

Halophilic Enzymes in Alien Environments

Alexandra Sonja Schindl

14263619

A thesis submitted to the University of Nottingham

for the degree of Doctorate of Philosophy

May , 2020

Contents

Abstract **VII**

Acknowledgements **VIII**

1 Introduction **1**

1.1 Industrial Enzymatic Synthesis and Reaction Engineering 1

1.2 Extremozymes 6

1.2.1 Thermophilic Proteins 10

1.2.2 Halophilic Proteins 13

1.2.3 Haloalkaliphilic Proteins 21

1.3 Archaea 22

1.3.1 *Haloferax volcanii* 25

1.4 Alcohol Dehydrogenases 28

1.4.1 *Haloferax volcanii* Alcohol Dehydrogenase 2 31

1.5 Ionic Liquids in Green Chemistry 34

1.5.1 Effects of Solvation of Ionic Liquid Ions on Proteins 36

1.5.2 Alcohol Dehydrogenases in Ionic Liquids 40

Hydrophilic ionic liquids 40

Water-immiscible ionic liquids (biphasic systems) 45

Water-miscible, amphiphilic ionic liquids 45

1.5.3 Molecular Dynamics Simulation 46

1.6 Nanoreactors 48

1.6.1 Micelles 49

1.7 Thesis Outline and Aims 51

2 Plasmid and Strain Construction **53**

2.1 Background . . . 53

2.2 Materials and Methods . . . 55

2.2.1 Materials . . . 55

Oligonucleotides . . . 55

Plasmids . . . 56

Haloferax volcanii strains . . . 58

Escherichia coli strains . . . 60

Chemicals and Enzymes . . . 60

Buffers and Solutions . . . 64

2.2.2 Methods . . . 65

DNA and Plasmid preparation . . . 65

Microbiological Methods for *Escherichia coli* . . . 68

Microbiological Methods for *Haloferax volcanii* . . . 69

Methodology for Gene Deletion in *Haloferax volcanii* . . . 72

Methodology for Gene Overexpression in *Haloferax volcanii* . . . 75

2.3 Results . . . 76

2.3.1 Plasmid Construction . . . 76

Genomic Clones and Deletion Constructs . . . 76

Episomal Plasmids for Overexpression of Tagged Proteins . . . 79

Episomal Plasmids for the Expression of β-galactosidase . . . 82

2.3.2 Strain Construction . . . 82

Masterstrain for Protein Overexpression . . . 84

tnaA-deleted Strains . . . 87

2.4 Discussion . . . 89

2.5 Conclusion and Future Perspectives . . . 89

3 Optimisation of Expression and Purification Conditions **91**

3.1 Background . . . 91

3.2 Materials and Methods . . . 94

3.2.1 Materials . . . 94

Haloferax volcanii strains . . . 94

Chemicals, Buffers and Media . . . 95
3.2.2 Methods . . . 99
Protein expression and purification . . . 99
Protein Characterisation Methods . . . 101
3.3 Results . . . 105
3.3.1 Analysis of Expression Constructs . . . 105
Purifications by Gravity Column . . . 105
Purifications by äkta Prime Plus . . . 108
3.3.2 Analysis of Tryptophanase A Deletion . . . 111
Blue-white Screening . . . 111
ONPG-Assay . . . 112
Comparison of Elution-profiles . . . 113
3.4 Discussion . . . 115
3.5 Conclusion and Future Perspectives . . . 117

4 Assays of ADH2 in Ionic Liquids 119
4.1 Background . . . 119
4.2 Material and Methods . . . 122
4.2.1 Material . . . 122
Ionic Liquids . . . 122
Preparation of Ionic Liquid Mixtures . . . 124
ADH2 Storage and Measurement Buffers and Solutions . . . 125
4.2.2 Methods . . . 126
Material for YADH assay . . . 127
4.3 Results . . . 127
4.3.1 Screening Activity of ADH2 in Ionic Liquid Mixtures . . . 127
4.3.2 Detailed Investigation of IL K . . . 128
Investigation of Solubility of NADPH . . . 128
Investigation of absorption curve . . . 130
4.3.3 Screening Activity of ADH2 in IL A, K and M . . . 130
4.3.4 Long-term stability of *Hv*ADH2 . . . 131
Long-term stability of ADH2 under storage conditions . . . 131

Long-term stability of ADH2 in IL A, K and M at RT 132
Influence of Zinc and Molarity on HvADH2 133
4.3.5 Yeast alcohol dehydrogenase in IL K 133
4.4 Discussion 134
4.5 Conclusion and Outlook 137

