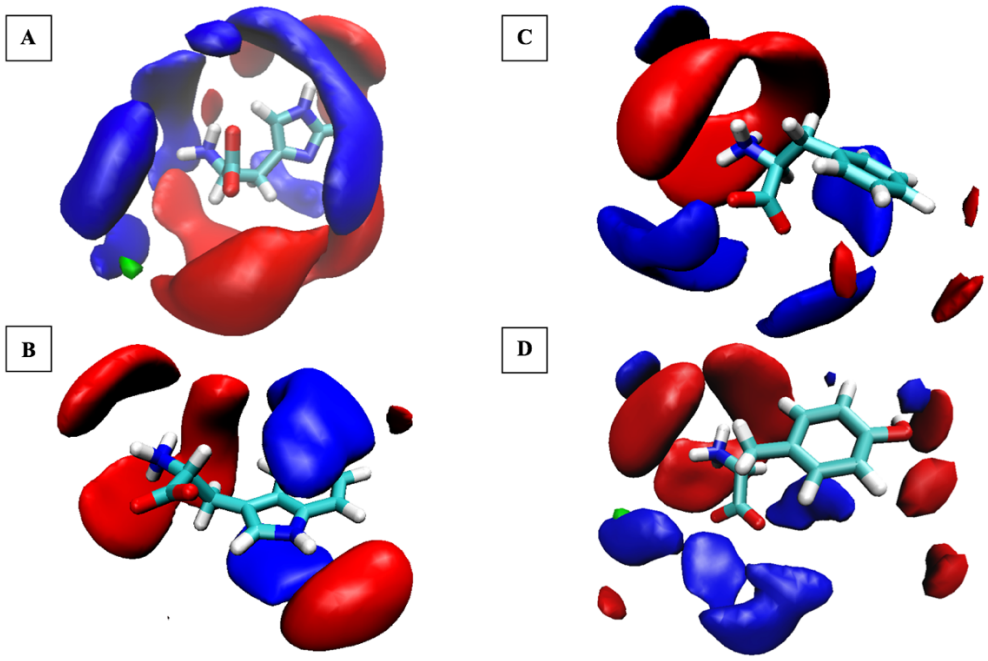


Figure 7.15. Spatial distribution functions around the central (A) histidine, (B) tryptophan (C) phenylalanine and (D) tyrosine at 300K $[C_2C_1Im][CH_3CO_2]$ IL in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$ and green= amino acid at iso-value=10 nm^{-3}.

Around the backbones of each of the amino acids, different patterns of interaction are can be clearly seen in the SDFs and the corresponding RDFs. The atom numbering used for the construction and explanation of RDFs are shown in Figure 7.1. The distribution of anion O1 atoms around the carboxyl group, backbone nitrogen and alpha-hydrogen atom on all amino acids are shown in figure 7.16. The distribution of anions around the backbone nitrogen differs for each amino acid. Phenylalanine and tyrosine, which are similar in structure and differ only with respect to an O-H group on the ring, show a similar pattern for anions distribution around the backbone nitrogen. They differ however, in respect to the peak heights, and the g(r) is larger for phenylalanine (nearly 18) than for tyrosine (around 14). This suggests that small electronic changes in the side chain can have a significant impact on the backbone arrangement, some of which may not be visualised at the iso-value of 10 nm^{-3} in SDFs. The peaks for tryptophan and histidine also occur at the same distance of 2.5 Å, however they are quite small relative to the phenylalanine and tyrosine, suggesting a much more significant impact induced by the heterocyclic rings, which is probably related to the neighbouring strong cationic density seen around the ring structures. (Figure 7.15)

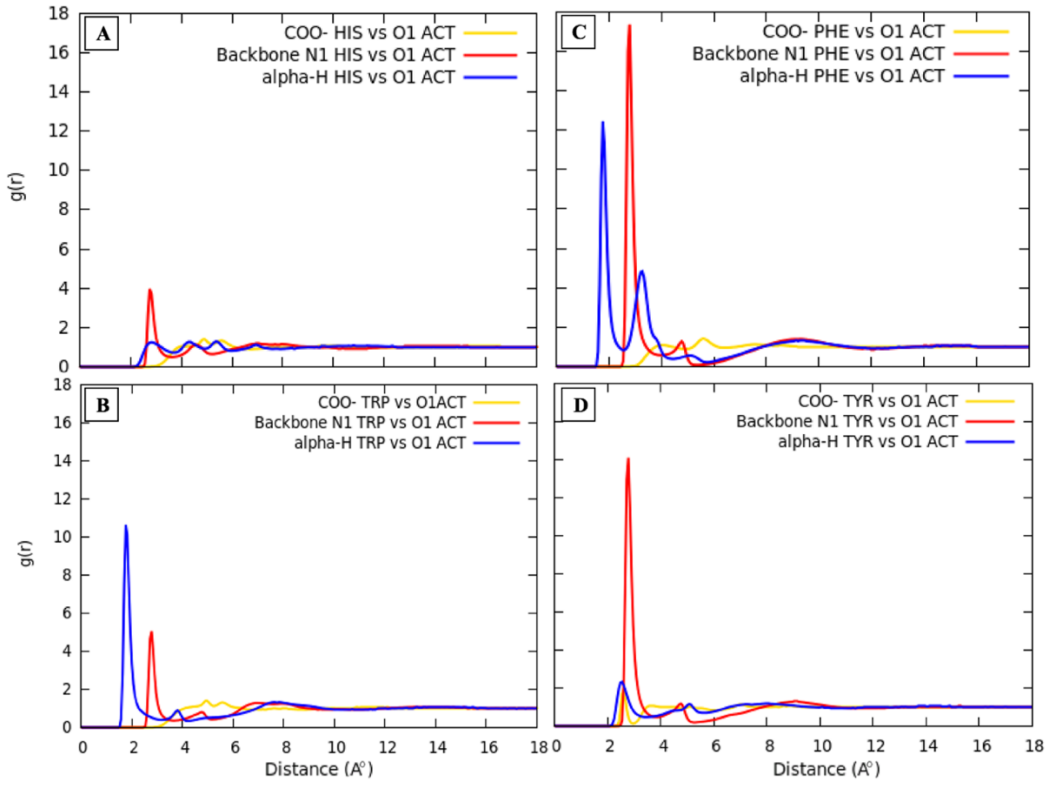


Figure 7. 16: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:[AA] (5:1) at 300K for O1 atom on acetate anion *vs* backbone nitrogen, carboxyl group and α-hydrogen on (A) Histidine (B) Tryptophan, (C) Phenylalanine and (D) tyrosine determined by MD simulation

The organisations and the peak heights for phenylalanine and tyrosine are comparable to those observed for lysine and aspartate in $[C_2C_1Im][CH_3CO_2]$. Nevertheless, the backbone nitrogen atomic site shows a more significant interaction with anions, with the overall order for backbone nitrogen interactions with anions in aromatic amino acids being phenylalanine > tyrosine >> tryptophan > histidine. The distribution for H1 ring hydrogen atom on $[C_2C_1Im]$ cation around the backbone carboxyl group and nitrogen are shown in figure 7.17. The electrostatic interactions are dominant around the carboxyl group with the closer peaks for $[C_2C_1Im]$ cation, highest being for phenylalanine and lowest for tyrosine, the trend followed here is phenylalanine > tryptophan > histidine > tyrosine. This demonstrates that the side chains have substantially different impacts on each side of the amino acid backbone, and this will be important in peptides as it suggests a directional interaction.

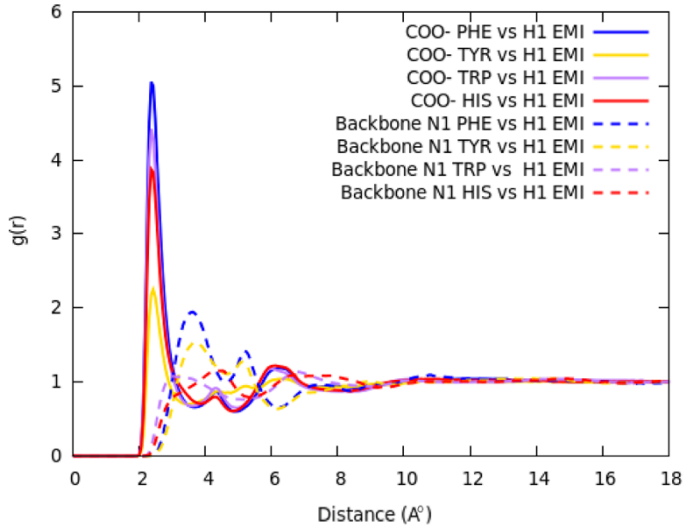


Figure 7. 17: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: [AA] (5:1) at 300K for H1 atom on imidazolium cation *vs* backbone nitrogen (dashed lines) and *vs* carboxyl group (solid lines) on aromatic amino acids determined by MD simulation

Histidine in its neutral form can have attractive cation-pi interactions that means that the imidazole in the side chain can interact strongly with the neighbouring cations and also is able to interact with the neighbouring aromatic amino acids, in this case another histidine molecule. This interaction is weaker than the interactions with the IL cations,[254] and that is reflected in figure 7.15. The behaviour is slightly different from other aromatic amino acids discussed above with an additional cationic shell appearing behind the backbone nitrogen. Comparing the figures 7.16 (A) and 7.17 for histidine show that the closest shells are formed by the cation around the carboxyl group around 2.5 Å, followed by the backbone nitrogen interacting with the anions around 4 Å,

clearly visible in the corresponding SDFs. As the temperature is raised, these cation interactions are most affected, as shown in the figure in appendix A.25.10.

Cation-anion and anion-anion interactions are shown in SDFs in appendix in figure A.25.11 and show a similar organisation to all previously discussed amino acids in pure $[C_2C_1Im][CH_3CO_2]$ except for tryptophan where cation-cation interactions are not that strong. For other aromatic amino acids, the magnitude of the cation-cation interaction relative to alanine is weak but the overall distribution of interactions is otherwise similar.

The domain analysis shows that the highest number of domains are formed for phenylalanine and the lowest for tryptophan. So this might indicate that anions play a major role in solvating the aromatic amino acids. Phenylalanine is able to interact well with both the ions of IL whereas histidine is more attractive to cations and hence the difference in the domain counts.

Table 7. 5: Domain analysis for aromatics systems in pure $[C_2C_1Im][CH_3CO_2]$ IL

Ionic Liquid	Amino acid	Temperature (K)	Domain count (average)		
			AA	$[CH_3CO_2]$	$[C_2C_1Im]$
$[C_2C_1Im][CH_3CO_2]$	Histidine	300K	11.66	182.92	5.64
	Phenylalanine		14.5	182.03	6.27
	Tryptophan		4.86	182.86	7.05
	Tyrosine		6.81	185.9	6.43

Self-diffusion coefficients for all aromatic amino acids are calculated following the same procedure followed for alanine amino acid and the MSD curves are shown in appendix in A.25.12. The diffusion coefficients are highest for histidine and the trend followed here is histidine > phenylalanine > tyrosine > tryptophan. Diffusion coefficients are helpful in a way that they help predicting about the fluidity and viscosity of the system at hand but they are not that straightforward to relate. The self-diffusion coefficients calculated for the amino acids solutes here are not independent from the effect of the nearby solutes. And they are interacting with the solvent ions as well.

Table 7. 6: Self diffusion coefficients calculated from MD simulations for aromatic amino acids in pure and IL/water mixtures of [C_2C_1Im][Cl] and [C_2C_1Im][CH_3CO_2] IL

System		Self-Diffusion coefficient = slope * 10/6 = D * 1e-10 m^2/s T=300K
[C_2C_1Im][CH_3CO_2]	histidine	0.54
	[C_2C_1Im]	0.58
	[CH_3CO_2]	0.55
	phenylalanine	0.14
	[C_2C_1Im]	0.15
	[CH_3CO_2]	0.14
	tryptophan	0.12
	[C_2C_1Im]	0.13
	[CH_3CO_2]	0.13
	tyrosine	0.12
	[C_2C_1Im]	0.14
	[CH_3CO_2]	0.13

7.4 Conclusions

Alanine is the simplest of the non-polar amino acids. The charged solutes have been analysed in the previous chapters and it has been shown that the nature and charge of the solute has varied effects on the structuring of two ionic liquids. Different roles of anions and cations were highlighted for two ionic liquids having the same imidazolium cation. The charge of the solute and the localisation of the charge also influenced the way the IL ions occupy certain regions and impact the overall organisation as seen by comparing arginine and lysine or aspartate and glutamate. The electrostatic interactions however had a dominant role to play. The temperature also had a key role to play that influenced these organisational patterns. So, in order to be able to see the effect of a neutral side chain in non-polar amino acid, zwitterionic form of alanine was selected. The molecular details through SDFs reveal that there are strong alanine-alanine interactions in [C_2C_1Im][Cl] IL at lower temperature, with anions dominating the regions around ammonium and cations are able to occupy the regions around the carboxylate group but this interaction completely disappeared for high concentration systems at the chosen iso-value. Anions were also able to interact through the methyl group, but position of the shell indicated the it might be resultant effect from the direct

electrostatic interactions of the cation and alanine molecule. RDFs comparisons revealed a prominent increase in the alanine-alanine interactions at a higher temperature, similar to the behaviour of lysine and aspartate in $[C_2C_1Im][Cl]$, however water had just a marginal effect. At lower temperatures, water was able to interact with both the ions of the ionic liquid and also reduced the inter-alanine interaction, but at higher temperature a similar aggregation effect was observed.

The investigation of these results reveals that the $[C_2C_1Im][CH_3CO_2]$ IL showed better solvating properties and that cations had more defined organisation around the alanine molecules relative to the chloride anions. A more stable organisation of the cations was observed in the presence of acetate anion; the strengths, however, were reduced. The values of self-diffusion coefficients were higher for amino acids in $[C_2C_1Im][Cl]$ IL relative to $[C_2C_1Im][CH_3CO_2]$ IL and is in line with what has been previously observed for other amino acids. In general, the species interacting more strongly in a given system have lower self-diffusion coefficient values. This lowering of D values in $[C_2C_1Im][CH_3CO_2]$ IL would suggest stronger association of these amino acids and the IL ions. IL ions can entrap the small solutes thus slowing down their motion. Aromatic amino acid interactions were analysed with $[C_2C_1Im][CH_3CO_2]$ IL. The cation-anion interactions were similar to those observed for other systems discussed, however, lower cation-cation interaction was observed for tryptophan. The cation favourable regions lie mostly around the aromatic ring and the closer shells are formed for cations, followed by the acetate anions. Histidine shows a relatively different organisation as compared to other amino acid in terms of anion dominance around the ammonium but that can be explained in terms of the effects of the strong ring organisation impacting further down on the overall organisation.

8

Structure and interactions of peptides in water and ionic liquids

8.1 Introduction

Proteins are a heterogeneous class of biological macromolecules and it is very important for them to maintain their secondary structure for their proper functioning. The secondary protein structures include α-helices, β-sheets, and coil regions and these structural elements are held together via a complex balance of hydrogen bonds, disulfide bridges, hydrophobic and ionic interactions.[156, 255] Solvent molecules can participate or compete with the amino acids units in most of these interactions. Proteins secondary structures are severely affected by changing solvent conditions like viscosity, pH value, buffer conditions, presence of (ionic) co-solvents, and temperature. The process by which proteins attain their functional structure has been the subject of numerous experimental and theoretical studies over the last few decades.[126, 127] Proteins, perform most of the biologically-relevant functions in living organisms, attain a unique tertiary structure in short time scales. The fastest folding time measured is in microseconds (μs),[256] which is now relatively easily accessible with computational simulation. Dis-functional states of proteins that fold into the incorrect structures are believed to be the cause of degenerative diseases.[257, 258] Recent theoretical advances and modern experimental techniques have made it possible to probe proteins at different stages during the folding process. Due to this, the nature of the underlying physical mechanisms and relevant interactions that determine the kinetics of folding, binding, function, and thermodynamic stability can be assessed.[259] ILs have been reported as useful solvents for biomolecules such as proteins, enzymes, and nucleic acids, such as DNA[260, 261] and have the ability to interact with both protein and water. This ability leads to varied functionalities of ILs towards proteins and enzyme structures, with enzyme function being able to be either stabilised or destabilised by the addition of

ionic liquids.[61, 155, 262, 263] The applications of ionic liquids have been comprehensively summarised in a review by Patel *et al.*[261] Both short peptides and small organic molecules have been shown to mimic many of the properties of proteins and biomaterials. The molecules have proven to be useful for the testing of the computational methods and also to observe and explain for example, folding of protein structures in atomic detail, or to explain the underlying molecular mechanism for solvation behaviour of natural polymers, such as glucose.[123, 264-266] Consequently, an investigation of smaller peptide systems is desirable for understanding the effect of ILs on well-defined secondary structures in order to better understand and predict the composite effect felt by more complex proteins. Therefore, in order to analyse the interactions and to observe how the organisation examined for the charged amino acids changes when in a peptide, and the subsequent impact on peptide secondary structure, peptide simulations were carried out for three different peptides.

For this purpose, short peptides with a defined folding configuration in water were identified from the literature. Protein G, which is an immunoglobulin-binding protein expressed in group C and G Streptococcal bacteria, a series of 8-mer peptide fragments with overlapping sequences from the original protein sequence were systematically generated by Ho, Bosco and Dill.[7] The 8-mers were chosen because this length appears adequate to identify elements of structure in PDB studies,[267] and because much longer fragments become too expensive for extensive computer simulations. In each case, Ho *et al.* determined whether the peptide converges to its native conformation in the folded protein or not.[7] A representative sequence was selected, whose native conformation was a hairpin structure, because it was found that this peptide segment spent most of the time near its respective native structures, relative to other 8-mers. Various peptides of Protein G has been studied previously to analyse the folding behaviour in water and aqueous ionic liquids.[155, 268]Along with these peptides, alanine-based peptides, namely penta-alanine (or $(Ala)_5$) and deca-alanine (or $(Ala)_{10}$) were chosen for the current study to analyse the structure and interactions in both water and ionic liquid and also to compare the results with the single amino acid systems discussed in chapter 7. Alanine is an attractive system as it provides a simple control, and is one of the simplest amino acids as all the side chains are non-polar,[269] in contrast to the amino acids already examined. The structure of polyalanine peptides appears to be exquisitely sensitive to its environment and peptide length.[269] Simulation studies, conducted by Hall and

coworkers[270, 271] using a coarse-grained model of the peptide and solvent, identified that polyalanine has a rich aggregation phase diagram in a water-like environment, with the coexistence of helical, random coil, and fibrillar and nonfibrillar β-sheets.

8.2 Materials and methods

8.2.1 Outline of simulation conditions

The three peptides were selected from literature LNGKTLKG (ACE(Leu-Asn-Gly-Lys-Thr-Leu-Lys-Gly)NME), the 5-mer AAAAA (ACE-(Ala-Ala-Ala-Ala-Ala)NME) referred to as penta-alanine and the 10-mer AAAAAAAAAA (ACE-(Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala-Ala)NME), referred to as deca-alanine, as explained in the Chap 3, section 3.5. All simulation were carried out at 300K for 1000ns. Table 8.1 shows the list of simulations that were carried out.

Table 8. 1: List of peptides simulations carried out in water or in $[C_2C_1Im][CH_3CO_2]$ IL

Sequence (one letter code)	Solvent media	Native structure	Peptide Force-field	Simulation time
LNGKTLKG	water	β-hairpin	FF99SB	1000 ns
LNGKTLKG	$[C_2C_1Im][CH_3CO_2]$		FF99SB	1000 ns
AAAAA	water	helix	FF99SB	1000 ns
AAAAA	$[C_2C_1Im][CH_3CO_2]$		FF99SB	1000 ns
AAAAA-AAAAA	water	helix	FF99SB	1000 ns
AAAAA-AAAAA	$[C_2C_1Im][CH_3CO_2]$		FF99SB	1000 ns

The parameters of simulation have already been discussed and the methodology followed for constructing the required peptide structures and running simulations has been explained in chapter 3, section 3.5. Packmol was used for creating peptide cubic boxes in $[C_2C_1Im][CH_3CO_2]$ IL and truncated octahedron boxes were used for peptide simulations in water. The 3-D images for these two simulation boxes and the starting structures for the three peptides are shown in Figure 8.1 and 8.2 respectively.

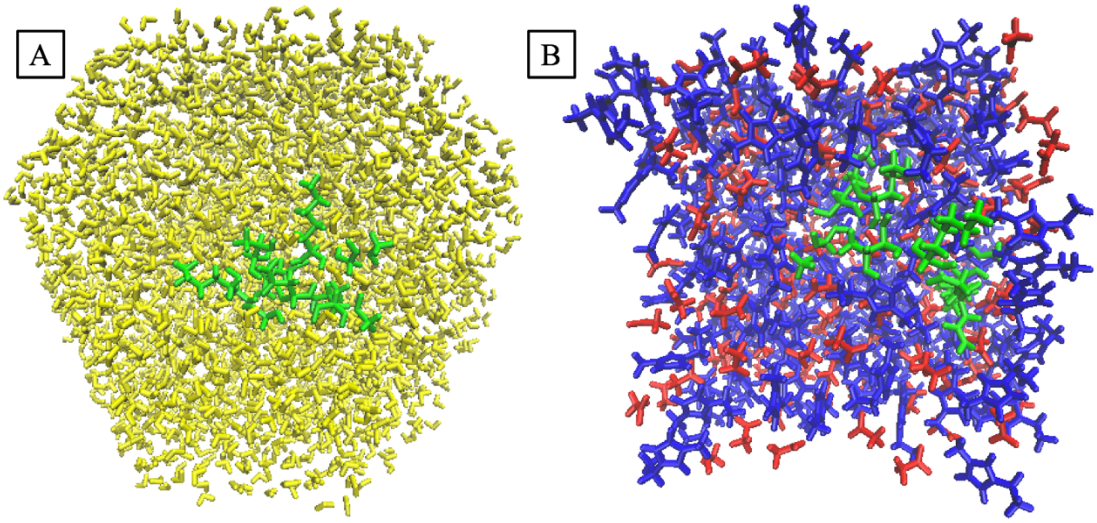


Figure 8. 1: Three-dimensional representation of peptide LNGKTLKG in (A) octahedron water box and (B) cubic $[C_2C_1Im][CH_3CO_2]$ IL box. Image taken from 1000ns simulation of deca-alanine peptide in two solvent media.

A truncated octahedron is more spherical than a cube and computationally less expensive for solvent molecules distant from the solute. The box shape itself influences the nature of the motions that the system can undergo by restricting sampling in certain directions, and thus affect the outcome of the simulation. For boxes with a relatively large amount of solvent compared with the size of the solute molecule, this effect will be small and can be neglected. However, when one is using a box containing only a minimal amount of solvent, the effect may well be significant.[272]

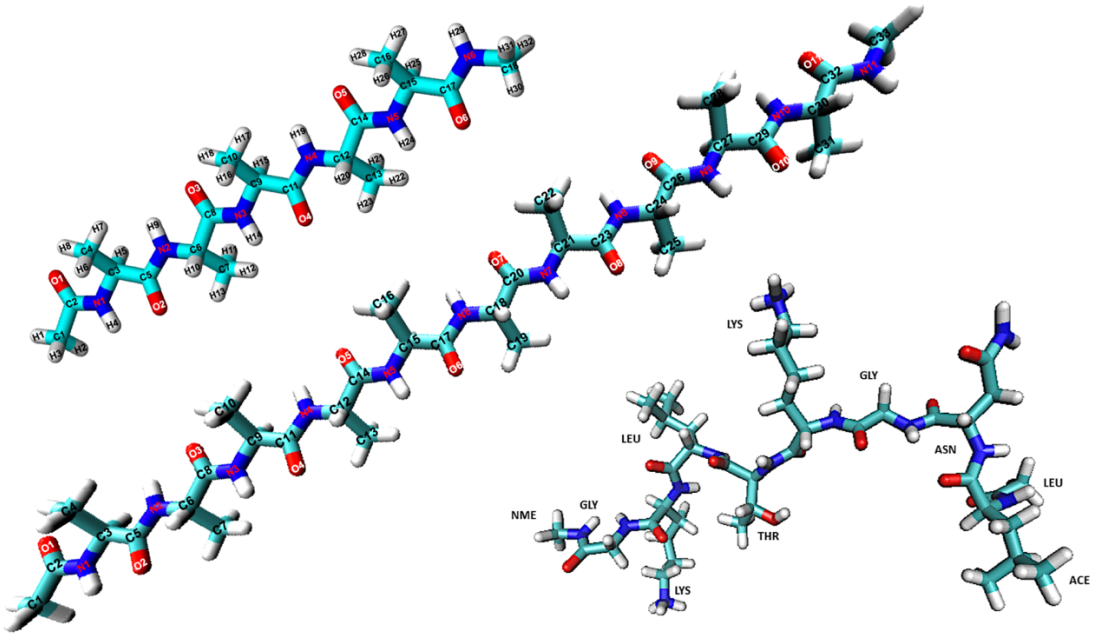


Figure 8.2. Starting structures of peptides penta-alanine, deca-alanine and G-protein fragment LNGKTLKG.

8.3 Results and discussions

8.3.1 Structure and interactions of Alanine (5-mer) peptide in pure $[C_2C_1Im][CH_3CO_2]$ IL and water at 300K

The structure of the alanine penta-peptide in ionic liquid was assessed through the radial distribution functions analysed by assessing the distribution of (a) the oxygen atoms of the acetate anions around the backbone nitrogen atoms and (b) the acidic proton on $[C_2C_1Im]$ cation around the oxygen atoms of the carbonyl groups of all alanine residues. The distribution pattern is quite similar for all the residues in the peptide, differing with respect to peak heights. Both cations and anions interact with the alanine residues in the peptide in a similar way to the interactions that are observed for the single amino acid simulations in the previous chapter 7. The interactions of ionic liquid ions for single amino acids were stronger relative to those observed here, consistent with no overall charge in the peptide models. In figure 8.3, it can be seen that the strongest interactions for acetate anions around the backbone nitrogen atoms occurs for alanine at the 2nd and the 5th positions, with a similar pattern for two residues in the middle. This indicates that the position of a residue is significant for the interactions observed. The cation distribution is not much affected by the position, but a slight difference is observed for alanine 2, where the O2 atom has a slightly stronger peak for interaction with the $[C_2C_1Im]$ cation. The SDFs are shown in the figure 8.4.

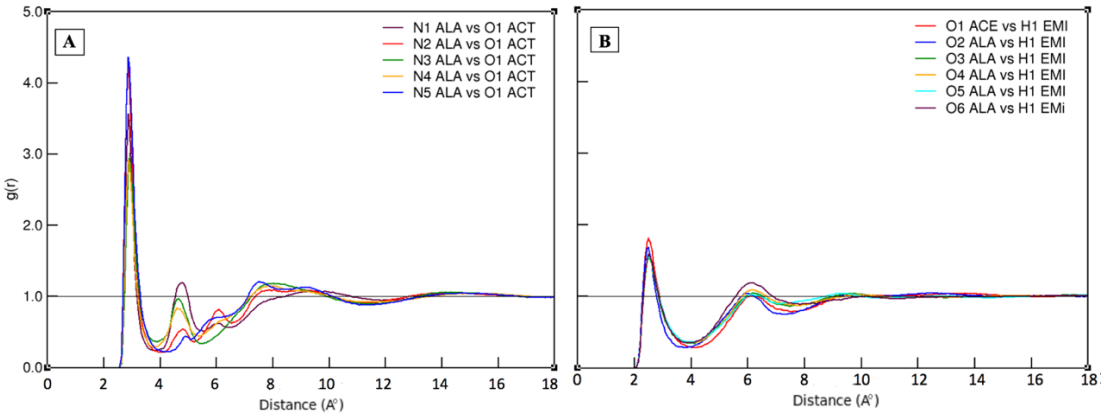


Figure 8.3. Centre-of-mass radial distribution functions for (A) acetate anion oxygen atoms around backbone nitrogen atoms and (B) acidic proton on imidazolium cation around oxygen atoms on alanine residues in alanine-5 peptide dissolved in pure $[C_2C_1Im][CH_3CO_2]$.

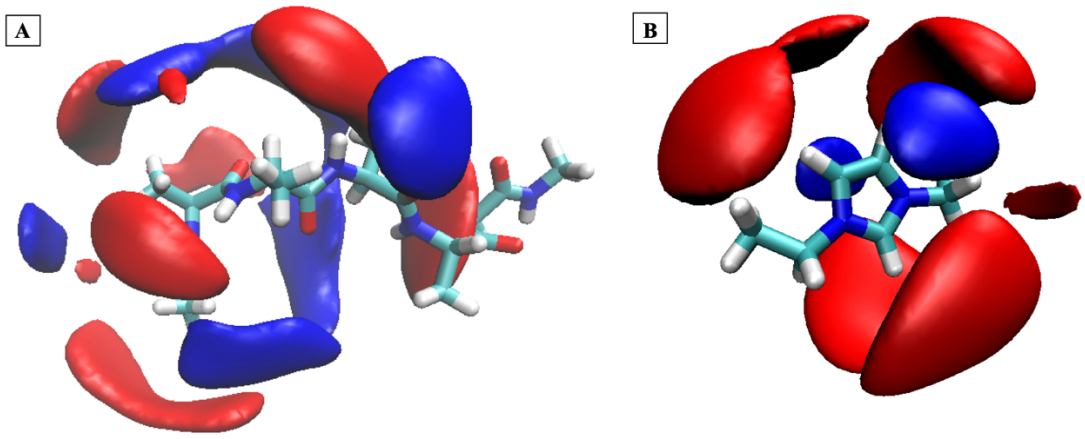


Figure 8.4. Spatial distribution functions around the central (A) alanine peptide and (B) $[C_2C_1Im]$ cation in peptide-$[C_2C_1Im][CH_3CO_2]$ solution at 300K. Blue = $[C_2C_1Im]^+$ and red = $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.

The RDFs tend to suggest that the original types of interaction of ions with alanine analysed individually is retained in the peptide sequence, however the strength of these interactions is reduced, in line with going from charged zwitterionic systems to polar, uncharged peptide bonds. This can be concluded from the SDFs as well by comparing the cation-anion interactions for alanine amino acids in Chapter 7, figure 7.11 (C) and the figure 8.4 (B), showing that the anions have stronger interactions with the cation relative to when single amino acids are involved. An added reason for this observation is that the single amino acids are more exposed to the surrounding fluid, and so there should be a decrease in the interaction strength for peptide components, which can be partially shielded through the flexibility of the chain. Despite these differences, these simulations show that the organisation around single amino acid molecules can help in predicting the interactions that will be established when these residues are incorporated into a peptide.

To compare the peptide simulations in ionic liquid relative to those carried out water, the structures of the peptides were analysed in both environments by constructing the Ramachandran plots and analysing the secondary structures over the entire simulation. The Ramachandran plot is important to structural biology as it describes a peptide backbone in the context of its backbone dihedral angles φ and ψ.[200]. Since its introduction, the Ramachandran plot has been a crucial tool to characterise protein backbone features. The φ and ψ values are usually constrained within specific areas of the plot due to steric hindrances that occur between adjacent atoms within a protein

structure, particularly for ordered structures such as helices and sheets. Ramachandran *et al.*[273] designed the plots to identify the permissible values of φ and ψ for all amino acids (except glycine and proline), so that the theoretical prediction of conformations of either protein or polypeptide backbones could be achieved.

A Ramachandran plot comparison for peptide simulations of penta-alanine in water and IL is shown in figure 8.5. Both systems were constructed by taking the same non-standard extended alanine pentapeptide, so this comparison depicts how the peptide folds, or tries to attain a defined secondary structure, in the two environments. The residue-wise Ramachandran plots are compared and shown in the Appendix in figure A.26.1. The top left and bottom left quarters of the figure represent φ and ψ values associated with the β-sheet and the right-handed α-helix structures. The more the region is populated, the more it is favourable for the amino acids within the peptide or protein to attain that structure, and thus that fold to form. Figure 8.5 (A) shows that the peptide is quite distributed in both these regions and some of the residues, especially the alanine residue in the middle, have a high tendency to populate the left-handed α-helix structure *i.e.* the top right quarter of the Ramachandran plot. The peptide is quite flexible in water and is able to be present in both main structural forms, but there is a slight dominance of β-sheet structure. In the case of $[C_2C_1Im][CH_3CO_2]$, the left-handed helix structure is completely lost or the residues are not flexible enough in the ionic liquid medium to attain this structure, possibly due to the presence of strong electrostatic interactions either restricting movement through cage-like effects or directly interacting with the peptide. These areas might also be restricted due to the difference in the viscosity of the solvent media that can have impact on how proteins fold. Water is less viscous, and a peptide can sample a greater range. Each alanine residue has a different behaviour in terms of the permissible values of φ and ψ angles.

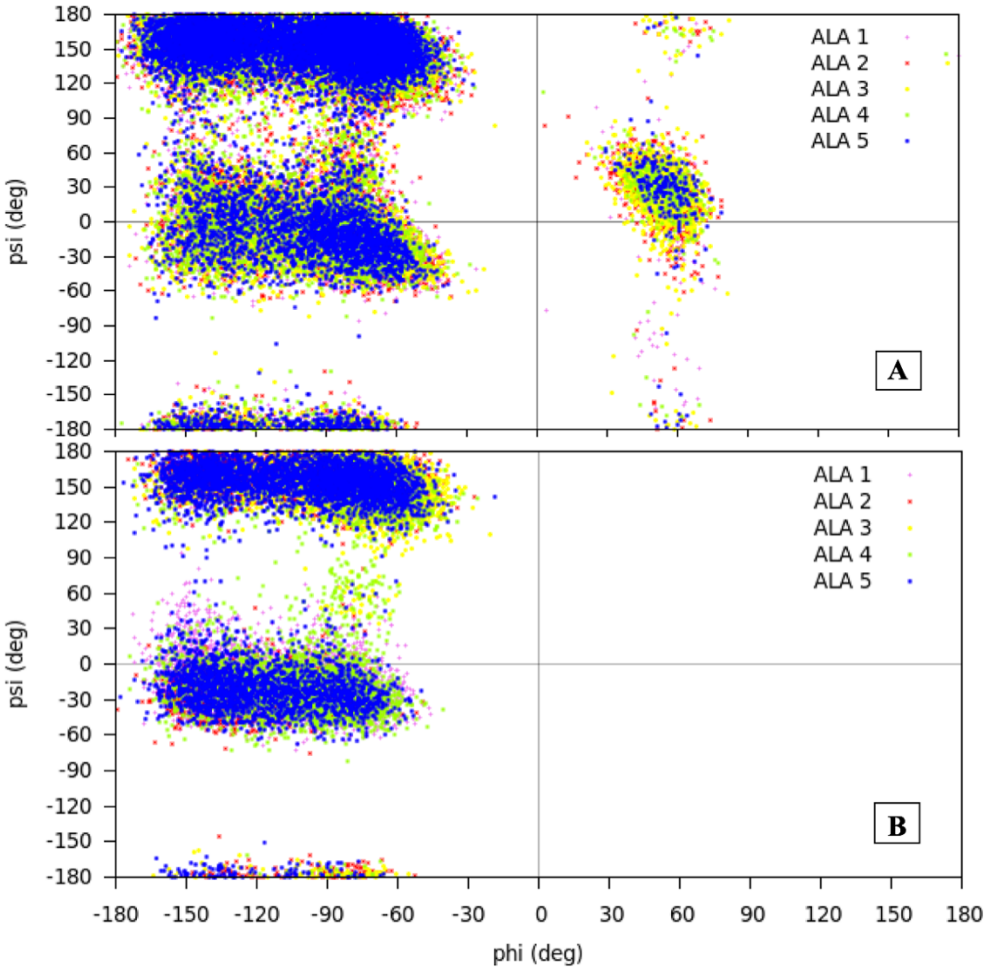


Figure 8.5. Ramachandran plots for penta-alanine or $(ALA)_5$ peptide in (A) water and (B) pure $[C_2C_1Im][CH_3CO_2]$.

Secondary structure analysis is shown in figure 8.6. This analysis indicates that, over the entire simulation time, the secondary structure fluctuates between bend, turn and α–helix with the dominance mostly of turns or bends in water, whereas in $[C_2C_1Im][CH_3CO_2]$ IL, these secondary structural elements are not achieved with a bend seen after 800ns simulation time.

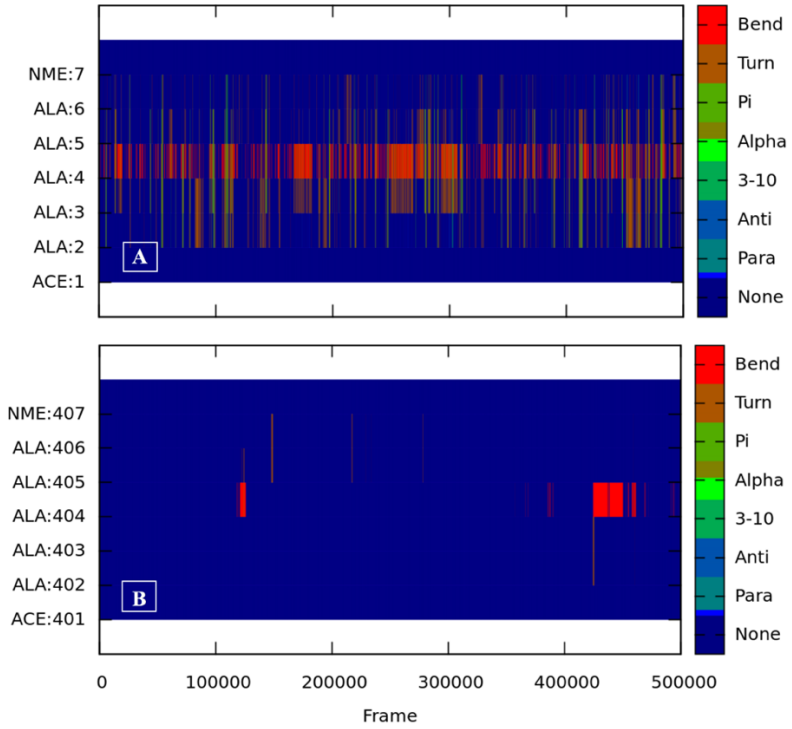


Figure 8.6. Secondary structure analysis of penta-alanine in (A) water and (B) pure $[C_2C_1Im][CH_3CO_2]$

8.3.2 Structure and interactions of Alanine (10-mer) peptide in pure $[C_2C_1Im][CH_3CO_2]$ IL and water at 300K

Related analyses were run for deca-alanine peptide. RDFs and SDFs in figures 8.7 and 8.8, respectively, show a similar pattern of interactions as noted for the penta-alanine. The organisation of cations and anions is consistent with that which has been observed for all the single amino acids and the penta-alanine at 300K and indicates that regardless of the nature of solute, $[C_2C_1Im][CH_3CO_2]$ tends to organise around the backbone nitrogen and carbonyl oxygens in a similar way, however, the strength of interactions varies.

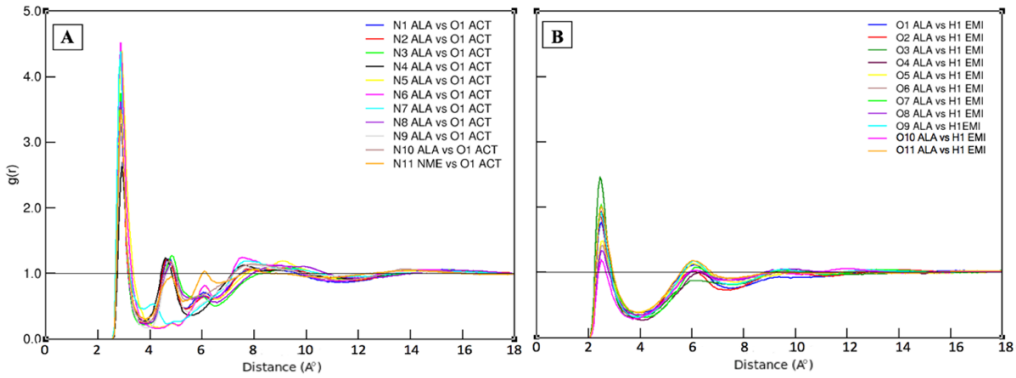


Figure 8.7. Centre-of-mass radial distribution functions for (A) acetate anion oxygen atoms around backbone nitrogen atoms and (B) acidic proton on imidazolium cation around oxygen atoms on alanine residues in $(ALA)_{10}$ peptide in pure $[C_2C_1Im][CH_3CO_2]$.

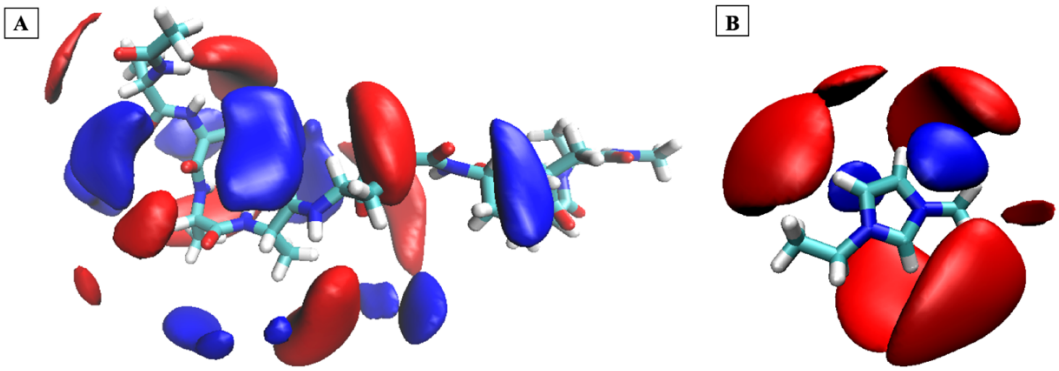


Figure 8.8. Spatial distribution functions around the central (A) $(ALA)_{10}$ peptide and (B) $[C_2C_1Im]$ cation in pure $[C_2C_1Im][CH_3CO_2]$ IL at 300K. Blue = $[C_2C_1Im]^+$ and red = $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.

Comparing the interactions for penta- and deca-alanine, the strength of interactions for both cations and anions are higher for the deca-alanine and the maximum interaction for the backbone nitrogen atoms occurs for the middle residues and lowest for the last three residues attached close to NME group. This is consistent with a cooperative action induced by neighbouring ionic organisation, effectively reinforcing a banded or shell-like structuring along the backbone. The Ramachandran plots for deca-alanine (Figure 8.9) also show an overall similar behaviour in $[C_2C_1Im][CH_3CO_2]$, relative to water, in terms of not occupying any space in the left-handed helix region. The residue-wise comparison for the Ramachandran plots is given in appendix in figure A.26.2. Similarly, the secondary structure analysis shows a highly restrained set of structures for the $[C_2C_1Im][CH_3CO_2]$ solvated peptide, relative to the much more flexible water-solvated peptide (Figure 8.10).

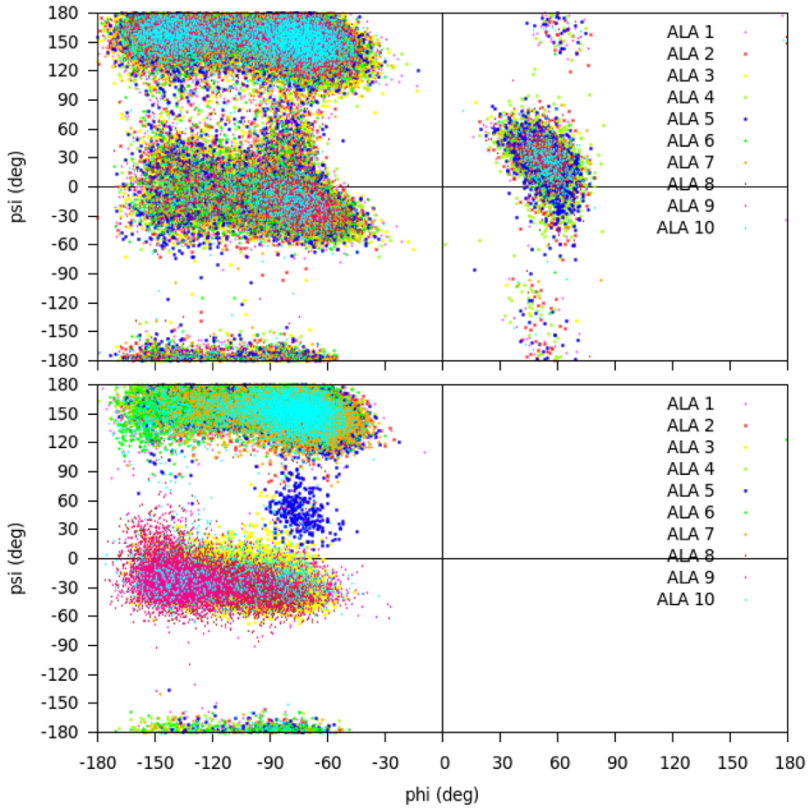


Figure 8.9. Ramachandran plots for penta ALA or $(ALA)_5$ peptide in (A) water and (B) pure $[C_2C_1Im][CH_3CO_2]$.

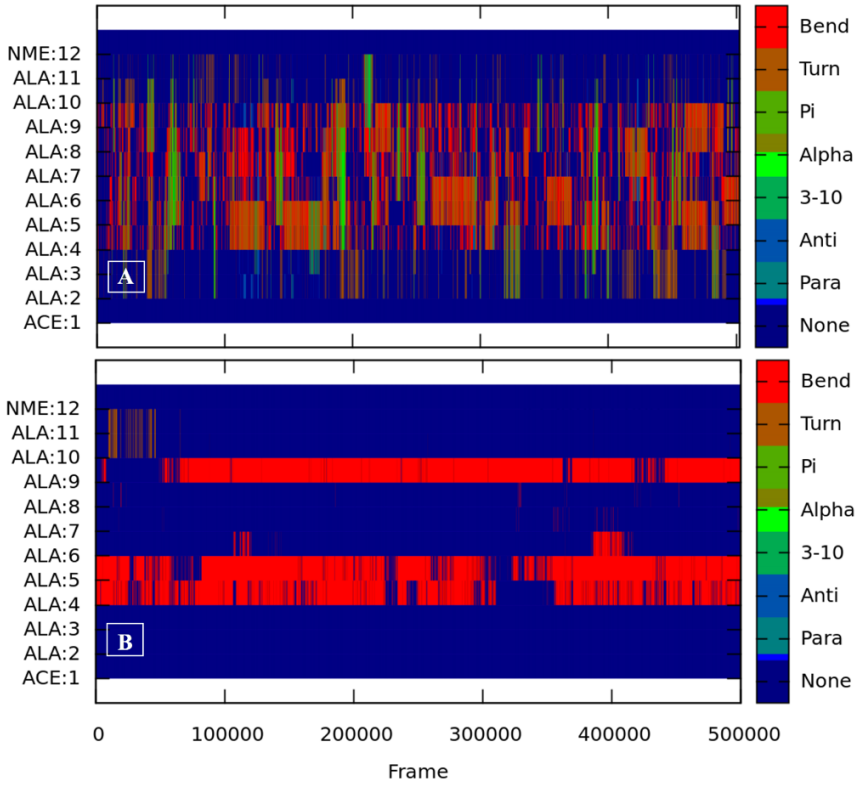


Figure 8.10. Secondary structure analysis of deca-alanine in (A) water and (B) pure $[C_2C_1Im][CH_3CO_2]$.

8.3.3 Structure and interactions of protein G peptide in pure $[C_2C_1Im][CH_3CO_2]$ IL and water at 300K

The protein G peptide sequence in the non-standard form was prepared and simulated in both water and $[C_2C_1Im][CH_3CO_2]$. The peptide sequence selected has a native β-hairpin structure and contains two lysine residues, so can provide insight into the interactions of lysine with $[C_2C_1Im][CH_3CO_2]$ within a peptide. These peptide results can be compared with the ordering observed for a single lysine amino acid, described in Chapter 5, section 5.4. The interactions with acetate anion around the backbone nitrogen are not significantly different for lysine residue in the middle of the peptide sequence relative to the lysine residue at 7th position close to NME. However, lysine at end of the peptide shows a balanced interaction by interacting with the IL cation acidic hydrogen equally well (Figure 8.11 and Appendix Figure A.26.4). The overall pattern of organisation (Figures 8.11 and 8.12) is similar to that observed for the single amino acid shown in Appendix A.14, although the strength of interactions is different. The SDF (Figure 8.12) is drawn by keeping the lys-4 atoms as the centre of the distribution and this confirms the similarity of the ordering, as is shown for single amino acid in chapter 5 (figure 5.4.2). There is clear interaction of the side chain ammonium with acetate, as well as the structuring along the backbone. The strengths of interactions are different depending on the position of the residue in the peptide sequence. This reflects what was observed for the alanine peptides, and strongly suggests that the cumulative organisation, or the lack of flexibility of internal amino acids, has an impact on the local strength of organisation. The pattern of organisation being similar to that seen for the single amino acids further supports that the single amino acid study in neat ionic liquids can help in predicting how the interactions will be seen when they are in a peptide.

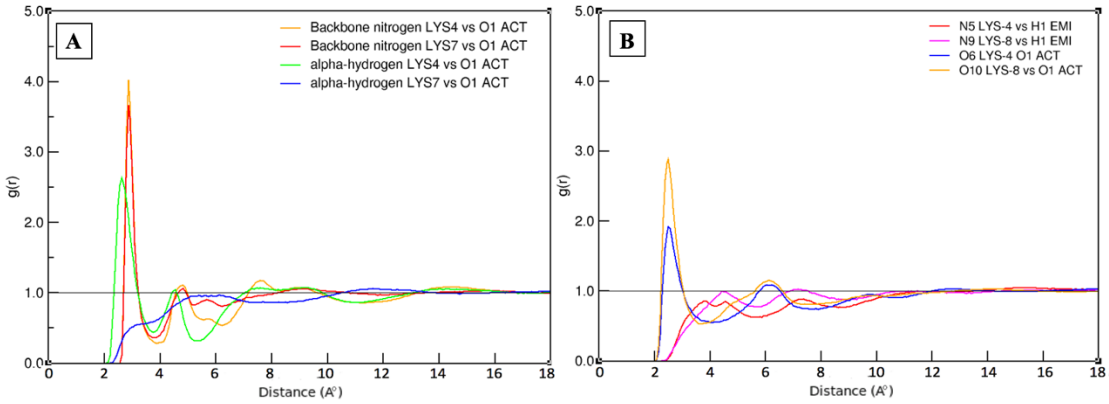


Figure 8.11. Centre-of-mass radial distribution functions for (A) acetate anion oxygen atoms around on two lysine residues' backbone nitrogen atoms and (B) acidic proton on imidazolium cation around two lysine residues' oxygen atoms on G-protein fragment LNGKTLKG in pure $[C_2C_1Im][CH_3CO_2]$ IL

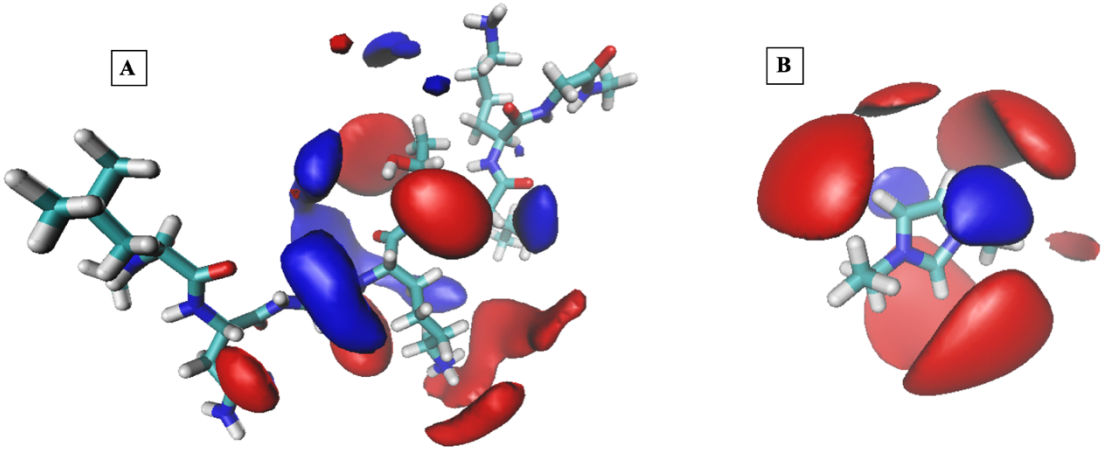


Figure 8.12. Spatial distribution functions around the central (A) G-protein fragment LNGKTLKG and (B) $[C_2C_1Im]$ cation in pure $[C_2C_1Im][CH_3CO_2]$ at 300K. Blue = $[C_2C_1Im]^+$ and red = $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}.