5 Molecular Dynamics Simulations 138
5.1 Background 138
5.2 Theoretical Background 141
5.2.1 Molecular Dynamics Simulations 141
5.2.2 Newton's Equation of Motion 141
5.2.3 Constant Temperature and Constant Pressure Dynamics 143
5.2.4 Particle-Mesh Ewald Electrostatics 145
5.2.5 Solvent Treatment 145
Implicit Solvent 146
Explicit Solvent 146
5.2.6 Periodic Boundary Conditions 147
5.2.7 Classical Force Fields 148
5.3 Applied Methods 150
5.3.1 Molecule number calculations for Packmol 153
5.3.2 Comparative Modelling 154
Sequence Alignment 156
Building the Model 157
Model Evaluation 163
5.3.3 Analytical Methods 164
5.4 Results 169
5.4.1 Obtaining the Homology Model 169
Swiss-Model 170
Phyre2 171
I-Tasser 172
Geometry Optimisation of the Tetramer 173
Verification of the Homology Model of the Tetramer 174

5.4.2 Coordination of ions . . . 175
Coordination of K^+ and Cl^- ions to the protein surface . . . 176
Coordination of ionic liquid ions to the protein surface . . . 177
5.4.3 Structural changes of *Hv*ADH2 in ionic liquid mixtures . . . 181
5.4.4 Docking into Lowest Energy Conformations . . . 187
5.5 Discussion . . . 190
5.6 Conclusion and Outlook . . . 197

6 Microcapsule Formation and Enzyme Encapsulation 199
6.1 Background . . . 199
6.2 Material and Methods . . . 201
Chemicals and Material . . . 201
Experimental . . . 201
6.2.1 Scanning Electron Microscopy . . . 202
6.2.2 Optical Microscopy . . . 202
6.2.3 Spectrophotometry . . . 202
6.3 Results . . . 203
6.3.1 Investigation of the Compostion of Microcapsules . . . 203
6.3.2 Structural Integrity of Microcapsules . . . 205
6.3.3 Activity Assays . . . 207
Activity assays of ADH2 in microcaspules prepared by evaporation 207
6.4 Discussion . . . 209
6.5 Conclusion and Outlook . . . 210

7 Conclusion and Future Perspectives 211
7.1 Concluding Remarks . . . 213

References 214

8 Appendix ii
8.1 Free Energy Landscapes at 0.15 M Ionic liquid . . . ii
8.1.1 Unscaled Free Energy Landscapes . . . iv
8.1.2 Free Energy Landscapes - Comparison of Concentrations . . . v

8.2 Relative Specific Activity Assays ILA - ILM vi
8.3 Relative Specific Activity Assays ILA, ILK and ILM xiii
8.4 RDFs of Potassium and Chloride in ILs xv
8.5 RMSDs of 0.15M ILC - ILJ, ILL xvi
8.6 Scripts for MD Simulations xix
8.7 Scripts for processing gromacs output xxiv

Abstract

Alcohol dehydrogenases catalyse the asymmetric reduction of ketones, yielding highly desirable pharmaceutical building blocks. Ketones are insufficiently water-soluble and catalysis provides meaningful yields only in non-aqueous media. Halophilic ADH2 from *Haloferax volcanii* is tolerant of organic solvents and this study targets the identification of ADH2 tolerance of ionic liquids. A proof of concept for creating a microenvironment through encapsulation is explored.

Genetic modification of the model organism *Haloferax volcanii* were achieved through and verified with standard genetic engineering procedures, including the use of auxotrophic selection markers for transformation. The overexpression and purification protocol was optimised to yield the highly active tetrameric form of the enzyme, confirming the importance of the quaternary organisation in halophilic proteins for enzymatic activity.
The protein was experimentally assayed in aqueous ionic liquid mixtures.Thirteen ionic liquid ion combinations were chosen to cover a broad diversity of electronic and H-bonding natures. The best system showed a 205 % increase in relative activity compared to native conditions.
A homology model of the ADH2 tetramer was built and MD simulations revealed ionic liquid induced changes in the secondary and tertiary structure. Specifically, coordination of ionic species disrupts secondary structural elements that are required for switching between the open and closed conformation of the enzyme.
Microcapsules were produced using solvent evaporation. Enzymatic activity was observed, however, the success of enzyme encapsulation could not be conclusively confirmed or dismissed.

Acknowledgements

Gotham. Maybe we should all pretend to be cuckoo from time to time. Thanks for sharing such amazing and nifty stories, Thorsten. And of course thank you for providing such a great lab, and SOPs for every technique, so thought-through and meticulous, that it made lab life feel like a walk on a sunny afternoon. Thanks for everything!
A huge thank you, Anna, for just the right amount of guidance I required, and with this, helping me restore my self-confidence, juuust a little bit. Because of this, you made it possible for me to see the wider picture and relevant scope of this work. Thanks!
Thank you Mischa for all the support and guidance and taking your time whenever I needed help and had questions. Thanks a lot!

I also want to thank all the people from the lab groups I was part of. Christof for instructing me on the modelling stuff, working together on the paper! was a pleasure. Athina and Shafaq for sharing their stories and their support and Damiano for having fun working together on his campervan. Valentine for working together and having fun during demonstrating for undergraduates and Ricardo for sharing and exchanging experience on MD simulations.
I want to thank Emanuele, Rebeca and Rachel, who were really supportive in finding my way around in boots. Happy to have them as friends and colleagues.
Many thanks to Laura! Further thanks go out to Adam, James, Sarah, Tom, Hannah, Rob, Lorenzo and all the people in the lab in QMC. Special thanks go to Hannah, Rebecca, Beccy, Patricia, Viccy, Chris and Ambika for all the help, pranks and memories in the lab.