In order to further analyse the role of IL cations and anions, radial distribution functions were drawn between the side-chain carbons of each residue in the peptide sequence and the tail carbon on the $[C_2C_1Im]^+$ cation (Figure 8.13). Comparing the two lysine residues reveals that the lysine residue at the end of peptide chain is able to interact well with $[C_2C_1Im]^+$ cation through the ring protons and the strength of interaction is higher relative to the lysine tail carbons at the middle of peptide.

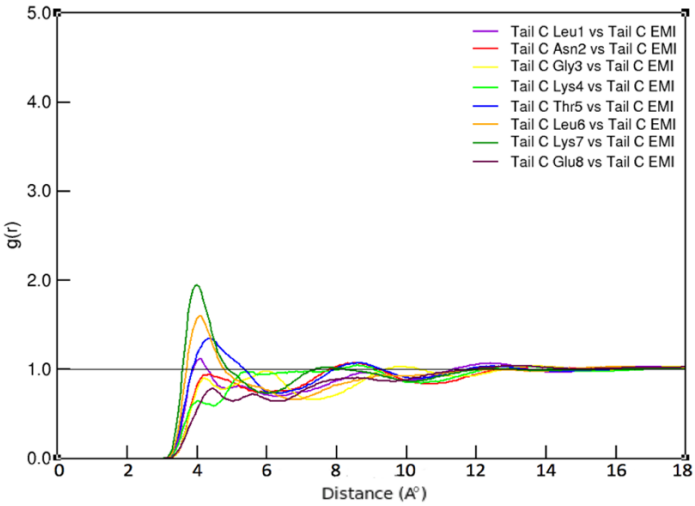


Figure 8.13. Centre-of-mass radial distribution functions for tail carbon on imidazolium cation around tail carbon atoms of each residue on G-protein fragment LNGKTLKG in pure $[C_2C_1Im][CH_3CO_2]$ IL

The Ramachandran plot for the LNGKTLKG peptide sequence is given in figure 8.14 and shows a similar trend in that the left-handed helix region is not accessed for the residues in ILs, as seen for alanine peptides. The two leucine residues and two glycine residues behave in similar fashion in terms of accessing more or less the same regions but the two lysine residues have different densities in different regions (appendix Figure A.26.6). This again might be related to the flexibility of the lysine residue at the end of the peptide chain than compared to the more restricted lysine molecule in the middle of the chain.

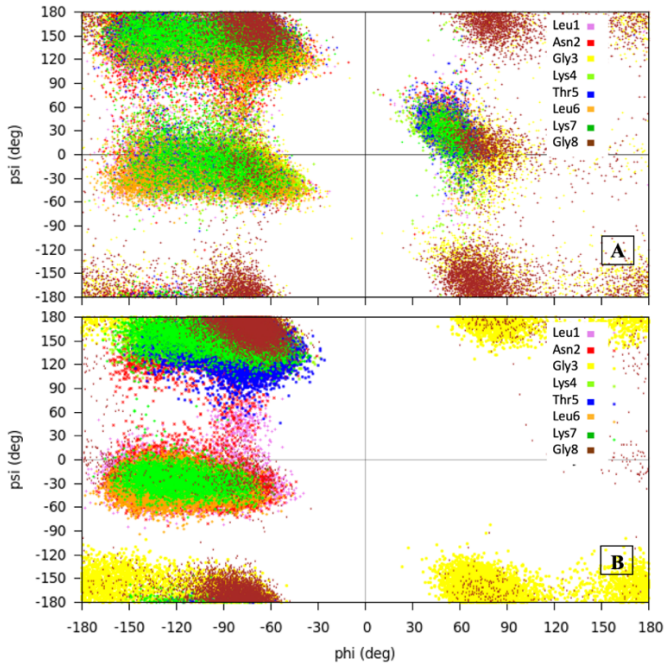


Figure 8.14. Ramachandran plots for protein G peptide sequence in (A) water and (B) pure $[C_2C_1Im][CH_3CO_2]$.

The secondary structure analysis for protein G peptide sequence was done both in water and in neat IL (Figure 8.15). Consistent with the alanine peptides, the flexibility and exchange with different structural forms is greatly diminished in the IL, relative to water. The secondary structure in water is domainated by turn and helix and the final frames indicate that the structure has converged to a turn that is consistent with literature but the peptide seems quite unstable, indicating probably longer simulations were required, whereas in $[C_2C_1Im][CH_3CO_2]$ only bends can be seen. As mentioned before, this could be related to the difference in viscosity of two different solvent media, that prohibits or restricts the movement of peptide in $[C_2C_1Im][CH_3CO_2]$ IL. A separate simulation was run for 1000 ns in $[C_2C_1Im][CH_3CO_2]$ IL to assess the effect on secondary structure. The final hairpin structure obtained in water simulations was taken as the starting structure for this simulation. It was observed that some of the secondary structure was lost in IL as shown in figure in appendix A.26.5. However, to investigate further and support these arguments, experimental evaluation of IL-peptide interactions is essential to explain the flexibility and permitted structural forms of these peptides in IL. Comparing the secondary structures and the Ramachandran plots for all three peptides indicate that $[C_2C_1Im][CH_3CO_2]$ IL has a characteristic effect on all the peptides and the left handed helix is not achievable in the 1000 ns (1 μs) of simulation time in this ionic liquid environment, and no definite secondary structure is obtained.

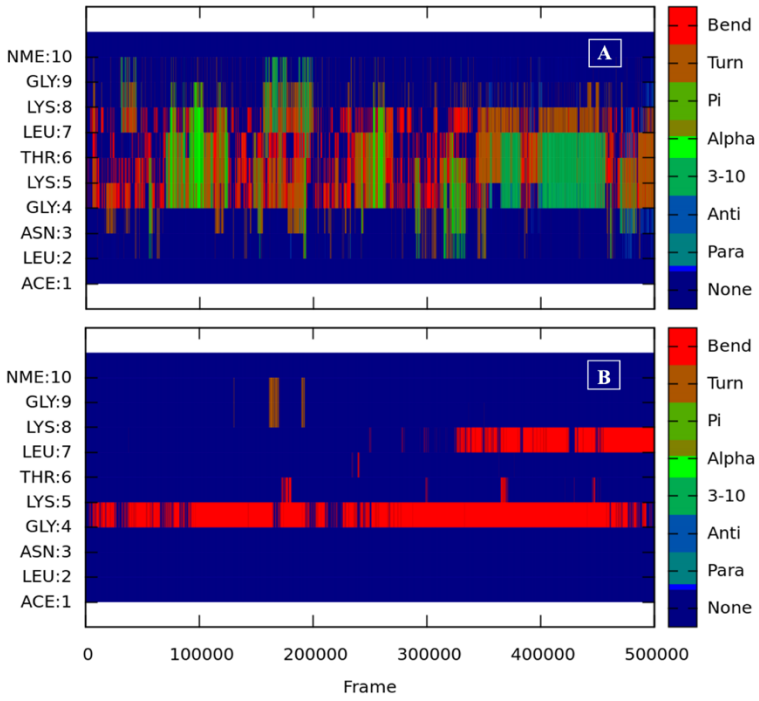


Figure 8.15. Secondary structure analysis of protein G peptide sequence in (A) water and (B) pure $[C_2C_1Im][CH_3CO_2]$.

The protein G peptide segment is inetresting in terms of the combination of hydrohopic aamino acids *e.g.* leu, and other polar residues such as lysine, aspargine and threonine. A general conclusion drawn from the Ramachandran plots can be that dominantly hydrophobic amino acids prefer the alpha-helix region, and the polar amino acids prefer the β-sheet region of the plot, but its is interesting to see that lysine behaves differently at the two different positions in the peptide sequence.

8.4 Conclusions

In these simulations, the peptide structures are able to interact with the $[C_2C_1Im][CH_3CO_2]$ IL through both electrostatic and hydrophobic interactions. The dominance is more for the electrostatic interactions between acetate anions and the backbone nitrogens and IL anion seems to play a major role in attenuating the peptide secondary struture by restricting the movement of the backbone atoms, as observed through ramachandran plots, secondary structure analysis and the interactions identified in the radial distribution functions. The pattern of organisation of cations and anions around the backbone is broadly the same for all the peptides considered in this study and indicate the characteristic features of electrostatic interactions with $[C_2C_1Im][CH_3CO_2]$, but the strengths of the interactions are dependent upon the nature of the residue involved, and its position in the peptide sequence. Clearly this is a multi-dimensional problem that will require more simulations to develop a fully predictive model for interaction with proteins and peptides, however, the initial elements and requirements for developing this understanding have been established in this study.

The results for the organisation around peptides were compared with the related simulations of single amino acids and it was shown that, although the strengths of interactions vary, the single amino acid systems can be very useful in understanding and predicting the key interactions of the comparable peptide-bound residues in ionic liquids. The peptide sequences were studied for the $[C_2C_1Im][CH_3CO_2]$, which has shown good solvation properties for other protein-based materials and it is shown here that the ionic liquid ions, especially the anion, is able to establish strong hydrogen bonds, potentially promoting structure breaking as these hydrogen bonds replace and restrict formation of intramolecular interactions between different residues in the peptides.

9

General Discussion, conclusions and suggestions for future work

9.1 Summary and conclusions

This chapter highlights and summarises the main findings and also include some future work that can be done based on this study.

The main objective of the study, initially, was to analyse the dissolution behaviour of biopolymers and protein-based materials such as keratin. Keratin, in particular is abundantly available in wool, feathers, hair etc. and has excellent insulating properties, on top of being biocompatible. This makes it an attractive target for reprocessing. Keratin has a characteristic three-dimensional stable structure that varies in respect to the composition of amino acids and depends upon the source from where its derived. A million of tons of these materials are wasted each year, which is not only economically ineffective because of loss of a valuable material that, if properly dissolved, can benefit the production of various high-value products. These materials have extremely low solubility in common organic solvents and if dissolved, can result in complete breaking of their structure. So, in view of the properties of such materials, the use of green and sustainable solvents for the dissolution of these materials is highly demanded. Ionic liquids have shown good solvation properties for cellulose, silks and even keratin but the research is still infancy. Ionic liquids are molten salts and have complex nature owing to the existence of a delicate balance between hydrogen bonding, columbic and electrostatic interactions between the constituent cations and anions. The current molecular dynamics simulation study has provided a comprehensive overview of the impact of differently charged solutes on the interactions and structuring in two imidazolium-based ILs, and also identify the specific role of cations and anions for

these solutes in $[C_2C_1Im][CH_3CO_2]$ $[C_2C_1Im][Cl]$ ionic liquids. This could have substantial impact in the future, guided design of ionic liquids that not only to maximise solubility in reprocessing but allow for controlled refolding to achieve or regenerate desirable properties in the final recovered material.

Firstly, the zwitterionic forms of the positively-charged amino acids lysine and arginine, and negatively-charged amino acids aspartate and glutamate were probed with the charge-scaled ions of $[C_2C_1Im][Cl]$ ionic liquid to better simulate the protein-relevant interactions in solution, and the dynamics of the systems, respectively. The experimental data[242] suggested the extremely low solubility of these three amino acids (lysine, arginine and aspartate) in this ionic liquid and this computational study was able to validate the experimental findings, in addition to highlighting the impacts of these solutes on the structuring in ionic liquids. The simulation run for two different concentration ratios of IL:AA, 200:1 and 5:1 at 300K revealed that even though the two amino acid pairs (lysine, arginine and aspartate, glutamate) have the same charge within each pair, the strengths of solute-solvent, solvent-solvent and solute-solute interactions were quite different, even when comparing the similar charged solutes e.g. lysine with arginine and aspartate with glutamate. This has further implications for organisation within proteins, as it illustrates the changes inferred on the organisation of ions with the amide backbone are highly dependent on the primary sequence, and in longer-range interactions that vary with absolute structure.

Lysine showed a stronger association with the IL anions than arginine, and aspartate showed strong interactions with IL cation. Cation-anion and cation-cation interactions are shown to have opposite trend for two sets of amino acids. For example, the basic structuring of the chloride anion around IL cation, and IL cation around a central IL cation, were similar, with anions occupying the regions through dominant electrostatic interactions around the aromatic rings, and cations displaying cation-pi interactions. However, in case of lysine and arginine the concentration increase had more impact on decreasing the cation-cation interactions, whereas for aspartate and glutamate, anions showed depletion around cations when concentration ratio was increased. The effects were marginal but still significant because the temperature increase further enhanced these effects. In this case, positively charged amino acids showed a complete loss of cation-cation contacts at high temperature in addition to depletion across $[C_2C_1Im]^+$

cation's methyl group, whereas for negatively charged amino acids, the cation-cation interactions were slightly reduced and did not disappear completely. Further investigation into the neutral alanine solutes revealed that temperature increase has a similar effect as for lysine, where cation-cation interactions are completely lost. Inter-amino acid interactions were found to be increased in both pure IL and water mixtures of $[C_2C_1Im][Cl]$ ionic liquid, however, more substantial increase was seen for negatively charged amino acids and neutral alanine.

The change in the nature of the IL anion, and especially hydrogen bonding ability going from chloride to acetate, had a major influence on the solvating ability for these amino acids, mainly analysed through domain analysis and compared with the experimental data for results validation. Acetate anions showed more structured shells around amino acids relative to chloride anions. Cations were observed to have stronger association with amino acids, and this indicates that the anion's nature and the difference in its hydrogen bond accepting ability can have a dramatic influence on the organisational pattern.

The organisation of these small model compounds then led to the investigation of peptide simulations where two alanine-based peptides (namely, penta- and deca-alanine) and a peptide segment from Protein G LNGKTLKG (ACE(Leu-Asn-Gly-Lys-Thr-Leu-Lys-Gly)NME), known to form a hairpin structure, were taken in their extended non-standard form. There were two objectives achieved here, firstly the simulations were run in water to inspect the secondary structure of the peptides and then in pure $[C_2C_1Im][CH_3CO_2]$ IL to observe the structural difference. The secondary structure that was mostly comprised of turns, bends and coils in some parts but not much included the helix structure, which was lost in $[C_2C_1Im][CH_3CO_2]$ IL. The peptide had two lysine molecules and analysing the structuring of the IL anions revealed that the organisational pattern was similar in terms of distance however, strength of interactions were different for each amino acid residue, indicating that the small amino acid models were able to correctly explain the organisation pattern but strengths are lower as compared to single amino acids.

9.2 Recommendations for future work

In view of the above findings, experimental or computational studies can be conducted by varying the nature of the cation to further insight into the nature of interactions. It was observed that cations bind differently with varied nature of solutes and also the organisation ability of cation in ionic liquids is not consistent when the nature of anion is changed. This provides intriguing areas that need to be further developed, for example if different cations can have a similar organisation (or similar cations a different organisation) for a specific anion, that would expect to shed more light on the microscopic detail of the role of specific anion and cations.

Infrared spectra have been proved to be extremely useful in identifying the non-covalent interactions such as hydrogen bonding, dispersion forces, columbic interactions and aid in explaining the structural details of ionic liquids.[274] Coupling future MD studies to experimental spectral studies, where possible, is recommended because the IR spectra, though with less accuracy can be generated through MD simulations that can be helpful in the validation of the organisational changes noted here. Neutron diffraction spectra have also proven helpful in this regard and are in use to explain the structuring in ionic liquids. They can provide quick validation of results by matching them with the computationally derived radial distribution functions.[208]

Experimental data identifying organisation in the ionic liquid-amino acid solutions were used to directly inform the simulation experiments to guide the initial selection of amino acids and was initially based on the charge and its abundance in wool keratin. As the project developed, attention was focused on probing the effects of various parameters such as temperature and concentration with respect to the varying nature of amino acid side chains. In future, additional emphasis could be applied to the impact of a broader range of ionic liquids on amino acids such as glutamate, for which it was observed that it was fairly soluble in both the ILs studied in the project, and for which the limits of interactions would help assess better how solubility/interactions could be tailored by judicious selection or use of binary systems. It would be interesting to visualise the structural differences or similarities for such AAs and their influence on the IL structuring. The uptake over the period of amino acid-based ILs, which links into this work indirectly, has been significant and there is much that could be done to link

in these two areas to develop a comprehensive idea of more complex interactions involving binary systems. In future, therefore, it would be interesting to further interrogate the behaviour of related solutes.

Finally, this work highlights the significant interplay between ions in more complex mixtures. A model that more comprehensively defines the complex dynamic equilibria seen, and can help pinpoint critical interacting sites along with the magnitude of interactions expected is badly needed. Both the protein models, and the single amino acid studies, however, suggest that even with this in hand, there are elements of ionic-cooperativity that need to be taken into account when identifying organisation around protein polymers, for example the selective reinforcement of some shell structures, or the ablation of others. Recognition that this is important arises from both observing the differences in the same residues along a backbone, and the 'remote' impacts of the side-chain organisation on dictating the organisation near to the backbone elements. Challenges in identifying how the microstructural organisation builds up either to enhance organisation at a site, or remove it, will provide a significant opportunity in ionic liquid chemistry with complex solutes for the near future.

References

1. Veselá, M.; Friedrich, J., Amino acid and soluble protein cocktail from waste keratin hydrolysed by a fungal keratinase of Paecilomyces marquandii. *Biotechnology and Bioprocess Engineering* **2009,** *14* (1), 84-90.
2. Moreira, F. G.; de Souza, C. G. M.; Costa, M. A. F.; Reis, S.; Peralta, R. M., Degradation of keratinous materials by the plant pathogenic fungus Myrothecium verrucaria. *Mycopathologia* **2007,** *163* (3), 153-160.
3. Aluigi, A.; Zoccola, M.; Vineis, C.; Tonin, C.; Ferrero, F.; Canetti, M., Study on the structure and properties of wool keratin regenerated from formic acid. *International Journal of Biological Macromolecules* **2007,** *41* (3), 266-273.
4. Horvath, A. L., Solubility of structurally complicated materials: 3. Hair. *TheScientificWorldJournal* **2009,** *9*, 255-271.
5. David M. Phillips, L. F. D., Deborah G. Conrady, Douglas M. Fox, Rajesh R. Naik,; Morley O. Stone, P. C. T., Hugh C. De Long, and Robert A. Mantz, Dissolution and Regeneration of Bombyx mori Silk Fibroin Using Ionic. *American Chemical Society* **2004**, 14350-14351.
6. Sun, N.; Rahman, M.; Qin, Y.; Maxim, M. L.; Rodríguez, H.; Rogers, R. D., Complete dissolution and partial delignification of wood in the ionic liquid 1-ethyl-3-methylimidazolium acetate. *Green Chemistry* **2009,** *11* (5), 646-655.
7. Ho, B. K.; Dill, K. A., Folding Very Short Peptides Using Molecular Dynamics. *PLoS Computational Biology* **2006,** *2* (4), e27.
8. Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M., PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009,** *30* (13), 2157-2164.
9. Case, D. A.; Cheatham Iii, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005,** *26* (16), 1668-1688.
10. Yu, L.; Dean, K.; Li, L., Polymer blends and composites from renewable resources. *Progress in polymer science* **2006,** *31* (6), 576-602.
11. Silva, N. H. C. S.; Vilela, C.; Marrucho, I. M.; Freire, C. S. R.; Pascoal Neto, C.; Silvestre, A. J. D., Protein-based materials: from sources to innovative sustainable materials for biomedical applications. *Journal of Materials Chemistry B* **2014,** *2* (24), 3715-3740.
12. Xue, Y.; Lofland, S.; Hu, X., Thermal Conductivity of Protein-Based Materials: A Review. **2019,** *11*, 456.
13. Schindl, A.; Hagen, M. L.; Muzammal, S.; Gunasekera, H. A. D.; Croft, A. K., Proteins in Ionic Liquids: Reactions, Applications, and Futures. *Frontiers in Chemistry* **2019,** *7* (347).
14. DeFrates, K. G.; Moore, R.; Borgesi, J.; Lin, G.; Mulderig, T.; Beachley, V.; Hu, X., Protein-Based Fiber Materials in Medicine: A Review. *Nanomaterials* **2018,** *8* (7), 457.
15. Shavandi, A.; Silva, T. H.; Bekhit, A. A.; Bekhit, A. E.-D. A., Keratin: dissolution, extraction and biomedical application. *Biomaterials Science* **2017,** *5* (9), 1699-1735.
16. Idris, A.; Vijayaraghavan, R.; Rana, U. A.; Patti, A. F.; MacFarlane, D. R., Dissolution and regeneration of wool keratin in ionic liquids. *Green Chem.* **2014,** *16* (5), 2857-2864.

17. Zheng, S.; Yi, N.; Zhang, S.; Zhang, X.; Wang, L., Highly Efficient Dissolution of Wool Keratin by Dimethylphosphate Ionic Liquids. *ACS Sustainable Chem. Eng.* **2015,** *3* (11), 2925-2932.
18. Attri, P.; Venkatesu, P., Refolding of urea-induced denaturation of model proteins by trimethylamine N-oxide. *Thermochimica Acta* **2011,** *526* (1-2), 143-150.
19. Hardy, J. G.; Römer, L. M.; Scheibel, T. R., Polymeric materials based on silk proteins. *Polymer* **2008,** *49* (20), 4309-4327.
20. Maksić, Z. B.; Kovačević, B., Towards the absolute proton affinities of 20 α-amino acids. *Chemical Physics Letters* **1999,** *307* (5–6), 497-504.
21. Harrison, A. G., The gas-phase basicities and proton affinities of amino acids and peptides. *Mass spectrometry reviews* **1997,** *16* (4), 201-217.
22. Dowling, L. M.; Crewther, W. G.; Inglis, A. S., The primary structure of component 8c-1, a subunit protein of intermediate filaments in wool keratin. Relationships with proteins from other intermediate filaments. *Biochemical Journal* **1986,** *236* (3), 695-703.
23. Li, J.; Li, Y.; Li, L.; Mak, A. F. T.; Ko, F.; Qin, L., Preparation and biodegradation of electrospun PLLA/keratin nonwoven fibrous membrane. *Polym Degrad Stabil* **2009,** *94* (10), 1800-1807.
24. Volkov, V.; Ferreira, A. V.; Cavaco-Paulo, A., On the Routines of Wild-Type Silk Fibroin Processing Toward Silk-Inspired Materials: A Review. *Macromolecular Materials and Engineering* **2015,** *300* (12), 1199-1216.
25. Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L., Silk-based biomaterials. *Biomaterials* **2003,** *24*, 401-16.
26. Vasconcelos, A.; Freddi, G.; Cavaco-Paulo, A., Biodegradable Materials Based on Silk Fibroin and Keratin. *Biomacromolecules* **2008,** *9* (4), 1299-1305.
27. Zhou, C.-Z.; Confalonieri, F.; Medina, N.; Zivanovic, Y.; Esnault, C.; Yang, T.; Jacquet, M.; Janin, J.; Duguet, M.; Perasso, R.; Li, Z.-G., Fine organization of Bombyx mori fibroin heavy chain gene. *Nucleic acids research* **2000,** *28* (12), 2413-2419.
28. Shen, Y.; Johnson, M. A.; Martin, D. C., Microstructural Characterization of Bombyx mori Silk Fibers. *Macromolecules* **1998,** *31* (25), 8857-8864.
29. Vepari, C.; Kaplan, D. L., Silk as a Biomaterial. *Prog Polym Sci* **2007,** *32*, 991-1007.
30. Zhou, C.-Z.; Confalonieri, F.; Jacquet, M.; Perasso, R.; Li, Z.-G.; Janin, J., Silk fibroin: Structural implications of a remarkable amino acid sequence. *Proteins: Structure, Function, and Bioinformatics* **2001,** *44* (2), 119-122.
31. Rouse, J. G.; Van Dyke, M. E., A Review of Keratin-Based Biomaterials for Biomedical Applications. *Materials* **2010,** *3* (2), 999-1014.
32. Tanabe, T.; Okitsu, N.; Yamauchi, K., Fabrication and characterization of chemically crosslinked keratin films. *Materials Science and Engineering: C* **2004,** *24* (3), 441-446.
33. Poole, A. J.; Church, J. S.; Huson, M. G., Environmentally Sustainable Fibers from Regenerated Protein. *Biomacromolecules* **2009,** *10* (1), 1-8.
34. McKittrick, J.; Chen, P. Y.; Bodde, S. G.; Yang, W.; Novitskaya, E. E.; Meyers, M. A., The Structure, Functions, and Mechanical Properties of Keratin. *JOM* **2012,** *64* (4), 449-468.
35. Onifade, A. A.; Al-Sane, N. A.; Al-Musallam, A. A.; Al-Zarban, S., A review: Potentials for biotechnological applications of keratin-degrading microorganisms and

their enzymes for nutritional improvement of feathers and other keratins as livestock feed resources. *Bioresource Technology* **1998,** *66* (1), 1-11.
36. Meng, Z. J.; Zheng, X. J.; Tang, K. Y.; Liu, J.; Qin, S. F., Dissolution of natural polymers in ionic liquids: A review. *E-Polymers* **2012**.
37. Khosa MA, U. A., A Sustainable Role of Keratin Biopolymer in Green Chemistry: A Review. *J Food Processing & Beverages.* **2013,** *1* (1).
38. Strnad, P.; Usachov, V.; Debes, C.; Gräter, F.; Parry, D. A. D.; Omary, M. B., Unique amino acid signatures that are evolutionarily conserved distinguish simple-type, epidermal and hair keratins. *Journal of Cell Science* **2011,** *124* (24), 4221-4232.
39. Shavandi, A.; Bekhit, A. E.-D. A.; Carne, A.; Bekhit, A., Evaluation of keratin extraction from wool by chemical methods for bio-polymer application. *J. Bioact. Compat. Polym.* **2017,** *32* (2), 163-177.
40. Liu, X.; Nie, Y.; Meng, X.; Zhang, Z.; Zhang, X.; Zhang, S., DBN-based ionic liquids with high capability for the dissolution of wool keratin. *RSC Adv.* **2017,** *7* (4), 1981-1988.
41. Idris, A.; Vijayaraghavan, R.; Rana, U. A.; Fredericks, D.; Patti, A. F.; MacFarlane, D. R., Dissolution of feather keratin in ionic liquids. *Green Chemistry* **2013,** *15* (2), 525.
42. Hameed, N.; Guo, Q., Natural wool/cellulose acetate blends regenerated from the ionic liquid 1-butyl-3-methylimidazolium chloride. **2009,** *78* (4), 999-1004.
43. Idris, A.; Vijayaraghavan, R.; Patti, A. F.; MacFarlane, D. R., Distillable Protic Ionic Liquids for Keratin Dissolution and Recovery. *ACS Sustainable Chemistry & Engineering* **2014,** *2* (7), 1888-1894.
44. Li, G.; Li, Y.; Chen, G.; He, J.; Han, Y.; Wang, X.; Kaplan, D. L., Silk-Based Biomaterials in Biomedical Textiles and Fiber-Based Implants. *Adv Healthc Mater* **2015**.
45. Zhao, H.; Xia, S.; Ma, P., Use of ionic liquids as 'green' solvents for extractions. *Journal of Chemical Technology & Biotechnology* **2005,** *80* (10), 1089-1096.
46. Farnoush Faridbod and Mohammad Reza Ganjali and Parviz Norouzi and Siavash Riahi and and Hamid, R., Application of Room Temperature Ionic Liquids in Electrochemical Sensors and Biosensors. **2011**.
47. Ghandi, K., A Review of Ionic Liquids, Their Limits and Applications. *Green and Sustainable Chemistry* **2014,** *Vol.04No.01*, 10.
48. Olivier-Bourbigou, H.; Magna, L.; Morvan, D., Ionic liquids and catalysis: Recent progress from knowledge to applications. *Applied Catalysis A: General* **2010,** *373* (1–2), 1-56.
49. Hallett, J. P.; Welton, T., Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2. *Chemical Reviews* **2011,** *111* (5), 3508-3576.
50. Welton, T., Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis. *Chemical reviews* **1999,** *99* (8), 2071-2084.
51. Wasserscheid, P.; Keim, W., Ionic Liquids—New "Solutions" for Transition Metal Catalysis. *Angewandte Chemie International Edition* **2000,** *39* (21), 3772-3789.
52. Brehm, M.; Weber, H.; Thomas, M.; Hollóczki, O.; Kirchner, B., Domain Analysis in Nanostructured Liquids: A Post-Molecular Dynamics Study at the Example of Ionic Liquids. *ChemPhysChem* **2015,** *16* (15), 3271-3277.
53. Hayes, R.; Warr, G. G.; Atkin, R., Structure and Nanostructure in Ionic Liquids. *Chemical Reviews* **2015,** *115*, 6357-6426.

54. Tietze, A. A.; Heimer, P.; Stark, A.; Imhof, D., Ionic liquid applications in peptide chemistry: synthesis, purification and analytical characterization processes. *Molecules* **2012,** *17* (4), 4158-85.
55. Domínguez-Pérez, M.; Tomé, L. I. N.; Freire, M. G.; Marrucho, I. M.; Cabeza, O.; Coutinho, J. A. P., Extraction of biomolecules using aqueous biphasic systems formed by ionic liquids and aminoacids. *Separation and Purification Technology* **2010,** *72*, 85-91.
56. Seduraman, A.; Wu, P.; Klahn, M., Extraction of tryptophan with ionic liquids studied with molecular dynamics simulations. *J Phys Chem B* **2012,** *116* (1), 296-304.
57. Kubisa, P., Application of ionic liquids as solvents for polymerization processes. **2004,** *29* (1), 3-12.
58. Sheldon, R., Catalytic reactions in ionic liquids. *Chemical Communications* **2001,** (23), 2399-2407.
59. Chowdhury, Z. Z.; Zain, S. M.; Abd Hamid, S. B.; Khalid, K., Catalytic role of ionic liquids for dissolution and degradation of biomacromolecules. *BioResources* **2014,** *9*, 1787-1823, 37 pp.
60. Plechkova, N. V.; Seddon, K. R., Applications of ionic liquids in the chemical industry. *Chemical Society Reviews* **2008,** *37* (1), 123-150.
61. Klahn, M.; Lim, G. S.; Seduraman, A.; Wu, P., On the different roles of anions and cations in the solvation of enzymes in ionic liquids. *Physical Chemistry Chemical Physics* **2011,** *13* (4), 1649-1662.
62. Shao, Q., On the influence of hydrated imidazolium-based ionic liquid on protein structure stability: A molecular dynamics simulation study. *The Journal of Chemical Physics* **2013,** *139* (11), 115102.
63. Rosatella, A. A.; Branco, L. C.; Afonso, C. A. M., Studies on dissolution of carbohydrates in ionic liquids and extraction from aqueous phase. *Green Chemistry* **2009,** *11* (9), 1406-1413.
64. Sang Hyun, L.; Minoru, M.; Jonathan, S. D.; Robert, J. L., Preparation of Biopolymer-Based Materials Using Ionic Liquids for the Biomedical Application. In *Ionic Liquid Applications: Pharmaceuticals, Therapeutics, and Biotechnology*, American Chemical Society: 2010; Vol. 1038, pp 115-134.
65. Plowman, J. E.; Clerens, S.; Lee, E.; Harland, D. P.; Dyer, J. M.; Deb-Choudhury, S., Ionic liquid-assisted extraction of wool keratin proteins as an aid to MS identification. *Anal. Methods* **2014,** *6* (18), 7305-7311.
66. Zhang, H.; Wu, J.; Zhang, J.; He, J., 1-Allyl-3-methylimidazolium Chloride Room Temperature Ionic Liquid: A New and Powerful Nonderivatizing Solvent for Cellulose. *Macromolecules* **2005,** *38* (20), 8272-8277.
67. Zhang, J.; Wu, J.; Yu, J.; Zhang, X.; He, J.; Zhang, J., Application of ionic liquids for dissolving cellulose and fabricating cellulose-based materials: state of the art and future trends. *Materials Chemistry Frontiers* **2017,** *1* (7), 1273-1290.
68. Fukaya, Y.; Hayashi, K.; Wada, M.; Ohno, H., Cellulose dissolution with polar ionic liquids under mild conditions: required factors for anions. *Green Chemistry* **2008,** *10* (1), 44-46.
69. Wang, M.; Zhao, T.; Wang, G.; Zhou, J., Blend films of human hair and cellulose prepared from an ionic liquid. *Textile Research Journal* **2014**.
70. Mäki-Arvela, P.; Anugwom, I.; Virtanen, P.; Sjöholm, R.; Mikkola, J. P., Dissolution of lignocellulosic materials and its constituents using ionic liquids—A review. **2010,** *32* (3), 175-201.

71. Jose, J. P.; Joseph, K., Advances in Polymer Composites: Macro- and Microcomposites – State of the Art, New Challenges, and Opportunities. In *Polymer Composites*, Wiley-VCH Verlag GmbH & Co. KGaA: 2012; pp 1-16.
72. Gupta, M. K.; Khokhar, S. K.; Phillips, D. M.; Sowards, L. A.; Drummy, L. F.; Kadakia, M. P.; Naik, R. R., Patterned Silk Films Cast from Ionic Liquid Solubilized Fibroin as Scaffolds for Cell Growth. *Langmuir* **2007,** *23*, 1315-1319.
73. Wang, X.; Partlow, B.; Liu, J.; Zheng, Z.; Su, B.; Wang, Y.; Kaplan, D. L., Injectable silk-polyethylene glycol hydrogels. *Acta Biomater* **2015,** *12*, 51-61.
74. Gunasekera, D. H. A. T.; Kuek, S.; Hasanaj, D.; He, Y.; Tuck, C.; Croft, A. K.; Wildman, R. D., Three dimensional ink-jet printing of biomaterials using ionic liquids and co-solvents. *Faraday Discussions* **2016,** *190*, 509-523.
75. Brandt, A.; Gräsvik, J.; Hallett, J. P.; Welton, T., Deconstruction of lignocellulosic biomass with ionic liquids. *Green Chemistry* **2013,** *15* (3), 550-583.
76. Cox, B. J.; Ekerdt, J. G., Depolymerization of oak wood lignin under mild conditions using the acidic ionic liquid 1-H-3-methylimidazolium chloride as both solvent and catalyst. **2012,** *118*, 584-588.
77. Pu, Y.; Jiang, N.; Ragauskas, A. J., Ionic Liquid as a Green Solvent for Lignin. *Journal of Wood Chemistry and Technology* **2007,** *27* (1), 23-33.
78. Rinaldi, R. A.-O.; Jastrzebski, R.; Clough, M. T.; Ralph, J.; Kennema, M.; Bruijnincx, P. A.-O.; Weckhuysen, B. M., Paving the Way for Lignin Valorisation: Recent Advances in Bioengineering, Biorefining and Catalysis. (1521-3773 (Electronic)).
79. Kadla, J. F.; Kubo, S., Lignin-based polymer blends: analysis of intermolecular interactions in lignin–synthetic polymer blends. *AIChE 2002* **2004,** *35* (3), 395-400.
80. Pinkert, A.; Marsh, K. N.; Pang, S.; Staiger, M. P., Ionic Liquids and Their Interaction with Cellulose. *Chemical Reviews* **2009,** *109* (12), 6712-6728.
81. Ohno, H.; Fukaya, Y., Task Specific Ionic Liquids for Cellulose Technology. *Chemistry Letters* **2008,** *38* (1), 2-7.
82. Lee, S. H.; Doherty, T. V.; Linhardt, R. J.; Dordick, J. S., Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. *Biotechnology and Bioengineering* **2009,** *102* (5), 1368-1376.
83. Fu, D.; Mazza, G.; Tamaki, Y., Lignin Extraction from Straw by Ionic Liquids and Enzymatic Hydrolysis of the Cellulosic Residues. *Journal of Agricultural and Food Chemistry* **2010,** *58* (5), 2915-2922.
84. Moniruzzaman, M.; Kamiya, N.; Goto, M., Activation and stabilization of enzymes in ionic liquids. *Organic & Biomolecular Chemistry* **2010,** *8*, 2887-2899.
85. Brandt, A.; Ray, M. J.; To, T. Q.; Leak, D. J.; Murphy, R. J.; Welton, T., Ionic liquid pretreatment of lignocellulosic biomass with ionic liquid-water mixtures. *Green Chemistry* **2011,** *13*, 2489-2499.
86. Forsyth, S. A.; MacFarlane, D. R.; Thomson, R. J.; von Itzstein, M., Rapid, clean, and mild O-acetylation of alcohols and carbohydrates in an ionic liquid. *Chemical Communications* **2002,** (7), 714-715.
87. Liu, Q.; Janssen, M. H. A.; van Rantwijk, F.; Sheldon, R. A., Room-temperature ionic liquids that dissolve carbohydrates in high concentrations. *Green Chemistry* **2005,** *7* (1), 39-42.
88. Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D., Dissolution of Cellose with Ionic Liquids. *Journal of the American Chemical Society* **2002,** *124* (18), 4974-4975.

89. Abushammala, H.; Mao, J., A Review on the Partial and Complete Dissolution and Fractionation of Wood and Lignocelluloses Using Imidazolium Ionic Liquids. P*olymers* 2020 , 1*2* , 195.: 2020.
90. Zavrel, M.; Bross, D.; Funke, M.; Büchs, J.; Spiess, A. C., High-throughput screening for ionic liquids dissolving (ligno-)cellulose. **2009,** *100* (9), 2580-2587.
91. Kilpeläinen, I.; Xie, H.; King, A.; Granstrom, M.; Heikkinen, S.; Argyropoulos, D. S., Dissolution of Wood in Ionic Liquids. *Journal of Agricultural and Food Chemistry* **2007,** *55* (22), 9142-9148.
92. Achinivu, E. C., Protic Ionic Liquids for Lignin Extraction-A Lignin Characterization Study. *International journal of molecular sciences* **2018,** *19* (2), 428.
93. Xu, A.; Wang, J.; Wang, H., Effects of anionic structure and lithium salts addition on the dissolution of cellulose in 1-butyl-3-methylimidazolium-based ionic liquid solvent systems. *Green Chemistry* **2010,** *12* (2), 268-275.
94. Vo, H. T.; Kim, C. S.; Ahn, B. S.; Kim, H. S.; Lee, H., Study on Dissolution and Regeneration of Poplar Wood in Imidazolium-Based Ionic Liquids. *Journal of Wood Chemistry and Technology* **2011,** *31* (2), 89-102.
95. Li, Y.; Liu, X.; Zhang, Y.; Jiang, K.; Wang, J.; Zhang, S., Why Only Ionic Liquids with Unsaturated Heterocyclic Cations Can Dissolve Cellulose: A Simulation Study. *ACS Sustainable Chemistry & Engineering* **2017,** *5* (4), 3417-3428.
96. Rinaldi, R., Instantaneous dissolution of cellulose in organic electrolyte solutions. *Chemical Communications* **2011,** *47* (1), 511-513.
97. Xu, A.; Zhang, Y.; Zhao, Y.; Wang, J., Cellulose dissolution at ambient temperature: Role of preferential solvation of cations of ionic liquids by a cosolvent. **2013,** *92* (1), 540-544.
98. Endo, T.; Hosomi, S.; Fujii, S.; Ninomiya, K.; Takahashi, K., Anion Bridging-Induced Structural Transformation of Cellulose Dissolved in Ionic Liquid. *The Journal of Physical Chemistry Letters* **2016,** *7* (24), 5156-5161.
99. Youngs, T. G. A.; Hardacre, C.; Holbrey, J. D., Glucose Solvation by the Ionic Liquid 1,3-Dimethylimidazolium Chloride: A Simulation Study. *The Journal of Physical Chemistry B* **2007,** *111* (49), 13765-13774.
100. Rabideau, B. D.; Agarwal, A.; Ismail, A. E., Observed Mechanism for the Breakup of Small Bundles of Cellulose Iα and Iβ in Ionic Liquids from Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2013,** *117* (13), 3469-3479.
101. Zhao, Y.; Liu, X.; Wang, J.; Zhang, S., Effects of Cationic Structure on Cellulose Dissolution in Ionic Liquids: A Molecular Dynamics Study. *ChemPhysChem* **2012,** *13* (13), 3126-3133.
102. Zhao, Y.; Liu, X.; Wang, J.; Zhang, S., Effects of anionic structure on the dissolution of cellulose in ionic liquids revealed by molecular simulation. **2013,** *94* (2), 723-730.
103. Goujon, N.; Wang, X.; Rajkowa, R.; Byrne, N., Regenerated silk fibroin using protic ionic liquids solvents: towards an all-ionic-liquid process for producing silk with tunable properties. *Chemical Communications* **2012,** *48*, 1278-1280.
104. Xie, H. B.; Li, S. H.; Zhang, S. B., Ionic liquids as novel solvents for the dissolution and blending of wool keratin fibers. *Green Chemistry* **2005,** *7* (8), 606-608.
105. Furth, M. E.; Atala, A.; Van Dyke, M. E., Smart biomaterials design for tissue engineering and regenerative medicine. *Biomaterials* **2007,** *28* (34), 5068-5073.
106. Reichl, S., Films based on human hair keratin as substrates for cell culture and tissue engineering. *Biomaterials* **2009,** *30* (36), 6854-6866.

107. Reichl, S.; Borrelli, M.; Geerling, G., Keratin films for ocular surface reconstruction. *Biomaterials* **2011,** *32* (13), 3375-3386.
108. Wang, Q.; Yang, Y.; Chen, X.; Shao, Z., Investigation of Rheological Properties and Conformation of Silk Fibroin in the Solution of AmimCl. *Biomacromolecules* **2012,** *13* (6), 1875-1881.
109. Phillips, D. M.; Drummy, L. F.; Naik, R. R.; Long, H. C. D.; Fox, D. M.; Trulove, P. C.; Mantz, R. A., Regenerated silk fiber wet spinning from an ionic liquid solution. *Journal of Materials Chemistry* **2005,** *15* (39), 4206-4208.
110. Robert, A. M.; Douglas, M. F.; Green, J. M.; Paul, A. F.; Hugh, C. D. L.; Paul, C. T., Dissolution of Biopolymers Using Ionic Liquids. **2007,** *62* (5-6), 275-280.
111. J. Dyson, P.; C. Grossel, M.; Srinivasan, N.; Vine, T.; Welton, T.; J. Williams, D.; J. P. White, A.; Zigras, T., Organometallic synthesis in ambient temperature chloroaluminate(III) ionic liquids. Ligand exchange reactions of ferrocene. *Journal of the Chemical Society, Dalton Transactions* **1997,** (19), 3465-3469.
112. Navone, L.; Speight, R., Understanding the dynamics of keratin weakening and hydrolysis by proteases. *PloS one* **2018,** *13* (8), e0202608-e0202608.
113. Greaves, T. L.; Weerawardena, A.; Krodkiewska, I.; Drummond, C. J., Protic Ionic Liquids: Physicochemical Properties and Behavior as Amphiphile Self-Assembly Solvents. *The Journal of Physical Chemistry B* **2008,** *112* (3), 896-905.
114. Achinivu, E. C.; Howard, R. M.; Li, G.; Gracz, H.; Henderson, W. A., Lignin extraction from biomass with protic ionic liquids. *Green Chemistry* **2014,** *16* (3), 1114-1119.
115. King, A. W. T.; Asikkala, J.; Mutikainen, I.; Järvi, P.; Kilpeläinen, I., Distillable Acid–Base Conjugate Ionic Liquids for Cellulose Dissolution and Processing. *Angewandte Chemie International Edition* **2011,** *50* (28), 6301-6305.
116. Wang, Y.-X.; Cao, X.-J., Extracting keratin from chicken feathers by using a hydrophobic ionic liquid. *Process Biochemistry* **2012,** *47* (5), 896-899.
117. Zhang, S.; Chen, Z.; Qi, X.; Deng, Y., Distinct influence of the anion and ether group on the polarity of ammonium and imidazolium ionic liquids. *New Journal of Chemistry* **2012,** *36* (4), 1043-1050.
118. Chen, J.; Vongsanga, K.; Wang, X.; Byrne, N., What Happens during Natural Protein Fibre Dissolution in Ionic Liquids. *Materials* **2014,** *7* (9), 6158-6168.
119. Rong Li, D. W., Preparation of Regenerated wool keratin films from wool keratin-ionic liquid solutions. *Journal of Applied Polymer Science* **2013**.
120. Tanford, C.; Kirkwood, J. G., Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *Journal of the American Chemical Society* **1957,** *79* (20), 5333-5339.
121. Kauzmann, W., Some factors in the interpretation of protein denaturation. *Advances in protein chemistry* **1959,** *14*, 1-63.
122. Prabhu, N.; Sharp, K., Protein-Solvent interactions. *Chemical reviews* **2006,** *106* (5), 1616-1623.
123. Lin Huang, J.; Noss, M. E.; Schmidt, K. M.; Murray, L.; Bunagan, M. R., The effect of neat ionic liquid on the folding of short peptides. *Chemical communications* **2011,** *47* (28), 8007-8009.
124. Plaquevent, J. C.; Levillain, J.; Guillen, F.; Malhiac, C.; Gaumont, A. C., Ionic liquids: new targets and media for alpha-amino acid and peptide chemistry. *Chem Rev* **2008,** *108* (12), 5035-60.
125. Anderson, J. L.; Ding, J.; Welton, T.; Armstrong, D. W., Characterizing Ionic Liquids On the Basis of Multiple Solvation Interactions. *Journal of the American Chemical Society* **2002,** *124*, 14247-14254.

126. Lozano, P.; De Diego, T.; Carrié, D.; Vaultier, M.; Iborra, J. L., Over-stabilization of Candida antarctica lipase B by ionic liquids in ester synthesis. *Biotechnol Lett* **2001,** *23* (18), 1529-1533.
127. De Diego, T.; Lozano, P.; Gmouh, S.; Vaultier, M.; Iborra, J. L., Fluorescence and CD spectroscopic analysis of the α-chymotrypsin stabilization by the ionic liquid, 1-ethyl-3-methylimidazolium bis[(trifluoromethyl)sulfonyl]amide. *Biotechnology and Bioengineering* **2004,** *88*, 916-924.
128. Seitkalieva, M. M.; Grachev, A. A.; Egorova, K. S.; Ananikov, V. P., Nanoscale organization of ionic liquids and their interaction with peptides probed by 13C NMR spectroscopy. *Tetrahedron* **2014,** *70* (36), 6075-6081.
129. Silva, M.; Figueiredo, A. M.; Cabrita, E. J., Epitope mapping of imidazolium cations in ionic liquid-protein interactions unveils the balance between hydrophobicity and electrostatics towards protein destabilisation. *Physical Chemistry Chemical Physics* **2014,** *16* (42), 23394-23403.
130. Constantinescu, D.; Weingärtner, H.; Herrmann, C., Protein Denaturation by Ionic Liquids and the Hofmeister Series: A Case Study of Aqueous Solutions of Ribonuclease A. *Angewandte Chemie International Edition* **2007,** *46* (46), 8887-8889.
131. Shim, Y.; Kim, H. J.; Jung, Y., Solvation of a Small Metal-Binding Peptide in Room-Temperature Ionic Liquids. *Bulletin of the Korean Chemical Society* **2012,** *33* (11), 3601-3606.
132. Giernoth, R.; Bankmann, D.; Schlorer, N., High performance NMR in ionic liquids. *Green Chemistry* **2005,** *7* (5), 279-282.
133. Mrestani-Klaus, C.; Richardt, A.; Wespe, C.; Stark, A.; Humpfer, E.; Bordusa, F., Structural Studies on Ionic Liquid/Water/Peptide Systems by HR-MAS NMR Spectroscopy. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2012,** *13* (7), 1836-1844.
134. Richardt, A.; Mrestani-Klaus, C.; Bordusa, F., Impact of ionic liquids on the structure of peptides proved by HR-MAS NMR spectroscopy. *Journal of Molecular Liquids* **2014,** *192* (0), 9-18.
135. Keaveney, S. T.; Harper, J. B.; Croft, A. K., Computational approaches to understanding reaction outcomes of organic processes in ionic liquids. *RSC advances* **2015,** *5* (45), 35709-35729.
136. Tietze, A. A.; Bordusa, F.; Giernoth, R.; Imhof, D.; Lenzer, T.; Maaß, A.; Mrestani-Klaus, C.; Neundorf, I.; Oum, K.; Reith, D.; Stark, A., On the Nature of Interactions between Ionic Liquids and Small Amino-Acid-Based Biomolecules. *ChemPhysChem* **2013,** *14* (18), 4044-4064.
137. Tomé, L. I. N.; Jorge, M.; Gomes, J. R. B.; Coutinho, J. A. P., Toward an Understanding of the Aqueous Solubility of Amino Acids in the Presence of Salts: A Molecular Dynamics Simulation Study. *The Journal of Physical Chemistry B* **2010,** *114* (49), 16450-16459.
138. Heyda, J.; Vincent, J. C.; Tobias, D. J.; Dzubiella, J.; Jungwirth, P., Ion Specificity at the Peptide Bond: Molecular Dynamics Simulations of N-Methylacetamide in Aqueous Salt Solutions. *The Journal of Physical Chemistry B* **2010,** *114* (2), 1213-1220.
139. Lund, M.; Vrbka, L.; Jungwirth, P., Specific Ion Binding to Nonpolar Surface Patches of Proteins. *Journal of the American Chemical Society* **2008,** *130* (35), 11582-11583.
140. Vrbka, L.; Jungwirth, P.; Bauduin, P.; Touraud, D.; Kunz, W., Specific Ion Effects at Protein Surfaces: A Molecular Dynamics Study of Bovine Pancreatic

Trypsin Inhibitor and Horseradish Peroxidase in Selected Salt Solutions. *The Journal of Physical Chemistry B* **2006,** *110* (13), 7036-7043.
141. Heyda, J.; Hrobárik, T.; Jungwirth, P., Ion-Specific Interactions between Halides and Basic Amino Acids in Water†. *The Journal of Physical Chemistry A* **2009,** *113* (10), 1969-1975.
142. Sagarik, K.; Dokmaisrijan, S., A theoretical study on hydration of alanine zwitterions. *Journal of Molecular Structure: THEOCHEM* **2005,** *718* (1–3), 31-47.
143. Fujita, T.; Watanabe, H.; Tanaka, S., Effects of salt addition on strength and dynamics of hydrophobic interactions. *Chemical Physics Letters* **2007,** *434* (1–3), 42-48.
144. Hess, B.; van der Vegt, N. F. A., Cation specific binding with protein surface charges. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106* (32), 13296-13300.
145. Vrbka, L.; Vondrášek, J.; Jagoda-Cwiklik, B.; Vácha, R.; Jungwirth, P., Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces. *Proceedings of the National Academy of Sciences* **2006,** *103* (42), 15440-15444.
146. Harper, J. B.; Lynden-Bell, R. M., Macroscopic and microscopic properties of solutions of aromatic compounds in an ionic liquid. *Molecular Physics* **2004,** *102*, 85-94.
147. Halverson, J. D.; Brandes, T.; Lenz, O.; Arnold, A.; Bevc, S.; Starchenko, V.; Kremer, K.; Stuehn, T.; Reith, D., ESPResSo++: A modern multiscale simulation package for soft matter systems. *Computer Physics Communications* **2013,** *184* (4), 1129-1149.
148. Karimi-Varzaneh, H. A.; Qian, H. J.; Chen, X. Y.; Carbone, P.; Muller-Plathe, F., IBIsCO: A Molecular Dynamics Simulation Package for Coarse-Grained Simulation. *Journal of Computational Chemistry* **2011,** *32* (7), 1475-1487.
149. Meier, K.; Choutko, A.; Dolenc, J.; Eichenberger, A. P.; Riniker, S.; van Gunsteren, W. F., Multi-Resolution Simulation of Biomolecular Systems: A Review of Methodological Issues. *Angewandte Chemie-International Edition* **2013,** *52* (10), 2820-2834.
150. Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R., GROMACS - A MESSAGE-PASSING PARALLEL MOLECULAR-DYNAMICS IMPLEMENTATION. *Computer Physics Communications* **1995,** *91* (1-3), 43-56.
151. Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T., DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. *Journal of Materials Chemistry* **2006,** *16* (20), 1911-1918.
152. Micaêlo, N. M.; Soares, C. M., Protein Structure and Dynamics in Ionic Liquids. Insights from Molecular Dynamics Simulation Studies. *The Journal of Physical Chemistry B* **2008,** *112* (9), 2566-2572.
153. Dasari, S.; Mallik, B. S., Biosolvation Nature of Ionic Liquids: Molecular Dynamics Simulation of Methylated Nucleobases in Hydrated 1-Ethyl-3-methylimidazolium Acetate. *ACS Omega* **2018,** *3* (7), 8344-8354.
154. Tomé, L. I. N.; Catambas, V. R.; Teles, A. R. R.; Freire, M. G.; Marrucho, I. M.; Coutinho, J. A. P., Tryptophan extraction using hydrophobic ionic liquids. *Separation and Purification Technology* **2010,** *72* (2), 167-173.
155. Lesch, V.; Heuer, A.; Tatsis, V. A.; Holm, C.; Smiatek, J., Peptides in the presence of aqueous ionic liquids: tunable co-solutes as denaturants or protectants? *Physical Chemistry Chemical Physics* **2015,** *17* (39), 26049-26053.

156. Haberler, M.; Schroder, C.; Steinhauser, O., Solvation studies of a zinc finger protein in hydrated ionic liquids. *Physical Chemistry Chemical Physics* **2011,** *13*, 6955-6969.
157. Klahn, M.; Lim, G. S.; Wu, P., How ion properties determine the stability of a lipase enzyme in ionic liquids: A molecular dynamics study. *Physical Chemistry Chemical Physics* **2011,** *13*, 18647-18660.
158. Paluch, A. S.; Vitter, C. A.; Shah, J. K.; Maginn, E. J., A comparison of the solvation thermodynamics of amino acid analogues in water, 1-octanol and 1-n-alkyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ionic liquids by molecular simulation. *Journal of Chemical Physics* **2012,** *137* (18).
159. Tomé, L. I. N.; Jorge, M.; Gomes, J. R. B.; Coutinho, J. A. P., Molecular Dynamics Simulation Studies of the Interactions between Ionic Liquids and Amino Acids in Aqueous Solution. *The Journal of Physical Chemistry B* **2012,** *116* (6), 1831-1842.
160. Fukumoto, K.; Yoshizawa, M.; Ohno, H., Room Temperature Ionic Liquids from 20 Natural Amino Acids. *Journal of the American Chemical Society* **2005,** *127* (8), 2398-2399.
161. Tarannum, A.; Rao, J. R.; Fathima, N. N., Choline-Based Amino Acid ILs–Collagen Interaction: Enunciating Its Role in Stabilization/Destabilization Phenomena. *The Journal of Physical Chemistry B* **2018,** *122* (3), 1145-1151.
162. Chevrot, G.; Fileti, E. E.; Chaban, V. V., Protein remains stable at unusually high temperatures when solvated in aqueous mixtures of amino acid based ionic liquids. **2016,** *22* (11), 258.
163. Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L., Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry : AABC* **2015,** *8*, 37-47.
164. Liu, H.; Maginn, E., A molecular dynamics investigation of the structural and dynamic properties of the ionic liquid 1-n-butyl-3-methylimidazolium bis(trifluoromethanesulfonyl)imide. *The Journal of Chemical Physics* **2011,** *135* (12), 124507.
165. Vettoretti, G.; Moroni, E.; Sattin, S.; Tao, J.; Agard, D. A.; Bernardi, A.; Colombo, G., Molecular Dynamics Simulations Reveal the Mechanisms of Allosteric Activation of Hsp90 by Designed Ligands. *Scientific Reports* **2016,** *6*, 23830.
166. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996,** *118* (45), 11225-11236.
167. K., V.; E., H.; C., A.; S., K.; S., Z.; J., S.; E., D.; O., G.; P., L.; I., V.; D., M. A., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2010,** *31* (4), 671-690.
168. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** *25* (9), 1157-1174.
169. Sprenger, K. G.; Jaeger, V. W.; Pfaendtner, J., The General AMBER Force Field (GAFF) Can Accurately Predict Thermodynamic and Transport Properties of Many Ionic Liquids. *The Journal of Physical Chemistry B* **2015,** *119* (18), 5882-5895.
170. Liu, H.; Sale, K. L.; Simmons, B. A.; Singh, S., Molecular Dynamics Study of Polysaccharides in Binary Solvent Mixtures of an Ionic Liquid and Water. *The Journal of Physical Chemistry B* **2011,** *115* (34), 10251-10258.

171. Jiang, W.; Yan, T.; Wang, Y.; Voth, G. A., Molecular Dynamics Simulation of the Energetic Room-Temperature Ionic Liquid, 1-Hydroxyethyl-4-amino-1,2,4-triazolium Nitrate (HEATN). *The Journal of Physical Chemistry B* **2008,** *112* (10), 3121-3131.
172. Liu, H.; Sale, K. L.; Holmes, B. M.; Simmons, B. A.; Singh, S., Understanding the Interactions of Cellulose with Ionic Liquids: A Molecular Dynamics Study. *The Journal of Physical Chemistry B* **2010,** *114* (12), 4293-4301.
173. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophysical journal* **2009,** *97* (3), 853-856.
174. Schmidt, J. R. P., W.F. , WebMO Enterprise, version 13.0; WebMO LLC: Holland, MI, USA, . 2013; pp http://www.webmo.net (accessed March, 2013).
175. Gaussian 03, R. C., Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT. 2004.
176. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
177. Kollman, U. C. S. a. P. A., "An approach to computing electrostatic charges for molecules," *J. Comp. Chem.,* **(1984),** *5 129-45.*
178. B. H. Besler, K. M. M. J., and P. A. Kollman,, Atomic charges derived from semiempirical methods. *J. Comp. Chem.,* **1990,** *11 431-39*.
179. Christopher I. Bayly, P. C., * Wendy D. Cornell,$ and Peter A. Kollman, A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model *J. Phys. Chem* **July 6, 1993,** *97* (40).

180. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **2006,** *25*, 247-260.
181. Tenney, C. M.; Massel, M.; Mayes, J. M.; Sen, M.; Brennecke, J. F.; Maginn, E. J., A Computational and Experimental Study of the Heat Transfer Properties of Nine Different Ionic Liquids. *Journal of Chemical & Engineering Data* **2014,** *59* (2), 391-399.
182. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of Chemical Physics* **2004,** *120* (20), 9665-9678.
183. Walker, R., and Munshi, Aditi. Amber tutorial "Simulating a DNA polyA-polyT decamer". https://ambermd.org/tutorials/basic/tutorial1/.
184. Mac Dowell, N.; Llovell, F.; Sun, N.; Hallett, J. P.; George, A.; Hunt, P. A.; Welton, T.; Simmons, B. A.; Vega, L. F., New Experimental Density Data and Soft-SAFT Models of Alkylimidazolium ([CnC1im]+) Chloride (Cl–), Methylsulfate ([MeSO4]−), and Dimethylphosphate ([Me2PO4]−) Based Ionic Liquids. *The Journal of Physical Chemistry B* **2014,** *118* (23), 6206-6221.
185. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013,** *9* (7), 3084-3095.
186. Brehm, M.; Kirchner, B., TRAVIS - A Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. *Journal of Chemical Information and Modeling* **2011,** *51* (8), 2007-2023.
187. Humphrey, W. D. A. a. S. K., "VMD - Visual Molecular Dynamics". *J. Molec. Graphics* **1996,** *vol. 14*, pp. 33-38.
188. Zielkiewicz, J., Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. *J Chem Phys* **2005,** *123* (10), 104501.
189. Thar, J.; Brehm, M.; Seitsonen, A. P.; Kirchner, B., Unexpected Hydrogen Bond Dynamics in Imidazolium-Based Ionic Liquids. *The Journal of Physical Chemistry B* **2009,** *113* (46), 15129-15132.
190. Brüssel, M.; Brehm, M.; Voigt, T.; Kirchner, B., Ab initio molecular dynamics simulations of a binary system of ionic liquids. *Physical Chemistry Chemical Physics* **2011,** *13* (30), 13617-13620.
191. Brehm, M.; Weber, H.; Pensado Alfonso, S.; Stark, A.; Kirchner, B., Liquid Structure and Cluster Formation in Ionic Liquid/Water Mixtures – An Extensive ab initio Molecular Dynamics Study on 1-Ethyl-3-Methylimidazolium Acetate/Water Mixtures – Part. In *Zeitschrift für Physikalische Chemie International journal of research in physical chemistry and chemical physics*, 2013; Vol. 227, p 177.
192. Galiński, M.; Lewandowski, A.; Stępniak, I., Ionic liquids as electrolytes. *Electrochimica Acta* **2006,** *51* (26), 5567-5580.
193. Valkenburg, M. E. V.; Vaughn, R. L.; Williams, M.; Wilkes, J. S., Thermochemistry of ionic liquid heat-transfer fluids. *Thermochimica Acta* **2005,** *425* (1–2), 181-188.
194. Kowsari, M. H.; Alavi, S.; Ashrafizaadeh, M.; Najafi, B., Molecular dynamics simulation of imidazolium-based ionic liquids. I. Dynamics and diffusion coefficient. *The Journal of Chemical Physics* **2008,** *129* (22), 224508.
195. Tsuzuki, S.; Shinoda, W.; Saito, H.; Mikami, M.; Tokuda, H.; Watanabe, M., Molecular Dynamics Simulations of Ionic Liquids: Cation and Anion Dependence of

Self-Diffusion Coefficients of Ions. *The Journal of Physical Chemistry B* **2009,** *113* (31), 10641-10649.
196. Cadena, C.; Zhao, Q.; Snurr, R. Q.; Maginn, E. J., Molecular Modeling and Experimental Studies of the Thermodynamic and Transport Properties of Pyridinium-Based Ionic Liquids. *The Journal of Physical Chemistry B* **2006,** *110* (6), 2821-2832.
197. Keaveney, S. T.; Schaffarczyk McHale, K. S.; Stranger, J. W.; Ganbold, B.; Price, W. S.; Harper, J. B., NMR Diffusion Measurements as a Simple Method to Examine Solvent–Solvent and Solvent–Solute Interactions in Mixtures of the Ionic Liquid [Bmim][N(SO2CF3)2] and Acetonitrile. *ChemPhysChem* **2016,** *17* (23), 3853-3862.
198. Andreussi, O.; Marzari, N., Transport properties of room-temperature ionic liquids from classical molecular dynamics. *The Journal of Chemical Physics* **2012,** *137* (4), 044508.
199. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22* (12), 2577-2637.
200. Ramakrishnan, C.; Ramachandran, G. N., Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophysical journal* **1965,** *5* (6), 909-933.
201. Ramakrishnan, C., Ramachandran and his Map. **2001,** *6* (10), 48.
202. Wiltgen, M.; Ranganathan, S.; Gribskov, M.; Nakai, K.; Schönbach, C., Algorithms for Structure Comparison and Analysis: Homology Modelling of Proteins. Academic Press: Oxford, 2019; pp 38-61.
203. Rogers, M. A.; Winter, H.; Langbein, L.; Wollschläger, A.; Praetzel-Wunder, S.; Jave-Suarez, L. F.; Schweizer, J., Characterization of Human KAP24.1, A Cuticular Hair Keratin-Associated Protein with Unusual Amino-Acid Composition and Repeat Structure. *Journal of Investigative Dermatology* **2007,** *127* (5), 1197-1204.
204. Corfield, M. C.; Robson, A., The amino acid composition of wool. *Biochemical Journal* **1955,** *59* (1), 62-68.
205. Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I., Protein Stability and Surface Electrostatics: A Charged Relationship. *Biochemistry* **2006,** *45* (9), 2761-2766.
206. Sokalingam, S.; Raghunathan, G.; Soundrarajan, N.; Lee, S.-G., A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. *PLOS ONE* **2012,** *7* (7), e40410.
207. Su, Y.; Doherty, T.; Waring, A. J.; Ruchala, P.; Hong, M., Roles of arginine and lysine residues in the translocation of a cell-penetrating peptide from (13)C, (31)P, and (19)F solid-state NMR. *Biochemistry* **2009,** *48* (21), 4587-4595.
208. Youngs, T. G. A.; Holbrey, J. D.; Mullan, C. L.; Norman, S. E.; Lagunas, M. C.; D'Agostino, C.; Mantle, M. D.; Gladden, L. F.; Bowron, D. T.; Hardacre, C., Neutron diffraction, NMR and molecular dynamics study of glucose dissolved in the ionic liquid 1-ethyl-3-methylimidazolium acetate. *Chemical Science* **2011,** *2* (8), 1594-1605.
209. Zhang, Y.; Maginn, E. J., A comparison of methods for melting point calculation using molecular dynamics simulations. *The Journal of Chemical Physics* **2012,** *136* (14), 144116.
210. Bowron, D. T.; D'Agostino, C.; Gladden, L. F.; Hardacre, C.; Holbrey, J. D.; Lagunas, M. C.; McGregor, J.; Mantle, M. D.; Mullan, C. L.; Youngs, T. G. A., Structure and Dynamics of 1-Ethyl-3-methylimidazolium Acetate via Molecular

Dynamics and Neutron Diffraction. *The Journal of Physical Chemistry B* **2010,** *114* (23), 7760-7768.
211. Hunt, P. A.; Ashworth, C. R.; Matthews, R. P., Hydrogen bonding in ionic liquids. *Chemical Society Reviews* **2015,** *44* (5), 1257-1288.
212. Bhargava, B. L.; Balasubramanian, S., Intermolecular structure and dynamics in an ionic liquid: A Car–Parrinello molecular dynamics simulation study of 1,3-dimethylimidazolium chloride. *Chemical Physics Letters* **2006,** *417* (4), 486-491.
213. Micaelo, N. M.; Baptista, A. M.; Soares, C. M., Parametrization of 1-Butyl-3-methylimidazolium Hexafluorophosphate/Nitrate Ionic Liquid for the GROMOS Force Field. *The Journal of Physical Chemistry B* **2006,** *110* (29), 14444-14451.
214. Skarmoutsos, I.; Dellis, D.; Matthews, R. P.; Welton, T.; Hunt, P. A., Hydrogen Bonding in 1-Butyl- and 1-Ethyl-3-methylimidazolium Chloride Ionic Liquids. *The Journal of Physical Chemistry B* **2012,** *116* (16), 4921-4933.
215. Kohagen, M.; Brehm, M.; Lingscheid, Y.; Giernoth, R.; Sangoro, J.; Kremer, F.; Naumov, S.; Iacob, C.; Kärger, J.; Valiullin, R.; Kirchner, B., How Hydrogen Bonds Influence the Mobility of Imidazolium-Based Ionic Liquids. A Combined Theoretical and Experimental Study of 1-n-Butyl-3-methylimidazolium Bromide. *The Journal of Physical Chemistry B* **2011,** *115* (51), 15280-15288.
216. Brehm, M.; Weber, H.; S. Pensado, A.; Stark, A.; Kirchner, B., *Liquid Structure and Cluster Formation in Ionic Liquid/Water Mixtures - An Extensive ab initio Molecular Dynamics Study on 1-Ethyl-3-Methylimidazolium Acetate/Water Mixtures - Part 2*. 2013; Vol. 227, p 177-203.
217. Brüssel, M.; Brehm, M.; Pensado, A. S.; Malberg, F.; Ramzan, M.; Stark, A.; Kirchner, B., On the ideality of binary mixtures of ionic liquids. *Physical Chemistry Chemical Physics* **2012,** *14* (38), 13204-13215.
218. Madeira Lau, R.; Sorgedrager, M. J.; Carrea, G.; van Rantwijk, F.; Secundo, F.; Sheldon, R. A., Dissolution of Candida antarctica lipase B in ionic liquids: effects on structure and activity. *Green Chemistry* **2004,** *6* (9), 483-487.
219. Fromm, J. R.; Hileman, R. E.; Caldwell, E. E. O.; Weiler, J. M.; Linhardt, R. J., Differences in the Interaction of Heparin with Arginine and Lysine and the Importance of these Basic Amino Acids in the Binding of Heparin to Acidic Fibroblast Growth Factor. *Archives of biochemistry and biophysics* **1995,** *323* (2), 279-287.
220. Giełdoń, A.; Bobrowski, M.; Bielicka-Giełdoń, A.; Czaplewski, C., Theoretical calculation of the physico-chemical properties of 1-butyl-4-methylpyridinium based ionic liquids. *Journal of Molecular Liquids* **2017,** *225*, 467-474.
221. Neves, C. M. S. S.; Kurnia, K. A.; Coutinho, J. A. P.; Marrucho, I. M.; Lopes, J. N. C.; Freire, M. G.; Rebelo, L. P. N., Systematic Study of the Thermophysical Properties of Imidazolium-Based Ionic Liquids with Cyano-Functionalized Anions. *The Journal of Physical Chemistry B* **2013,** *117* (35), 10271-10283.
222. Seddon Kenneth, R.; Stark, A.; Torres, M.-J., Influence of chloride, water, and organic solvents on the physical properties of ionic liquids. In *Pure and Applied Chemistry*, 2000; Vol. 72, p 2275.
223. Pauling, L.; Corey, R. B., TWO HYDROGEN-BONDED SPIRAL CONFIGURATIONS OF THE POLYPEPTIDE CHAIN. *Journal of the American Chemical Society* **1950,** *72*, 5349-5349.
224. Neves, C. M. S. S.; Lemus, J.; Freire, M. G.; Palomar, J.; Coutinho, J. A. P., Enhancing the adsorption of ionic liquids onto activated carbon by the addition of inorganic salts. *Chemical engineering journal (Lausanne, Switzerland : 1996)* **2014,** *252*, 305-310.

225. Jacquemin, J.; Husson, P.; Padua, A. A. H.; Majer, V., Density and viscosity of several pure and water-saturated ionic liquids. *Green Chemistry* **2006,** *8* (2), 172-180.
226. Rodríguez, H.; Brennecke, J. F., Temperature and Composition Dependence of the Density and Viscosity of Binary Mixtures of Water + Ionic Liquid. *Journal of Chemical & Engineering Data* **2006,** *51* (6), 2145-2155.
227. Wang, Y.; Voth, G. A., Unique Spatial Heterogeneity in Ionic Liquids. *Journal of the American Chemical Society* **2005,** *127* (35), 12192-12193.
228. Canongia Lopes, J. N. A.; Pádua, A. A. H., Nanostructural Organization in Ionic Liquids. *The Journal of Physical Chemistry B* **2006,** *110* (7), 3330-3335.
229. Pensado, A. S.; Brehm, M.; Thar, J.; Seitsonen, A. P.; Kirchner, B., Effect of Dispersion on the Structure and Dynamics of the Ionic Liquid 1-Ethyl-3-methylimidazolium Thiocyanate. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2012,** *13* (7), 1845-1853.
230. Malberg, F.; Pensado, A. S.; Kirchner, B., The bulk and the gas phase of 1-ethyl-3-methylimidazolium ethylsulfate: dispersion interaction makes the difference. *Physical Chemistry Chemical Physics* **2012,** *14* (35), 12079-12082.
231. Canongia Lopes, J. N.; Costa Gomes, M. F.; Pádua, A. A. H., Nonpolar, Polar, and Associating Solutes in Ionic Liquids. *The Journal of Physical Chemistry B* **2006,** *110* (34), 16816-16818.
232. Weber, H.; Hollóczki, O.; Pensado, A. S.; Kirchner, B., Side chain fluorination and anion effect on the structure of 1-butyl-3-methylimidazolium ionic liquids. *The Journal of Chemical Physics* **2013,** *139* (8), 084502.
233. Menjoge, A.; Dixon, J.; Brennecke, J. F.; Maginn, E. J.; Vasenkov, S., Influence of Water on Diffusion in Imidazolium-Based Ionic Liquids: A Pulsed Field Gradient NMR study. *The Journal of Physical Chemistry B* **2009,** *113* (18), 6353-6359.
234. Tokuda, H.; Hayamizu, K.; Ishii, K.; Susan, M. A. B. H.; Watanabe, M., Physicochemical Properties and Structures of Room Temperature Ionic Liquids. 2. Variation of Alkyl Chain Length in Imidazolium Cation. *The Journal of Physical Chemistry B* **2005,** *109* (13), 6103-6110.
235. Brehm, M.; Sebastiani, D., Simulating structure and dynamics in small droplets of 1-ethyl-3-methylimidazolium acetate. *The Journal of Chemical Physics* **2018,** *148* (19), 193802.
236. Fendt, S.; Padmanabhan, S.; Blanch, H. W.; Prausnitz, J. M., Viscosities of Acetate or Chloride-Based Ionic Liquids and Some of Their Mixtures with Water or Other Common Solvents. *Journal of Chemical & Engineering Data* **2011,** *56* (1), 31-34.
237. Kim, H.; Ahn, Y.; Kwak, S.-Y., Comparing the influence of acetate and chloride anions on the structure of ionic liquid pretreated lignocellulosic biomass. *Biomass and Bioenergy* **2016,** *93*, 243-253.
238. Zhang, Y.; Chan, J. Y. G., Sustainable chemistry: imidazolium salts in biomass conversion and CO2 fixation. *Energy & Environmental Science* **2010,** *3* (4), 408-417.
239. Shiflett, M. B.; Yokozeki, A., Phase Behavior of Carbon Dioxide in Ionic Liquids: [emim][Acetate], [emim][Trifluoroacetate], and [emim][Acetate] + [emim][Trifluoroacetate] Mixtures. *Journal of Chemical & Engineering Data* **2009,** *54* (1), 108-114.
240. Jaeger, V.; Burney, P.; Pfaendtner, J., Comparison of Three Ionic Liquid-Tolerant Cellulases by Molecular Dynamics. *Biophysical Journal* **2015,** *108* (4), 880-892.

241. Nie, Y.; Zhang, S.; Zhang, X.; Bai, L.; He, H.; Huang, J. Preparation method of ionic liquid keratin spinning solution. CN103710775A, 2014.
242. Hogg, S., The Development of keratin dissolution in ionic liquids. 2015.
243. Fujita, K.; MacFarlane, D. R.; Forsyth, M., Protein solubilising and stabilising ionic liquids. *Chemical Communications* **2005,** (38), 4804-4806.
244. Ghosh, A.; Clerens, S.; Deb-Choudhury, S.; Dyer, J. M., Thermal effects of ionic liquid dissolution on the structures and properties of regenerated wool keratin. *Polym. Degrad. Stab.* **2014,** *108*, 108-115.
245. Zhang, T.; Datta, S.; Eichler, J.; Ivanova, N.; Axen, S. D.; Kerfeld, C. A.; Chen, F.; Kyrpides, N.; Hugenholtz, P.; Cheng, J.-F.; Sale, K. L.; Simmons, B.; Rubin, E., Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance. *Green Chemistry* **2011,** *13* (8), 2083-2090.
246. Nordwald, E. M.; Kaar, J. L., Stabilization of enzymes in ionic liquids via modification of enzyme charge. *Biotechnology and Bioengineering* **2013,** *110*, 2352-2360.
247. Betts, M. J. a. R., R. B., Amino Acid Properties and Consequences of Substitutions. In *Bioinformatics for Geneticists*.
248. Cammarata, L.; Kazarian, S. G.; Salter, P. A.; Welton, T., Molecular states of water in room temperature ionic liquids. *Physical Chemistry Chemical Physics* **2001,** *3* (23), 5192-5200.
249. Anjana, R.; Vaishnavi, M. K.; Sherlin, D.; Kumar, S. P.; Naveen, K.; Kanth, P. S.; Sekar, K., Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. *Bioinformation* **2012,** *8* (24), 1220-1224.
250. Dougherty, D. A., Cation-π Interactions Involving Aromatic Amino Acids. *The Journal of Nutrition* **2007,** *137* (6), 1504S-1508S.
251. Capela, E. V.; Quental, M. V.; Coutinho, J. A. P.; Freire, M. G., Effective separation of aromatic and aliphatic amino acids mixtures using ionic-liquid-based aqueous biphasic systems. *Green chemistry : an international journal and green chemistry resource : GC* **2017,** *19* (8), 1850-1854.
252. van Spronsen, F. J.; van Rijn M Fau - Bekhof, J.; Bekhof J Fau - Koch, R.; Koch R Fau - Smit, P. G.; Smit, P. G., Phenylketonuria: tyrosine supplementation in phenylalanine-restricted diets. (0002-9165 (Print)).
253. Hanke, C. G.; Johansson, A.; Harper, J. B.; Lynden-Bell, R. M., Why are aromatic compounds more soluble than aliphatic compounds in dimethylimidazolium ionic liquids? A simulation study. *Chemical Physics Letters* **2003,** *374*, 85-90.
254. Liao, S.-M.; Du, Q.-S.; Meng, J.-Z.; Pang, Z.-W.; Huang, R.-B., The multiple roles of histidine in protein interactions. *Chemistry Central Journal* **2013,** *7* (1), 44.
255. van Rantwijk, F.; Sheldon, R. A., Biocatalysis in Ionic Liquids. *Chemical Reviews* **2007,** *107* (6), 2757-2785.
256. Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Calderone, T. L.; Oas, T. G., The energy landscape of a fast-folding protein mapped by Ala→Gly Substitutions. *Nature Structural Biology* **1997,** *4*, 305.
257. Dobson, C. M., Protein folding and misfolding. *Nature* **2003,** *426*, 884.
258. Chaudhuri, T. K.; Paul, S., Protein-misfolding diseases and chaperone-based therapeutic approaches. *The FEBS Journal* **2006,** *273* (7), 1331-1349.
259. García, A. E.; Sanbonmatsu, K. Y., Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins: Structure, Function, and Bioinformatics* **2000,** *42* (3), 345-354.

260. Kumar, A.; Bisht, M.; Venkatesu, P., Biocompatibility of ionic liquids towards protein stability: A comprehensive overview on the current understanding and their implications. *International Journal of Biological Macromolecules* **2017,** *96*, 611-651.
261. Patel, R.; Kumari, M.; Khan, A. B., Recent Advances in the Applications of Ionic Liquids in Protein Stability and Activity: A Review. *Applied Biochemistry and Biotechnology* **2014,** *172* (8), 3701-3720.
262. Haberler, M.; Schröder, C.; Steinhauser, O., Hydrated Ionic Liquids with and without Solute: The Influence of Water Content and Protein Solutes. *Journal of Chemical Theory and Computation* **2012,** *8* (10), 3911-3928.
263. Lim, G. S.; Klähn, M., On the Stability of Proteins Solvated in Imidazolium-Based Ionic Liquids Studied with Replica Exchange Molecular Dynamics. *The Journal of Physical Chemistry B* **2018,** *122* (39), 9274-9288.
264. Groß, A.; Hashimoto, C.; Sticht, H.; Eichler, J., Synthetic Peptides as Protein Mimics. *Frontiers in bioengineering and biotechnology* **2016,** *3*, 211-211.
265. Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H., Folding dynamics and pathways of the trp-cage miniproteins. *Biochemistry* **2014,** *53* (38), 6011-6021.
266. D'Agostino, C.; Mantle, M. D.; Mullan, C. L.; Hardacre, C.; Gladden, L. F., Diffusion, Ion Pairing and Aggregation in 1-Ethyl-3-Methylimidazolium-Based Ionic Liquids Studied by 1H and 19F PFG NMR: Effect of Temperature, Anion and Glucose Dissolution. *ChemPhysChem* **2018,** *19* (9), 1081-1088.
267. Bystroff, C.; Baker, D., Prediction of local structure in proteins using a library of sequence-structure motifs1. *Journal of Molecular Biology* **1998,** *281* (3), 565-577.
268. Jens, S., Aqueous ionic liquids and their effects on protein structures: an overview on recent theoretical and experimental results. *Journal of Physics: Condensed Matter* **2017,** *29* (23), 233001.
269. Soto, P.; Baumketner, A.; Shea, J.-E., Aggregation of polyalanine in a hydrophobic environment. *The Journal of Chemical Physics* **2006,** *124* (13), 134904.
270. Nguyen, H. D.; Hall, C. K., Phase diagrams describing fibrillization by polyalanine peptides. *Biophysical journal* **2004,** *87* (6), 4122-4134.
271. Nguyen, H. D.; Marchut, A. J.; Hall, C. K., Solvent effects on the conformational transition of a model polyalanine peptide. *Protein Science* **2008,** *13* (11), 2909-2924.
272. Wassenaar, T. A.; Mark, A. E., The effect of box shape on the dynamic properties of proteins simulated under periodic boundary conditions. *Journal of Computational Chemistry* **2005,** *27* (3), 316-325.
273. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V., Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **1963,** *7* (1), 95-99.
274. Yamada, T.; Tominari, Y.; Tanaka, S.; Mizuno, M., Infrared Spectroscopy of Ionic Liquids Consisting of Imidazolium Cations with Different Alkyl Chain Lengths and Various Halogen or Molecular Anions with and without a Small Amount of Water. *The Journal of Physical Chemistry B* **2017,** *121* (14), 3121-3129.

APPENDICES

Appendix A. 1: Karl Fischer measurement

Table A. 1: Karl Fischer measurements for water in Ionic liquids taken from Steven Maxwell Hogg's work

Ionic Liquid	Karl Fischer Measurement	per gram of IL	per litre of IL
$[C_2C_1Im][CH_3CO_2]$	1292.50 ppm	0.00129250 g	1.29 g
$[C_2C_1Im][Cl]$	16175.00 ppm	0.01617500 g	16.1 g

Appendix A. 2: Calculation of number of water molecules

Table A. 2: Calculation for number of water molecules from Karl Fischer measurements

Ionic liquid	Mol.wt	Amount of water in 1g IL g/g	IL:H_20 (molar ratio)	IL:H_2O with 40 AA molecules
$[C_2C_1Im][CH_3CO_2]$	170.21	0.00129250	(1/170.21) : (0.00129250/18) 0.005875 : 0.00007180 (0.005875 : 0.00007180)/0.005875 1mole : 0.012232 moles 200: 2.44 2/202 = 0.0099% H_2O in IL-H_2O mixture	200 ($[C_2C_1Im]^+$ cations) 200 ($[CH_3CO_2]^-$ anions) 2 (H_2O molecules) 40 AA molecules
$[C_2C_1Im][Cl]$	146.62	0.016175	(1/146.62) : (0.01617500/18) 0.00682 : 0.0008986 1 : 0.13176 200 : 26.4 26/226 = 0.115% H2O in IL-H_2O mixture	200 ($[C_2C_1Im]^+$ cations) 200 (Cl^- ions) 26 (H_2O molecules) 40 AA molecules

Appendix A. 3: Input Files for Lysine Molecule QM calculations

Appendix A.3. 1: Geometry Optimisation

```
%CHK=output.chk
#P B3LYP/6-31G(d) OPT(EstmFC)* Geom=Connectivity

LYS-charged
```

```
1 1
C
N 1 B1
H 2 B2 1 A1
H 2 B3 1 A2 3 D1
H 2 B4 1 A3 3 D2
C 1 B5 2 A4 3 D3
O 6 B6 1 A5 2 D4
O 6 B7 1 A6 2 D5
C 1 B8 2 A7 3 D6
C 9 B9 1 A8 2 D7
C 10 B10 9 A9 1 D8
C 11 B11 10 A10 9 D9
N 12 B12 11 A11 10 D10
H 13 B13 12 A12 11 D11
H 13 B14 12 A13 11 D12
H 13 B15 12 A14 11 D13
H 12 B16 11 A15 10 D14
H 12 B17 11 A16 10 D15
H 11 B18 10 A17 9 D16
H 11 B19 10 A18 9 D17
H 10 B20 9 A19 1 D18
H 10 B21 9 A20 1 D19
H 9 B22 1 A21 2 D20
H 9 B23 1 A22 2 D21
H 1 B24 2 A23 3 D22

B1 1.4989406
B2 1.0343283
B3 1.0222622
B4 1.0197121
B5 1.5594258
B6 1.2550353
B7 1.2652073
B8 1.5339741
B9 1.5210495
B10 1.5191891
B11 1.5253627
B12 1.4876218
B13 1.0185108
B14 1.0392732
B15 1.0396309
B16 1.1286834
B17 1.1275009
B18 1.1261779
B19 1.1242964
B20 1.1239507
B21 1.1236448
B22 1.1234186
B23 1.1232180
B24 1.1290206
A1 107.46650
A2 109.43790
A3 111.35082
A4 108.32749
A5 117.37713
A6 117.18540
A7 112.19769
A8 112.53235
A9 112.30498
A10 113.83166
A11 113.66852
A12 110.52053
A13 110.55505
A14 110.49656
```

```
A15 108.76555
A16 110.22497
A17 110.28902
A18 109.16335
A19 108.13739
A20 109.74979
A21 110.95507
A22 108.26646
A23 107.97712
D1 117.74021
D2 -120.51003
D3 -38.988364
D4 -141.12542
D5 42.577906
D6 82.530877
D7 -151.06575
D8 92.444469
D9 -144.69457
D10 80.568344
D11 162.03152
D12 -76.584892
D13 40.117870
D14 -159.80304
D15 -40.779452
D16 -20.976121
D17 96.220327
D18 -147.47737
D19 -31.604407
D20 -28.803469
D21 88.669075

1 2 1.0 6 1.0 9 1.0 25 1.0
2 1 1.0 3 1.0 4 1.0 5 1.0
3 2 1.0
4 2 1.0
5 2 1.0
6 1 1.0 7 2.0 8 1.0
7 6 2.0
8 6 1.0
9 1 1.0 10 1.0 23 1.0 24 1.0
10 9 1.0 11 1.0 21 1.0 22 1.0
11 10 1.0 12 1.0 19 1.0 20 1.0
12 11 1.0 13 1.0 17 1.0 18 1.0
13 12 1.0 14 1.0 15 1.0 16 1.0
14 13 1.0
15 13 1.0
16 13 1.0
17 12 1.0
18 12 1.0
19 11 1.0
20 11 1.0
21 10 1.0
22 10 1.0
23 9 1.0
24 9 1.0
25 1 1.0
```

*EstmFC key word used for optimisation here denotes that estimated force constants are used to ensure the complex job finishes in time. It doesn't affect the force constants derived later for structures from forcefields. This key word was not always used, and for other structures, only the keyword OPT was used.

Appendix A.3. 2: Resp Charge Derivation

```
%chk=output.chk
%nprocshared=4
%mem=128mb
#P HF/6-31G* Pop=MK iop(6/33=2,6/42=6,6/50=1) guess=read geom=check

Lysine charged

1 1
```

Appendix A. 4: Gaussian optimisation output files for IL ions and amino acids

$[C_2C_1Im]^+$ cation

```
1\1\GINC-WS-PSC\FOpt\RB3LYP\6-31G(d)\C6H11N2(1+)\SHAFAQ\16-Jul-
2015\0\\#P               B3LYP/6-31G*               Opt\\EMIM\\1,1\C,-
0.0154616294,0.,0.0000137284\N,-
0.0068019496,0.,1.4723308183\C,1.1128073636,0.,2.2819314408\C,0.67483
68423,0.,3.5729037839\N,-0.7077747382,0.,3.5367522738\C,-
1.0946807386,0.,2.256010805\H,-2.1137356392,0.,1.9038145618\C,-
1.5826261283,0.,4.7394207947\C,-3.0685790469,0.,4.4087108485\H,-
3.6275677599,0.,5.3484550771\H,-
3.3658070498,0.8944121684,3.8516372033\H,-3.3658070498,-
0.8944121684,3.8516372033\H,-1.3109577659,-
0.8856645258,5.3208776216\H,-
1.3109577659,0.8856645258,5.3208776216\H,1.2216151548,0.,4.5033501311
\H,2.1125467987,0.,1.8756873946\H,1.0171578068,0.,-0.3480428779\H,-
0.5197802513,-0.8959292737,-0.3674514343\H,-
0.5197802513,0.8959292737,-0.3674514343\\Version=EM64L-
G09RevB.01\State=1-A'\HF=-344.5480968\RMSD=2.823e-09\RMSF=3.444e-
05\Dipole=0.1734904,0.,-0.4033156\Quadrupole=1.7188567,-
9.4768388,7.7579821,0.,-4.1334011,0.\PG=CS [SG(C6H5N2),X(H6)]\\@
```

Acetate anion

```
1\1\GINC-COMPUTE065\FOpt\RB3LYP\6-31G(d)\CH3CO2(1-)\ENXSM15\13-Jun-
2017\0\\#P  B3LYP/6-31G(d)  OPT(EstmFC)  Geom=Connectivity\\acetate\\-
1,1\C,-0.0346511582,0.,0.0421921251\C,-
0.0459552666,0.,1.6176848255\O,1.0962671601,0.,2.1441416296\O,-
1.1807087194,0.,2.1570515301\H,-1.0477995468,0.,-
0.3802000874\H,0.5089959076,0.8803652228,-
0.3316850115\H,0.5089959076,-0.8803652228,-
0.3316850115\\Version=ES64L-G16RevA.03\State=1-A'\HF=-
228.4979111\RMSD=2.721e-09\RMSF=7.668e-05\Dipole=0.0135354,0.,-
1.2532329\Quadrupole=-3.0043663,3.8742145,-0.8698482,0.,-
0.0566209,0.\PG=CS [SG(C2H1O2),X(H2)]\\@

1\1\GINC-COMPUTE046\FOpt\RB3LYP\6-31G(d)\CH3CO2(1-)\ENXSM15\07-May-
2019\0\\#P B3LYP/6-31G(d) OPT Geom=Connectivity\\CH3CO2(-1)\\-1,1\C,-
0.0346511582,0.,0.0421921251\C,-
0.0459552666,0.,1.6176848255\O,1.0962671601,0.,2.1441416296\O,-
1.1807087194,0.,2.1570515301\H,-1.0477995468,0.,-
0.3802000874\H,0.5089959076,0.8803652228,-
0.3316850115\H,0.5089959076,-0.8803652228,-
0.3316850115\\Version=ES64L-G16RevA.03\State=1-A'\HF=-
228.4979111\RMSD=2.721e-09\RMSF=7.668e-05\Dipole=0.0135354,0.,-
```

1.2532329\Quadrupole=-3.0043663,3.8742145,-0.8698482,0.,-
0.0566209,0.\PG=CS [SG(C2H1O2),X(H2)]\\@

Lysine

1\1\GINC-COMPUTE018\FOpt\RB3LYP\6-31G(d)\C6H15N2O2(1+)\ENXSM15\16-
May2019\0\\#P B3LYP/6-31G(d) OPT(EstmFC) SCRF=(PCM,Solvent=Water)
Geom=Connectivity\\LYS-charged\\1,1\C,0.0359485934,-
0.0059748732,0.0385157426\N,-
0.2097540102,0.0046349411,1.5304455285\H,0.752725029,0.0071321284,1.9
332040893\H,-0.6845144092,0.8562878396,1.8431638343\H,-0.7463928079,-
0.7978718385,1.8715933567\C,1.4394134063,0.6333060401,-
0.153592769\O,1.6524539034,1.1072550559,-
1.3066799286\O,2.2131913037,0.5625966137,0.8267257546\C,-
0.0367140044,-1.4603901985,-0.4767742602\C,0.5965649695,-
1.656362124,-1.8705887465\C,2.0961164091,-2.0435489108,-
1.817375751\C,2.9214453595,-1.5399349134,-
3.008388593\N,3.5174204473,-0.1809875786,-
2.7186063099\H,3.806139359,0.2953096242,-
3.5772391876\H,2.8323594142,0.4434869045,-
2.1832437054\H,4.3532935388,-0.2727859637,-
2.1348349254\H,3.7567548832,-2.2020365596,-
3.2407774549\H,2.3102636282,-1.4316193819,-
3.9066267747\H,2.5604208207,-1.6870822396,-
0.889341925\H,2.1822331964,-3.1345683037,-
1.7842104638\H,0.0516503731,-2.4457586515,-
2.3983049645\H,0.4627656396,-0.7406552736,-
2.4541913704\H,0.4618564289,-2.120937966,0.2450639178\H,-
1.0916938538,-1.7547083106,-0.4978664982\H,-
0.7264218107,0.6100261568,-0.4382455995\\Version=ES64L-
G16RevA.03\State=1-A\HF=-497.5014899\RMSD=8.317e-09\RMSF=2.908e-
05\Dipole=-0.5825441,-1.1680808,-0.0930212\Quadrupole=7.0997756,-
15.8076447,8.7078691,-3.1880188,-25.6799035,1.9559422\PG=C01

Arginine

1\1\GINC-COMPUTE012\FOpt\RB3LYP\6-31G(d)\C6H15N4O2(1+)\ROOT\29-Sep-
2020\0\\#P B3LYP/6-31G(d) OPT(EstmFC) SCRF=(PCM,Solvent=Water)
Geom=check guess=read\\ARG\\1,1\C,2.2839088963,-0.2539992197,-
0.5647230023\C,1.5859932879,-1.5457624335,-
0.0949523708\C,0.0837947542,-1.6064531044,-0.4467302763\C,-
0.8000227844,-1.1047120747,0.7087526429\N,-2.2056306697,-
0.9395268421,0.3340220545\C,-
2.8124776472,0.2171972762,0.0336091402\N,-4.1545036919,0.2372065356,-
0.0759149448\H,-4.6131924815,0.9927681553,-0.56469596\H,-
4.6987263287,-0.6000462892,0.075661924\N,-2.1249988188,1.351639017,-
0.1023959217\H,-2.6370014195,2.2125014694,-0.2375827622\H,-
1.1062623511,1.3780325127,-0.2851907129\H,-2.7706688013,-
1.7774170134,0.2891908094\H,-0.7705072769,-
1.8199952125,1.5367099921\H,-0.4436360972,-
0.1519469058,1.1032036367\H,-0.116424838,-1.010145888,-
1.3409574063\H,-0.1929628087,-2.6412110864,-
0.6747282153\H,2.1067977344,-2.3948782605,-
0.5507807395\H,1.7145524788,-
1.6445878718,0.991923692\N,3.6611483193,-
0.1810991276,0.0567549879\H,4.3575330797,0.2208745474,-
0.576390349\H,3.5142779737,0.4882264254,0.8505142568\H,4.0198779874,-
1.0795289893,0.3914283062\C,1.5599018828,1.0471816603,-

0.0999479127\O,0.5569551883,1.3734601313,-
0.7849935284\O,2.0407396649,1.5971853039,0.9204527828\H,2.3971040008,
-0.2532629908,-1.6491659954\\Version=ES64L-G16RevA.03\State=1-A\HF=-
607.027585\RMSD=5.770e-09\RMSF=6.008e-06\Dipole=-0.6875369,-
2.7884441,-0.4626563\Quadrupole=45.6494628,-19.7157053,-25.9337575,-
8.9655775,-0.6999883,-3.9645634\PG=C01 [X(C6H15N4O2)]\\@

Aspartate

1\1\GINC-LOGIN01\FOpt\RB3LYP\6-31G(d)\C4H6N1O4(1-)\ENXSM15\13-Nov-
2015\0\\#P B3LYP/6-31G(d) OPT(EstmFC) SCRF=(PCM,Solvent=Water)
Geom=Connectivity\\C4H6O4N(-1)\\-
1,1\C,0.0489360429,0.0721034747,0.0789727454\C,0.0749090885,0.0112660
357,1.6172064553\N,1.4927597056,-
0.2034479056,2.0648922814\H,1.6743586121,0.2086974727,2.9887378837\H,
2.1039046559,0.2572463286,1.3138087222\H,1.602328169,-
1.2299422922,2.1510730502\C,-0.7900686125,-
1.170810222,2.1431898613\O,-2.0292134314,-
0.9958913447,2.1212633663\O,-0.1418942959,-
2.1990407945,2.4905272458\H,-
0.2696662725,0.959051521,2.0429953687\C,1.0454043294,1.0679904733,-
0.5682909417\O,0.6668889413,1.7059471361,-
1.570612479\O,2.2141946975,1.1267869451,-0.0433971488\H,-
0.9688229664,0.311461406,-0.23711753\H,0.2827392412,-0.92994186,-
0.3132086121\\Version=AM64L-G03RevD.01\State=1-A\HF=-
511.8649734\RMSD=4.868e-09\RMSF=5.240e-
05\Thermal=0.\Dipole=1.9123822,0.0840468,2.082323\PG=C01
[X(C4H6N1O4)]\\@

Glutamate

1\1\GINC-COMPUTE079\FOpt\RB3LYP\6-31G(d)\C5H8N1O4(1-)\ENXSM15\28-Sep-
2020\0\\#P B3LYP/6-31G(d) OPT SCRF=(PCM,Solvent=Water) Geom=Check
guess=read\\Glutamic_acid\\-1,1\C,1.4587826175,-
1.0579524119,0.2175823136\C,0.0248327448,-0.9797271433,-
0.3400669612\C,-0.9400675807,-0.0398913757,0.3964128477\N,-
0.4753007888,1.3856577776,0.3014712569\H,-
0.5430451731,1.8774180225,1.1934176971\H,0.5547715003,1.4170865096,-
0.070977625\H,-1.1720556898,1.8045401797,-0.3516750195\C,-
2.3759885854,-0.1031268096,-0.2267828591\O,-2.9811578118,-
1.1871665272,-0.0914945752\O,-2.7485365637,0.9488568639,-
0.8237469422\H,-0.9893992803,-
0.3177999181,1.4527646278\H,0.053341977,-0.7004165767,-
1.4021503472\H,-0.4409926966,-1.9690507405,-
0.2911648049\C,2.4535800748,0.0405366815,-
0.2374923915\O,1.9785827798,1.194718704,-0.5412164434\O,3.661969427,-
0.26366957,-0.2641852552\H,1.9048229542,-2.0199357987,-
0.0568653519\H,1.4288978731,-
1.0479754779,1.3180235947\\Version=ES64L-G16RevA.03\State=1-A\HF=-
551.1689795\RMSD=2.123e-09\RMSF=1.785e-05\Dipole=-
0.2648379,0.7336439,1.829561\Quadrupole=-
37.3401144,16.5543343,20.7857802,-6.3161741,-
1.8281853,4.4934599\PG=C01 [X(C5H8N1O4)]\\@

Alanine

1\1\GINC-NODE025\FOpt\RHF\6-31G(d)\C3H7N1O2\ENXSM15\27-Feb-2018\0\\#P
HF/6-31G*   Opt\\ALA\\0,1\C,-0.1178034469,-
0.0682121766,0.0267417277\N,-
0.0181384974,0.0511652182,1.5306850106\H,0.7075195262,0.8307805472,1.
5610815734\H,-0.8733246278,0.3531888015,1.9650798263\H,0.3050215519,-
0.7789208074,1.9950347798\C,-1.551797186,0.0190075238,-
0.4617670721\H,-1.5358635457,0.0112683971,-1.5420545145\H,-
2.1576356888,-0.8117279578,-0.1115470135\H,-
2.0148178057,0.9499494332,-0.1470969043\C,0.7982245198,1.0926358305,-
0.504637041\O,1.3831861214,1.6893637224,0.4201213564\O,0.8065840507,1
.232012599,-1.7026573369\H,0.3254829797,-1.0154937967,-
0.2468645274\\Version=AM64L-G03RevD.01\State=1-A\HF=-
321.8258899\RMSD=6.139e-09\RMSF=1.959e-05\Thermal=0.\Dipole=-
1.8273557,-2.22107,2.6434247\PG=C01 [X(C3H7N1O2)]\\@

Phenylalanine

1\1\GINC-NODE070\FOpt\RB3LYP\6-31G(d)\C9H11N1O2\ENXSM15\16-Oct-
2016\0\\#P   B3LYP/6-31G*   OPT   SCRF=(PCM,Solvent=Water)
Geom=Connectivity\\Phenylalanine\\0,1\C,0.070541936,0.2266798057,-
0.0820380551\C,0.2106944246,-
0.4063922578,1.3050094358\C,1.5853267098,-
0.2541444334,1.9252223671\C,2.6941249373,-
0.9216638557,1.3795401089\C,3.9626479481,-
0.7773799657,1.9427830656\C,4.145540554,0.0392003066,3.0641131187\C,3
.0505404768,0.7055646846,3.6169807446\C,1.780558401,0.5578042677,3.05
02902324\H,0.9285292987,1.0746817546,3.4935028259\H,3.1804057335,1.34
09056396,4.4920790291\H,5.1354068222,0.1516734974,3.5039487576\H,4.80
98725417,-1.3071786803,1.5093320115\H,2.5612117438,-
1.5674320509,0.5109257581\H,-0.0449358797,-
1.462880511,1.169835503\H,-
0.5529974972,0.0178232356,1.9714216411\N,0.2310374439,1.7294207526,-
0.0484858304\H,0.0341834787,2.1398249508,0.8764916815\H,-
0.5256621463,2.0429629324,-0.7113971834\H,1.1567291464,2.0611009365,-
0.3554481636\C,-1.3167225401,-0.0381963845,-0.766466474\O,-
1.8548694698,0.985014099,-1.2830063297\O,-1.7231259399,-
1.2154427913,-0.7489783804\H,0.8533549325,-0.1482249085,-
0.752256227\\Version=AM64L-G03RevD.01\State=1-A\HF=-
554.8100429\RMSD=3.163e-09\RMSF=1.558e-
04\Thermal=0.\Dipole=3.6715134,2.6784751,1.7918657\PG=C01
[X(C9H11N1O2)]\\@

Histidine

1\1\GINC-COMPUTE079\FOpt\RB3LYP\6-31G(d)\C7H11N3O2\ENXSM15\28-Sep-
2020\0\\#P   B3LYP/6-31G(d)   OPT   SCRF=(PCM,Solvent=Water)
Geom=Connectivity\\his-neu\\0,1\C,-0.0902607443,-
0.0290911158,0.0469474084\C,-
0.021489351,0.1033095246,1.5743085593\C,1.3687349807,-
0.2178769357,2.1547048903\C,1.3878656203,-
0.1536935195,3.6536550265\C,2.0884945737,0.7098144427,4.4632154998\N,
1.756757331,0.361661999,5.7557485881\C,0.8831929064,-
0.6789510368,5.6825746163\N,0.6346549742,-
1.01975433,4.4317972833\H,0.4672105598,-
1.1460010754,6.5652026467\H,2.102851276,0.7964528563,6.6002841176\H,2
.7743935267,1.5146339518,4.2456986123\H,1.65846526,-

1.2263287651,1.8261278763\H,2.1031797667,0.4735558548,1.7354046234\H,
-0.3184093534,1.1192965468,1.8668932002\H,-0.7475860297,-
0.5856836079,2.0242149374\N,-1.5284076695,0.0402339716,-
0.4269231685\H,-1.827673472,-0.783859099,-0.9517626255\H,-
2.2016964744,0.198875123,0.3265717053\H,-1.469777675,0.8918794843,-
1.0678498623\C,0.6391264017,1.1186215791,-
0.7419083913\O,1.8682282464,1.2160167244,-0.581449043\O,-
0.1303537986,1.8248306297,-1.4639820066\H,0.3116956713,-
0.9953110966,-0.2672250924\\Version=ES64L-G16RevA.03\State=1-A\HF=-
588.0885805\RMSD=5.513e-09\RMSF=3.393e-06\Dipole=-2.6572176,-
1.4919168,2.3677953\Quadrupole=7.8091879,-
9.997066,2.1878782,3.4600291,20.3290299,20.1852637\PG=C01
[X(C7H11N3O2)]\\@

Tyrosine

1\1\GINC-NODE064\FOpt\RB3LYP\6-31G(d)\C9H11N1O3\ENXSM15\17-Oct-
2016\0\\#P   B3LYP/6-31G*   OPT(EstmFC)   SCRF=(PCM,Solvent=Water)
Geom=Connectivity\\tyrosine\\0,1\C,-
0.5270824208,0.1700372743,0.1036929156\C,0.071843773,0.0255126663,1.5
142291202\N,1.5745392927,-
0.1116376074,1.4522119707\H,2.04566023,0.4151061391,2.2028378061\H,1.
7131337205,-
1.1382497482,1.6417917833\H,1.9958994203,0.1754586268,0.5588409289\C,
-0.4466098368,-1.2751752363,2.2120401721\O,0.4171809808,-
2.1929845144,2.3355369529\O,-1.6505829633,-
1.2776209272,2.5329960553\H,-
0.1600748029,0.905271093,2.1216651634\C,-0.0512094623,1.3935152477,-
0.6513336805\C,-0.4654783492,2.6809951761,-0.27541298\C,-
0.0219445394,3.8123417492,-0.9564334569\C,0.8592649019,3.6772319342,-
2.0395211795\C,1.2838828602,2.4017345285,-
2.4289191206\C,0.8270487467,1.2781164586,-
1.7389961801\H,1.1554330336,0.2892173105,-
2.0613470295\H,1.9641842596,2.3026557565,-
3.2727259235\O,1.324567356,4.7438971618,-
2.7506550477\H,0.9779641008,5.5836480678,-2.362137845\H,-
0.3585430185,4.8050361761,-0.6558817721\H,-
1.154301667,2.8046250654,0.5607939074\H,-0.300258751,-0.7383986794,-
0.4706566707\H,-
1.6132522622,0.1918617225,0.2466953227\\Version=AM64L-
G03RevD.01\State=1-A\HF=-630.0352568\RMSD=6.803e-09\RMSF=2.077e-
04\Thermal=0.\Dipole=2.9716843,3.7863479,-1.4614349\PG=C01
[X(C9H11N1O3)]\\@

Tryptophan

1\1\GINC-NODE074\FOpt\RB3LYP\6-31G(d)\C11H12N2O2\ENXSM15\24-Oct-
2016\0\\#P   B3LYP/6-31G(d)   OPT   SCRF=(PCM,Solvent=Water)
Geom=Connectivity\\tryptophan\\0,1\C,-0.0508338482,0.3966219959,-
0.0837517243\C,0.1498234513,-
0.2584362531,1.3024927029\N,1.5794932503,-
0.0308645928,1.7377959667\H,2.2490041527,0.1014708298,0.9643030543\H,
1.9300953514,-
0.7961526103,2.3335899826\H,1.501678142,0.8314964346,2.3316221008\C,-
0.7191838299,0.4222609637,2.4114489847\O,-
1.8900103448,0.0214504398,2.538313353\O,-
0.1181196172,1.3446532528,3.0422791958\H,-0.0285064902,-
1.3357188144,1.2650645026\C,-1.3742580637,0.1044826396,-

```
0.725755051\C,-1.7695356454,-1.1166957157,-1.3852642422\C,-
1.1319119189,-2.3457991444,-1.641625309\C,-1.8230095054,-
3.3353835329,-2.3330491093\C,-3.145943796,-3.1246462756,-
2.780398866\C,-3.8024408964,-1.9199543866,-2.5471764274\C,-
3.1041192925,-0.9262110244,-1.8496252207\N,-
3.4865982455,0.3468053205,-1.4856912446\C,-
2.4477237275,0.9590386264,-0.8192382836\H,-
2.5539034513,1.9741654031,-0.4523705479\H,-
4.3894115648,0.7788947713,-1.691414011\H,-4.8214933268,-
1.7529274947,-2.8932651587\H,-3.6612729427,-3.9182727513,-
3.3190215313\H,-1.3384371432,-4.2893858596,-2.534776754\H,-
0.1109363843,-2.5227098208,-1.304452014\H,0.7600453138,0.0475428252,-
0.7417344339\H,0.0796254497,1.481006996,0.0307689673\\Version=AM64L-
G03RevD.01\State=1-A\HF=-686.384232\RMSD=3.643e-09\RMSF=1.068e-
04\Thermal=0.\Dipole=3.1722197,-0.5079161,-2.2102522\PG=C01
[X(C11H12N2O2)]\\@
```

Appendix A.4. 1: Gaussian frequency analysis (shortened) output file for lysine

1\1\GINC-COMPUTE053\Freq\RB3LYP\6-31G(d)\C6H15N2O2(1+)\ROOT\16-May-2019\0\\#P B3LYP/6-31G(d) FREQ SCRF=(PCM,Solvent=Water) guess=read geom=check\\Lysine charged\\1,1\C,0.0359485934,-0.0059748732,0.0385157426\N,-0.2097540102,0.0046349411,1.5304455285\H,0.752725029,0.0071321284,1.9332040893\H,……
……
……
0.7264218107,0.6100261568,-0.4382455995\\Version=ES64L-G16RevA.03\State=1-A\HF=-497.5014899\RMSD=5.605e-09\RMSF=2.905e-05\ZeroPoint=0.2300239\Thermal=0.2409779\Dipole=-0.5825439,-1.1680815,-0.0930193\DipoleDeriv=-0.0234989,-0.2325302,
……
……
0.0047011,-0.0077025\Polar=100.4715496,3.1079118,90.3363979,-8.8684993,1.4632833,110.9158064\Quadrupole=7.0997784,-15.8076508,8.7078724,-3.18802,-25.6799039,1.9559435\PG=C01 [X(C6H15N2O2)]\NImag=0\\0.52625092,-0.06804507,0.48527807,0.08513579,-0.05810817,0.45745662,-0.09055664,0.00570034,0.00425018,0.70118639,-0.00115112,……...……
……
…0.00001281,-0.00000428,0.00000420,0.00000039,-0.00000340,-0.00001186,-0.00000842,-0.00001151,-0.00000394,-0.00000391,-0.00000408,0.00001006,-0.00000494,-0.00000486,0.00000879,-0.00000309,-0.00001134,0.00000406,-0.00001334,-0.00000672,0.00001432,0.00000540,0.00000606,-0.00000084,0.00000928\\\@

Appendix A. 5: Library, mol2 and frcmod files for lysine molecule

Appendix A.5. 1: Lysine mol2 file

```
@<TRIPOS>MOLECULE
LYS
   25    24     1     0     0
SMALL
resp


@<TRIPOS>ATOM
      1 C1         -1.7050    0.1870    0.4980 c3        1 LYS      0.245992
      2 N1         -2.8650   -0.1320   -0.4310 n4        1 LYS     -0.652717
```

```
      3 H1          -2.3750   -0.7450   -1.1450 hn        1 LYS      0.381547
      4 H2          -3.5930   -0.6910    0.0250 hn        1 LYS      0.381547
      5 H3          -3.3030    0.6870   -0.8620 hn        1 LYS      0.381547
      6 C2          -0.7130   -0.9820    0.2600 c         1 LYS      0.687997
      7 O1           0.1290   -1.1590    1.1810 o         1 LYS     -0.671423
      8 O2          -0.7940   -1.5280   -0.8630 o         1 LYS     -0.671423
      9 C3          -1.0730    1.5440    0.0940 c3        1 LYS     -0.231238
     10 C4           0.4340    1.6770    0.4250 c3        1 LYS     -0.026637
     11 C5           1.3710    1.1140   -0.6830 c3        1 LYS     -0.042693
     12 C6           2.6310    0.3800   -0.1930 c3        1 LYS      0.028161
     13 N2           2.3920   -1.1050   -0.0370 n4        1 LYS     -0.281554
     14 H4           3.1940   -1.5750    0.3940 hn        1 LYS      0.285598
     15 H5           1.4830   -1.2910    0.5580 hn        1 LYS      0.285598
     16 H6           2.2220   -1.5470   -0.9460 hn        1 LYS      0.285598
     17 H7           3.4580    0.4960   -0.8970 hx        1 LYS      0.076569
     18 H8           2.9620    0.7460    0.7830 hx        1 LYS      0.076569
     19 H9           0.8190    0.4440   -1.3540 hc        1 LYS      0.048050
     20 H10          1.7060    1.9440   -1.3140 hc        1 LYS      0.048050
     21 H11          0.6540    2.7400    0.5630 hc        1 LYS      0.059711
     22 H12          0.6310    1.1890    1.3840 hc        1 LYS      0.059711
     23 H13         -1.2050    1.6890   -0.9880 hc        1 LYS      0.090338
     24 H14         -1.6320    2.3450    0.5890 hc        1 LYS      0.090338
     25 H15         -2.0650    0.1990    1.5270 hx        1 LYS      0.064765
@<TRIPOS>BOND
     1    1    2 1
     2    1    6 1
     3    1    9 1
     4    1   25 1
     5    2    3 1
     6    2    4 1
     7    2    5 1
     8    6    7 1
     9    6    8 1
    10    9   10 1
    11    9   23 1
    12    9   24 1
    13   10   11 1
    14   10   21 1
    15   10   22 1
    16   11   12 1
    17   11   19 1
    18   11   20 1
    19   12   13 1
    20   12   17 1
    21   12   18 1
    22   13   14 1
    23   13   15 1
    24   13   16 1
@<TRIPOS>SUBSTRUCTURE
     1 LYS         1 TEMP              0 ********    0 ROOT
```

Appendix A.5. 2:Library file

```
!!index array str
 "LYS"
!entry.LYS.unit.atoms table  str name  str type  int typex  int resx  int
flags  int seq  int elmnt  dbl chg
 "C1" "c3" 0 1 131072 1 6 0.1967936
 "N1" "n4" 0 1 131072 2 7 -0.5221736
 "H1" "hn" 0 1 131072 3 1 0.3052376
 "H2" "hn" 0 1 131072 4 1 0.3052376
 "H3" "hn" 0 1 131072 5 1 0.3052376
 "C2" "c" 0 1 131072 6 6 0.5503976
 "O1" "o" 0 1 131072 7 8 -0.5371384
```

```
 "O2" "o" 0 1 131072 8 8 -0.5371384
 "C3" "c3" 0 1 131072 9 6 -0.1849904
 "C4" "c3" 0 1 131072 10 6 -0.0213096
 "C5" "c3" 0 1 131072 11 6 -0.0341544
 "C6" "c3" 0 1 131072 12 6 0.0225288
 "N2" "n4" 0 1 131072 13 7 -0.2252432
 "H4" "hn" 0 1 131072 14 1 0.2284784
 "H5" "hn" 0 1 131072 15 1 0.2284784
 "H6" "hn" 0 1 131072 16 1 0.2284784
 "H7" "hx" 0 1 131072 17 1 0.0612552
 "H8" "hx" 0 1 131072 18 1 0.0612552
 "H9" "hc" 0 1 131072 19 1 0.03844
 "H10" "hc" 0 1 131072 20 1 0.03844
 "H11" "hc" 0 1 131072 21 1 0.0477688
 "H12" "hc" 0 1 131072 22 1 0.0477688
 "H13" "hc" 0 1 131072 23 1 0.0722704
 "H14" "hc" 0 1 131072 24 1 0.0722704
 "H15" "hx" 0 1 131072 25 1 0.051812

!entry.LYS.unit.atomspertinfo table  str pname  str ptype  int ptypex  int
pelmnt  dbl pchg
"C1" "c3" 0 -1 0.0
 "N1" "n4" 0 -1 0.0
 "H1" "hn" 0 -1 0.0
 "H2" "hn" 0 -1 0.0
 "H3" "hn" 0 -1 0.0
 "C2" "c" 0 -1 0.0
 "O1" "o" 0 -1 0.0
 "O2" "o" 0 -1 0.0
 "C3" "c3" 0 -1 0.0
 "C4" "c3" 0 -1 0.0
 "C5" "c3" 0 -1 0.0
 "C6" "c3" 0 -1 0.0
 "N2" "n4" 0 -1 0.0
 "H4" "hn" 0 -1 0.0
 "H5" "hn" 0 -1 0.0
 "H6" "hn" 0 -1 0.0
 "H7" "hx" 0 -1 0.0
 "H8" "hx" 0 -1 0.0
 "H9" "hc" 0 -1 0.0
 "H10" "hc" 0 -1 0.0
 "H11" "hc" 0 -1 0.0
 "H12" "hc" 0 -1 0.0
 "H13" "hc" 0 -1 0.0
 "H14" "hc" 0 -1 0.0
 "H15" "hx" 0 -1 0.0
!entry.LYS.unit.boundbox array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.LYS.unit.childsequence single int
2
!entry.LYS.unit.connect array int
 0
 0
!entry.LYS.unit.connectivity table  int atom1x  int atom2x  int flags
 1 2 1
 1 6 1
 1 9 1
 1 25 1
 2 3 1
 2 4 1
 2 5 1
```

```
 6 7 1
 6 8 1
 9 10 1
 9 23 1
 9 24 1
 10 11 1
 10 21 1
 10 22 1
 11 12 1
 11 19 1
 11 20 1
 12 13 1
 12 17 1
 12 18 1
 13 14 1
 13 15 1
 13 16 1
!entry.LYS.unit.hierarchy table  str abovetype  int abovex  str belowtype  int
belowx
 "U" 0 "R" 1
 "R" 1 "A" 1
 "R" 1 "A" 2
 "R" 1 "A" 3
"R" 1 "A" 4
 "R" 1 "A" 5
 "R" 1 "A" 6
 "R" 1 "A" 7
 "R" 1 "A" 8
 "R" 1 "A" 9
 "R" 1 "A" 10
 "R" 1 "A" 11
 "R" 1 "A" 12
 "R" 1 "A" 13
 "R" 1 "A" 14
 "R" 1 "A" 15
 "R" 1 "A" 16
 "R" 1 "A" 17
 "R" 1 "A" 18
 "R" 1 "A" 19
 "R" 1 "A" 20
 "R" 1 "A" 21
 "R" 1 "A" 22
 "R" 1 "A" 23
 "R" 1 "A" 24
 "R" 1 "A" 25
!entry.LYS.unit.name single str
 "LYS"
!entry.LYS.unit.positions table  dbl x  dbl y  dbl z
 -1.705000 0.187000 0.498000
 -2.865000 -0.132000 -0.431000
 -2.375000 -0.745000 -1.145000
 -3.593000 -0.691000 0.025000
 -3.303000 0.687000 -0.862000
 -0.713000 -0.982000 0.260000
 0.129000 -1.159000 1.181000
 -0.794000 -1.528000 -0.863000
 -1.073000 1.544000 0.094000
 0.434000 1.677000 0.425000
 1.371000 1.114000 -0.683000
 2.631000 0.380000 -0.193000
 2.392000 -1.105000 -0.037000
 3.194000 -1.575000 0.394000
 1.483000 -1.291000 0.558000
 2.222000 -1.547000 -0.946000
 3.458000 0.496000 -0.897000
```

```
 2.962000 0.746000 0.783000
 0.819000 0.444000 -1.354000
 1.706000 1.944000 -1.314000
 0.654000 2.740000 0.563000
 0.631000 1.189000 1.384000
 -1.205000 1.689000 -0.988000
 -1.632000 2.345000 0.589000
 -2.065000 0.199000 1.527000
!entry.LYS.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int
c5x  int c6x
 0 0 0 0 0 0
!entry.LYS.unit.residues table  str name  int seq  int childseq  int startatomx
str restype  int imagingx
 "LYS" 1 26 1 "?" 0
!entry.LYS.unit.residuesPdbSequenceNumber array int
 0
!entry.LYS.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.LYS.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
```

Appendix A.5. 3: frcmod file for lysine

```
remark goes here
MASS

BOND

ANGLE

DIHE

IMPROPER
c3-o -c -o           1.1           180.0          2.0           General
improper torsional angle (1 general atom type)

NONBON
```

Appendix A. 6: Input files for Packmol, co-ordinates and MD Simulation

Appendix A.6. 1: Packmol input file for building [C_2C_1Im][Cl]:[LYS] box for high concentration (5:1) system

```
tolerance 2.0
filetype pdb
output box-lys-charged.pdb

structure EMI.pdb
  number 200
  inside box 0. 0. 0. 45. 45. 45.
end structure

structure CL.pdb
  number 240
  inside box 0. 0. 0. 45. 45. 45.
end structure

structure lys-charged.pdb
  number 40
  inside box 0. 0. 0. 45. 45. 45.
end structure
```

Appendix A.6. 2: Input file for co-ordinate files (commands.in)

```
source leaprc.gaff
loadamberparams frcmod.ionsjc_spce
loadoff CL-trial.lib
loadoff EMI.lib
loadoff lys-charged.lib
loadamberparams emim.frcmod
loadamberparams lys-charged.frcmod
loadamberparams CL-trial.frcmod
box= loadpdb box-lys-charged.pdb
setbox box centers
saveamberparm box box.prmtop box.inpcrd
```

Appendix A.6. 3: Input file for first energy minimisation stage

```
EMIM-Cl-LYS initial minimization
 &cntrl
  imin= 1,
  maxcyc=20000,
  ntpr=100,
  ntwx=100,
  ncyc=2500,
  ntb=1,
  ntr=1,
  restraint_wt=5.0,
  restraintmask=':LYS'
/
```

Appendix A.6. 4: Input file for second energy minimisation stage

```
minimization 2: no restraints
 &cntrl
  imin   = 1,
  maxcyc = 20000,
  ncyc   = 500,
  ntb    = 1,
  ntr    = 0,
  cut    = 10.0
 /
```

Appendix A.6. 5: Input file for first and second MD simulation run

First stage, Heating

```
Heating
 &cntrl
  imin=0,
  irest=0,
  ntx=1,
  nstlim=50000, dt=0.001,
  ntc=2,
  ntf=2,
  cut=10.0,
  ntb=1, ntp=0,
  ntr=0,
  ntt=3, ig=-1, gamma_ln=1.0,
  ioutfm=1,
  tempi=0.0,temp0=300.0,
  ntpr=500, ntwx=500, ntwr=500,
 /
```

Second stage, Equilibration

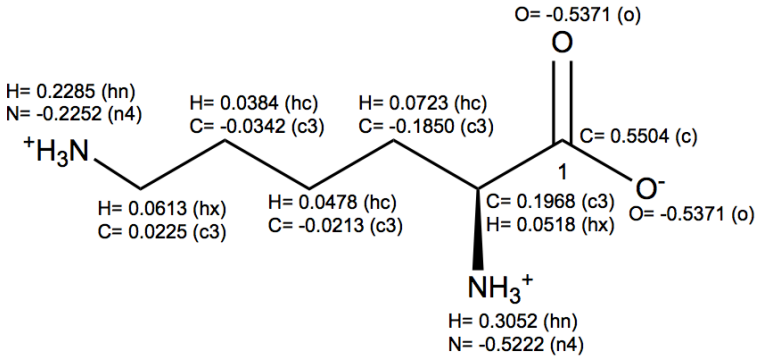
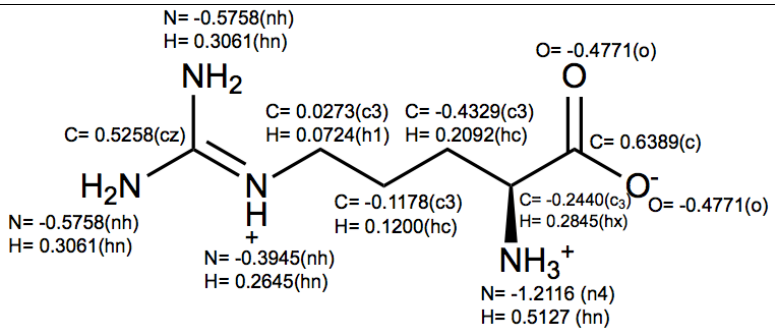
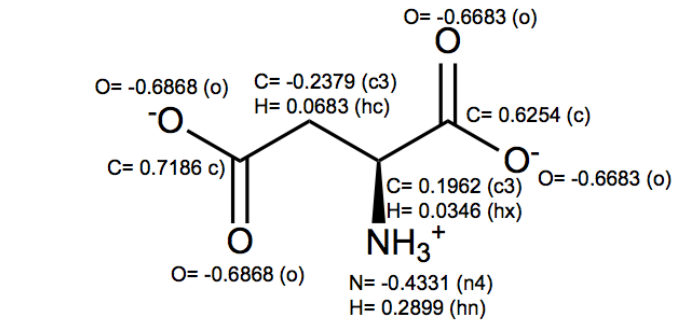
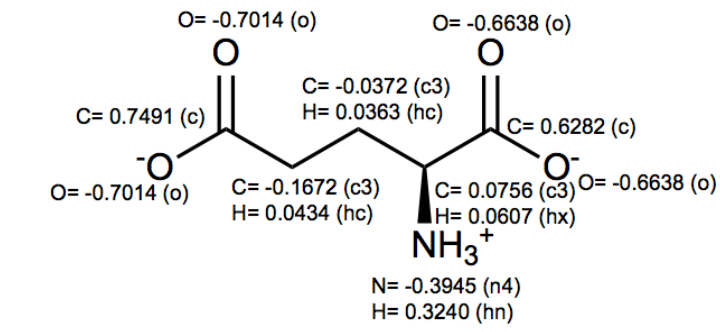
```
Equilibration at constt. pressure
  &cntrl
   imin=0,
   ntpr=1000, ntwx=1000, ntwr=1000,
   ntx=5,
   pres0=1.0,
   irest=1,
```

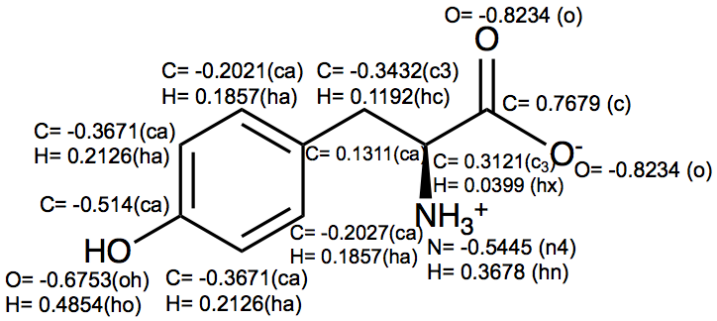
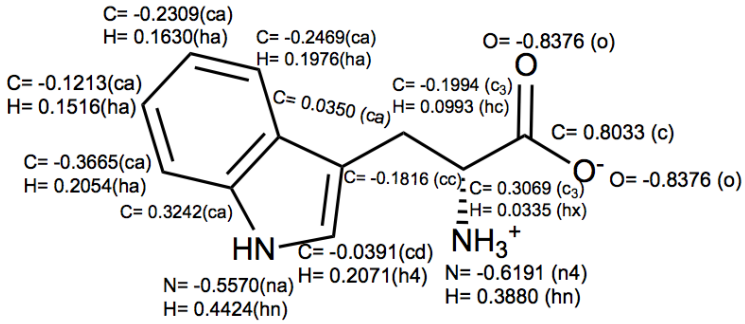
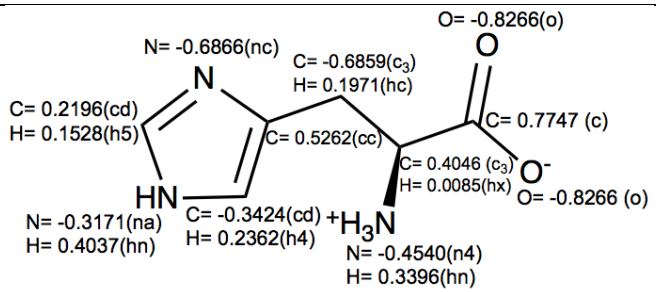
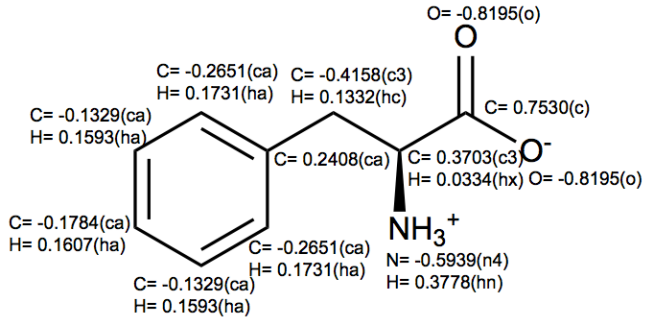
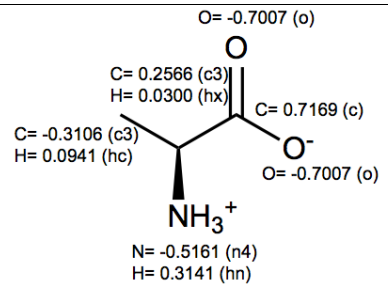
```
   cut=10.0,
   taup=2.0,
   tempi=300.0, temp0=300.0,
   ntt=3, gamma_ln=1.0,
   ntb=2,
   ntp=1,
   ntc=2,
   ntf=2,
   ioutfm=1,
   nstlim=500000, dt=.002, ig=-1,
  /
```

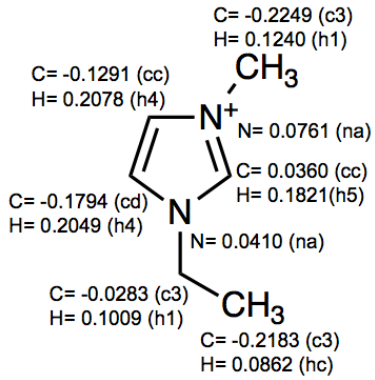
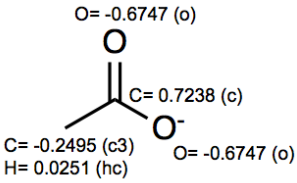
Appendix A.6. 6: Input file for third MD simulation production run

```
Production for box
  &cntrl
   imin=0, ntpr=1000, ntwx=1000,
   ntx=5, irest=1,
   tempi=300.0, temp0=300.0, ntt=3,
   gamma_ln=5.0, ntb=2, ntp=1, taup=1.0, ioutfm=1, nstlim=50000000,
   ntwr=1000, dt=.002, ig=-1,
   ntf=2, ntc=2, cut=10.0
/
```

Appendix A. 7: Chem draw structures for AAs and IL ions with scaled charges

AA or IL ion	Scaled atomic charges
LYS	
ARG	
ASP	
GLU	

TYR	
TRP	
HIS	
PHE	
ALA	

$[C_2C_1Im]^+$ EMIM	
$[CH_3CO_2]^-$ OAc^- ACT	

The scaled atomic charges are shown for each molecule in the table above, along with the atom types generated by antechamber in brackets. All Lennard-Jones parameters and force constants are taken from GAFF forcefield and are not manually derived. Based on the atom types the LJ parameters from GAFF forcefield are given below,

Atom type	distance $r_0=\sqrt[6]{2}\sigma$ Å	Well depth ε kcal/mol	Reference
h1	1.387	0.0157	Veenstra et al JCC,8,(1992),963
h2	1.287	0.0157	Veenstra et al JCC,8,(1992),963
h3	1.187	0.0157	Veenstra et al JCC,8,(1992),963
h4	1.409	0.015	Spellmeyer, one electrowithdr. neighbor
h5	1.359	0.015	Spellmeyer, two electrowithdr. neighbor
ha	1.459	0.015	Spellmeyer
hc	1.487	0.0157	OPLS
hn	0.6	0.0157	!Ferguson base pair geom.
ho	0	0	OPLS Jorgensen, JACS,110,(1988),1657
hx	1.1	0.0157	Veenstra et al JCC,8,(1992),963
o	1.6612	0.21	OPLS
oh	1.721	0.2104	OPLS
c	1.908	0.086	OPLS
c1	1.908	0.21	cp C DLM 11/2007 well depth from OPLS replacing 0.0860
c2	1.908	0.086	sp2 atom in the middle of C=CD-CD=C
c3	1.908	0.1094	OPLS
ca	1.908	0.086	OPLS
cc	1.908	0.086	OPLS
cd	1.908	0.086	OPLS
cz	1.908	0.086	OPLS
n	1.824	0.17	OPLS
n4	1.824	0.17	OPLS
na	1.824	0.17	OPLS
nh	1.824	0.17	OPLS
cl	1.948	0.265	Fox, JPCB,102,8070,(98),flex.mdl CHCl3

These values are then combined to provide parameters for all pairs of atom types in the system, at the time a *.prmtop file is written. The image of .prmtop (shortened form) for high concentration system of lysine/[C_2C_1Im][Cl] at 300K file is shown below.

```
%FLAG LENNARD_JONES_ACOEF
%FORMAT(5E16.8)
  1.04308023E+06  9.95480466E+05  9.44293233E+05  9.24822270E+05  8.82619071E+05
  8.19971662E+05  5.98885646E+04  5.46147253E+04  5.30987710E+04  2.43828624E+03
  9.71708117E+04  8.96776989E+04  8.61541883E+04  4.33325458E+03  7.51607703E+03
  6.78771368E+04  6.20665997E+04  6.01816484E+04  2.82099197E+03  4.98586848E+03
  3.25969625E+03  7.18621074E+04  6.58473870E+04  6.37148278E+04  3.03448006E+03
  5.34045360E+03  3.50301067E+03  3.76169105E+03  3.52723645E+06  3.52778674E+06
  3.12734028E+06  2.86107279E+05  4.24433254E+05  3.17807134E+05  3.31255111E+05
  8.25305789E+06  2.56678134E+03  2.12601181E+03  2.27577561E+03  4.90239937E+01
  1.07193646E+02  5.94667300E+01  6.63368273E+01  2.45929350E+04  1.39982777E-01
  6.47841731E+05  6.06829342E+05  5.74393458E+05  3.23283631E+04  5.44261042E+04
  3.69471530E+04  3.93690817E+04  2.52848425E+06  1.02595236E+03  3.79876399E+05
  2.27401052E+04  2.01791425E+04  2.01619733E+04  7.50063100E+02  1.41077189E+03
  8.79040886E+02  9.55000044E+02  1.32974259E+05  9.14716912E+00  1.12780457E+04
  2.01823541E+02
%FLAG LENNARD_JONES_BCOEF
%FORMAT(5E16.8)
  6.75612247E+02  7.36907417E+02  8.01323529E+02  5.99015525E+02  6.53361429E+02
  5.31102864E+02  9.85097219E+01  1.05031585E+02  8.73413012E+01  1.20953369E+01
  1.26919150E+02  1.36131731E+02  1.12529845E+02  1.63092814E+01  2.17257828E+01
  1.06076943E+02  1.13252061E+02  9.40505980E+01  1.31591746E+01  1.76949863E+01
  1.43076527E+01  1.07908863E+02  1.15327881E+02  9.56748258E+01  1.34932874E+01
  1.81057616E+01  1.46638650E+01  1.50233639E+01  7.26401607E+02  8.11089111E+02
  6.44046702E+02  1.25890352E+02  1.55090356E+02  1.34202904E+02  1.35459444E+02
  6.49662513E+02  2.06278363E+01  2.09604198E+01  1.82891803E+01  1.73473071E+00
  2.59456373E+00  1.93248820E+00  2.01792524E+00  3.73323452E+01  9.37598976E-02
  6.26720080E+02  6.77220874E+02  5.55666448E+02  8.51921330E+01  1.11805549E+02
  9.21192136E+01  9.40124296E+01  7.23919256E+02  1.53505284E+01  5.64885984E+02
  6.13981767E+01  6.45756063E+01  5.44372326E+01  6.78541827E+00  9.41257003E+00
  7.42992380E+00  7.65648470E+00  8.68088256E+01  7.57919667E-01  5.08951803E+01
  3.56012899E+00
```

These atom pair parameters are used to calculate the ACOEFF and BCOEFF values in the .prmtop file (as mentioned in the image above)

$$ACOEFF = \epsilon r_0^{12}; BCOEFF = 2\epsilon r_0^6$$

The A and B coefficients are then used in sander's internal version of the VDW equation,

$$E_{VDW} = \frac{ACOEFF}{(r/\mathring{A})^{12}} - \frac{BCOEFF}{(r/\mathring{A})^6}$$

Appendix A. 8: Graphs from heating and equilibration MD stages for density, temperature and energy for all Amino acid simulations

Appendix A.8. 1: Density graphs from equilibration MD stage for all Amino acid-IL systems and IL/water systems

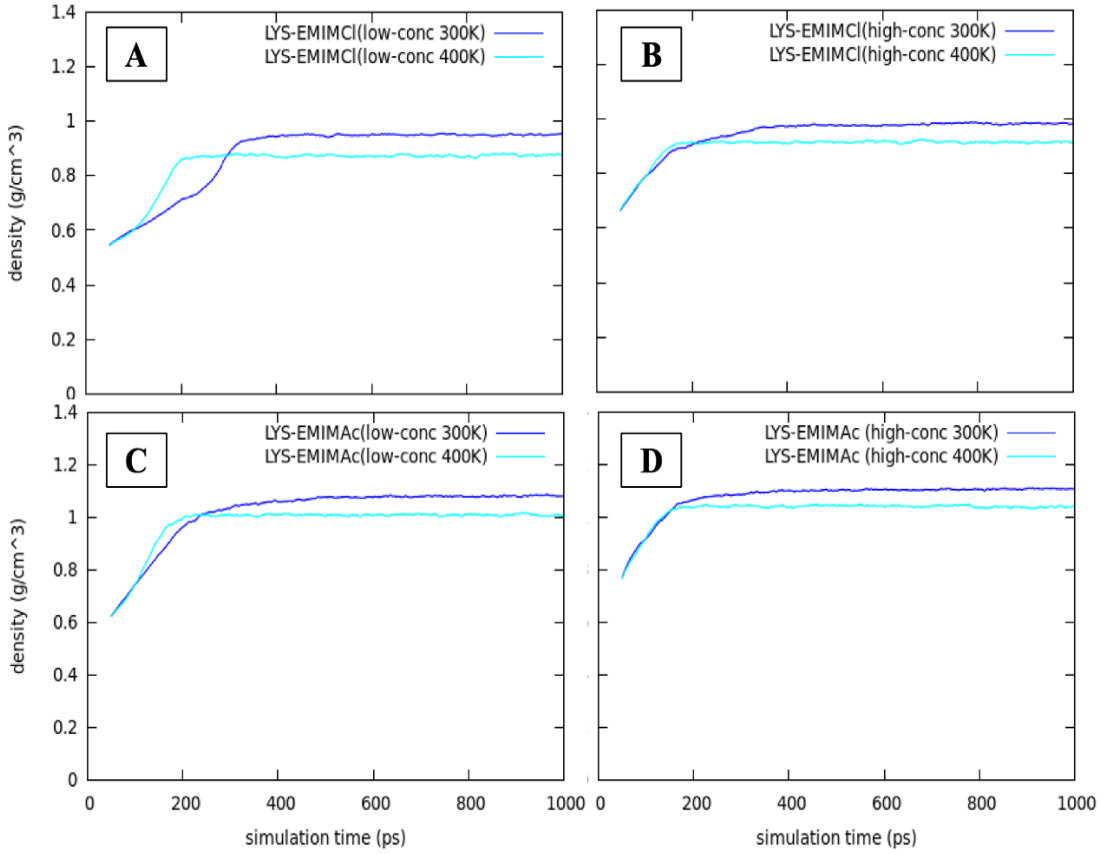


Figure A.8. 1: Density equilibration graphs for lysine systems. (A) low concentration Lys/[C_2C_1Im][Cl], (B) high concentration Lys/[C_2C_1Im][Cl], (C) low concentration Lys/[C_2C_1Im][CH_3CO_2] and (D) high concentration Lys/[C_2C_1Im][CH_3CO_2]. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

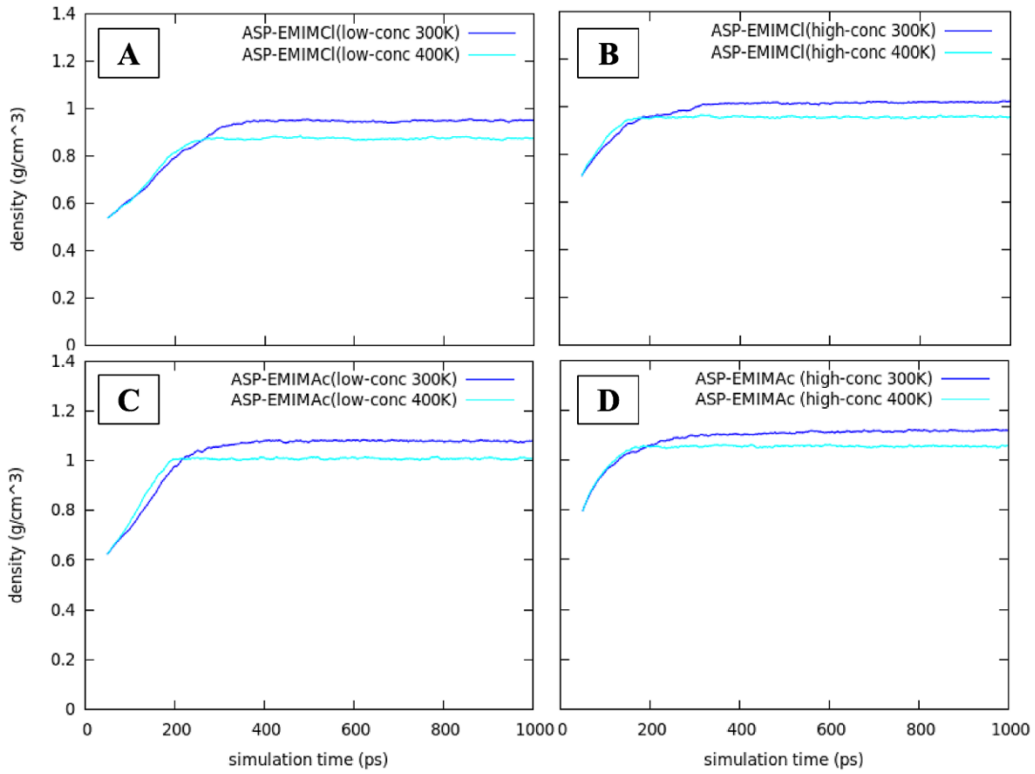


Figure A.8. 2: Density equilibration graphs for aspartate systems. (A) low concentration Asp/[C_2C_1Im][Cl], (B) high concentration Asp/[C_2C_1Im][Cl], (C) low concentration Asp/[C_2C_1Im][CH_3CO_2] and (D) high concentration Asp/[C_2C_1Im][CH_3CO_2]. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

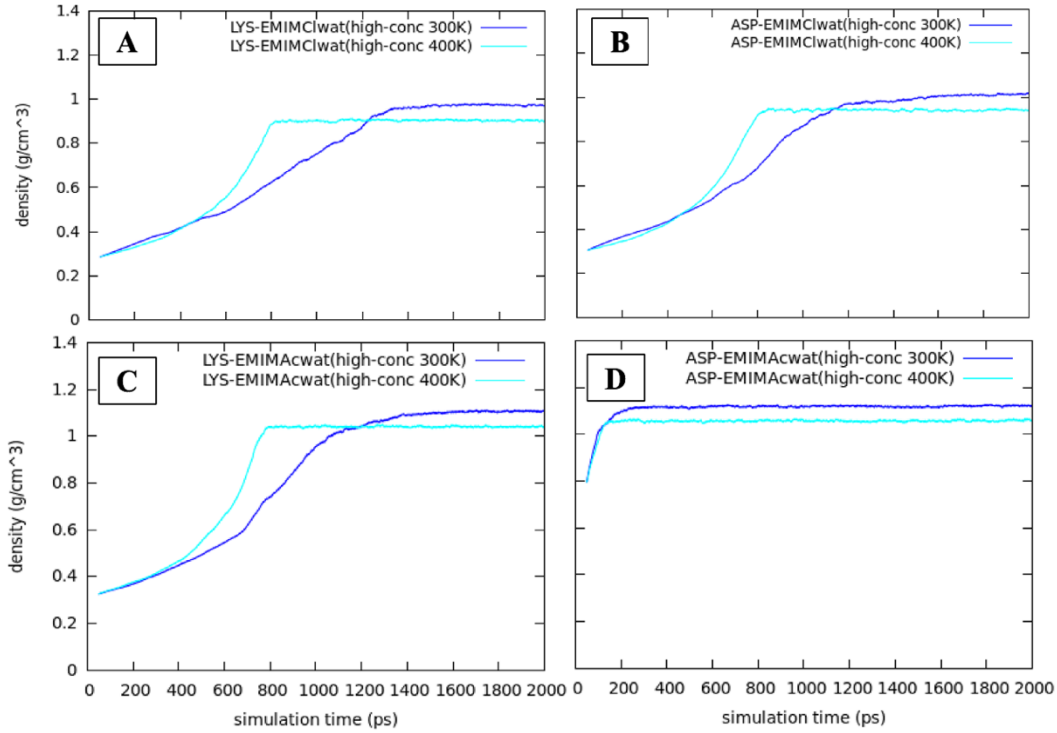


Figure A.8. 3: Density equilibration graphs for IL-wat/AA systems. (A) high concentration Lys/[C_2C_1Im][Cl]-H_2O, (B) high concentration Asp/[C_2C_1Im][Cl]-H_2O, (C) high concentration Lys/[C_2C_1Im][CH3CO2]-H_2O and (D) high concentration Asp/[C_2C_1Im][CH_3CO_2]-H_2O. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

Appendix A.8. 2: Box dimensions for lysine and aspartate systems in pure IL and IL/water mixtures of $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ ILs

System	Initial box size Å^3	Final box sizes Å^3			
		T=300K		T=400K	
		200:1	5:1		5:1
lysine $[C_2C_1Im][Cl]$	91125	51393	61842	56115	66642
lysine $[C_2C_1Im][Cl]$- wat			63241		68565
aspartate $[C_2C_1Im][Cl]$		52048	63579	56688	68820
aspartate $[C_2C_1Im][Cl]$-wat			64388		70351
lysine $[C_2C_1Im][CH_3CO_2]$		52200	63438	56415	67072
lysine $[C_2C_1Im][CH_3CO_2]$- wat			63574		67475
aspartate $[C_2C_1Im][CH_3CO_2]$		52654	64884	56188	68594
aspartate $[C_2C_1Im][CH_3CO_2]$-wat			64845		68545

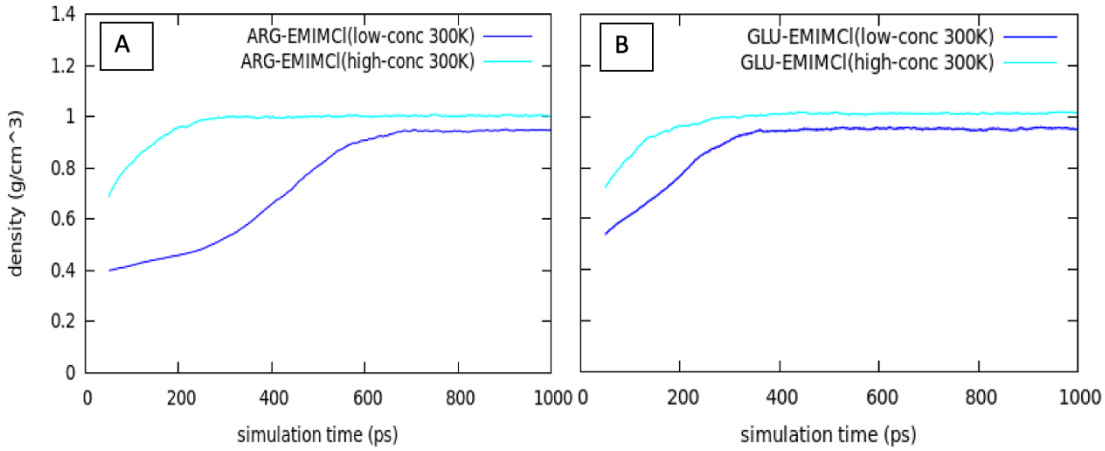


Figure A.8. 4: Density equilibration graphs for (A) arginine (Arg/$[C_2C_1Im][Cl]$) systems and (B) glutamate (Glu/$[C_2C_1Im][Cl]$) systems at 300K. Dark blue line shows a low concentration system and light blue line shows high concentration system.

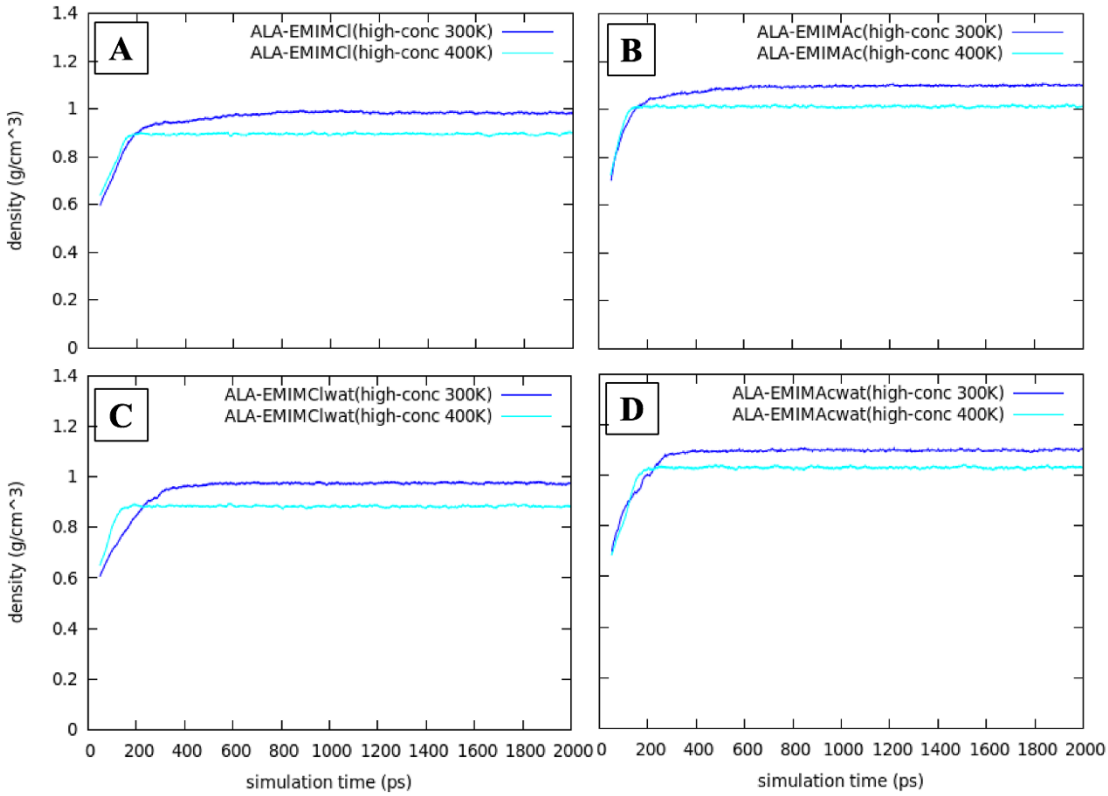


Figure A.8. 5: Density equilibration graphs for alanine systems. (A) high concentration Ala/$[C_2C_1Im][Cl]$, (B) high concentration Ala/$[C_2C_1Im][CH_3CO_2]$ (C) high concentration Ala/$[C_2C_1Im][Cl]$-H_2O and (D) high concentration Lys/$[C_2C_1Im][CH_3CO_2]$-H_2O. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

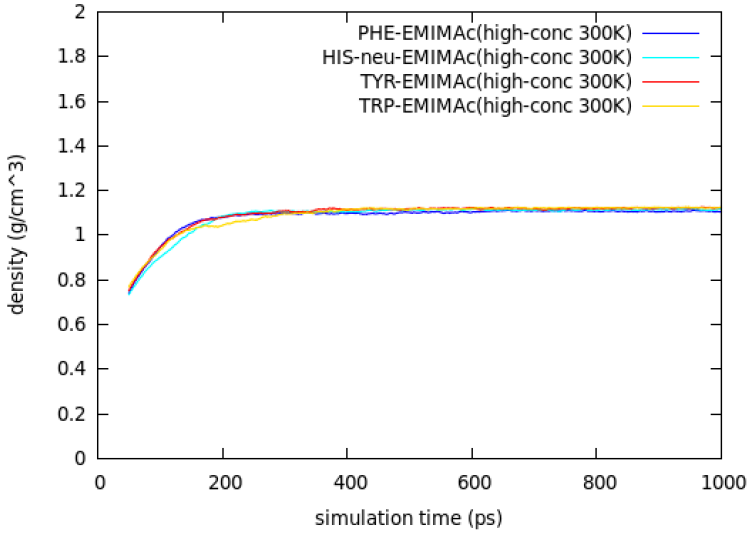


Figure A.8. 6: Density equilibration graphs for aromatic high concentration systems at 300K. Phe/$[C_2C_1Im][CH_3CO_2]$ (dark blue), His/$[C_2C_1Im][CH_3CO_2]$ (light blue), Tyr/$[C_2C_1Im][CH_3CO_2]$ (red) and Trp/$[C_2C_1Im][CH_3CO_2]$ (yellow)

Appendix A.8. 3: Density graphs from equilibration MD stage for all Peptides (LNGKTLKG, ALA-5 and ALA-10) in $[C_2C_1Im][CH_3CO_2]$ IL and water systems

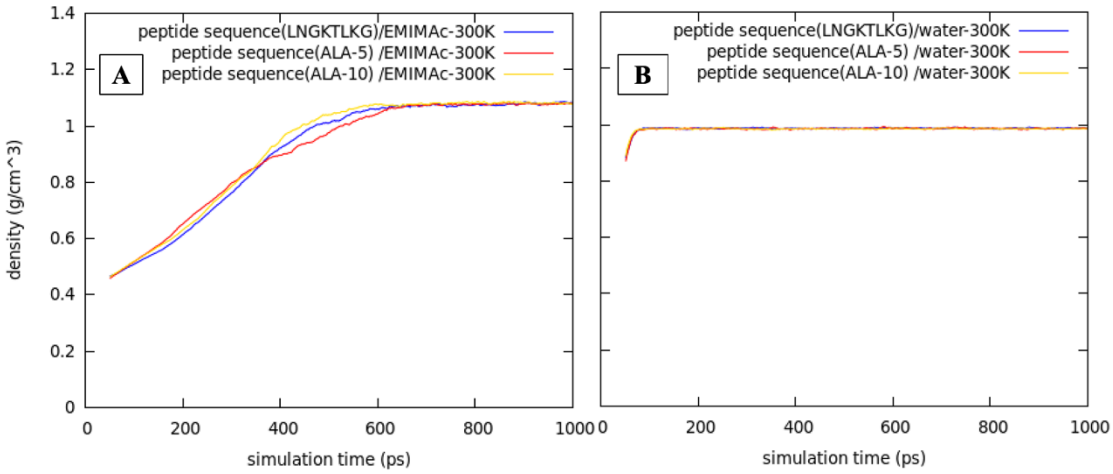


Figure A.8. 7: Density equilibration graphs for peptide systems. (A) Peptides in $[C_2C_1Im][CH_3CO_2]$ IL at 300K (B) Peptides in water at 300K. Dark blue line is for LNGKTLKG peptide, red line is for AAAAA peptide and yellow line is for AAAAA-AAAAA peptide

Appendix A.8. 4: Temperature graphs from heating and equilibration stages of MD simulation for all Amino acid- IL and IL/water systems

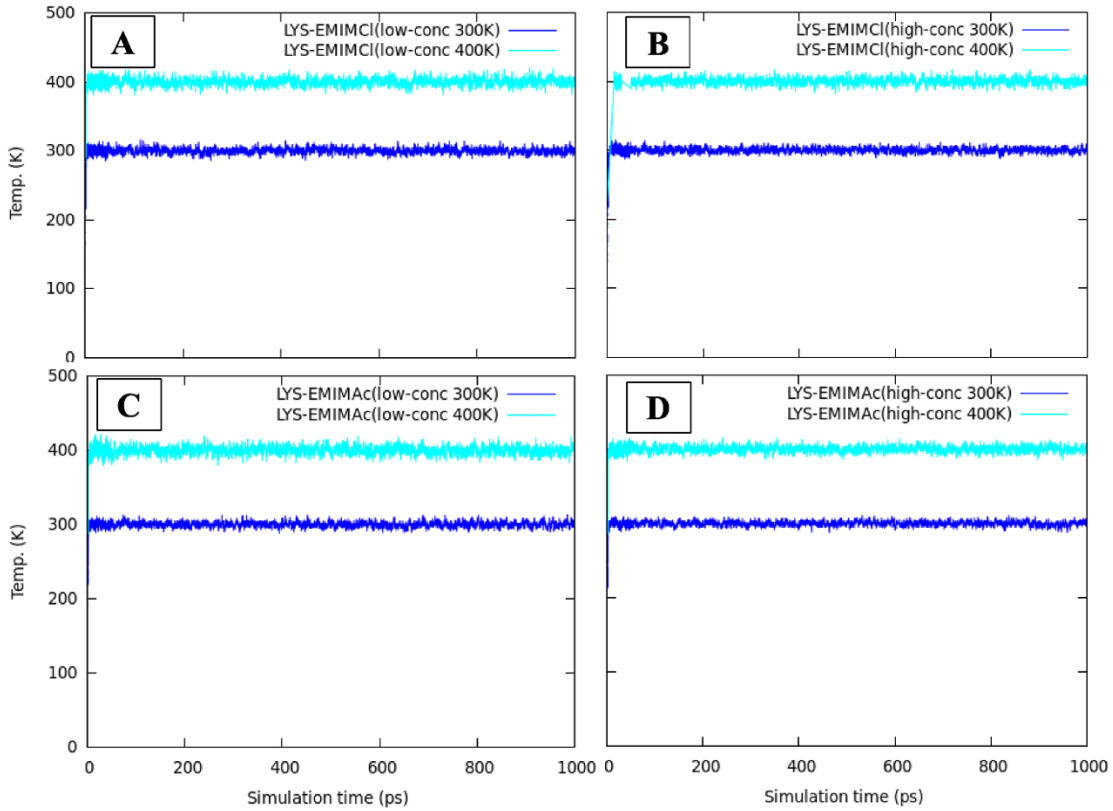


Figure A.8. 8: Temperature graphs from heating and equilibration stages for lysine systems. (A) low concentration Lys/$[C_2C_1Im][Cl]$, (B) high concentration Lys/$[C_2C_1Im][Cl]$, (C) low concentration Lys/$[C_2C_1Im][CH_3CO_2]$ and (D) high concentration Lys/$[C_2C_1Im][CH_3CO_2]$. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

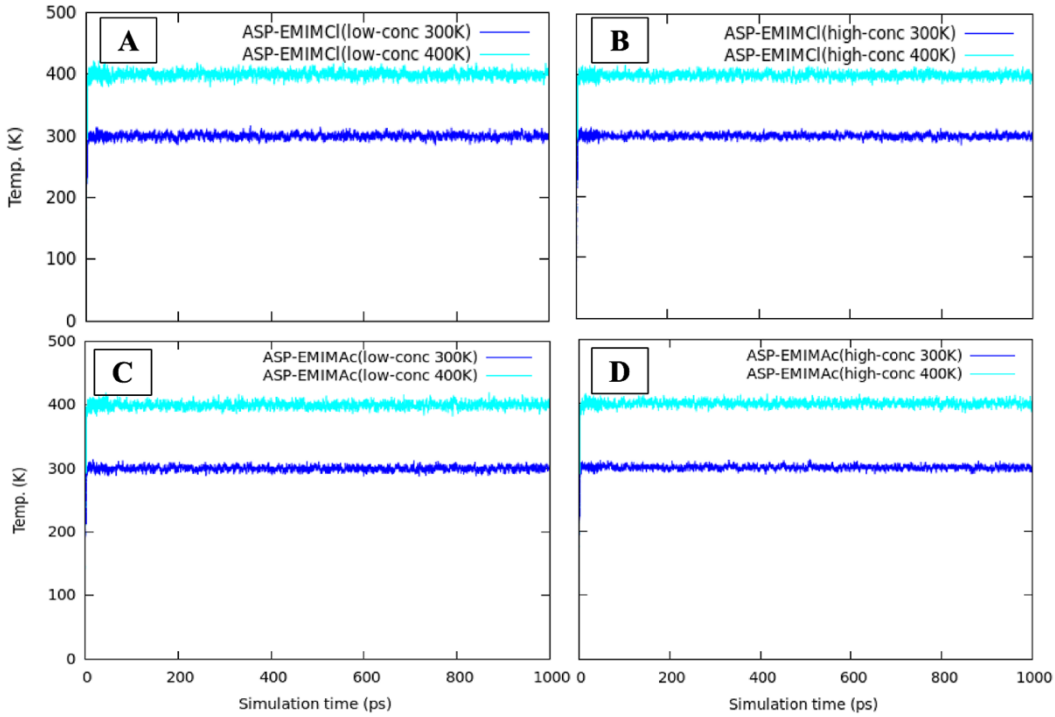


Figure A.8. 9: Temperature graphs from heating and equilibration stages for aspartate systems. (A) low concentration Asp/[C_2C_1Im][Cl], (B) high concentration Asp/[C_2C_1Im][Cl], (C) low concentration Asp/[C_2C_1Im][CH_3CO_2] and (D) high concentration Asp/[C_2C_1Im][CH_3CO_2]. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

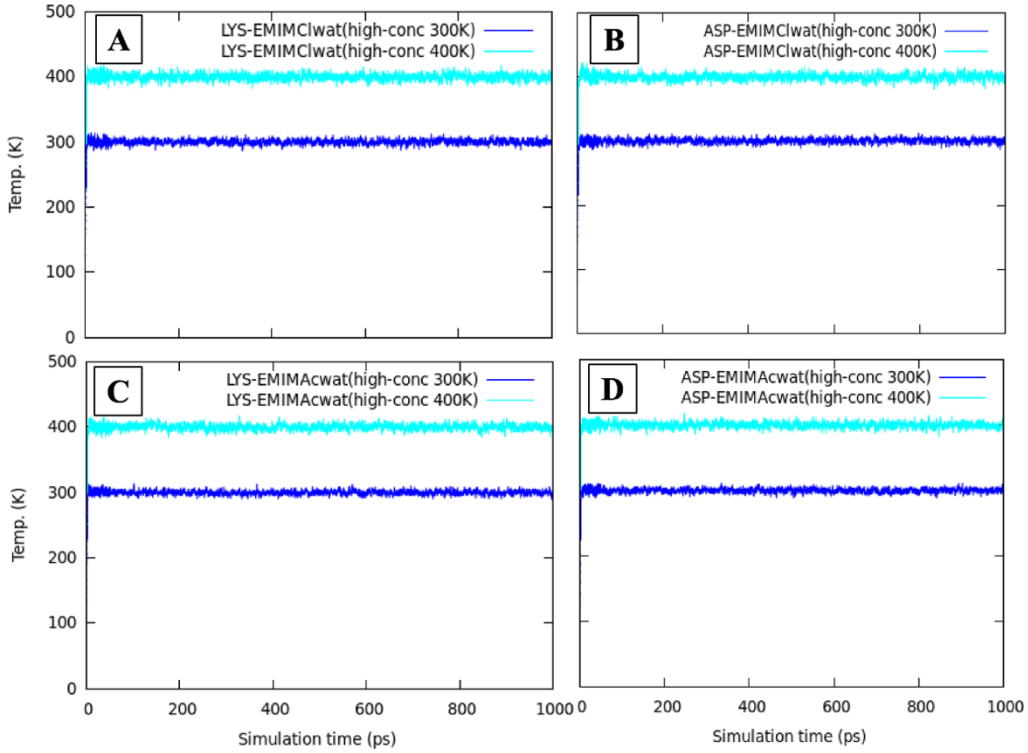


Figure A.8. 10: Temperature graphs from heating and equilibration stages for IL-wat/AA systems. (A) high concentration Lys/[C_2C_1Im][Cl]-H_2O, (B) high concentration Asp/[C_2C_1Im][Cl]-H_2O, (C) high concentration Lys/[C_2C_1Im][CH_3CO_2]-H_2O and (D) high concentration Asp/[C_2C_1Im][CH_3CO_2]-H_2O. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

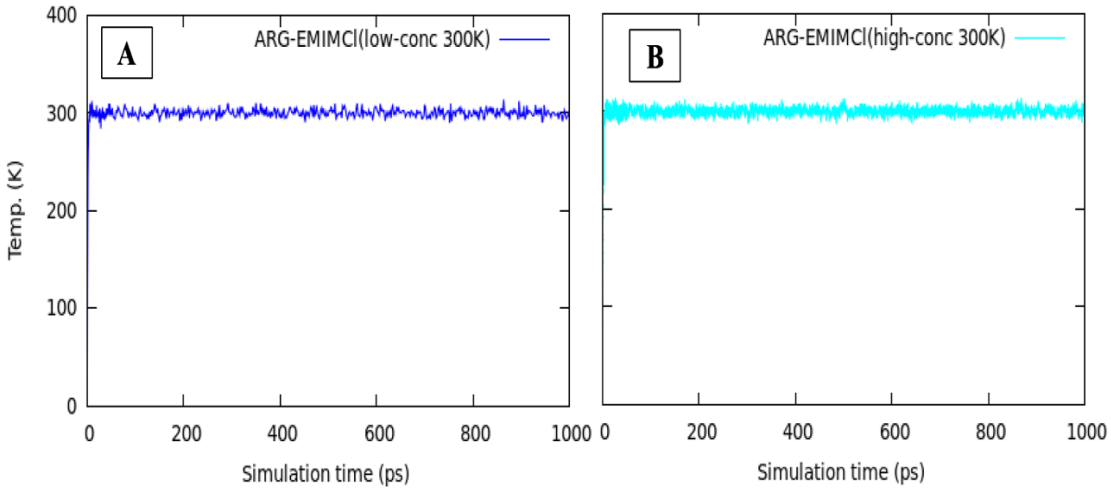


Figure A.8. 11: Temperature graphs from heating and equilibration stages for arginine (Arg/[C_2C_1Im][Cl]) systems at 300K. (A) low concentration system is shown by a dark blue line (B) high concentration system is shown by light blue line.

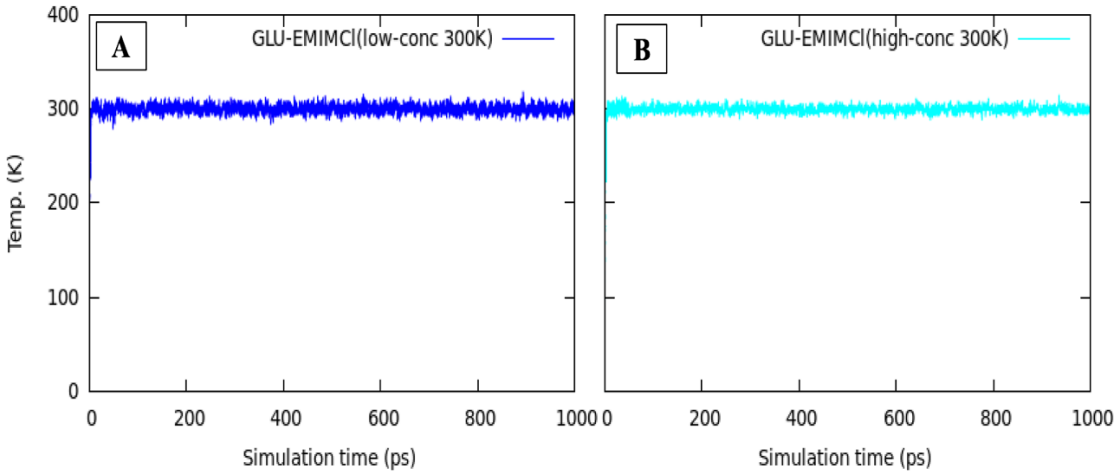


Figure A.8. 12: Temperature graphs from heating and equilibration stages for glutamate (Glu/[C_2C_1Im][Cl]) systems at 300K. (A) low concentration system is shown by a dark blue line (B) high concentration system is shown by light blue line.

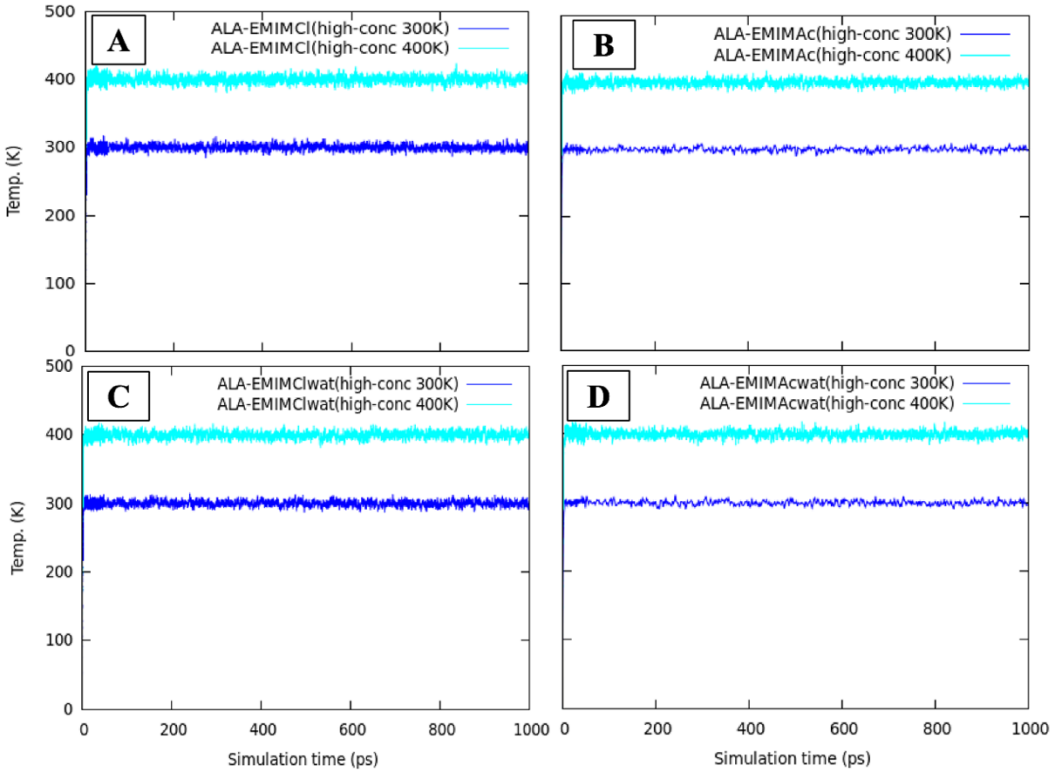


Figure A.8. 13: Temperature graphs from heating and equilibration stages for alanine (A) high concentration Ala/[C_2C_1Im][Cl], (B) high concentration Ala/[C_2C_1Im][CH_3CO_2] (C) high concentration Ala/[C_2C_1Im][Cl]-H_2O and (D) high concentration Lys/[C_2C_1Im][CH_3CO_2]-H_2O. Dark blue line in each case shows the systems at 300K and light blue line shows the systems at 400K.

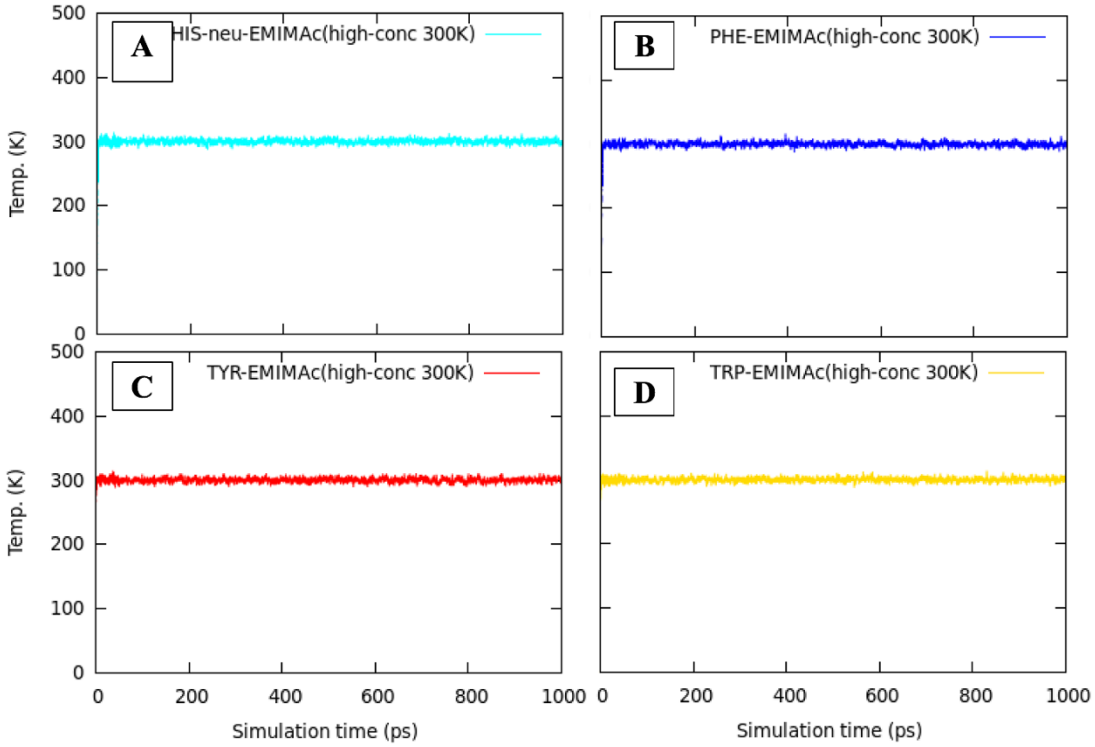


Figure A.8. 14: Temperature graphs from heating and equilibration stages for aromatic high concentration systems at 300K. (A) His/[C_2C_1Im][CH_3CO_2] (light blue), (B) Phe/[C_2C_1Im][CH_3CO_2] (dark blue), (C) Tyr/[C_2C_1Im][CH_3CO_2] (red) and (D) Trp/[C_2C_1Im][CH_3CO_2] (yellow)

Appendix A.8. 5: Temperature graphs from heating and equilibration stages of MD simulation for (LNGKTLKG, ALA-5 and ALA-10) in [C_2C_1Im][CH_3CO_2] IL and water systems

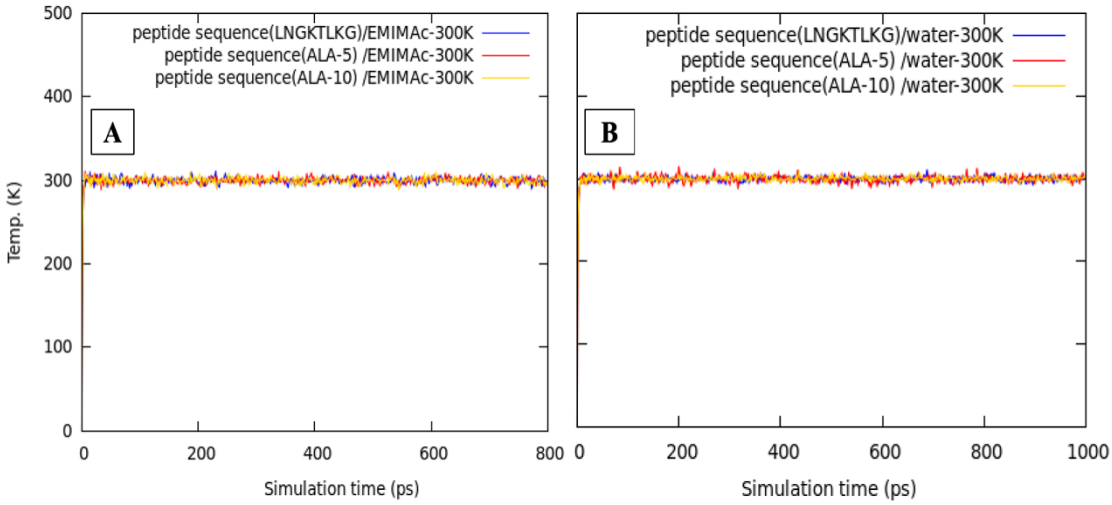


Figure A.8. 15: Temperature equilibration graphs for peptide systems. (A) Peptides in [C_2C_1Im][CH_3CO_2] IL at 300K (B) Peptides in water at 300K. Dark blue line is for LNGKTLKG peptide, red line is for AAAAA peptide and yellow line is for AAAAA-AAAAA peptide

Appendix A.8. 6: Energy graphs from heating and equilibration stages of MD simulations for all AA-IL and IL/water systems

Note: Coloured line in each graph shows, except for aromatics, dark blue line = total kinetic energy (EKTOT) for the systems at 300K, red line = total kinetic energy (EKTOT) for the systems at 400K, yellow line =total potential energy (EPTOT) for the systems at 300K, green line = total potential energy (EPTOT) for the systems at 400K, black line for total energy (ETOT) for the systems at 300K and light blue line = total energy(ETOT) for the systems at 400K.

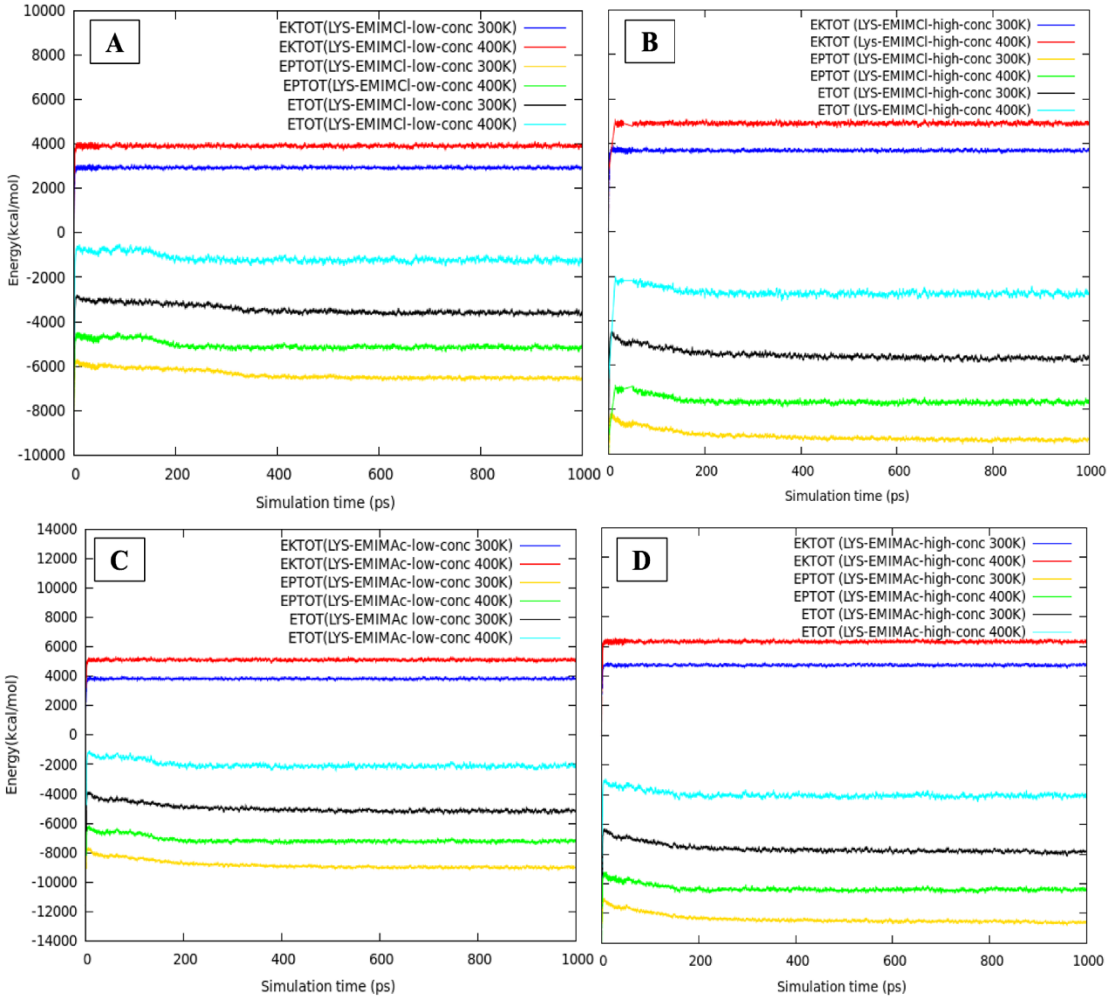


Figure A.8. 16: Energy graphs from heating and equilibration stages of MD simulations for lysine systems (A) low concentration Lys/[C_2C_1Im][Cl], (B) high concentration Lys/[C_2C_1Im][Cl], (C) low concentration Lys/[C_2C_1Im][CH_3CO_2] and (D) high concentration Lys/[C_2C_1Im][CH_3CO_2].

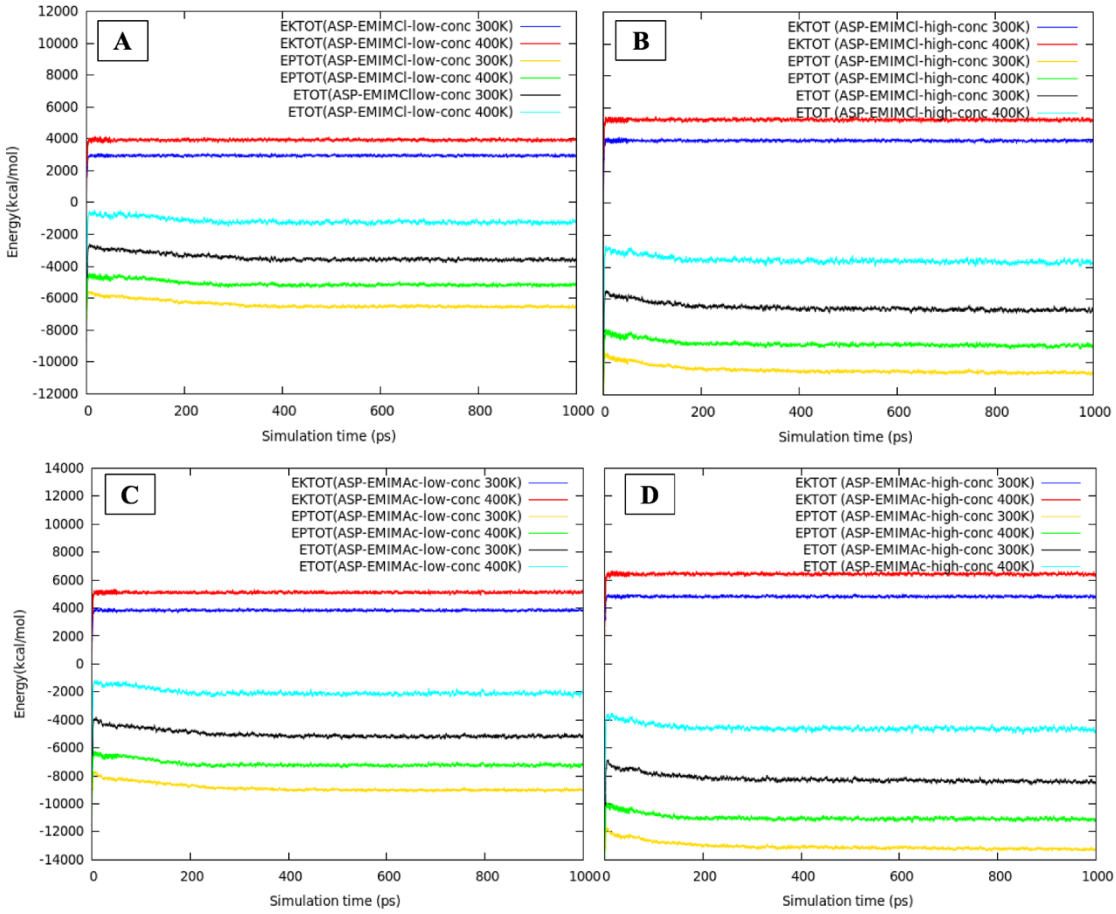


Figure A.8. 17: Energy graphs from heating and equilibration stages of MD simulations for aspartate systems. (A) low concentration Asp/[C_2C_1Im][Cl], (B) high concentration Asp/[C_2C_1Im][Cl], (C) low concentration Asp/[C_2C_1Im][CH_3CO_2] and (D) high concentration Asp/[C_2C_1Im][CH_3CO_2].

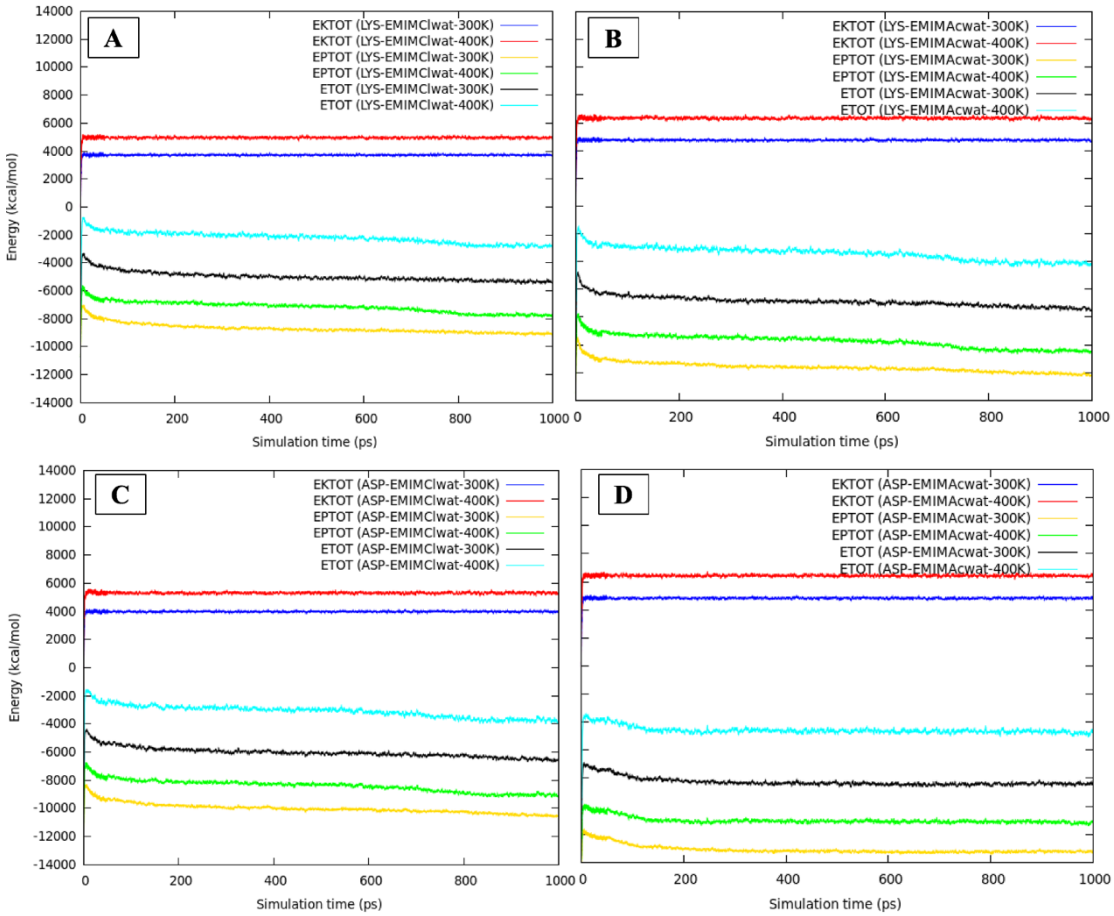


Figure A.8. 18: Energy graphs from heating and equilibration stages of MD simulations for IL-wat/AA systems (A) high concentration Lys/[C_2C_1Im][Cl]-H_2O, (B) high concentration Lys/[C_2C_1Im][CH_3CO_2]-H_2O, (C) high concentration Asp/[C_2C_1Im][Cl]-H_2O and (D) high concentration Asp/[C_2C_1Im][CH_3CO_2]-H_2O.

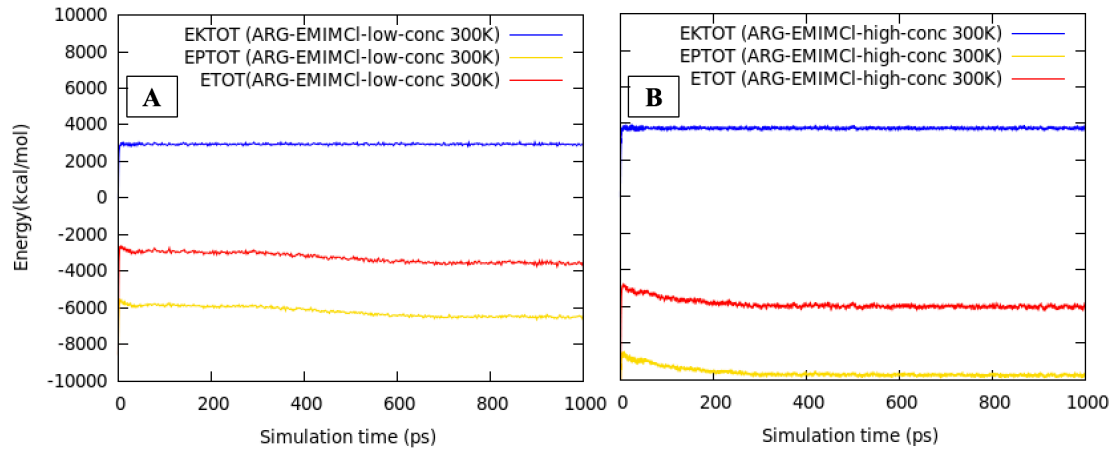


Figure A.8. 19: Energy graphs from heating and equilibration stages of MD simulations for arginine (Arg/[C_2C_1Im][Cl]) systems at 300K. (A) low concentration and (B) high concentration

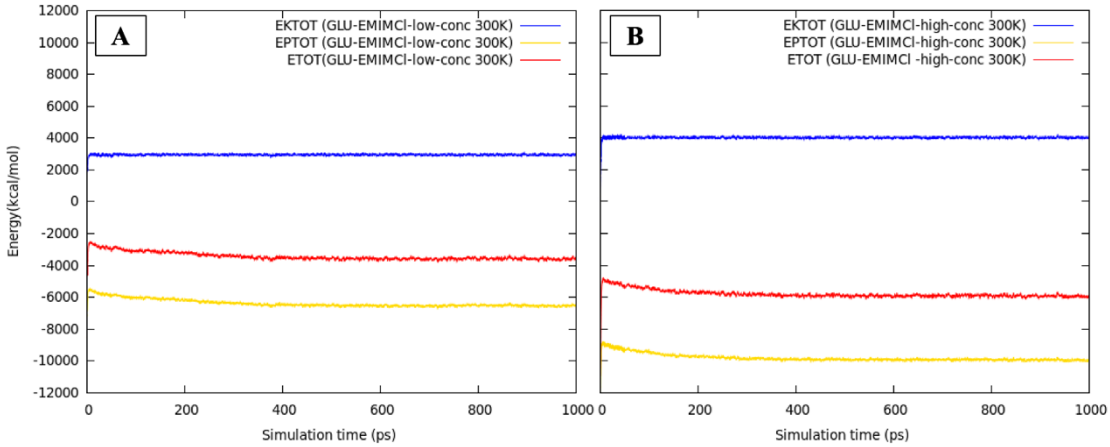


Figure A.8. 20: Energy graphs from heating and equilibration stages of MD simulations for glutamate (Glu/$[C_2C_1Im][Cl]$) systems at 300K. (A) low concentration and (B) high concentration

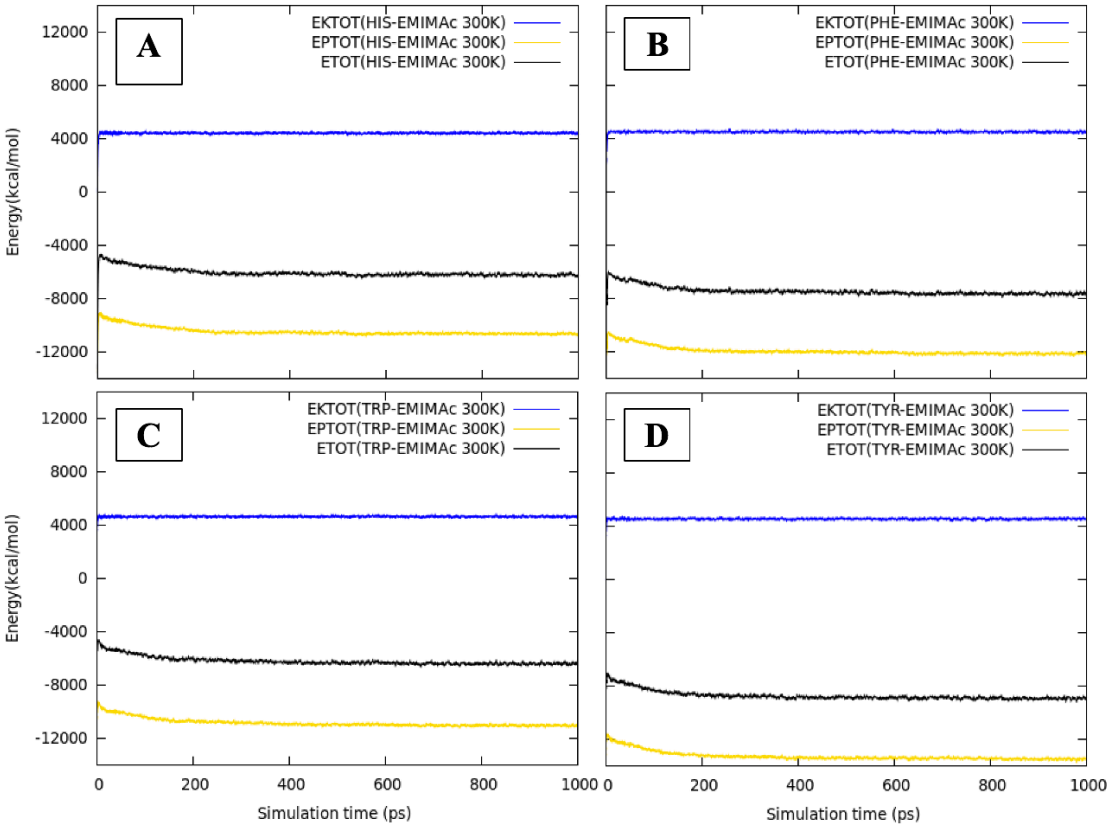


Figure A.8. 21: Energy graphs from heating and equilibration stages of MD simulations for aromatic high concentration systems at 300K. (A) His/$[C_2C_1Im][CH_3CO_2]$, (B) Phe/$[C_2C_1Im][CH_3CO_2]$, (C) Trp/$[C_2C_1Im][CH_3CO_2]$ and (D) Tyr/$[C_2C_1Im][CH_3CO_2]$. Coloured line in each graph shows, dark blue line = total kinetic energy (EKTOT) for the system, yellow line =total potential energy (EPTOT) for the systems, black line for total energy (ETOT) for the systems.

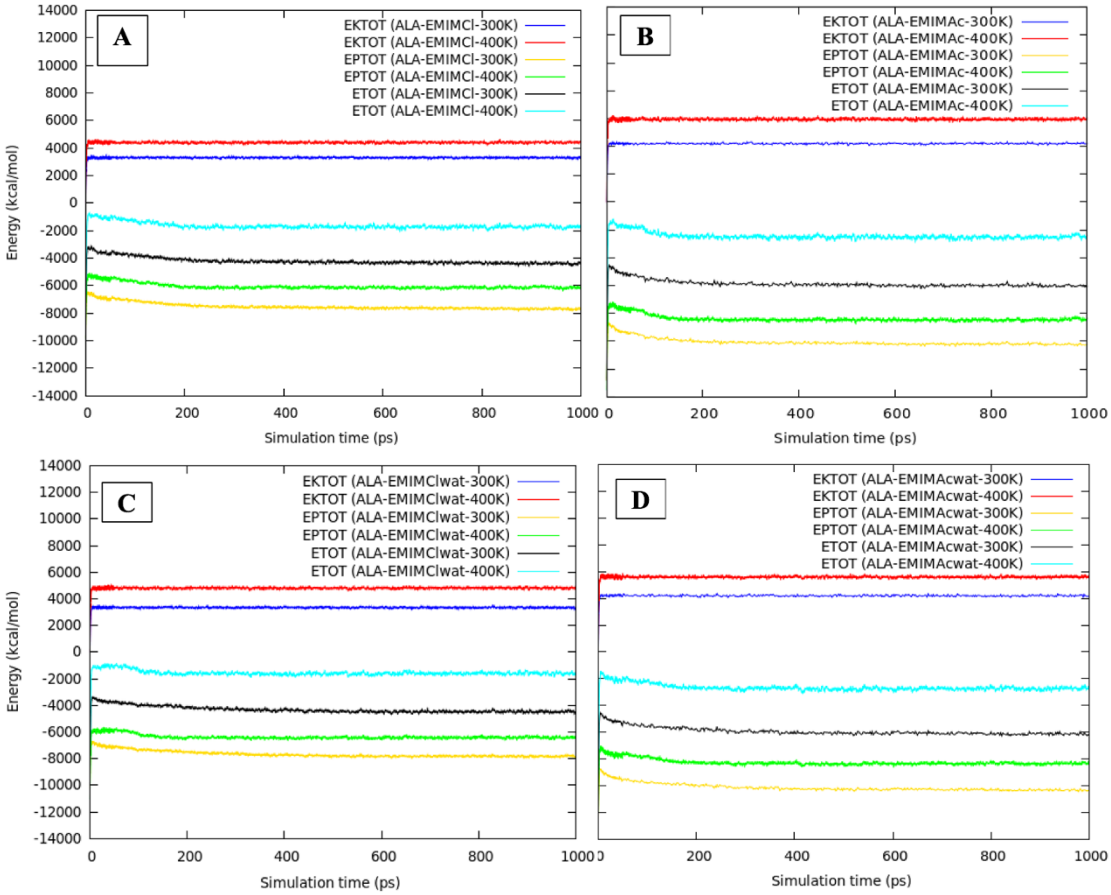


Figure A.8. 22: Energy graphs from heating and equilibration stages of MD simulations for alanine systems (A) high concentration Ala/[C_2C_1Im][Cl], (B) high concentration Ala/[C_2C_1Im][CH_3CO_2], (C) high concentration Ala/[C_2C_1Im][Cl]-H_2O and (D) high concentration Lys/[C_2C_1Im][CH_3CO_2]-H_2O.

Appendix A.8. 7: Energy graphs from heating and equilibration stages of MD simulations for all peptides (LNGKTLKG, ALA-5 and ALA-10) in $[C_2C_1Im][CH_3CO_2]$ IL and water systems

Note: Coloured line in each graph shows, dark blue line = total kinetic energy (EKTOT) for the systems in $[C_2C_1Im][CH_3CO_2]$ IL, red line = total kinetic energy (EKTOT) for the systems in water, yellow line =total potential energy(EPTOT) for the systems in $[C_2C_1Im][CH_3CO_2]$ IL, green line = total potential energy(EPTOT) for the systems in water, black line for total energy(ETOT) for the systems in $[C_2C_1Im][CH_3CO_2]$ IL and light blue line = total energy (ETOT) for the systems in water.

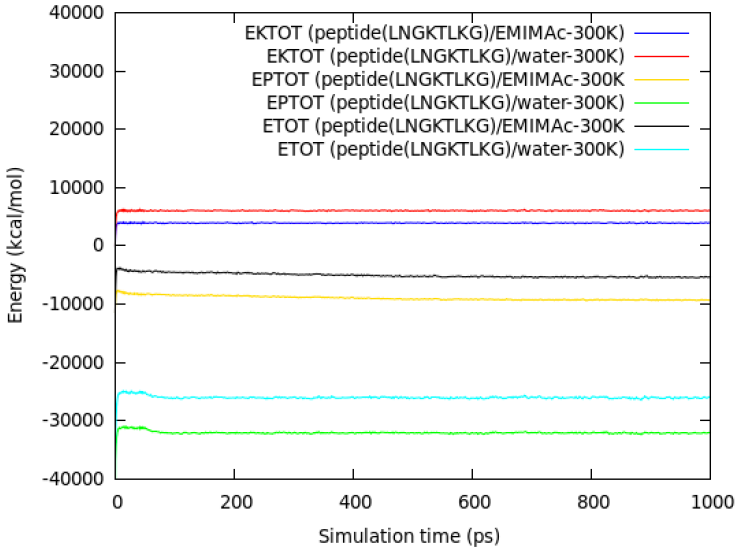


Figure A.8. 23: Energy equilibration graphs for peptide LNGKTLKG systems in water and $[C_2C_1Im][CH_3CO_2]$ IL at 300K.

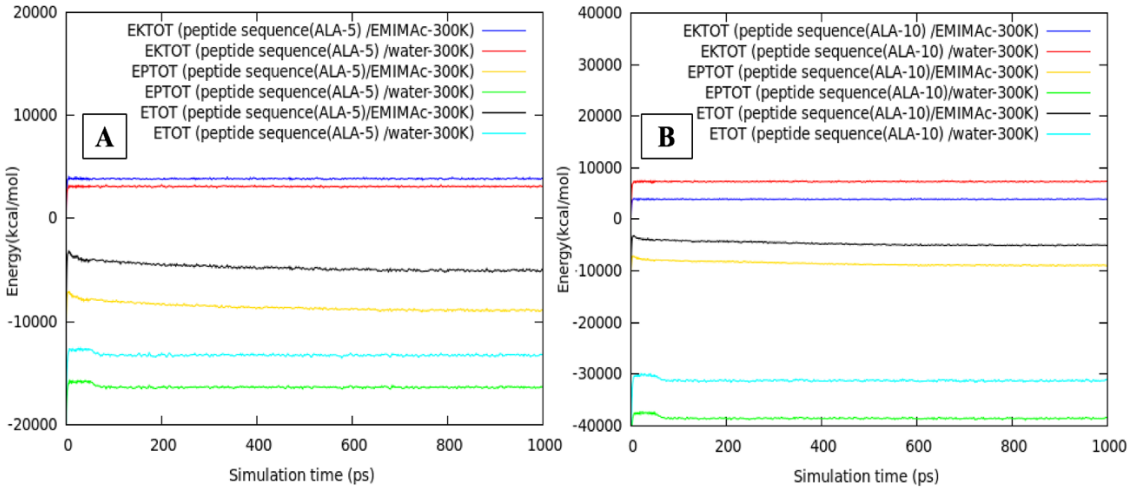


Figure A.8. 24: Energy equilibration graphs for peptide systems in water and $[C_2C_1Im][CH_3CO_2]$ IL at 300K. (A) AAAAA peptide in water and IL (B) AAAAA-AAAAA in water and IL

Appendix A. 9: Cation-Anion interactions for positively and negatively charged amino acids in pure ILs

Appendix A.9. 1: Site-site radial distribution function of $[C_2C_1Im]$ cation and chloride anion around LYS, ARG, ASP and GLU at 300K and for LYS and ASP at 400K low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$

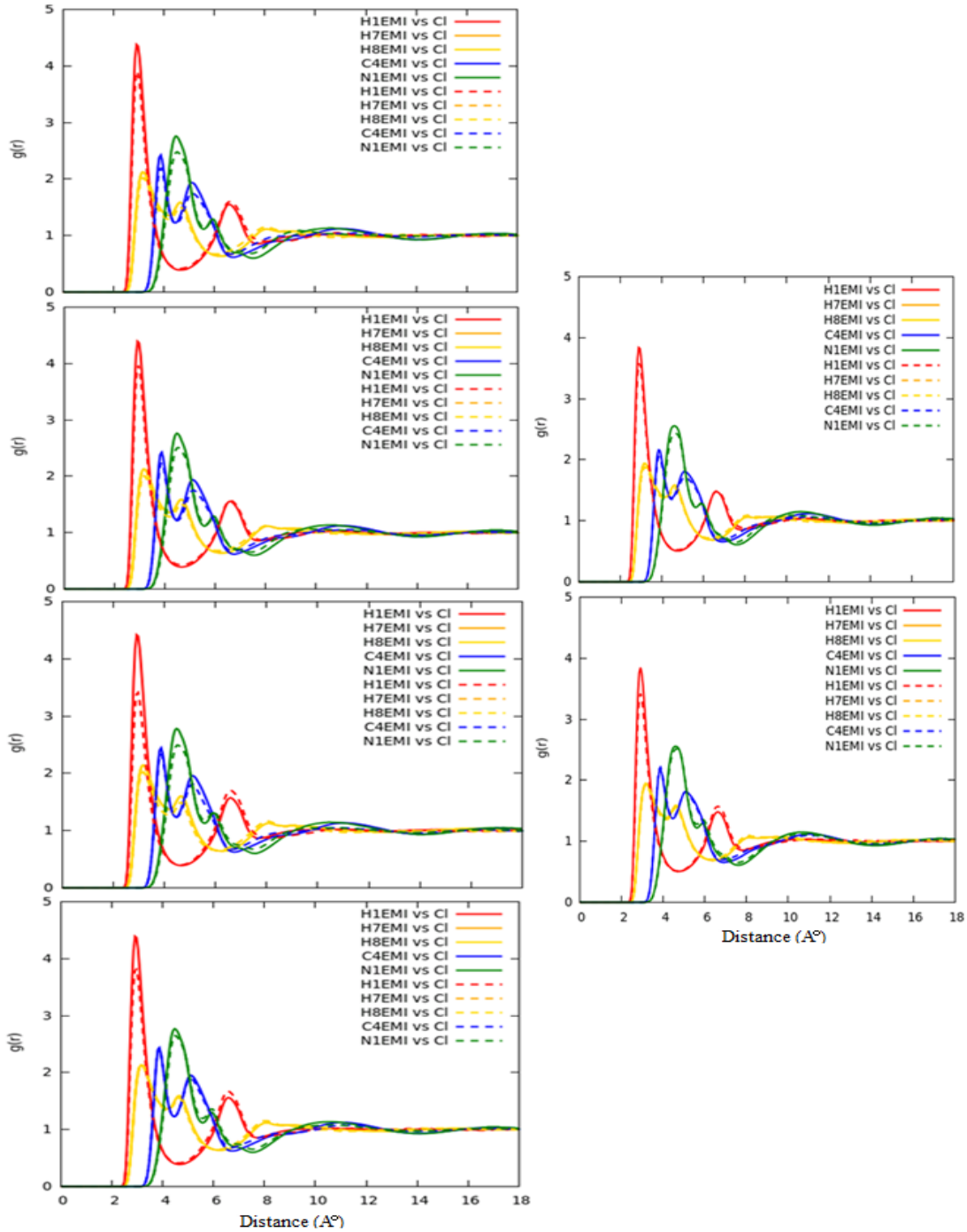


Figure A.9.1: Centre-of-mass radial distribution functions of $[C_2C_1Im][Cl]$: AA 200:1(solid) *vs* 5:1 (dashed) for Lys, Arg, Asp and Glu(top to bottom) respectively at 300K (left) and Lys and Asp (top to bottom) respectively at 400K (right) for H1,H7,H8,N1,C4 atoms of $[C_2C_1Im]^+$ *vs* Cl^-.

Appendix A.9. 2: Centre-of-mass radial distribution function around $[C_2C_1Im]^+$ cation atomic sites and O1 atom of $[CH_3CO_2]^-$ anion for LYS in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K

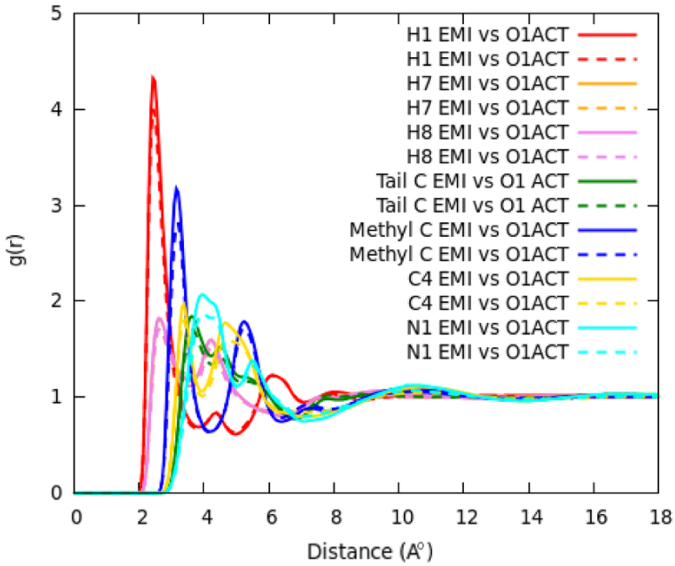


Figure A.9. 2: Centre-of-mass radial distribution function of atomic sites on $[C_2C_1Im]^+$ cation and O1 atom on $[CH_3CO_2]^-$ anion for LYS in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K

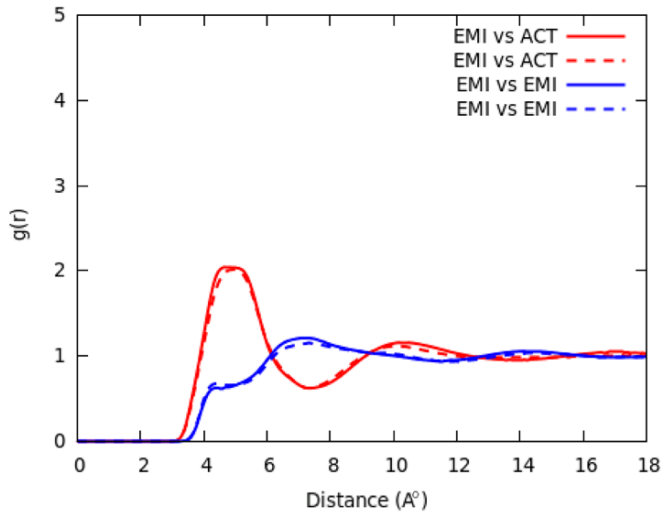


Figure A.9. 3: Centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:[lysine] for overall solvent-solvent interactions for lysine molecule for low concentration (solid lines) and high concentration (dashed lines) systems at 300K determined by MD simulation

Appendix A. 10: Effect of concentration ratio of IL:AA on solute-solvent interactions for lysine and arginine systems at 300K in $[C_2C_1Im][Cl]$ ionic liquid

Appendix A.10. 1: Centre-of-mass radial distribution function of $[C_2C_1Im]$ cation around carboxylate group of lysine and arginine molecule for Low concentration *vs* asp high concentration systems in $[C_2C_1Im]$ [Cl] at 300 K

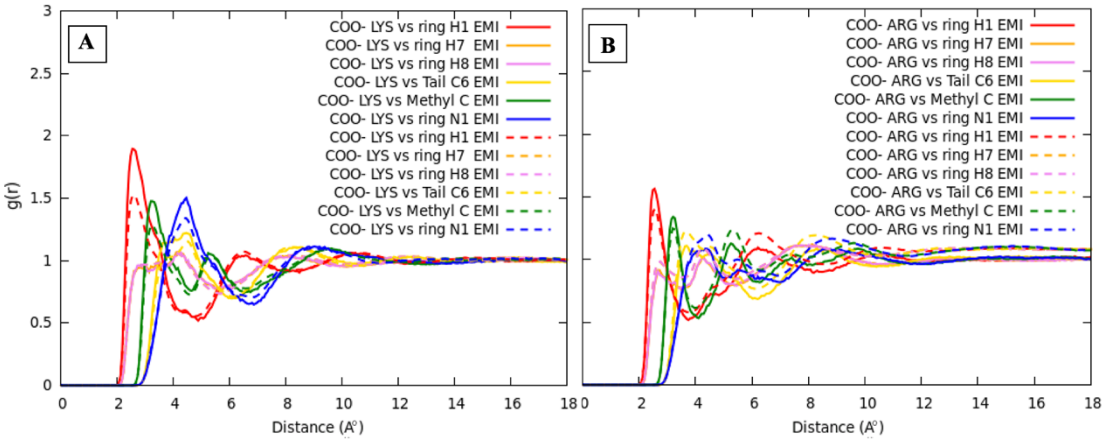


Figure A.10.1: Centre-of-mass radial distribution functions for lysine and arginine in pure $[C_2C_1Im][Cl]$ at 300K for H1,H7,H8,N1,C1,C4,C6 atomic sites on $[C_2C_1Im]^+$ cation vs carboxyl groups (COO^-) on (A) lysine molecule and (B) arginine molecule. Low concentration (200:1) are shown by solid lines and high concentration (5:1) by dashed lines, determined by MD simulation

Appendix A.10. 2: Site-site radial distribution function of $[C_2C_1Im]$ cation around backbone Nitrogen atom of lysine and arginine for low *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K

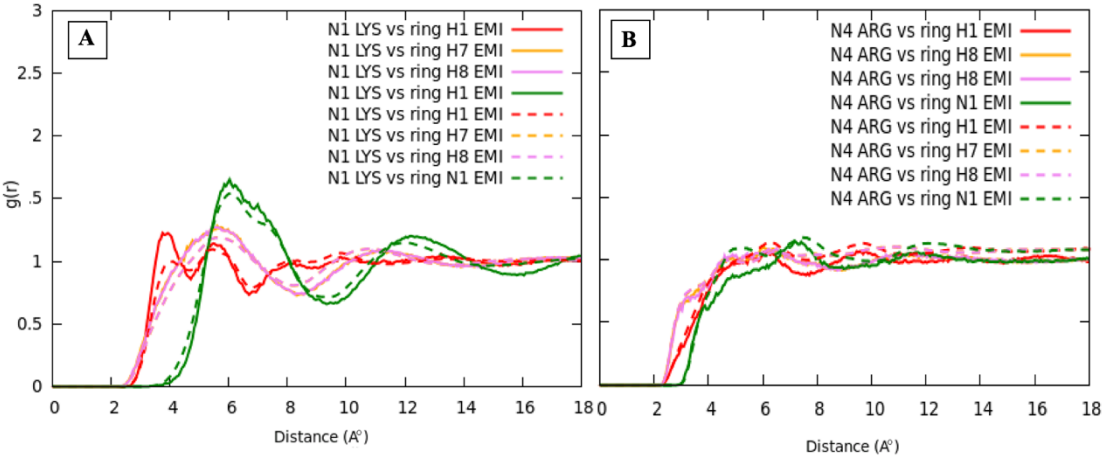


Figure A.10. 2: Comparison of site-site radial distribution function of $[C_2C_1Im]^+$ cation atomic sites H1, H7, H8 and N1 around backbone nitrogen atoms on (A) lysine and (B) arginine for low (IL:AA (200:1) solid lines) *vs* high (IL:AA (5:1) dashed lines) concentration systems in $[C_2C_1Im][Cl]$ at 300 K

Appendix A.10. 3: Site-site radial distribution function of [C_2C_1Im] cation around alpha-hydrogen and Nitrogen atom of lysine and arginine for Low concentration *vs* high concentration systems in [C_2C_1Im][Cl] at 300 K

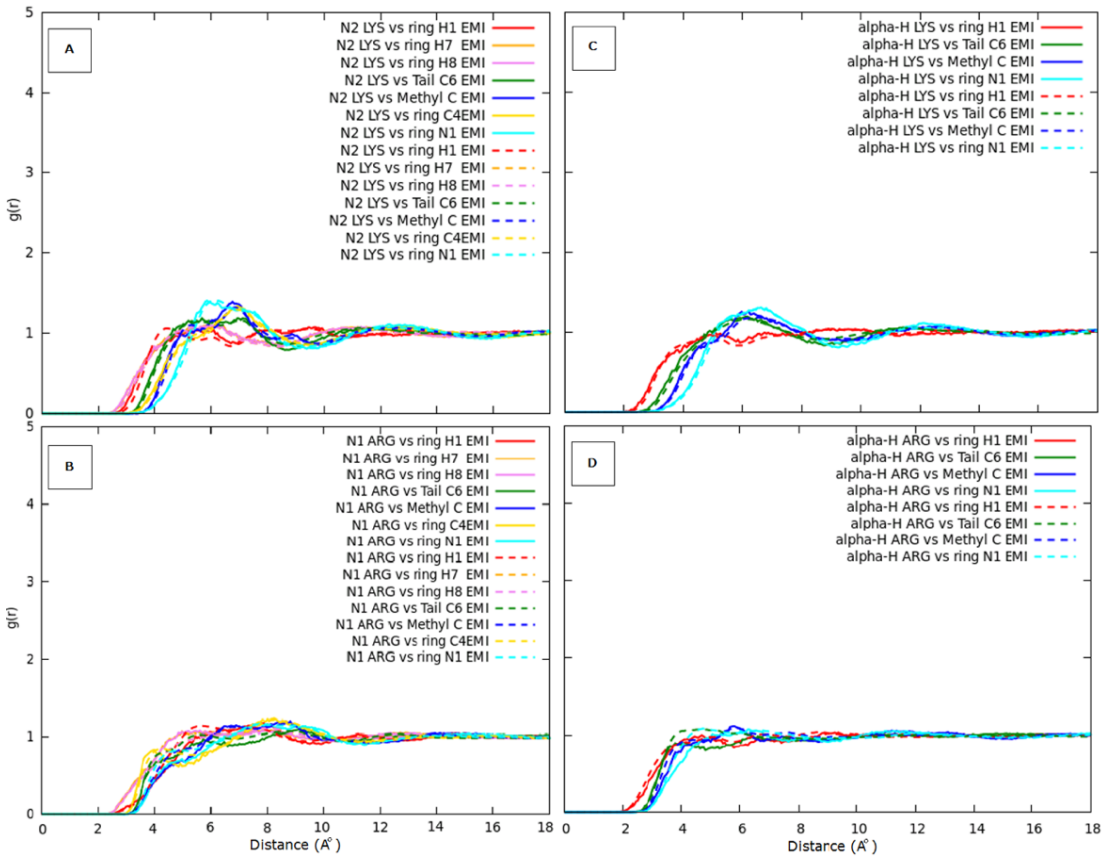


Figure A.10. 3: Comparison of centre-of-mass radial distribution function of $[C_2C_1Im]^+$ cation atomic sites H1, H7, H8, N1, C4, C6, C1 around (A,C) backbone nitrogen and a-hydrogen on lysine and (B) backbone nitrogen and α-hydrogen arginine for low concentration (solid line) *vs* high concentration systems (dashed lines) in [C_2C_1Im][Cl] at 300 K.

Appendix A.10. 4: Spatial distribution function of $[C_2C_1Im]^+$ cation around backbone nitrogen and carboxyl group of arginine molecule for low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K

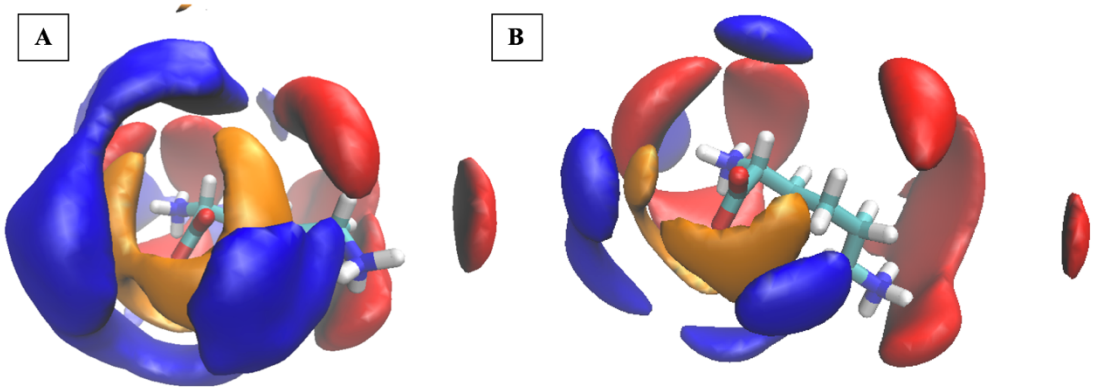


Figure A.10. 4: Spatial distribution functions around the central lysine molecule (A) low concentration system *vs* (B)high concentration system, in $[C_2C_1Im][Cl]$ IL at 300K. Blue= $[C_2C_1Im]^+$, red= Cl^- and orange= H1 atoms on $[C_2C_1Im]^+$ cation at iso-value=10 nm^{-3}

Appendix A.10. 5: Site-site radial distribution function of chloride anion around carboxylate group of lysine and arginine for low *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K

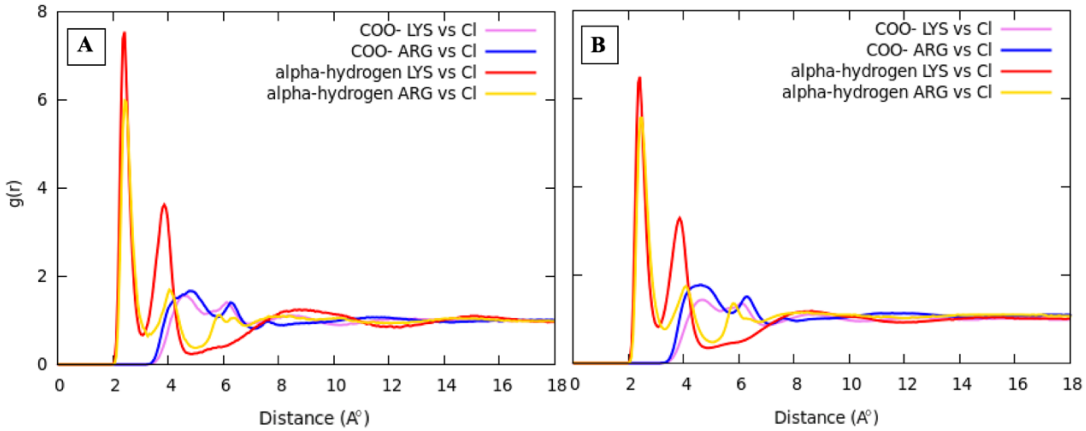


Figure A.10. 5: Site-site radial distribution function of $[Cl]^-$ anion around α-hydrogen and backbone nitrogen atom of lysine and arginine for (A) low concentration *vs* (B) high concentration systems in $[C_2C_1Im][Cl]$ at 300 K

Appendix A. 11: Effect of Temperature on Interactions for Lysine amino acid in $[C_2C_1Im][Cl]$ Ionic liquid at 300K *vs* 400K

Appendix A.11. 1: Site-site radial distribution function of cation atomic sites around carbonyl group on lysine molecule for low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K *vs* 400K

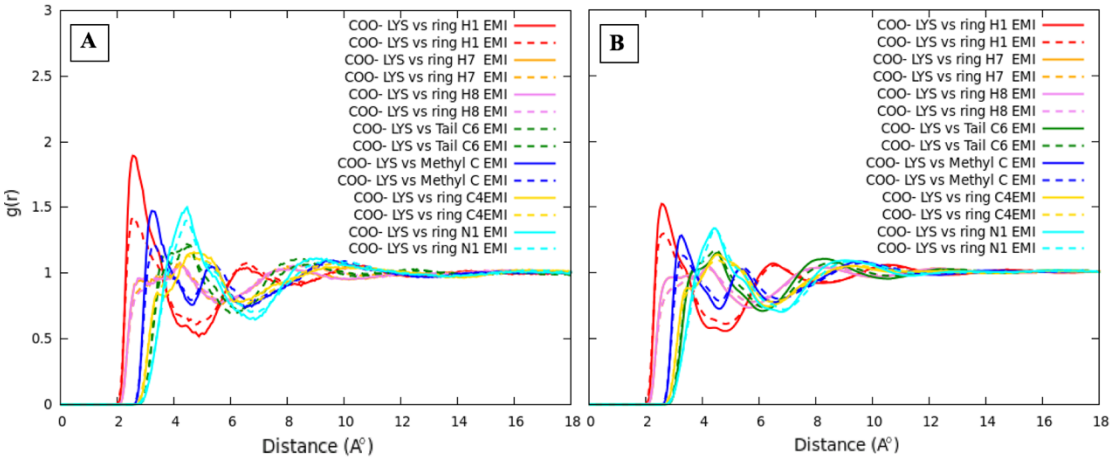


Figure A.11.1: Site-site radial distribution function of carbonyl group on lysine molecule around cation atomic sites for (A) low concentration *vs* (B) high concentration systems in $[C_2C_1Im][Cl]$ at 300K (solid lines) *vs* 400K (dashed lines)

Appendix A.11. 2: Site-site radial distribution function of chloride anion around backbone nitrogen atom on lysine molecule for low concentration *vs* high concentration systems in $[C_2C_1Im][Cl]$ at 300 K *vs* 400K

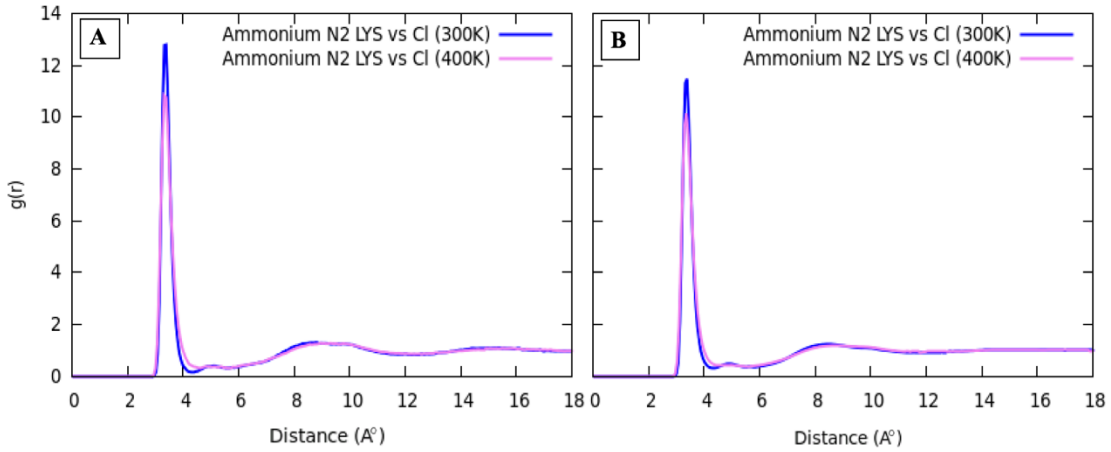


Figure A.11. 2: Site-site radial distribution function of chloride anion around ammonium N2 on lysine molecule (A) low concentration *vs* (B) high concentration systems in $[C_2C_1Im][Cl]$ at 300 K and 400K

Appendix A.11. 3: Overall Centre-of-mass radial distribution function of solute-IL, IL-IL and solute-solute interactions in pure [C_2C_1Im][Cl] at 300 K *vs* 400K

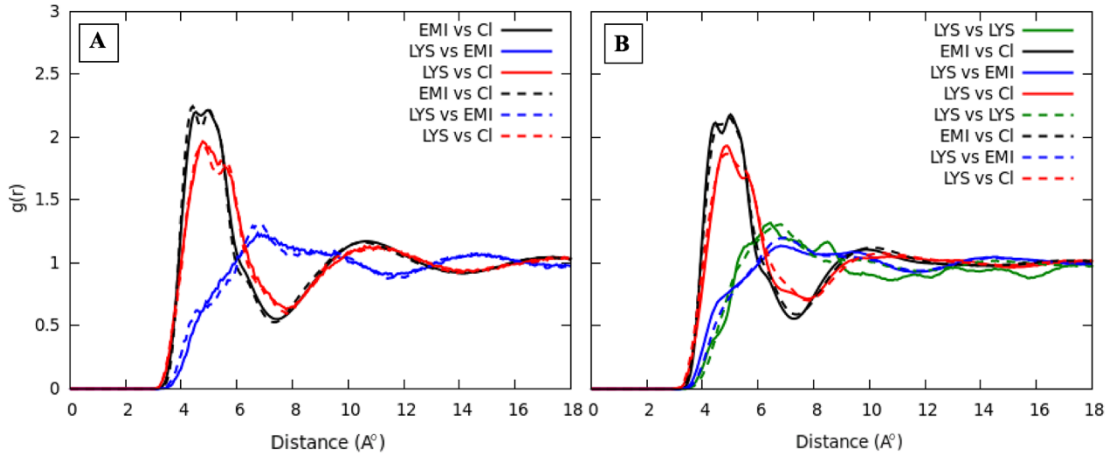


Figure A.11. 3: Overall centre-of-mass radial distribution functions for lys-lys, lys-cation, lysine-anion and cation-anion interactions for (A) low concentration *vs* (B) high concentration systems in [C_2C_1Im][Cl] at 300K (solid lines) and 400K (dashed lines)

Appendix A. 12: Effect on interactions for lysine high concentration systems at 300K *vs* 400K in [C_2C_1Im][Cl] IL/ water mixture

Appendix A.12. 1 Overall Centre-of-mass radial distribution function of solute-IL, IL-IL and solute-solute interactions in pure [C_2C_1Im][Cl] and [C_2C_1Im][Cl]/H_2O mixture

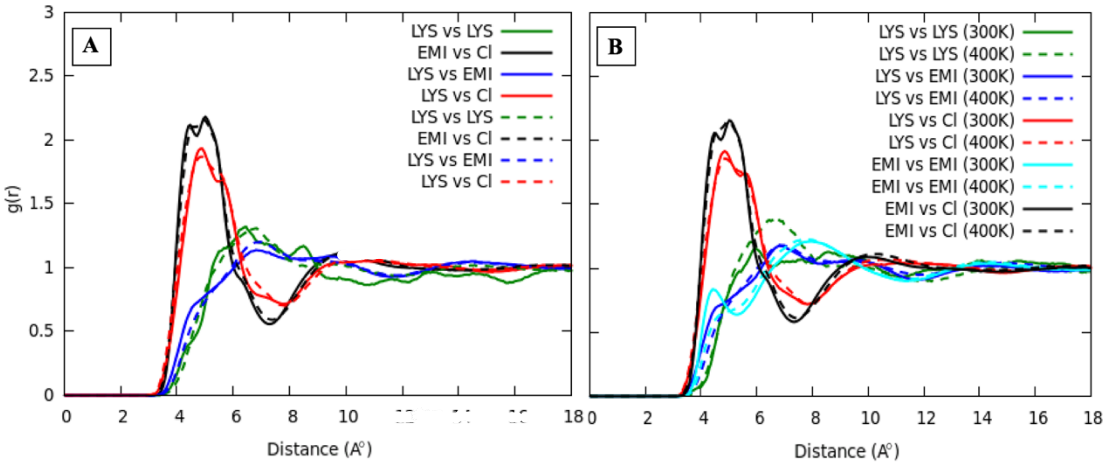


Figure A.12. 1: Centre-of-mass radial distribution functions for [C_2C_1Im][Cl]:LYS for IL-IL, IL-solute and solute-solute interactions for (A) high concentration lysine systems in pure [C_2C_1Im][Cl] and (B) high concentration lysine systems in [C_2C_1Im][Cl]/water mixture 300K(solid lines) *vs* 400K (dashed lines)

Appendix A.12. 2: Spatial distribution function for $[C_2C_1Im][Cl]$:LYS high concentration system for solute-solvent interactions in pure and $[C_2C_1Im][Cl]/H_2O$ mixture

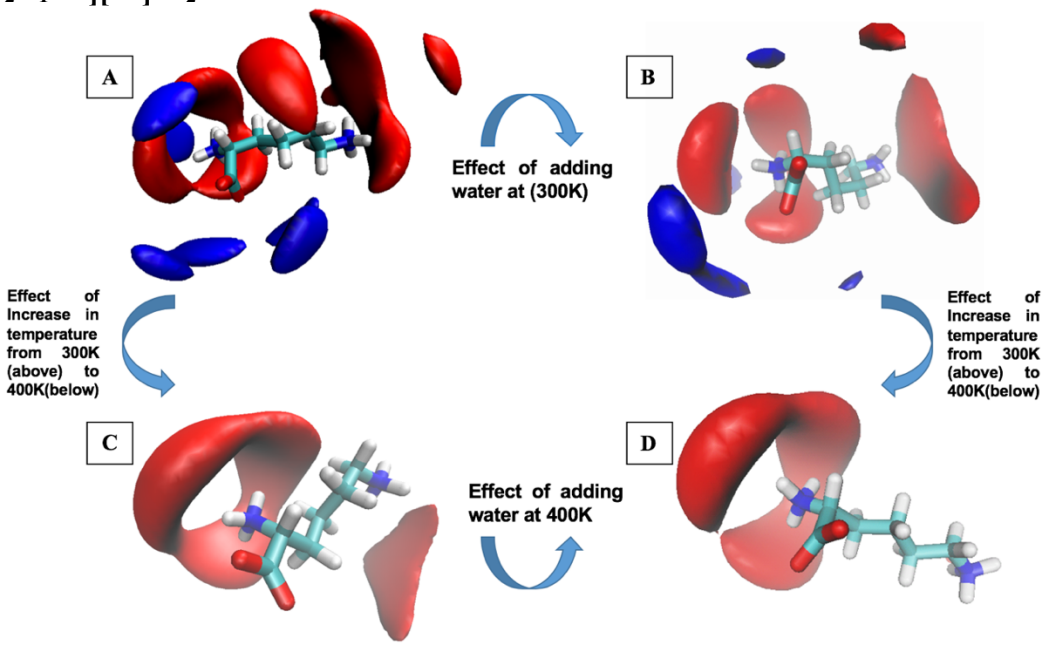


Figure A.12. 2: Spatial distribution functions around the central lysine molecule in pure and $[C_2C_1Im][Cl]/H_2O$ mixture (A) pure IL at 300K (B) IL/WAT mix at 300K (C) pure IL at 400K (D) IL/WAT mix at 400K in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- at iso-value=10 nm^{-3}

Appendix A.12. 3: Spatial distribution function for cation-anion and cation-cation interactions for lysine high concentration systems in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]/H_2O$ mixture

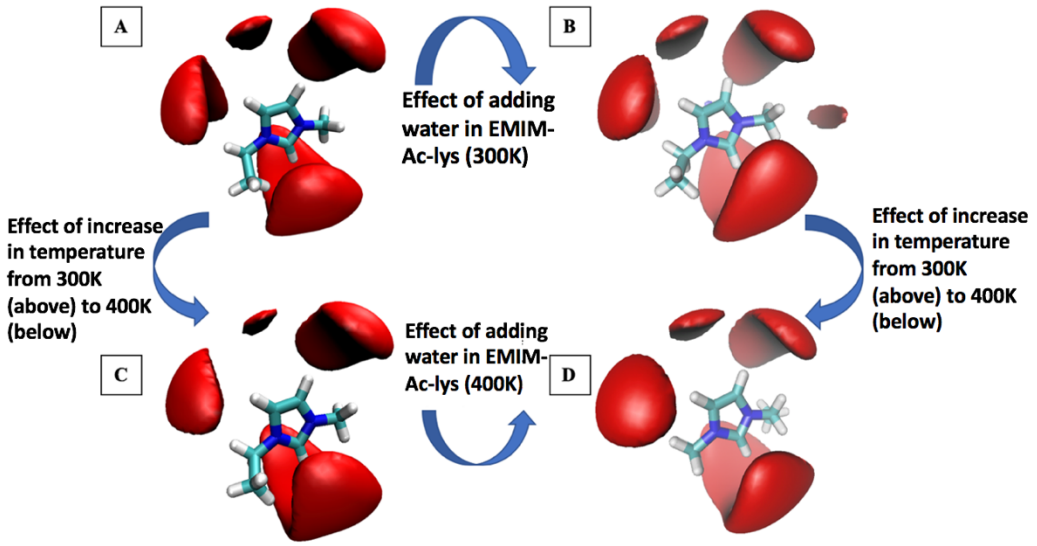


Figure A.12. 3: Spatial distribution functions for cation-cation and cation-anion interactions around the central $[C_2C_1Im]^+$ cation for lysine high concentration systems in (A,C) pure IL at 300K and 400K respectively and in (B,D) IL/H_2O mix at 300K and 400K respectively in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1im]^+$ and red= Cl^- or $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}

Appendix A.12. 4: Snapshots of lysine high concentration systems in pure $[C_2C_1Im][Cl]$ IL and $[C_2C_1Im][Cl]$-water mixture, after 100 ns simulation

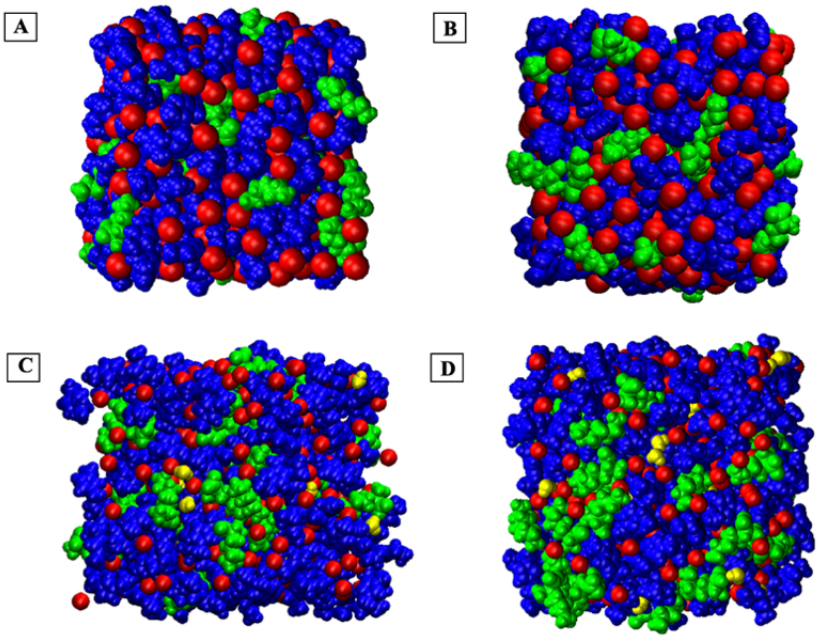


Figure A.12. 4: Snapshots of lysine high concentration systems in $[C_2C_1Im][Cl]$ IL after 100 ns simulation for (A,C) at 300K for pure and IL/WAT mix, respectively (B,D) at 400K for pure and IL/WAT mix, respectively. Blue= $[C_2C_1Im]^+$, red= Cl^-, green= lysine and yellow= water.

Appendix A. 13: Diffusion coefficients, MSD and log-log plots for lysine in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ at 400K

Appendix A.13.1: MSD curves and log-log plots for lysine in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ at 400K

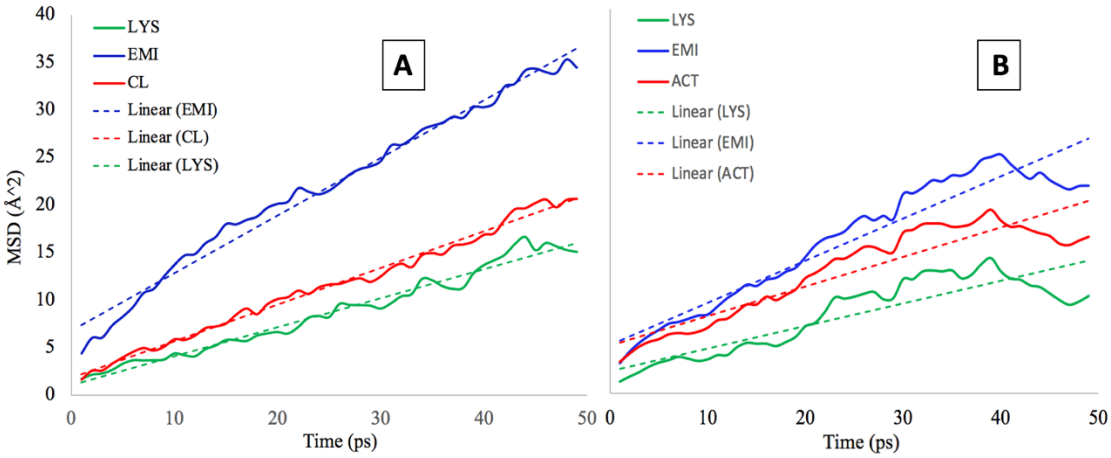


Figure A.13.1: MSD curves for high concentration systems for lysine in (A) $[C_2C_1Im][Cl]$ and (B) $[C_2C_1Im][CH_3CO_2]$ at 400K

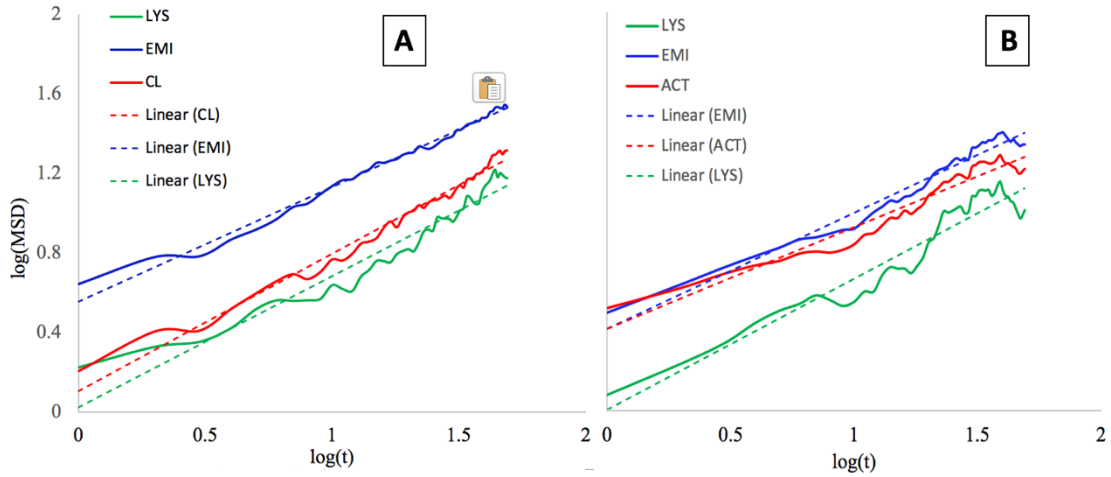


Figure A.13.2: log-log plots for high concentration lysine systems in (A) $[C_2C_1Im][Cl]$ at 400K and (B) $[C_2C_1Im][CH_3CO_2]$at 400K

Appendix A.13.2: Self-Diffusion coefficients values calculated from 100ns simulation for $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixture

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s			
		T=300K		T=400K	
		200:1	5:1	200:1	5:1
lysine-charged ($[C_2C_1Im][Cl]$/	lysine	1.94	2.32	5.91	2.29
	$[C_2C_1Im]$	1.99	2.41	12.51	6.50
	CL	1.82	2.36	10.91	5.01
lysine-charged ($[C_2C_1Im][Cl]$/wat)	lysine		0.34		2.43
	$[C_2C_1Im]$		0.52		7.65
	CL		0.412		6.58
	water		1.67		34.94

Appendix A. 14: Effect of concentration ratio of IL:AA on interactions for lysine and systems at 300K in $[C_2C_1Im][CH_3CO_2]$ ionic liquid

Appendix A.14. 1: Site-site radial distribution function of atomic sites on lysine molecule around $[C_2C_1Im]^+$ cation and acetate anion in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K

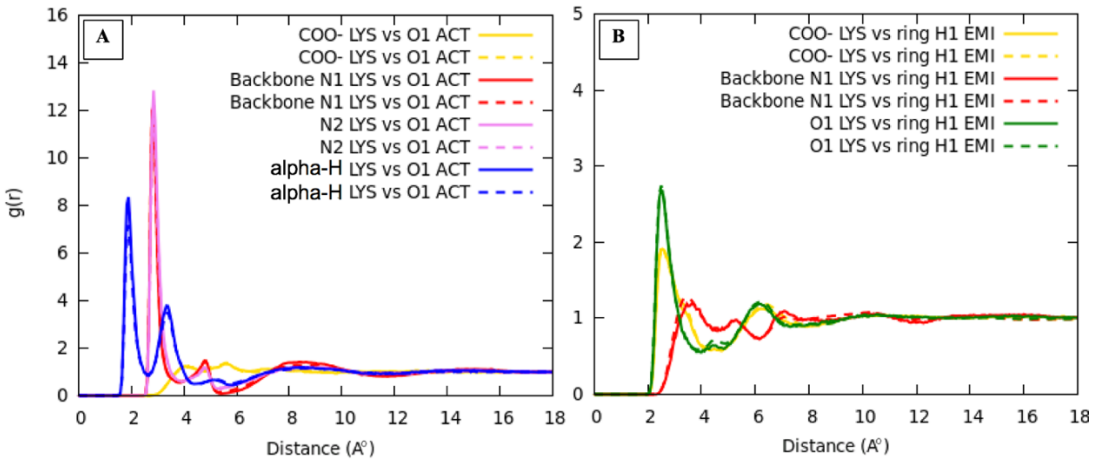


Figure A.14.1: Site-site radial distribution function of $[C_2C_1Im][CH_3CO_2]$:lys for anion O1 atom around (A) COO^-, backbone N1, ammonium N2 and α-hydrogen H15 on lys molecule and cation H1 atom distribution around (B) COO^-, O1 atom and backbone N1, for high concentration systems (dashed lines) *vs* low concentration (solid lines) at 300K

Appendix A.14. 2: Overall solute-solute and solute-solvent interactions for lysine systems at 300K determined by MD simulation [C_2C_1Im][CH_3CO_2] IL at 300 K

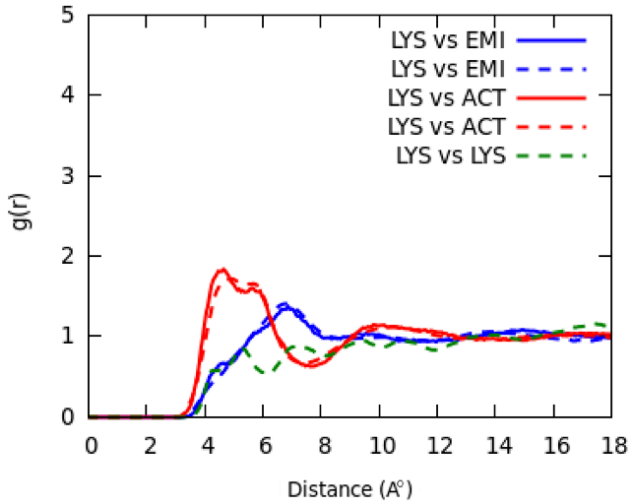


Figure A.14.2: Centre-of-mass radial distribution functions for [C_2C_1Im][CH_3CO_2]:lys for overall solute-solute and solute-solvent interactions for lysine molecule for low concentration (solid lines) *vs* high concentration (dashed lines) systems at 300K determined by MD simulation

Appendix A.14. 3: Centre-of-mass radial distribution function of atomic sites on lysine molecule around $[C_2C_1Im]^+$ cation in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K

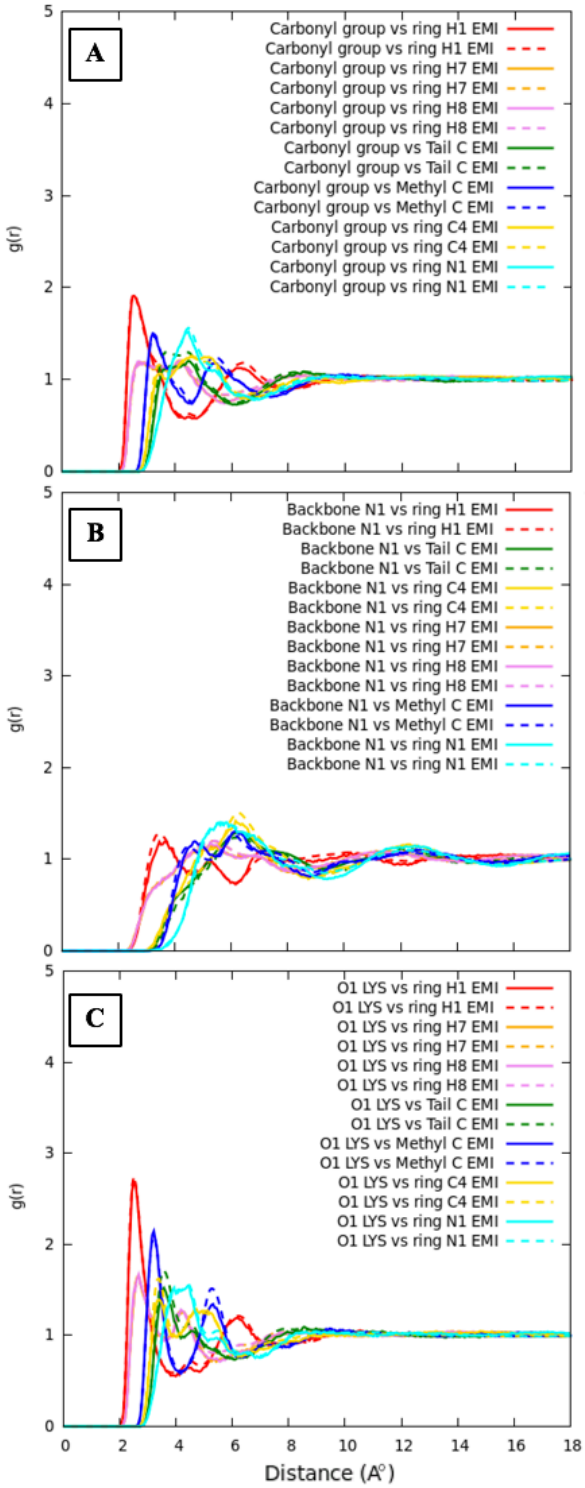


Figure A.14.3: Centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:lys at 300K for H1,H7,H8,N1,C1,C4 and C6 atomic sites on $[C_2C_1Im]^+$ cation around (A) N1 backbone nitrogen atom (B) COO^- carbonyl group (C) O1 oxygen atom on lysine molecule, (low concentration (200:1) solid lines) and (high concentration (5:1) dashed lines) determined by MD simulation

Appendix A. 15: Effect of Temperature on Interactions for Lysine amino acid in $[C_2C_1Im][CH_3CO_2]$ Ionic liquid at 300K *vs* 400K

Appendix A.15. 1: Centre-of-mass radial distribution of solute-solute, solute-solvent and solvent-solvent interactions for LYS in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K

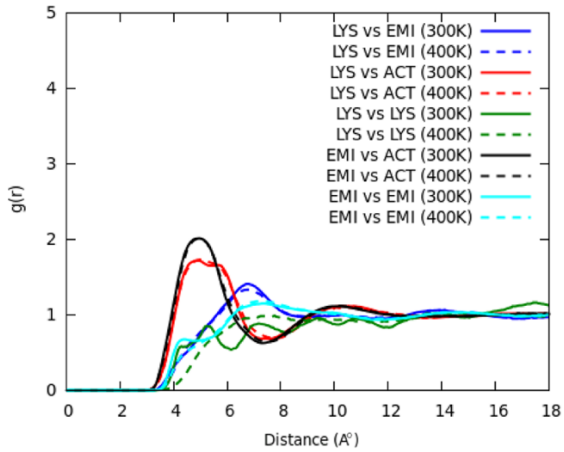


Figure A.15.1: Centre-of-mass radial distribution function for solute-solute, solute-solvent and solvent-solvent interactions for lysine high concentration systems in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K (solid lines) *vs* 400K (dashed lines)

Appendix A.15. 2: Site-site radial distribution around lysine atomic sites in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K

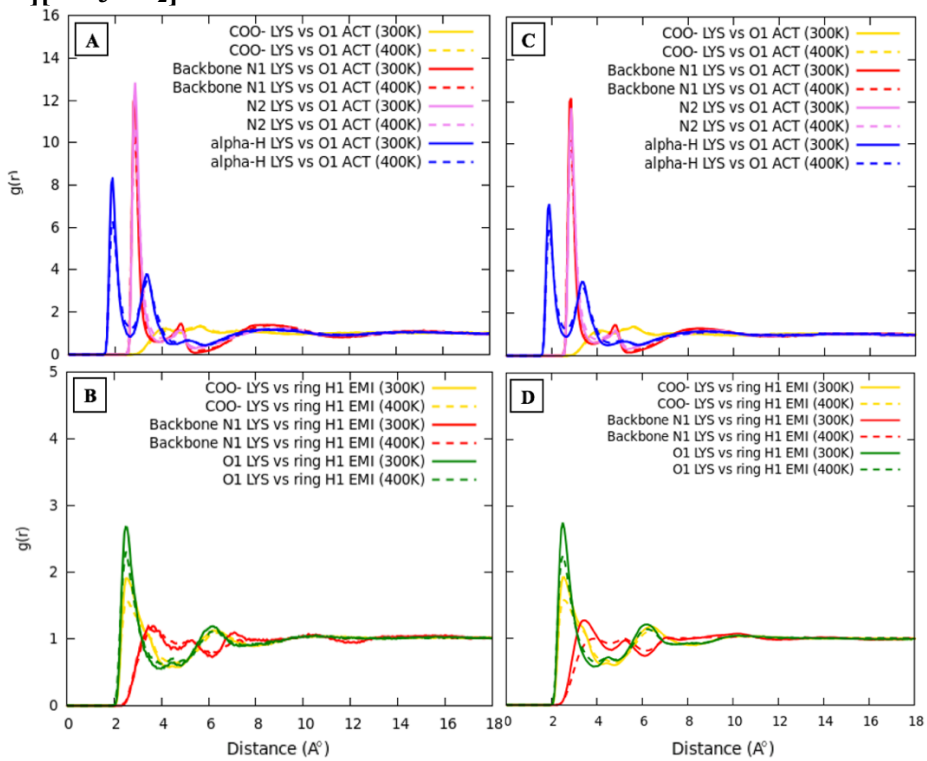


Figure A.15. 2: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: lysine low concentration ((200:1 (left A,B)) *vs* high concentration ((5:1) (right C,D)) at 300K (solid lines) *vs* 400K (dashed lines) for (A,C) acetate anion O1 atom around backbone nitrogen, carbonyl group, terminal nitrogen and α-hydrogen (B,D) H1 atom on $[C_2C_1Im]^+$ cation around backbone nitrogen, carbonyl group and O1 atom determined by MD simulation

Appendix A.15. 3: Spatial distribution functions comparing cation-anion, cation-cation interactions for lysine in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K

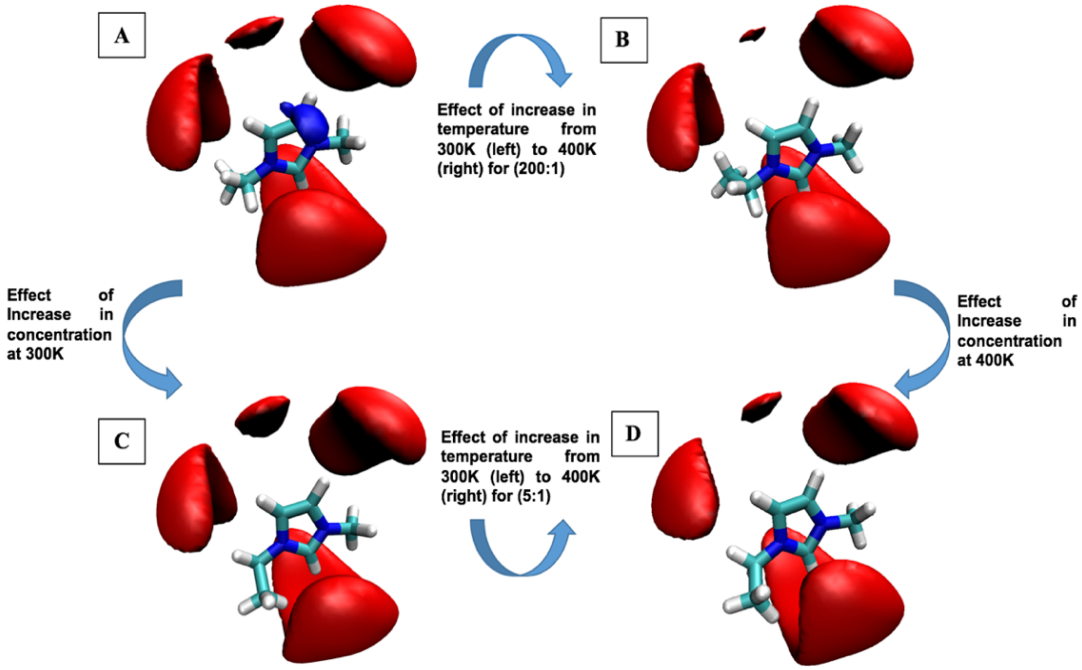


Figure A.15. 3: Spatial distribution functions for cation-cation and cation-anion interactions around the central $[C_2C_1Im]^+$ cation for lysine low concentration system at (A) 300K and (B) 400K, and for lysine high concentration system at (C) 300K and (D) 400K in $[C_2C_1Im][CH_3CO_2]$ IL in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}

Appendix A. 16: Effect of concentration and temperature on interactions for lysine systems at 300K *vs* 400K in $[C_2C_1Im][CH_3CO_2]$ IL/H_2O mixture

Appendix A.16. 1: Site-site radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: lysine around alpha-hydrogen atom in pure IL and IL/water-mix

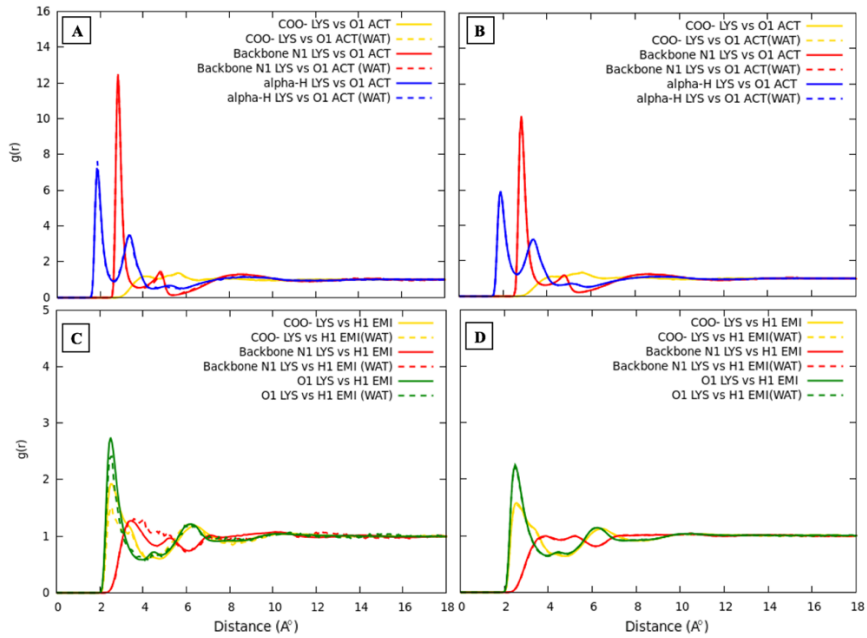


Figure A.16.1: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:lysine high concentration (5:1) systems with water (dashed lines) and without water (solid lines) at 300K (A,C) *vs* 400K (B,D). (A,B) show acetate anion (O1 atom) distribution around backbone nitrogen, carbonyl group and α-hydrogen on lys. (B,D) show $[C_2C_1Im]^+$ cation (H1 atom) distribution around backbone nitrogen, carbonyl group and O1 atom lys *vs* H1 determined by MD simulation

Appendix A.16. 2: Centre-of-mass radial distribution of solute-solute, solute-solvent and solvent-solvent interactions for lysine in $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K in pure IL and IL/water-mix

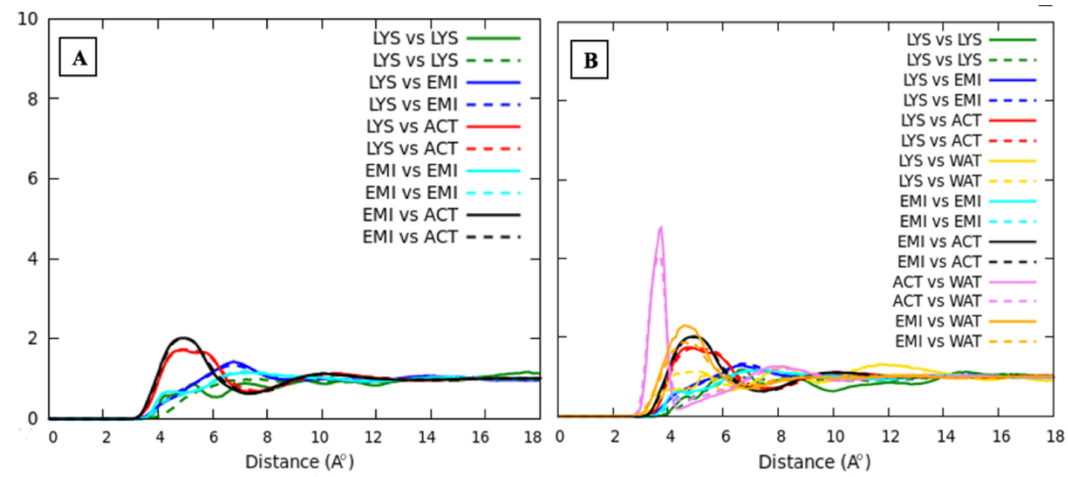


Figure A.16. 2: Comparison of centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: lysine high concentration (5:1) systems for solute-solvent, solute-solute and solvent-solvent interactions in (A) pure IL *vs* (B) IL/water mix at 300K (solid lines) and 400K (dashed lines) determined by MD simulation

Appendix A.16. 3: Spatial distribution function for $[C_2C_1Im][CH_3CO_2]$:lys high concentration system for solute-solvent interactions in pure and $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture

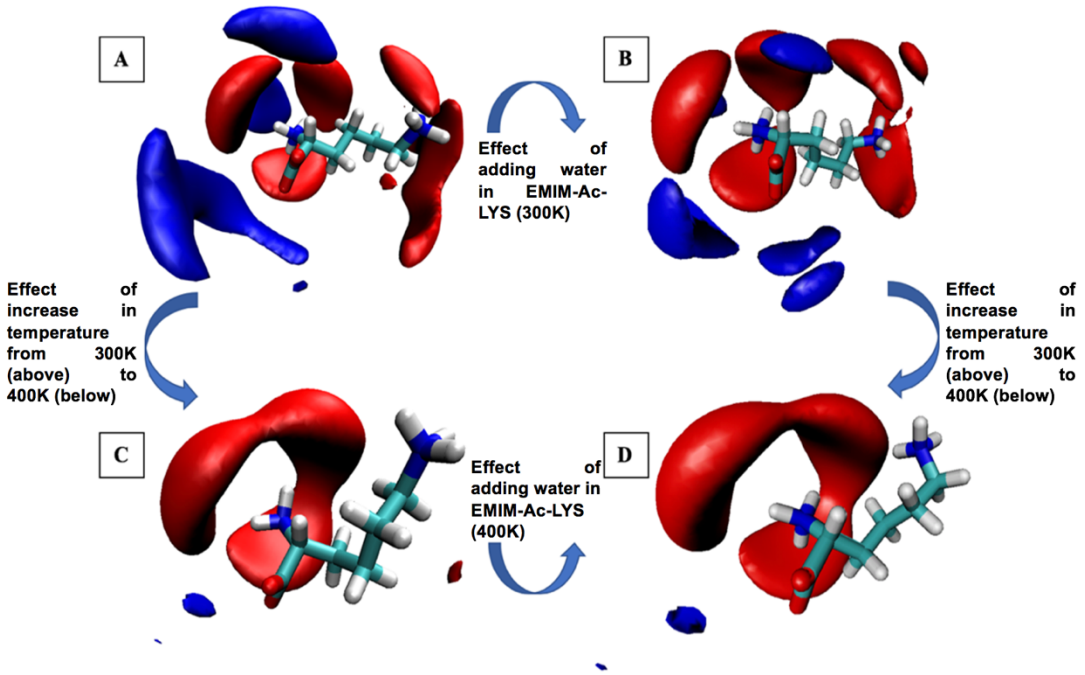


Figure A.16. 3: Spatial distribution functions around the central lysine molecule in pure and $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture (A) pure IL at 300K (B) IL/WAT mix at 300K (C) pure IL at 400K (D) IL/WAT mix at 400K in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$at iso-value=10 nm^{-3}

Appendix A.16. 4: Spatial distribution function for cation-anion and cation-cation interactions in high concentration systems of lysine in pure and $[C_2C_1Im][CH_3CO_2]$/water mixture

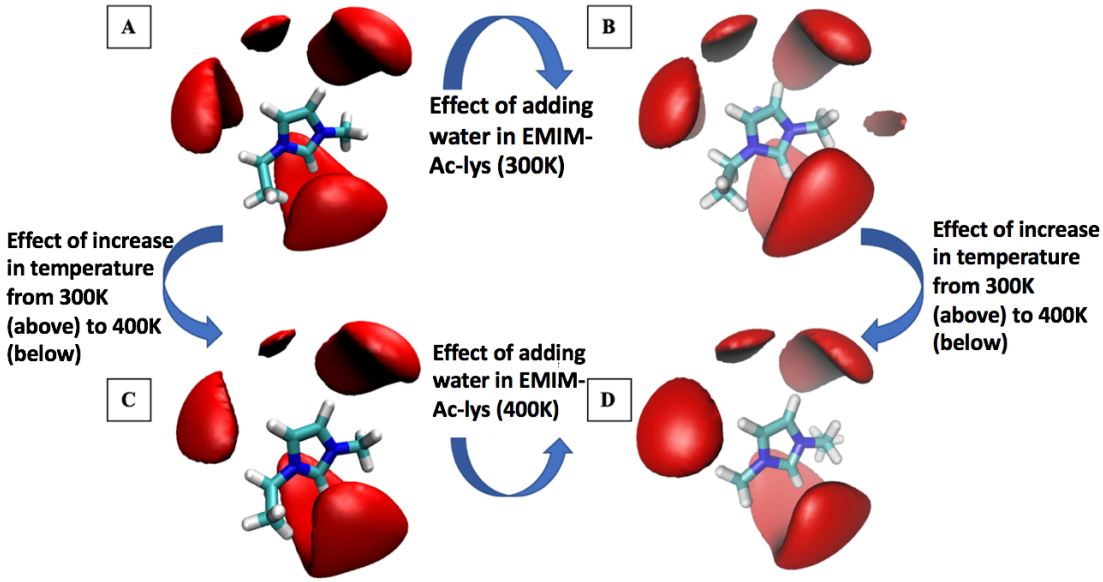


Figure A.16. 4: Spatial distribution functions for cation-cation and cation-anion interactions around the central $[C_2C_1Im]+$ cation for (A,C) lysine high concentration system at 300K and 400K respectively (pure IL) and(B,D) lysine high concentration system at 300K and 400K respectively (IL/H2O mix) in a view to keep the acidic hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$ at iso-value=10 nm^{-3}

Appendix A.16. 5: Self-diffusion coefficients values from entire 100ns simulation for pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 300K and 400K

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s			
		T=300K		T=400K	
		200:1	5:1	200:1	5:1
lysine-charged ($[C_2C_1Im][CH_3CO_2]$)	lysine	1.53	0.27	3.90	1.92
	$[C_2C_1Im]$	1.17	0.29	4.14	4.38
	$[CH_3CO_2]$	1.07	0.27	3.32	3.13
lysine-charged ($[C_2C_1Im[CH_3CO_2]$)/wat	lysine		0.23		0.76
	$[C_2C_1Im]$		0.29		2.99
	$[CH_3CO_2]$		0.22		1.78
	water		0.14		8.65

Appendix A. 17: Comparison of lysine systems in $[C_2C_1Im][Cl]$ *vs* $[C_2C_1Im][CH_3CO_2]$

Appendix A.17. 1:Spatial distribution functions for comparison between pure and water mixtures of $[C_2C_1Im][Cl]$and $[C_2C_1Im][CH_3CO_2]$at 300K and 400K

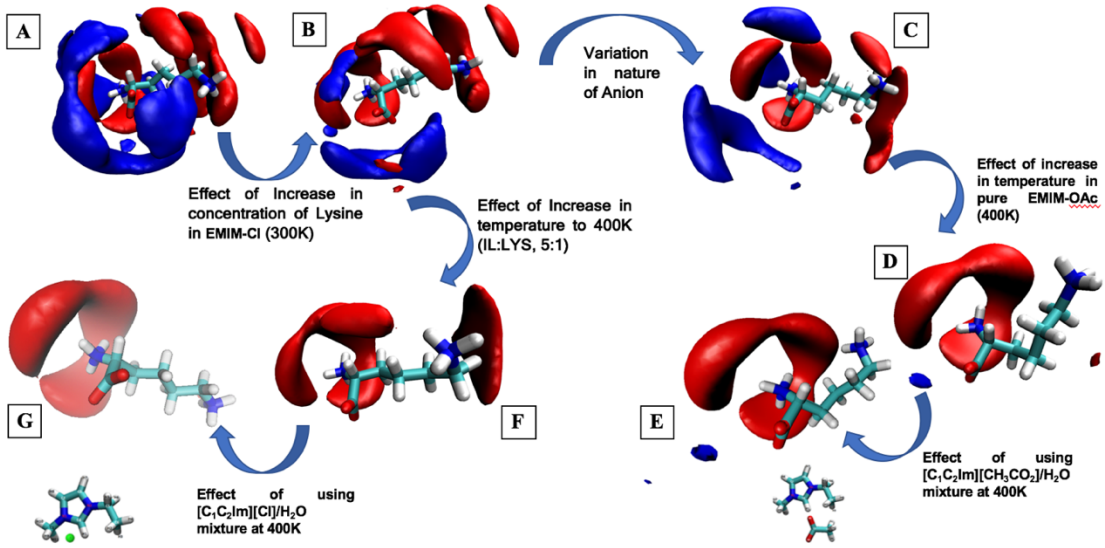


Figure A.17.1: Spatial distribution functions around the central lysine molecule in pure and water mixtures of $[C_2C_1Im][Cl]$ *vs* $[C_2C_1Im][CH_3CO_2]$ (A) pure $[C_2C_1Im][Cl]$ (200:1) at 300K (B) pure $[C_2C_1Im][Cl]$ (5:1) at 300K (C) pure $[C_2C_1Im][CH_3CO_2]$ (5:1) at 300K (D) pure $[C_2C_1Im][CH_3CO_2]$ (5:1) at 300K (D) pure $[C_2C_1Im][CH_3CO_2]$ (5:1) at 400K (E) $[C_2C_1Im][CH_3CO_2]/H_2O$ mix (5:1) at 400K (F) pure $[C_2C_1Im][Cl]$ (5:1) at 400K (D) $[C_2C_1Im][Cl]$/H2O mix (5:1) at 400K in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$ and red= Cl^- or $[CH_3CO_2]^-$ at iso-value=10nm^{-3}

Appendix A. 18: Input files for aspartate

Appendix A.18. 1: Geometry optimisation

```
%CHK=output.chk
%nprocshared=8
%mem=256mb
#P B3LYP/6-31G(d) OPT(EstmFC) SCRF=(PCM,Solvent=Water) Geom=Connectivity

C4H6O4N(-1)

-1 1
C
C 1 B1
N 2 B2 1 A1
H 3 B3 2 A2 1 D1
H 3 B4 2 A3 1 D2
H 3 B5 2 A4 1 D3
C 2 B6 1 A5 3 D4
O 7 B7 2 A6 1 D5
O 7 B8 2 A7 1 D6
H 2 B9 1 A8 3 D7
C 1 B10 2 A9 3 D8
O 11 B11 1 A10 2 D9
O 11 B12 1 A11 2 D10
H 1 B13 2 A12 3 D11
```

```
H 1 B14 2 A13 3 D12

B1 1.5480080
B2 1.5228279
B3 1.1149344
B4 1.1177279
B5 1.1165384
B6 1.5335792
B7 1.2079254
B8 1.3210043
B9 1.1169173
B10 1.5232127
B11 1.3160148
B12 1.2055132
B13 1.1149631
B14 1.1135456
A1 109.03236
A2 110.05951
A3 112.90887
A4 113.03100
A5 112.48357
A6 124.77194
A7 110.42304
A8 108.73247
A9 113.66503
A10 112.13708
A11 121.52532
A12 110.77147
A13 109.00031
D1 -176.65862
D2 67.398650
D3 -57.421701
D4 -125.60157
D5 -84.988034
D6 100.10986
D7 116.87855
D8 -32.958539
D9 -91.404974
D10 74.322773
D11 -157.21157
D12 85.952075

1 11 1.0 2 1.0 14 1.0 15 1.0
2 1 1.0 7 1.0 3 1.0 10 1.0
3 2 1.0 4 1.0 5 1.0 6 1.0
4 3 1.0
5 3 1.0
6 3 1.0
7 2 1.0 8 2.0 9 1.0
8 7 2.0
9 7 1.0
10 2 1.0
11 12 1.0 1 1.0 13 2.0
12 11 1.0
13 11 2.0
14 1 1.0
15 1 1.0
```

Appendix A.18. 2: RESP charge derivation

```
%CHK=output.chk
%nprocshared=8
%mem=256mb
#P  HF/6-31G(d)  OPT  SCRF=(PCM,Solvent=Water)Pop=MK  iop(6/33=2,6/42=6,6/50=1)
guess=read geom=check

ASP

-1 1
```

Appendix A.18.3: Distribution of [C_2C_1Im][Cl] IL ions around central aspartate molecule at 300K

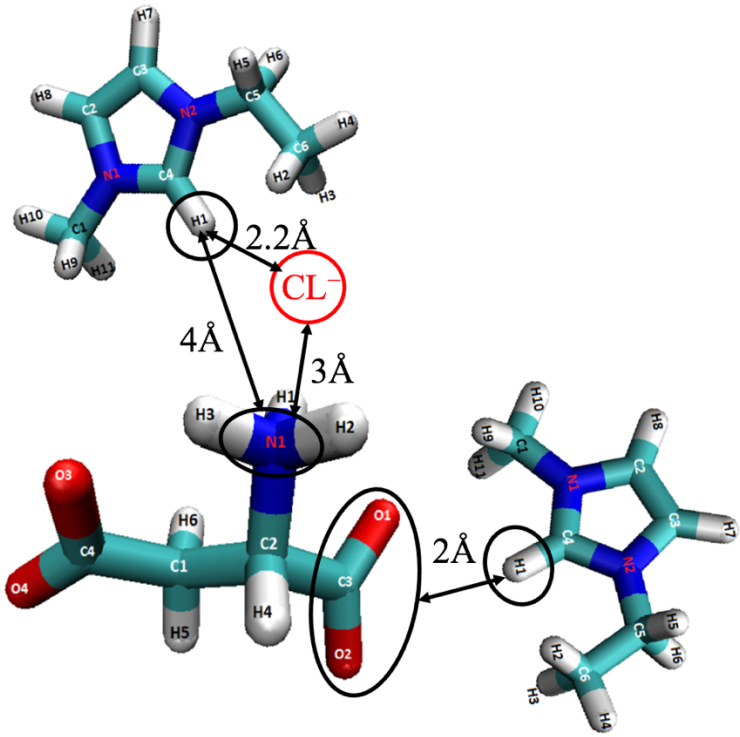


Figure A.18. 1: Distribution of [C_2C_1Im][Cl] IL ions around central aspartate molecule for high concentration system at 300K

Appendix A. 19: Effect of concentration ratio of IL:AA on interactions for aspartate and glutamate systems at 300K in [C_2C_1Im][Cl] ionic liquid

Appendix A.19.1: Centre-of mass radial distribution functions for [C_2C_1Im][Cl]: [AA] for overall solute-solute and solute-solvent interactions for low *vs* high concentration systems at 300K

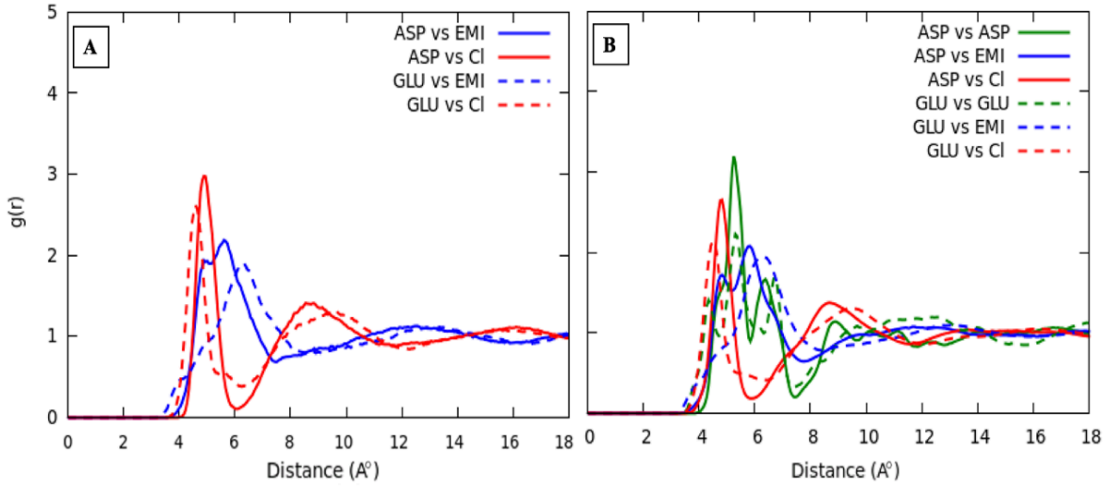


Figure A.19. 1: Centre-of-mass radial distribution functions for [C_2C_1Im][Cl]: [AA] for overall solute-solute and solute-solvent interactions for aspartate (solid lines) and glutamate (dashed lines) for (A) low concentration *vs* (B) high concentration systems at 300K determined by MD simulation

Appendix A.19. 2: Centre-of-mass radial distribution functions for [C_2C_1Im][Cl] for overall cation-anion and cation-cation interactions for aspartate and glutamate at 300K.

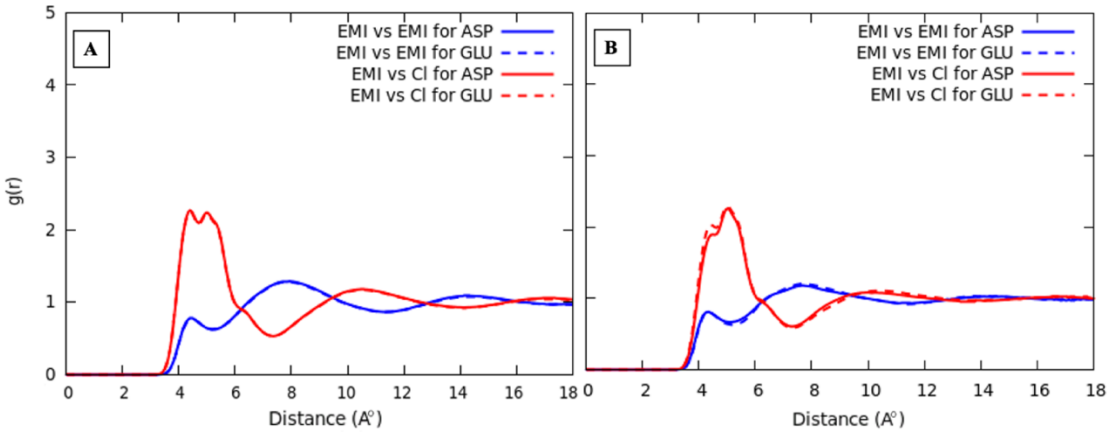


Figure A.19. 2: Centre-of-mass radial distribution functions for [C_2C_1Im][Cl] for overall cation-anion and cation-cation interactions for aspartate (solid lines) and glutamate (dashed lines) for (A) low concentration *vs* (B) high concentration systems at 300K determined by MD simulation

Appendix A. 20: Effect of Temperature on Interactions for aspartate amino acid in $[C_2C_1Im][Cl]$ Ionic liquid at 300K *vs* 400K

Appendix A.20. 1: Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: [ASP] for overall solute-solute and solute-solvent interactions for aspartate at 300K *vs* 400K

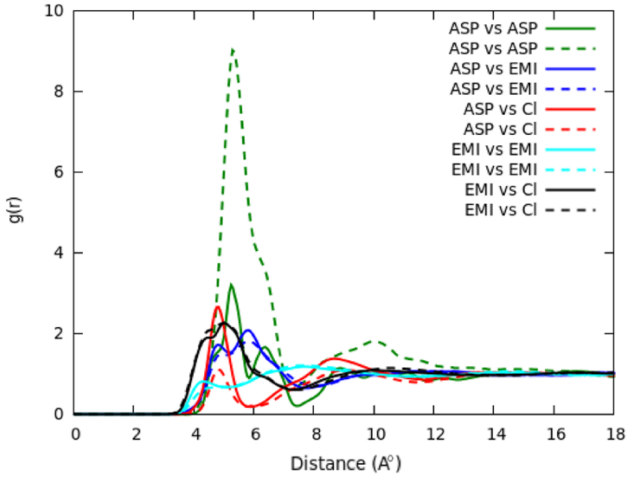


Figure A.20. 1: Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$: [ASP] for overall solute-solute and solute-solvent interactions for aspartate at 300K (solid lines) and 400K (dashed lines) for high concentration systems determined by MD simulation

Appendix A. 21: Effect of temperature on interactions for aspartate high concentration systems at 300K *vs* 400K in pure and IL/water mixture of $[C_2C_1Im][Cl]$

Appendix A.21. 1: Comparison for spatial distribution functions for aspartate interactions around central aspartate molecule for pure and $[C_2C_1Im][Cl]$ IL/water mixture

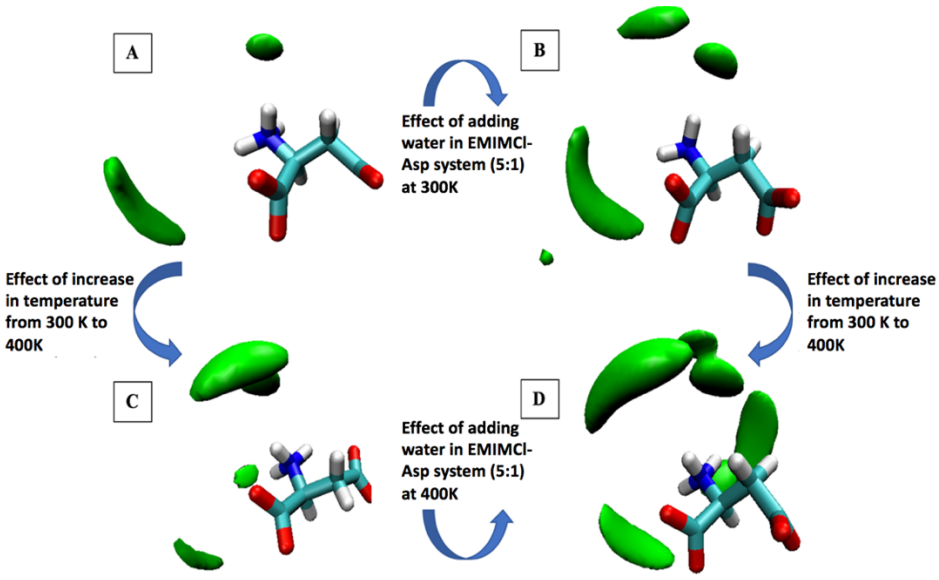


Figure A.21. 1: Spatial distribution functions around the central aspartate molecule in (A) pure $[C_2C_1Im][Cl]$ at 300K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 300K (C) pure $[C_2C_1Im][Cl]$ at 400K (B) $[C_2C_1Im][Cl]/H_2O$ mixture at 400K. Green= aspartate at iso-value=10 nm^{-3}

Appendix A.21. 2: Comparison for centre-of-mass radial distribution functions for interactions around central aspartate molecule for pure and $[C_2C_1Im][Cl]$ IL/water mixture

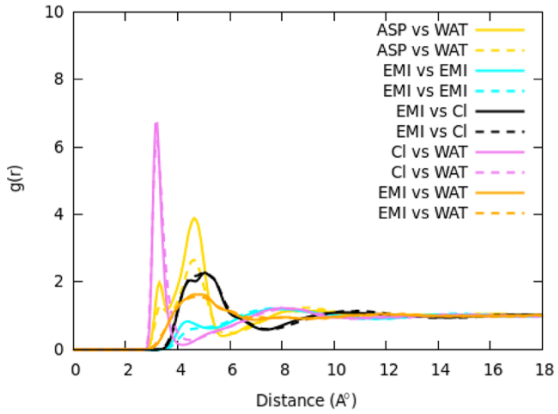


Figure A.21. 2: Site-site radial distribution functions for interactions around central aspartate molecule for (A) pure $[C_2C_1Im][Cl]$ IL and (B) $[C_2C_1Im][Cl]$ IL/water mixture at 300K (solid lines) *vs* 400K (dashed lines)

Appendix A.21. 3: Mean square displacement (MSD) curves and log-log plots for aspartate in pure $[C_2C_1Im][Cl]$ and $[C_2C_1Im][CH_3CO_2]$ at 400K

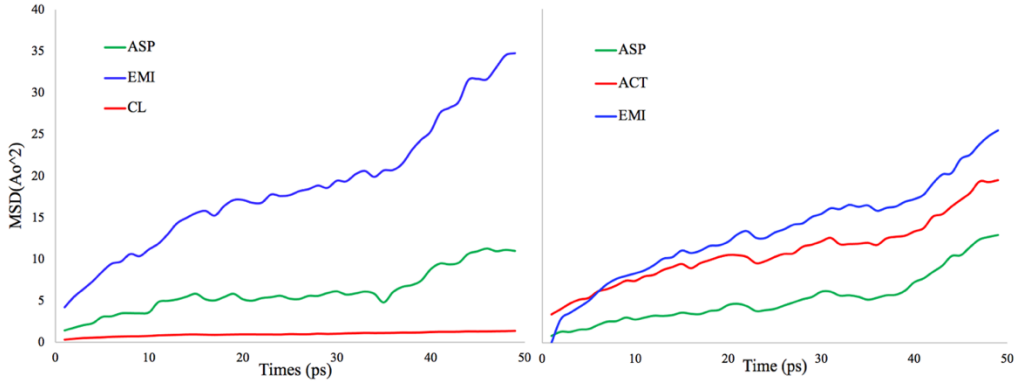


Figure A.21. 3: MSD curves for high concentration systems for aspartate in (A) $[C_2C_1Im][Cl]$ *vs* (B) $[C_2C_1Im][CH_3CO_2]$ at 400K

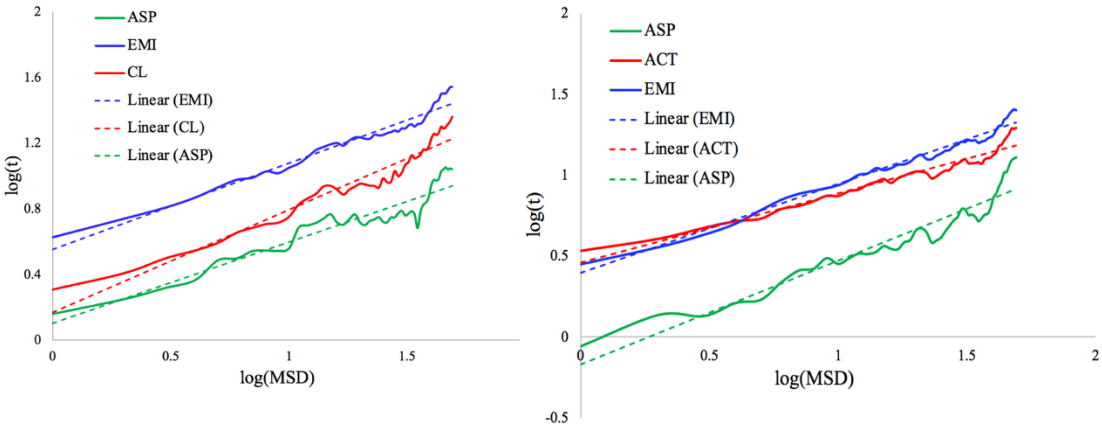


Figure A.21. 4: log-log plots for high concentration aspartate systems in (A) $[C_2C_1Im][Cl]$ *vs* (B) $[C_2C_1Im][CH_3CO_2]$at 400K

Appendix A.21. 4: Self-Diffusion coefficients values calculated from 100ns simulation for Aspartate in $[C_2C_1Im][Cl]$ and $[C_2C_1Im][Cl]$/water mixture

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s			
		T=300K		T=400K	
		200:1	5:1	200:1	5:1
aspartate $[C_2C_1Im][Cl]$	aspartate	0.08	0.52	12.32	4.39
	$[C_2C_1Im]$	0.32	0.58	18.03	9.10
	Cl	0.26	0.54	17.29	9.24
aspartate $[C_2C_1Im][Cl]$/wat	aspartate		3.53		2.65
	$[C_2C_1Im]$		3.62		8.44
	Cl		3.63		8.44
	water		4.79		26.13

Appendix A. 22: Effect of concentration ratio of IL:AA on interactions for aspartate systems at 300K in $[C_2C_1Im][CH_3CO_2]$ ionic liquid

Appendix A.22. 1: Centre-of-mass radial distribution function of atomic sites on aspartate molecule around $[C_2C_1Im]^+$ cation (H1 atom) in $[C_2C_1Im][CH_3CO_2]$ at 300 K

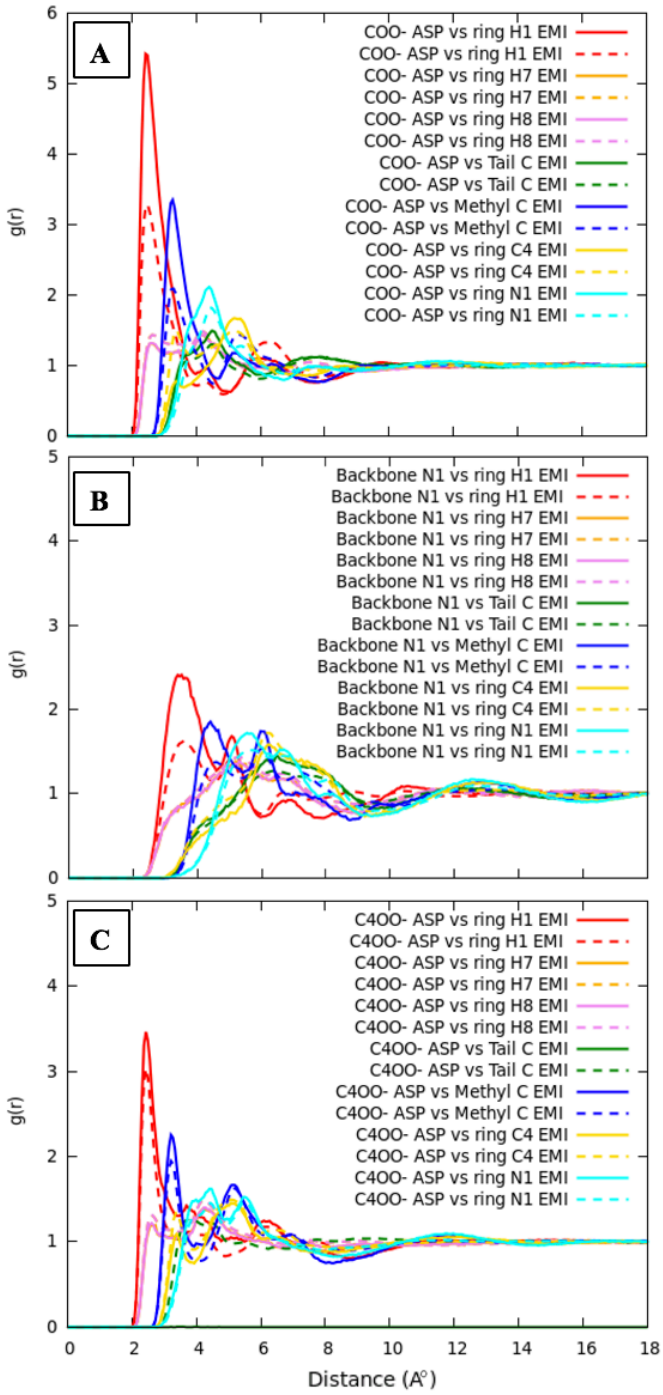


Figure A.22.1: Centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:[asp] at 300K for H1,H7,H8,N1,C1,C4,C6 atomic sites on $[C_2C_1Im]^+$ cation *vs* (A) COO^- carboxyl group (B) N1 backbone nitrogen atom and (C) COO^- group on C4 atom on aspartate molecule, (low concentration (200:1) solid lines) and (high concentration (5:1) dashed lines) determined by MD simulation

Appendix A.22.2: Overall mass-centred distribution of cation, anion and aspartate molecule around the central aspartate molecule

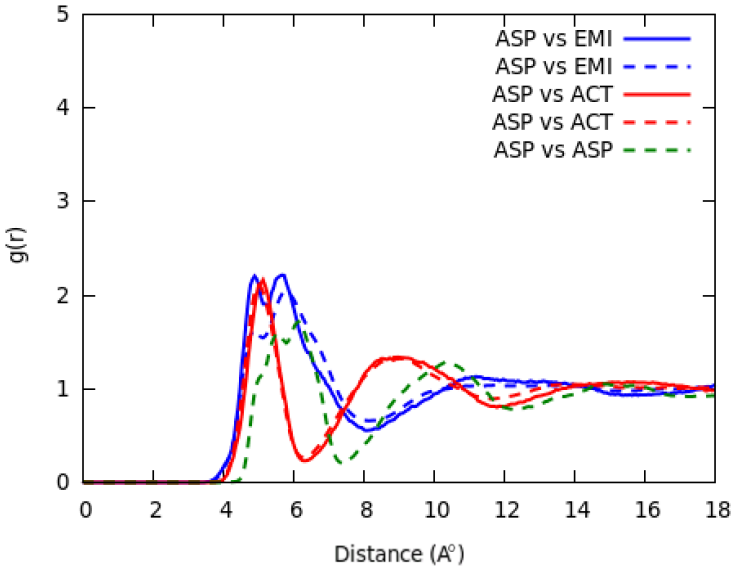


Figure A.22. 1: Centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$:[asp] for overall solute-solute and solute-solvent interactions for aspartate molecule for low concentration (solid lines) and high concentration (dashed lines) systems at 300K determined by MD simulation

Appendix A.22. 3: Centre-of-mass radial distribution functions for overall cation-anion and cation-cation interactions in pure $[C_2C_1Im][Cl]$ for aspartate systems at 300K

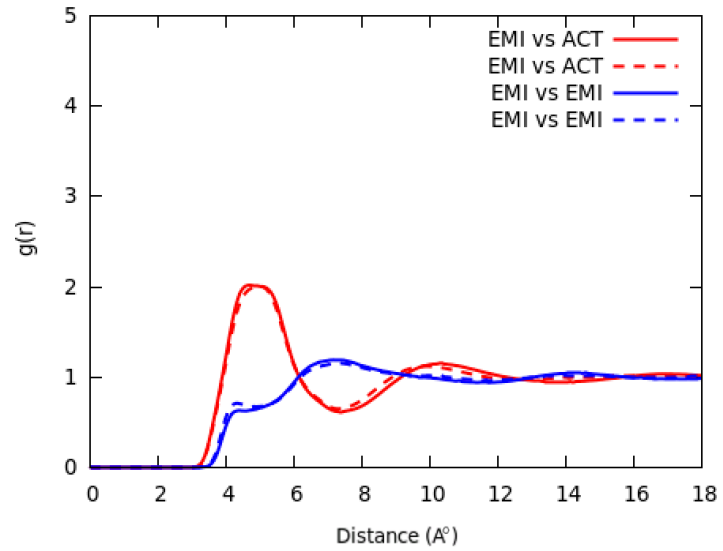


Figure A.22. 2: Centre-of-mass radial distribution function for overall cation-anion and cation-cation interactions around mass-centred $[C_2C_1Im]^+$ cation for asp systems in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K, low concentration (solid lines) *vs* high concentration (dashed lines).

Appendix A.22. 4: Centre-of-mass radial distribution function of $[CH_3CO_2]^-$ anion (O1 atom) around $[C_2C_1Im]^+$ cation atomic sites for aspartate systems in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K

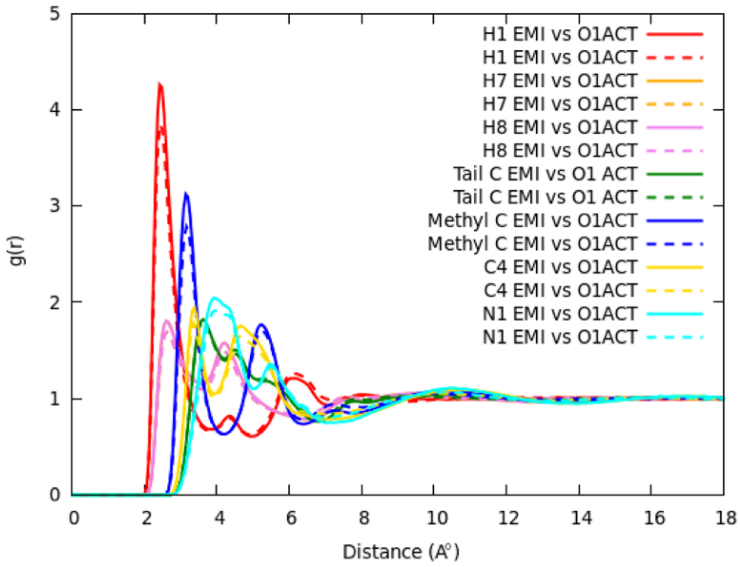


Figure A.22. 3: Centre-of-mass radial distribution function of O1 atom on $[CH_3CO_2]^-$ anion around (H1,H7,H8,N1,C4,C1,C6) atomic sites on $[C_2C_1Im]+$ cation for aspartate systems in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K

Appendix A.22. 5: Spatial distribution function for anion-anion and anion-cation interactions in pure $[C_2C_1Im][CH_3CO_2]$ for high concentration aspartate systems

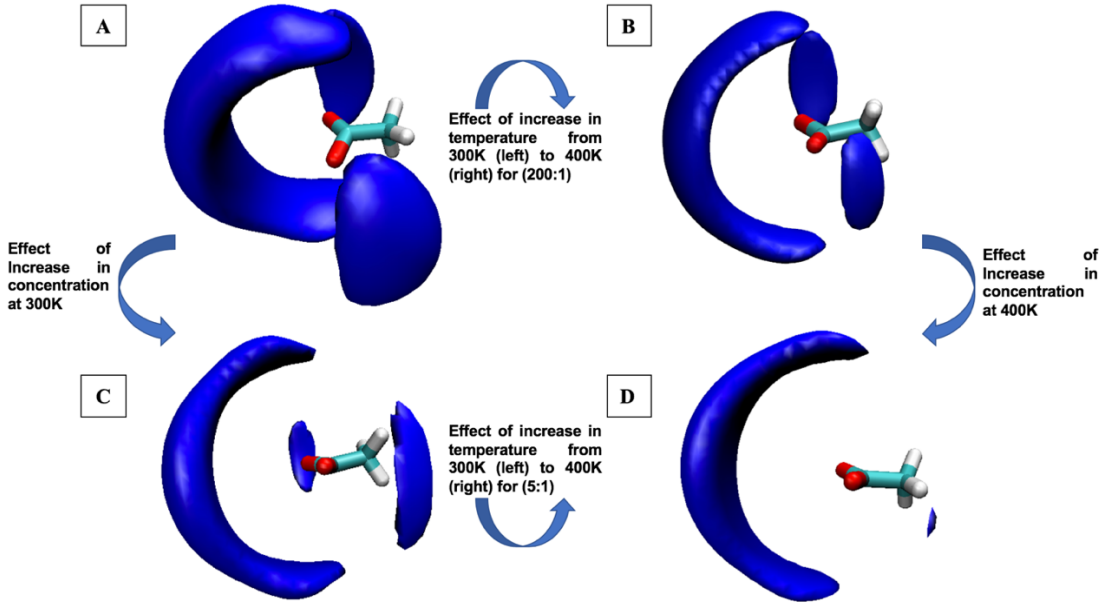


Figure A.22.4: Spatial distribution functions around the central anion in pure $[C_2C_1Im][CH_3CO_2]$ IL (A) asp low concentration system at 300K (B) asp low concentration system at 400K (C) asp high concentration system at 300K (D) asp high concentration system at 400K in a view to keep carbonyl carbon pointing left. Blue= $[C_2C_1Im]^+$ iso-value=10 nm^{-3}

Appendix A. 23: Effect of Temperature on Interactions for Aspartate amino acid in $[C_2C_1Im][CH_3CO_2]$ Ionic liquid at 300K *vs* 400K

Appendix A.23. 1: Centre-of-mass radial distribution functions for overall solute-solute and solute-solvent interactions for aspartate in $[C_2C_1Im][CH_3CO_2]$ at 300K *vs* 400K

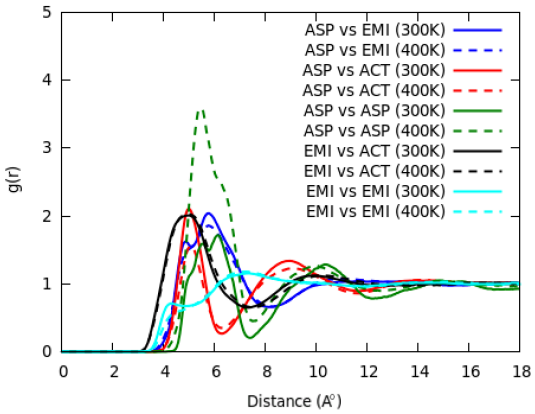


Figure A.23. 1: Centre-of-mass radial distribution functions for $[C_2C_1Im][Cl]$:[ASP] (5:1) for overall solute-solute and solute-solvent interactions for aspartate at 300K (solid lines) and 400K (dashed lines) determined by MD simulation

Appendix A.23. 2: Centre-of-mass radial distribution function around $[C_2C_1Im]^+$ cation atomic sites and $[CH_3CO_2]^-$ anion for aspartate systems in $[C_2C_1Im][CH_3CO_2]$ IL at 300 K *vs* 400K

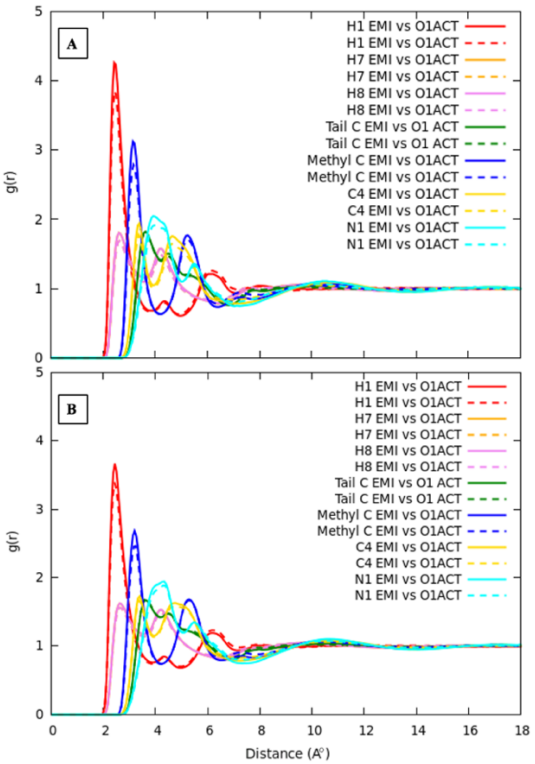


Figure A.23. 2: Centre-of-mass radial distribution function around (H1,H7,H8,N1,C4,C1,C6) atomic sites on $[C_2C_1Im]^+$ cation and O1 atom on $[CH_3CO_2]^-$ anion for asp in $[C_2C_1Im][CH_3CO_2]$ IL. (A) low *vs* (B) high concentration system. Solid and dashed lines show systems at 300K and 400K respectively.

Appendix A. 24: Effect of temperature on interactions for aspartate high concentration systems at 300K *vs* 400K in pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$ IL/ water mixture

Appendix A.24. 1: Comparison for centre-of-mass radial distribution functions for interactions around central aspartate molecule for pure and $[C_2C_1Im][CH_3CO_2]$ IL/water mixture

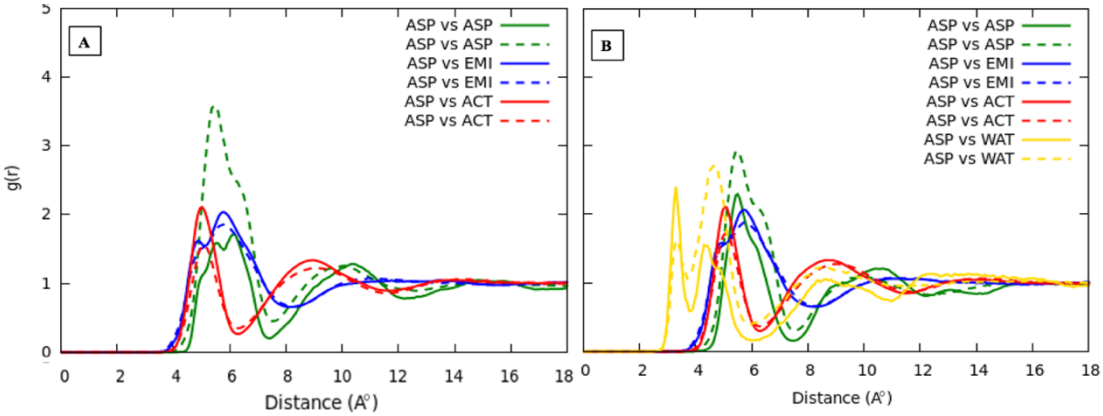


Figure A.24. 1: Centre-of-mass radial distribution functions for interactions around aspartate molecule for (A) pure $[C_2C_1Im][CH_3CO_2]$ IL and (B) $[C_2C_1Im][CH_3CO_2]$ IL/water mixture at 300K (solid lines) *vs* 400K (dashed lines)

Appendix A.24. 2: Comparison for centre-of-mass radial distribution functions for interactions around central aspartate molecule for pure and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ IL

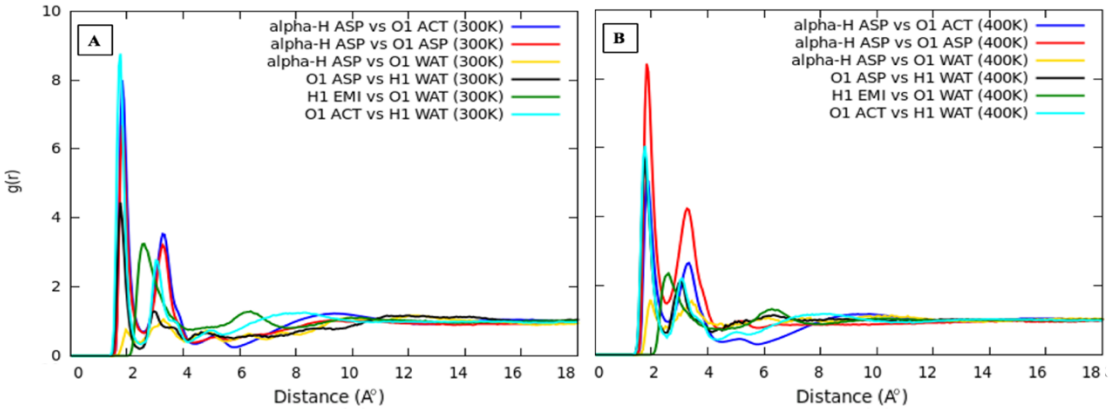


Figure A.24. 2: Centre-of-mass radial distribution functions for interactions around aspartate molecule for pure $[C_2C_1Im][CH_3CO_2]$ IL and $[C_2C_1Im][CH_3CO_2]$ IL/water mixture at (A) 300K *vs* (B) 400K

Appendix A.24. 3: Comparison for centre-of-mass radial distribution functions for interactions around [C_2C_1Im] cation for pure and [C_2C_1Im][CH_3CO_2] IL/water mixture

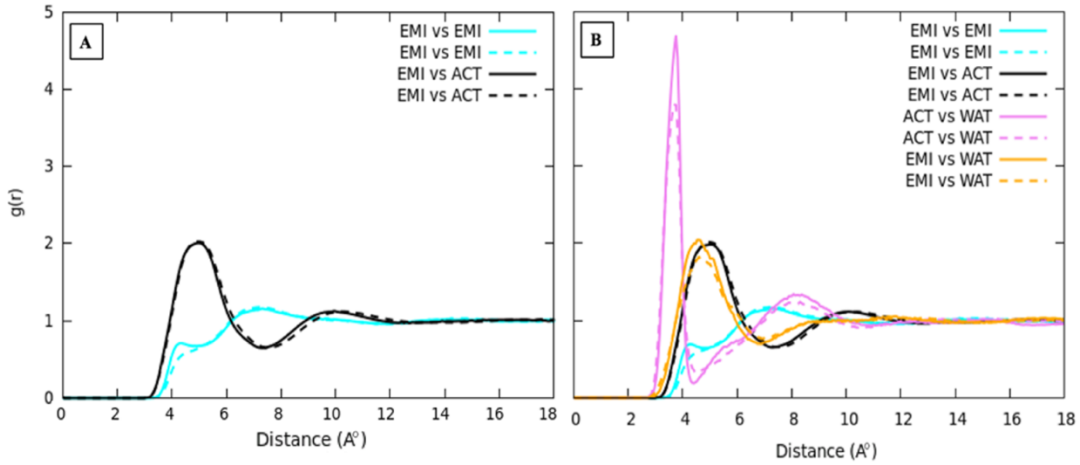


Figure A.24. 3: Centre-of-mass radial distribution functions for interactions around cation for (A) pure [C_2C_1Im][CH_3CO_2] IL and (B) [C_2C_1Im][CH_3CO_2] IL/water mixture at 300K (solid lines) *vs* 400K (dashed lines)

Appendix A.24. 4: Snapshots of aspartate high concentration systems in pure [C_2C_1Im][CH_3CO_2] IL and IL/water mix respectively at 300K and 400K

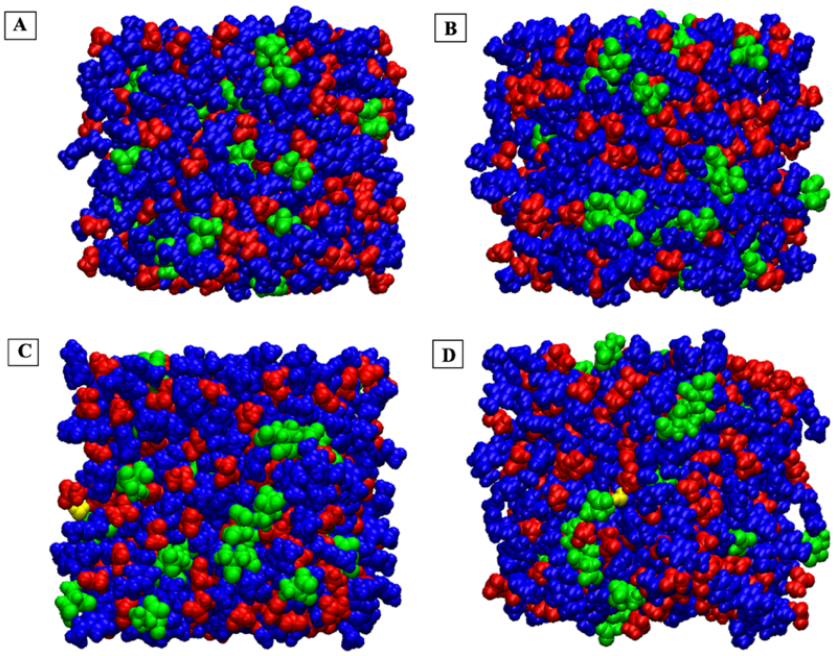


Figure A.24. 4: Snapshots of aspartate high concentration systems in [C_2C_1Im][CH_3CO_2] IL after 100 ns simulation for (A,C) 300K for pure and IL/water mix respectively (B,D) 400K for pure and IL/water mix respectively

Appendix A.24. 5: Self-Diffusion coefficients values calculated from 100ns simulation for aspartate in pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]$ /H_2O mixture at 300K and 400K

System		Self-Diffusion coefficient= slope * 10/6 = D * 1e-10 m^2/s			
		T=300K		T=400K	
		200:1	5:1	200:1	5:1
aspartate $[C_2C_1Im][CH_3CO_2]$	aspartate	0.84	0.34	4.34	2.06
	$[C_2C_1Im]$	1.19	0.36	8.00	4.76
	$[CH_3CO_2]$	1.13	0.34	7.14	4.28
aspartate $[C_2C_1Im][CH_3CO_2]$/wat	aspartate		2.16		0.93
	$[C_2C_1Im]$		2.22		3.31
	$[CH_3CO_2]$		2.19		2.6
	water		3.19		5.98

Appendix A. 25: Effect of temperature on interactions for alanine amino acid systems at 300K *vs* 400K in pure and IL/water mixture of $[C_2C_1Im][Cl]$

Appendix A.25. 1: Centre-of-mass radial distribution for overall interactions between solute and solvent molecules

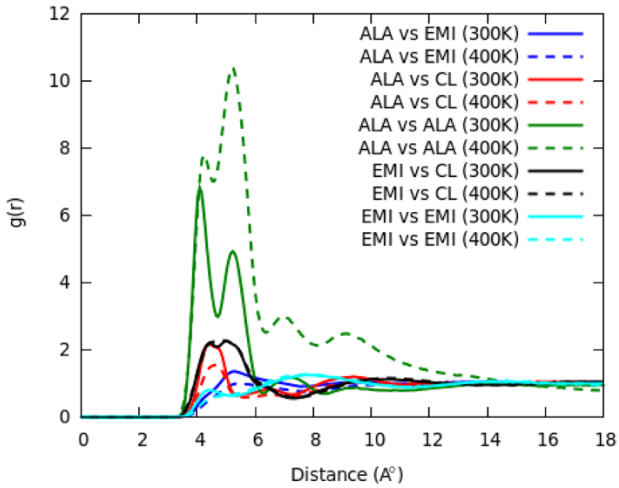


Figure A.25. 1: Centre-of-mass radial distribution functions for interactions around alanine molecule for pure $[C_2C_1Im][Cl]$ IL 300K (solid lines) *vs* 400K (dashed lines)

Appendix A.25. 2: Centre-of-mass radial distribution for interaction between atomic sites on [C_2C_1Im] cation and chloride anions

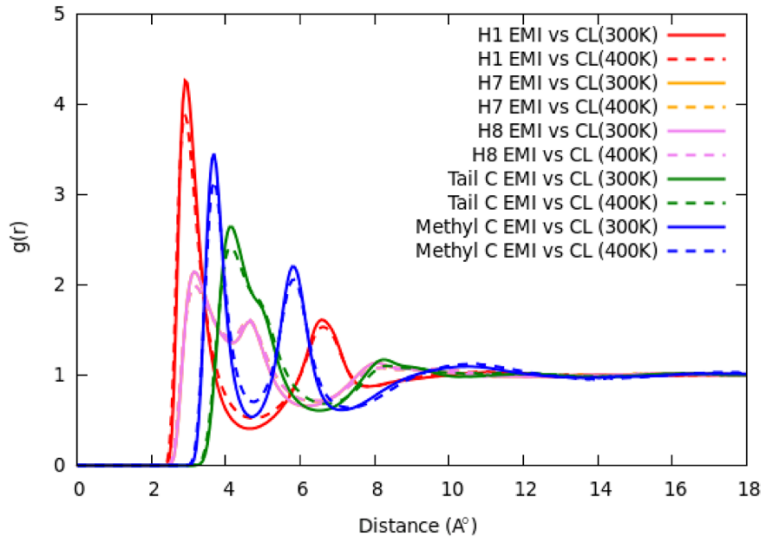


Figure A.25. 2: Centre-of-mass radial distribution for [C_2C_1Im][Cl]:ALA (5:1) systems for ring hydrogen H1,H7,H8 on [C_2C_1Im]$^+$ cation *vs* chloride anion at 300K (solid lines) *vs* 400K (dashed lines)

Appendix A.25. 3: Comparison of solute-solvent, solute-solute and solvent-solvent interactions for alanine systems in pure [C_2C_1Im][Cl] IL and [C_2C_1Im][Cl]/water mixture

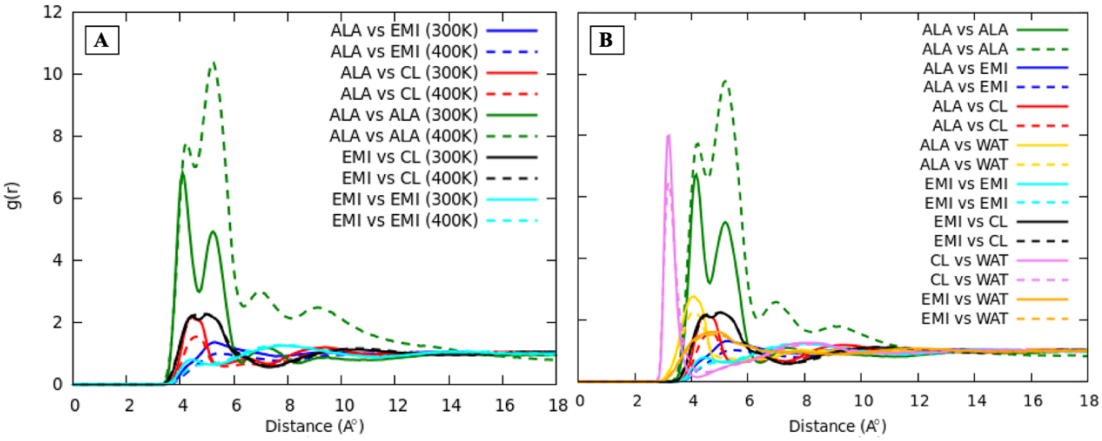


Figure A.25. 3: Centre-of-mass radial distribution functions for interactions among solute and solvent for [C_2C_1Im][Cl]:ALA (5:1) at 300K (solid lines) and 400K (dashed lines) for (A) pure IL and (B) IL/water mixture.

Appendix A.25. 4: Comparison of solute-solvent, solute-solute and solvent-solvent interactions for pure $[C_2C_1Im][Cl]$ IL and $[C_2C_1Im][Cl]$/water mixture

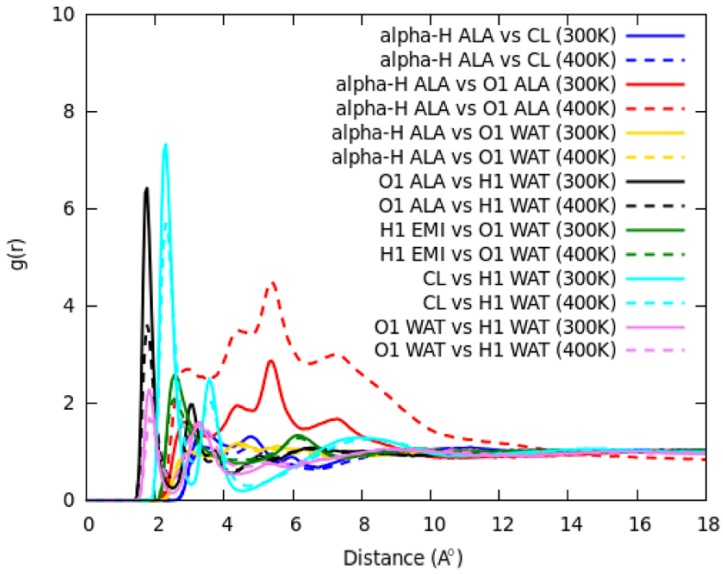


Figure A.25. 4: Centre-of-mass radial distribution functions between lysine, $[Cl]^-$ anion, $[C_2C_1Im]^+$ cation and water atomic sites for alanine systems in $[C_2C_1Im][Cl]/H_2O$ mixture at 300K (solid lines) *vs* 400K (dashed lines)

Appendix A.25. 5: Centre-of-mass radial distribution for interaction between atomic sites on $[C_2C_1Im]^+$ cation and chloride anions in pure *vs* IL/water mixture of $[C_2C_1Im][Cl]$ IL

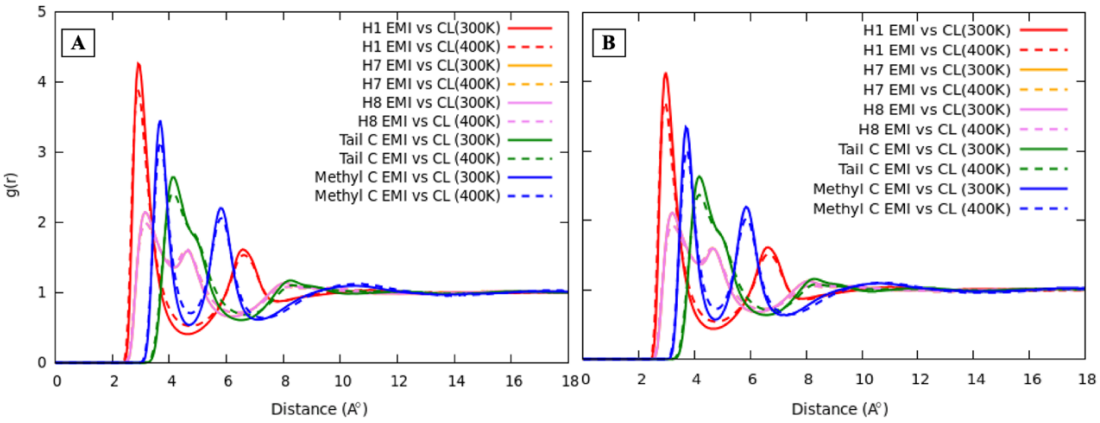


Figure A.25. 5: Centre-of-mass radial distribution for $[C_2C_1Im][Cl]$:ALA (5:1) systems for ring hydrogen H1,H7,H8 on $[C_2C_1Im]^+$ cation *vs* chloride anion at 300K (solid lines) *vs* 400K (dashed lines) for (A) pure IL and (B) IL/water mixture

Appendix A.25. 6: Comparison of centre-of-mass radial distribution for overall interactions between solute and solvent molecules for alanine systems in pure and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ IL

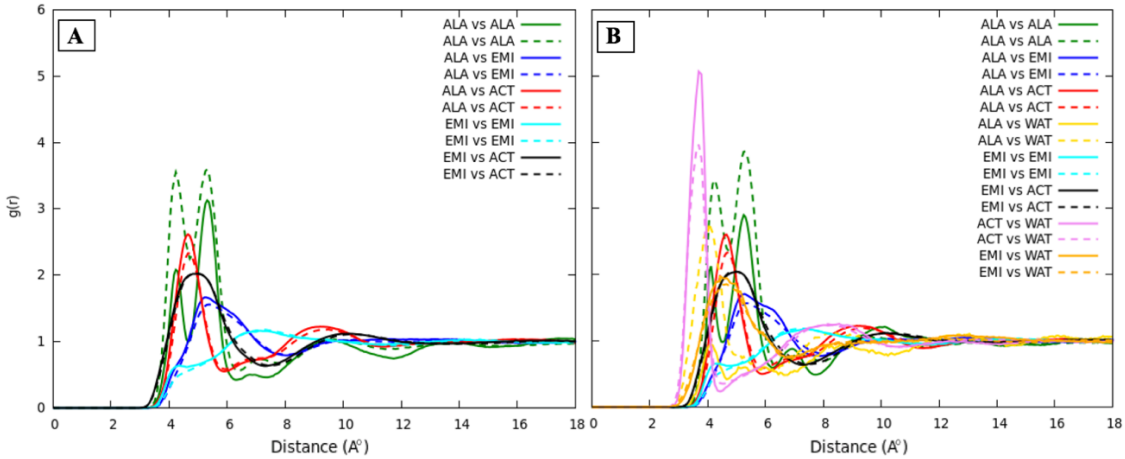


Figure A.25. 6: Centre-of-mass radial distribution functions for interactions among solute and solvent for $[C_2C_1Im][CH_3CO_2]$:ALA (5:1) at 300K (solid lines) and 400K (dashed lines) for (A) pure IL and (B) IL/water mixture.

Appendix A.25. 7: Comparison of distribution of acetate anions and $[C_2C_1Im]^+$ cations around alanine atomic sites for pure IL and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ as a function of temperature

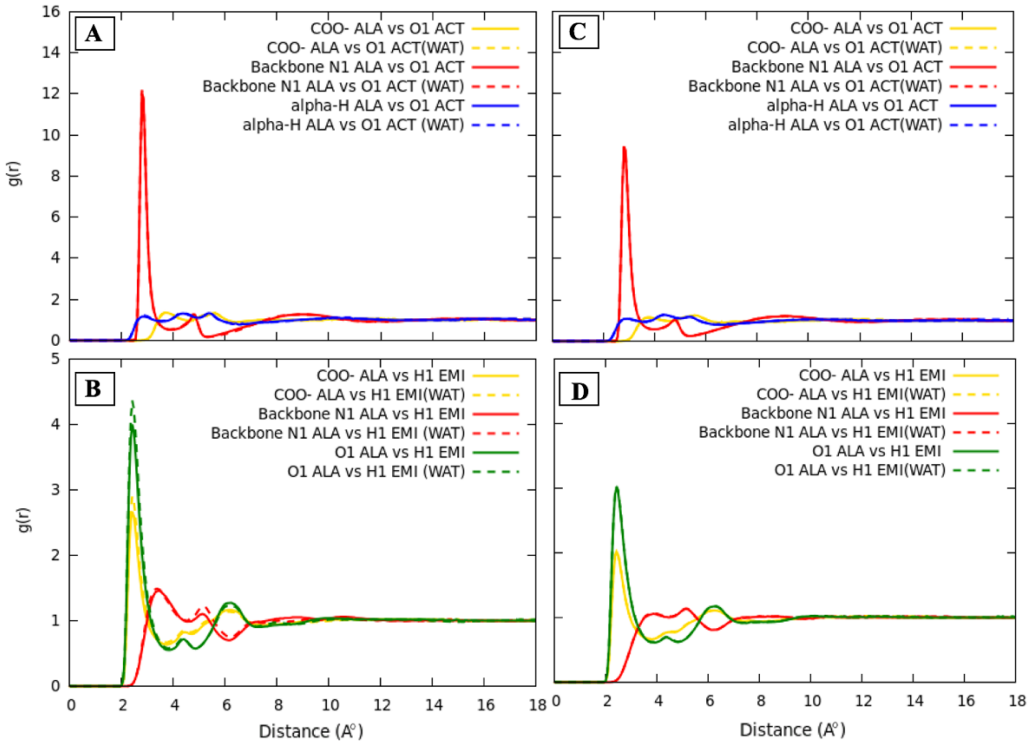


Figure A.25. 7: Centre-of-mass radial distribution functions for $[C_2C_1Im][CH_3CO_2]$: ala (5:1) for (A,C) acetate anion (O1 atom) distribution around backbone nitrogen, carboxyl group and alpha-hydrogen at 300K and 400K respectively (B) H1 $[C_2C_1Im]^+$ distribution around backbone nitrogen, carboxyl groups and O1 cation at 300K and 400K respectively, determined by MD simulation

Appendix A.25. 8: Snapshots for alanine systems in pure IL and IL/water mixture of $[C_2C_1Im][CH_3CO_2]$ at 300K *vs* 400K

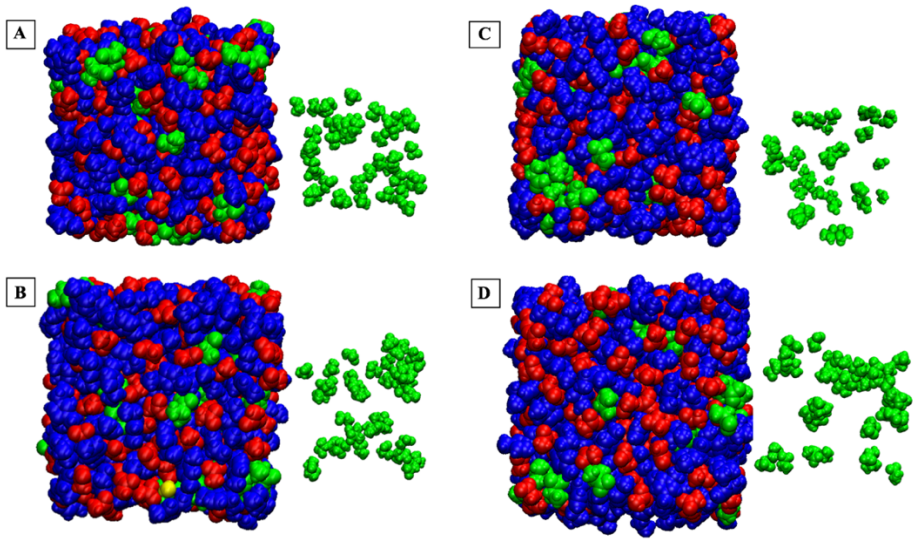


Figure A.25. 8: Snapshots for alanine systems in (A,C) pure $[C_2C_1Im][CH_3CO_2]$ IL at 300K *vs* 400K and in (B,D) $[C_2C_1Im][CH_3CO_2]$/water mixture at 300K *vs* 400K respectively from 100ns MD simulations

Appendix A.25. 9: log-log plots for MSD curves for the calculation of diffusive regime for alanine systems in pure $[C_2C_1Im][Cl]$, $[C_2C_1Im][Cl]/H_2O$ mixture, pure $[C_2C_1Im][CH_3CO_2]$ and $[C_2C_1Im][CH_3CO_2]/H_2O$ mixture at 400K

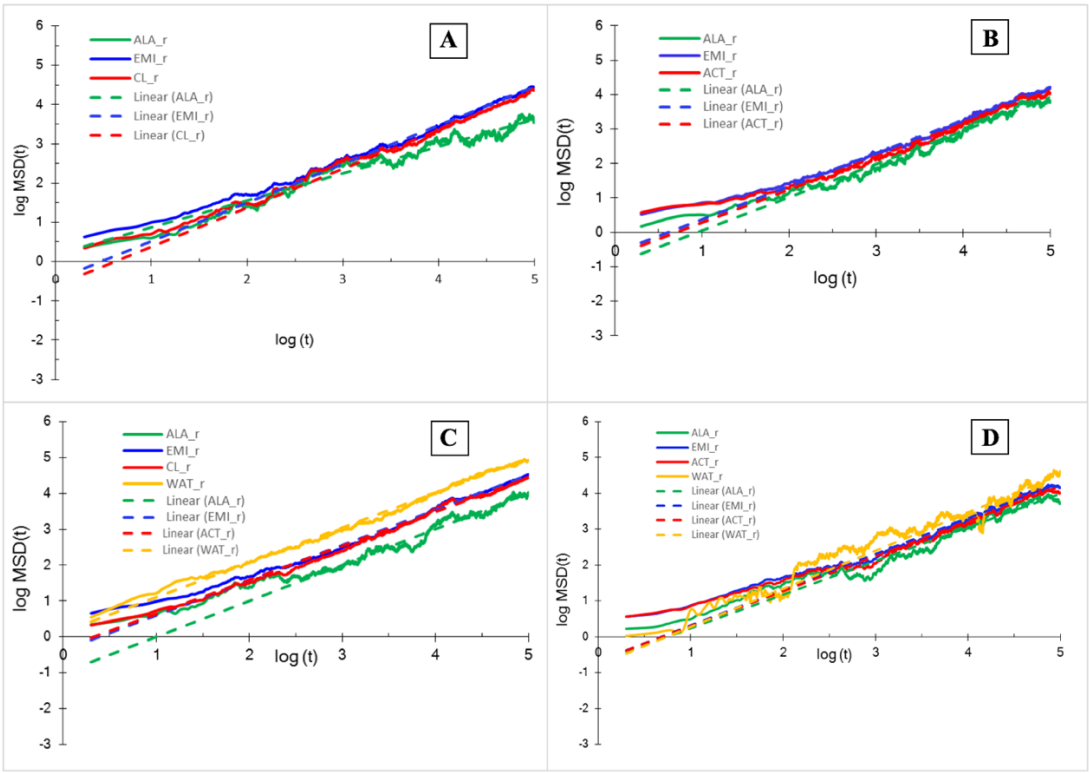


Figure A.25. 9: log-log plots for IL:AA 5:1 (A) pure $[C_2C_1Im][Cl]$, (B) pure $[C_2C_1Im][CH_3CO_2]$ IL (C) $[C_2C_1Im][Cl]$/water mixture, (D) $[C_2C_1Im][CH_3CO_2]$/water mixture at 400K

Appendix A.25. 10: Effect of temperature on the cation and anion interactions with histidine molecule

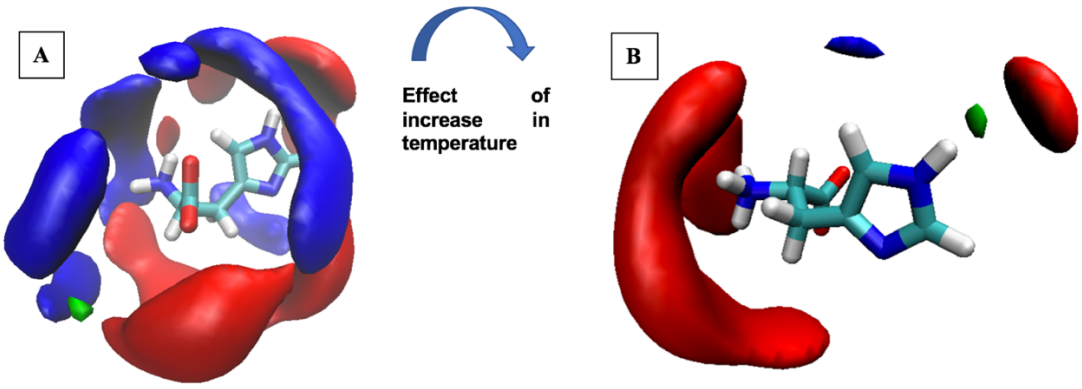


Figure A.25. 10: Spatial distribution functions around the central histidine in pure $[C_2C_1Im][CH_3CO_2]$ IL at (A) 300K (B) 400K in a view to keep the backbone carbon pointing downwards. Blue= $[C_2C_1Im]^+$, red= $[CH_3CO_2]^-$ and green= histidine at iso-value=10 nm^{-3}

Appendix A.25. 11: Cation-anion and cation-cation interactions for aromatic amino acids systems in $[C_2C_1Im][CH_3CO_2]$ IL

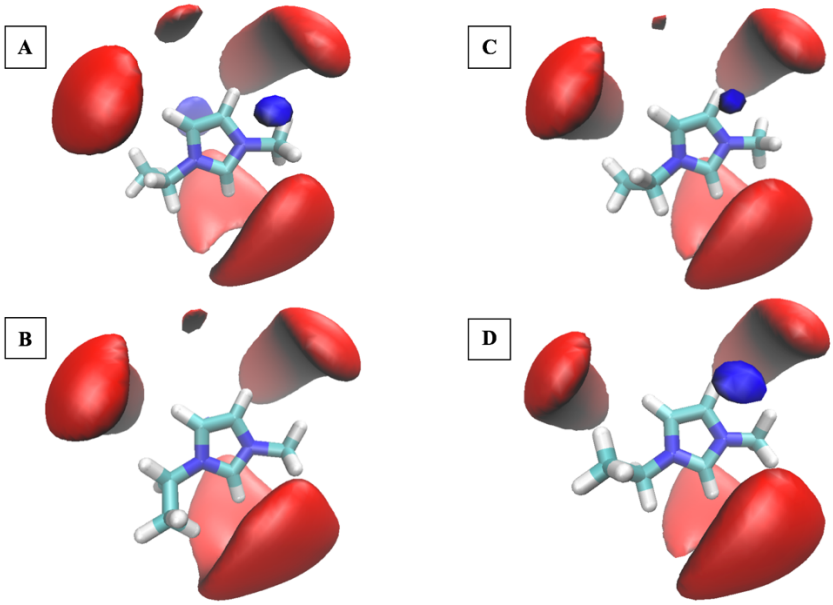


Figure A.25. 11: Spatial distribution functions around the central $[C_2C_1Im]^+$ cation for aromatic systems in pure $[C_2C_1Im][CH_3CO_2]$ IL for A) histidine (B) phenyl-alanine (C) tryptophan and (D) tyrosine at 300K in a view to keep ring hydrogen H1 pointing downwards. Blue= $[C_2C_1Im]^+$ and red= $[CH_3CO_2]^-$at iso-value=10 nm^{-3}

Appendix A.25. 12: MSD curves for aromatics systems in pure $[C_2C_1Im][CH_3CO_2]$ IL at 300K

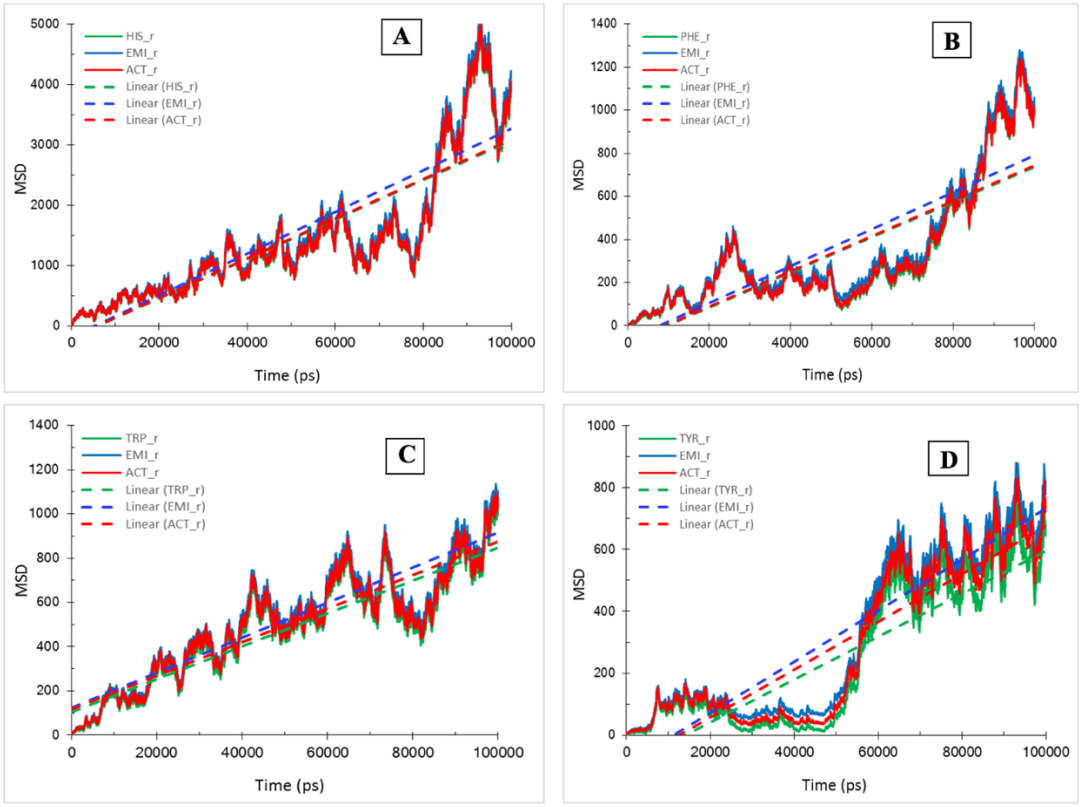


Figure A.25. 12: MSD curves for aromatic amino acids in pure $[C_2C_1Im][CH_3CO_2]$ IL for (A) histidine, (B) phenyl-alanine, (C) tryptophan (D) tyrosine at 300K

Appendix A. 26: Peptide simulations in water and $[C_2C_1Im][CH_3CO_2]$ ionic liquid

Appendix A.26. 1: Input file for simulation of peptide sequence from Protein G in water

The steps followed to generate box with water for sequence from protein G are shown below,

```
source
/media/RAIDstorage/software/amber14/dat/leap/cmd/oldff/leaprc.ff99SB
MOL= sequence { ACE LEU ASN GLY LYS THR LEU LYS GLY NME }
edit MOL
charge MOL
addions MOL Cl– 2
check MOL
loadamberparams frcmod.ionsjc_tip3p
saveamberparm MOL MOL-nowat.prmtop MOL-nowat.inpcrd
solvateoct MOL TIP3PBOX 10
savepdb MOL MOL-wat.pdb
saveamberparm MOL MOL-wat.prmtop MOL-wat.inpcrd
```

The value of 10 above in the commands show that a 10 Å buffer of water is created around the peptide solute. Libraries were saved for each peptide fragment by following the above procedure and corresponding pdb structures and topology files were generated.

The steps followed to generate box with $[C_2C_1Im][CH_3CO_2]$ for sequence from protein G are shown below,

```
source
/media/RAIDstorage/software/amber14/dat/leap/cmd/oldff/leaprc.ff99SB
source leaprc.gaff
loadamberparams frcmod.ionsjc_spce
loadoff ACT.lib
loadoff EMI.lib
loadoff MOL.lib
loadamberparams emim.frcmod
loadamberparams acetate.frcmod
box= loadpdb box-ALA10-IL.pdb
setbox box centers
saveoff box box.lib
savepdb box box.pdb
saveamberparm box box.prmtop box.inpcrd
```

Appendix A.26. 2: Ramachandran plots comparison for penta alanine peptide in $[C_2C_1Im][CH_3CO_2]$ IL *vs* Water

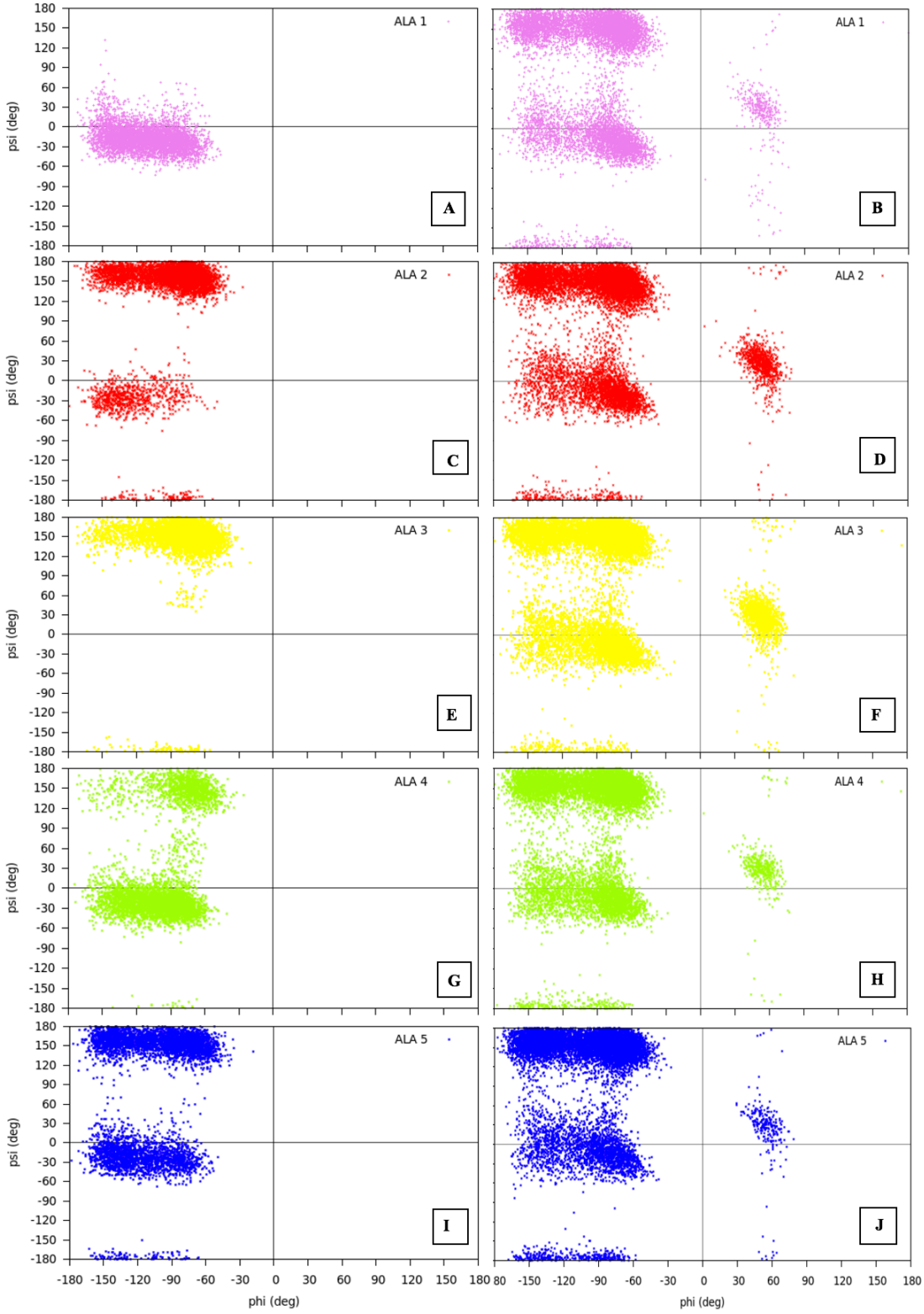


Figure A.26. 1: Ramachandran plots for penta alanine peptide. Residue number wise comparison (A,C,E,G,I) in $[C_2C_1Im][CH_3CO_2]$ IL *vs* (B,D,F,H,J) in water at 300K

Appendix A.26. 3: Ramachandran plots comparison for deca alanine peptide in $[C_2C_1Im][CH_3CO_2]$ IL *vs* water

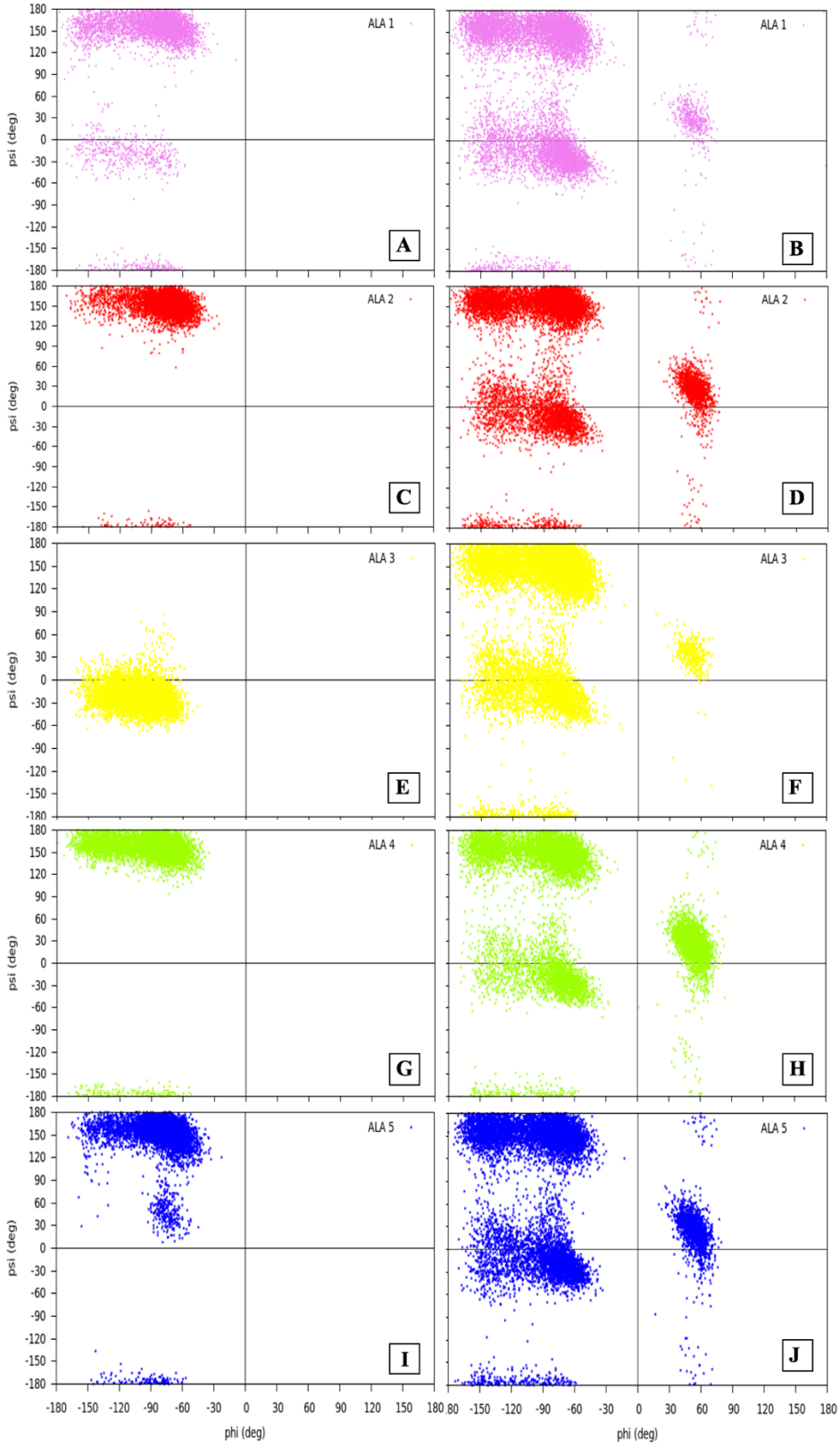


Figure A.26. 2: Ramachandran plots for deca alanine peptide. Residue number wise comparison (A,C,E,G,I) in $[C_2C_1Im][CH_3CO_2]$ IL *vs* (B,D,F,H,J) in water at 300K

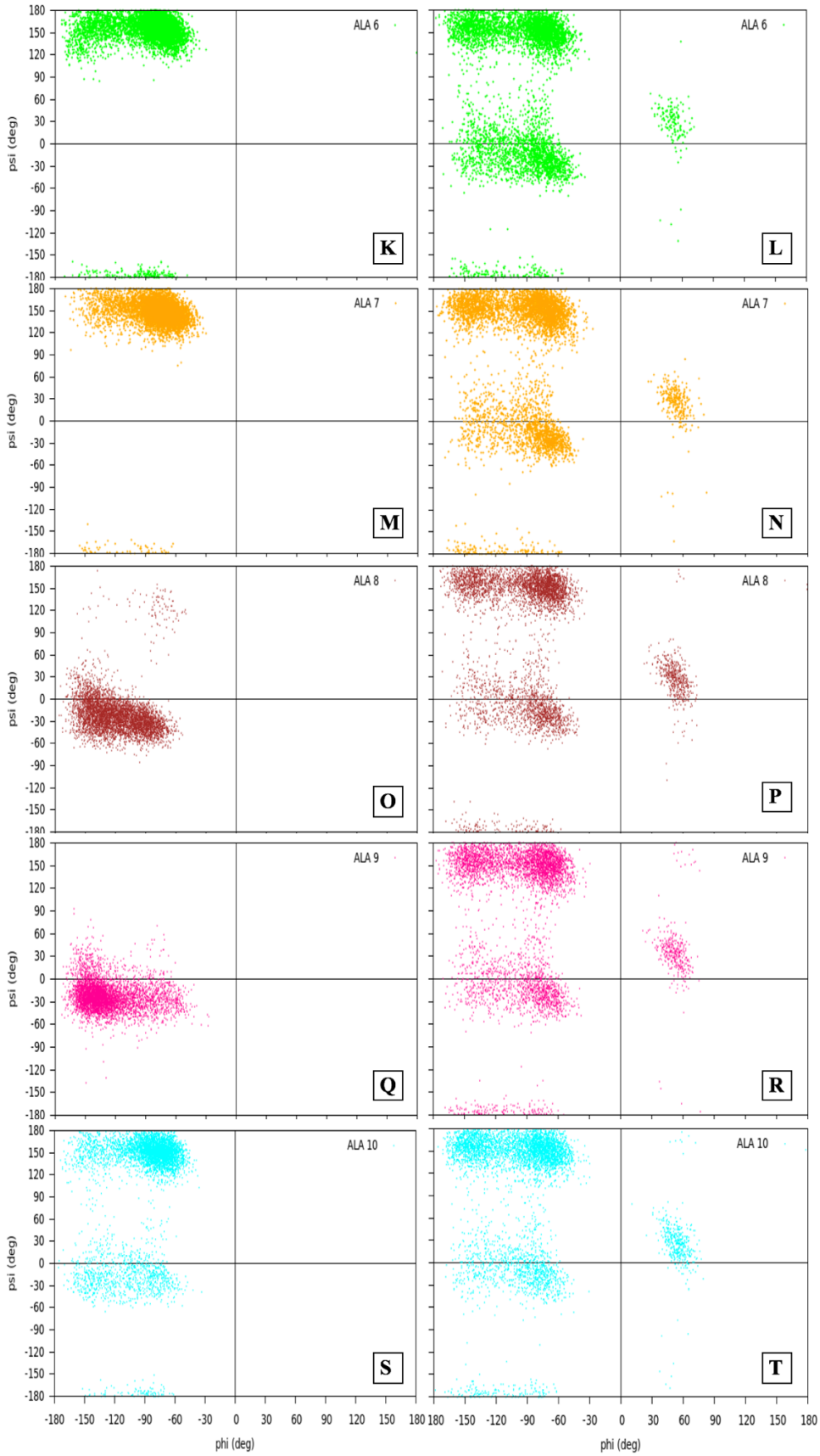


Figure A.26. 3: Ramachandran plots for deca alanine peptide. Residue number wise comparison (K,M,O,Q,S) in $[C_2C_1Im][CH_3CO_2]$ IL *vs* (L,N,P,R,T) in water at 300K

Appendix A.26. 4: Radial distribution functions for cation and anion distribution around backbone atoms of residues in protein G sequence

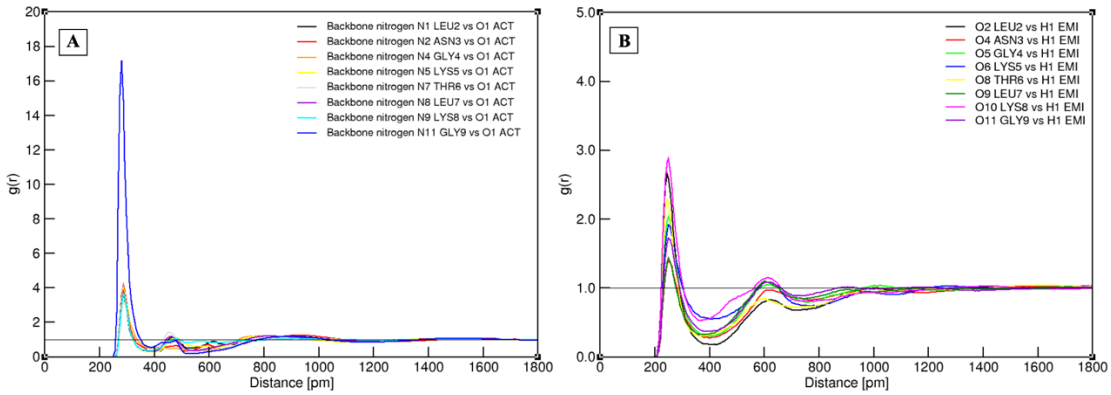


Figure A.26. 4: Centre-of-mass radial distribution functions for (A) acetate anion (O1 atom) around backbone nitrogen atoms and (B) H1 atom on $[C_2C_1Im]^+$ cation around oxygen atoms on alanine residues in $(ALA)_{10}$ peptide in pure $[C_2C_1Im][CH_3CO_2]$ IL

Appendix A.26. 5: Ramachandran plots and secondary structures comparison for peptide segment of protein G in $[C_2C_1Im][CH_3CO_2]$ IL *vs* water

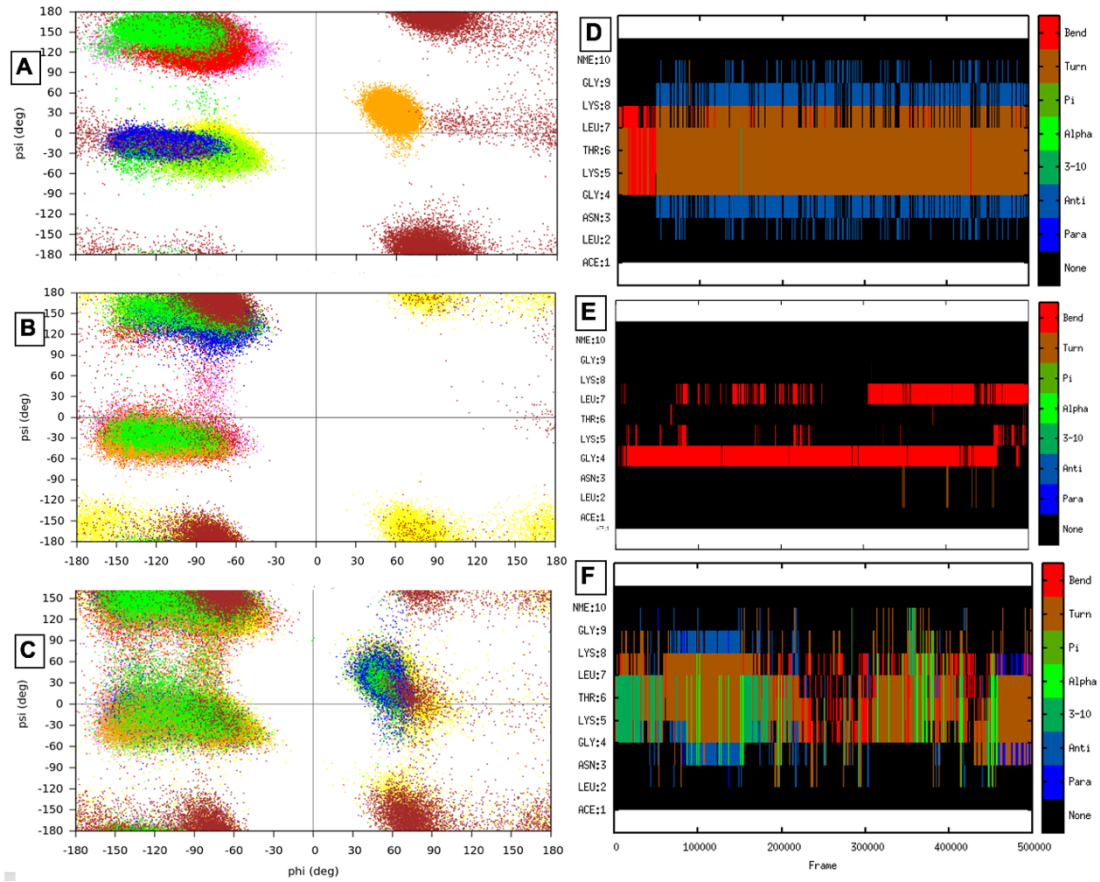


Figure A.26. 5: Ramachandran plots and secondary structures for protein G peptide sequence respectively, in (A,D) water and then in $[C_2C_1Im][CH_3CO_2]$ IL, (B,E) pure $[C_2C_1Im][CH_3CO_2]$ IL and (C,F) water.

Appendix A.26. 6: Ramachandran plots comparison for peptide segment of protein G in $[C_2C_1Im][CH_3CO_2]$ IL *vs* water

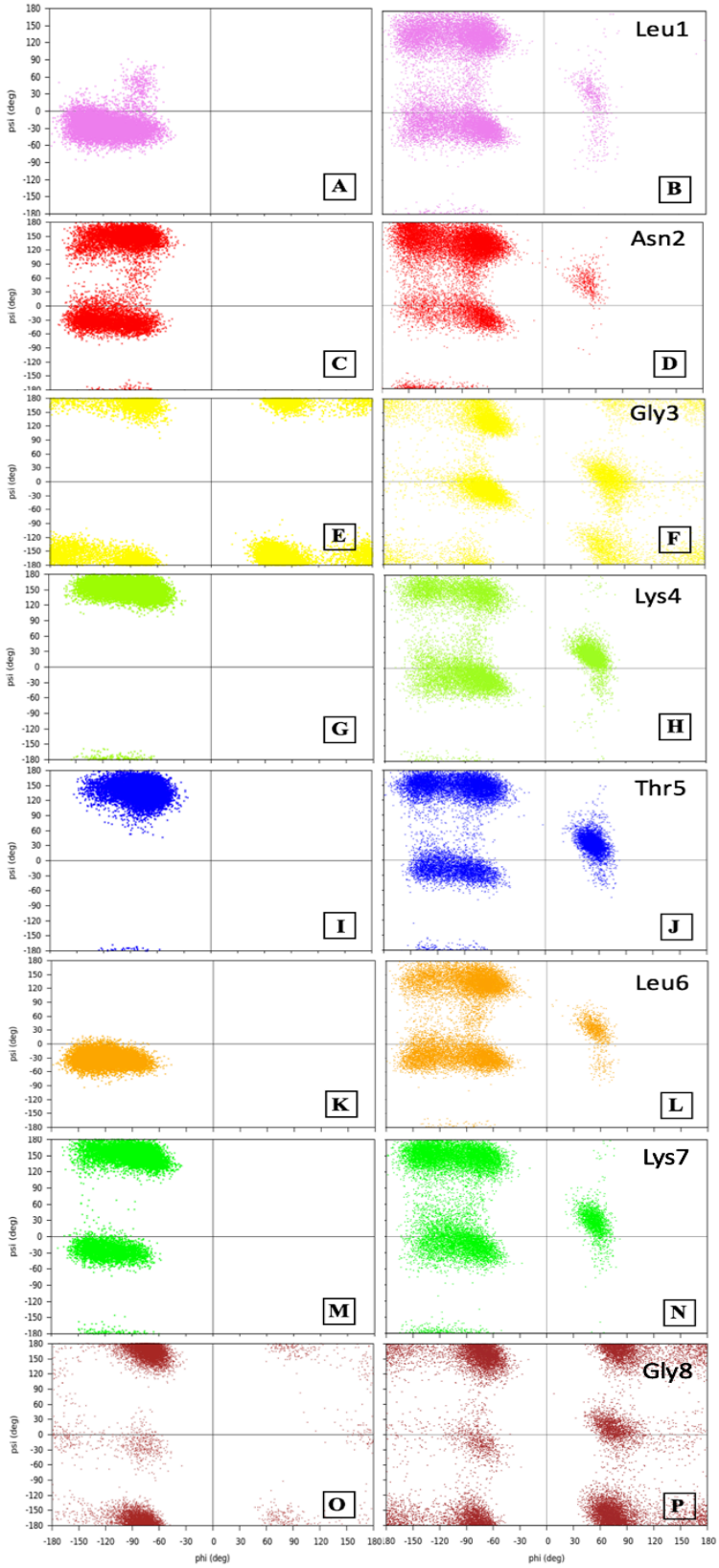


Figure A.26. 6:Ramachandran plots for protein G peptide sequence. Residue number wise comparison (A,C,E,G,I,K,M,O) in $[C_2C_1Im][CH_3CO_2]$ IL *vs* (B,D,F,H,J,L,N,P) in water at 300K