Further special thanks go to all the people, whos friendships I take away with me from Nottingham and whom I am very grateful for the time we've spent together and with the adventures we had, helping me keep my sanity in the most stressiest times of my PhD; namely, Nathan, Mat, Ben and Irene, Sam and Steve, Steve Baileys, Tom, Zoe, Andreas, James and Jesse and Josh.
Most of all I want to thank Rupert, who endured my emotional hick-ups during writing up and was really, really supportiv. Danke du rattigste aller Ratten.
Finally, I want to thank my Dad, who is always there for me.

Sometimes I think it is a great mistake to have matter that can think and feel. It complains so. By the same token, though, I suppose that boulders and mountains and moons could be accused of being a little too phlegmatic.

- The Sirens of Titan -

Kurt Vonnegut

List of Abbreviations, Acronyms and Initialisations

5-FOA	5-fluoroorotic Acid
Å	Ångstrom
ADH	Alcohol dehydrogenase
Amp	Ampicillin
ATP	Adenosine 5'triphosphate
BCA	Bradford concentration assay
bp	Base pair
CMC	Critical micelle concentration
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
dNTP	Deoxynucleotide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FEL	Free energy landscape
GDH	Glucose dehydrogenase
Hv	*Haloferax volcanii*
IL	Ionic liquid

kb	Kilobase
LB	Lysogeny broth
Leu	Leucine
Mb	Megabase
MD	Molecular dynamics
MDT	monomer-dimer-tetramer
MDTT	monomer-dimer-trimer-tetramer
NADP	Nicotinamide adenine dinucleotide phosphate
NBD	Nucleotide binding domain
ONPG	ortho-nitrophenyl-para-galactoside
PAGE	Polyacrylamide gel electrophoresis
PBC	Periodic boundary conditions
PCR	Polymerase chain reaction
pI	Isoelectric point
PMSF	Phenylmethanesulfonylfluoride
RDF	Radial distribution function
RESP	Restrained electrostatic potential
R_{gyr}	Radius of gyration
$r_{(g)}$	Radius of gyration
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SBD	Substrtae binding domain
SDS	Sodium dodecyl sulphate
SDSA	Synthesis-dependent strand annealing

ssDNA	Single-stranded deoxyribose nucleic acid
SSPE	Saline sodium phosphate EDTA
SW	Salt water
TAE	Tris/Acetic acid/EDTA
TBE	Tris/Borate/EDTA
TE	Tris/EDTA
TEMED	Tetramethylethylenediamine
Thy	Thymidine
T_m	Melting temperature
Trp	Tryptophan
Ura	Uracil
v/v	Volume per volume
w/v	Weight per volume
WT	Wild-type
X-gal	5-Bromo-4-chloro-3-indolyl--D-galactopyranoside
Δ	G Gibbs free energy
ω-TA	ω-Transaminase
ΔC_p	Heat capacity

Ionic Liquid Abbreviations

BF_4	Tetra-fluoro-borate
Br	Bromide
BMIM	Butyl-methyl-imidazolium
Cl	Chloride
DiMIM	1,3-Dimethylimidazolium

EMIM	Ethyl-methyl-imidazolium
EMP	Ethyl-methyl-phosphate
ES	Ethyl-Sulfate
I	Iodide
$MDEGSO_4$	Diethyleneglycolmonomethylether-Sulfate
MMIM	Methyl-methyl-imidazolium
Me_3S	Trimethylsulfonium
Me_3SO	Trimethylsulfoxonium
Me_4N	Tetra-methyl-ammonium
$MeOEtSO_4$	(2-methoxyethyl)sulfate
$MeSO_3$	Methanesulfonate
$MeSO_4$	Methylsulfate
$N_{1,1,8,8}$	N,N-dimethyl-N-octyloctan-1-aminium
$N_{4,1,1,1}$	N,N,N-trimethyl-1-Butanaminium
NTf_2	Bis(trifluoromethylsulfonyl)imide
OAc	Acetate
OSO_3Me	Methylsulfate
OTf	Trifluoromethanesulfonate
$P_{6,6,6,14}$	Trihexyltetradecylphosphonium
PF_6	Hexa-fluoro-phosphate
Tetrakis	Tetrakis(hydroxymethyl)phosphonium

Register of Ionic Liquids used for this thesis

Table 1: Ionic Liquids used in this Study

Name	**abridged**
Trimethylsulfonium Methylsulfate	IL A $[Me_3S][OSO_3Me]$
Tetrakis(hydroxymethyl)phosphonium Chloride	IL B [Tetrakis][Cl]
Choline Chloride	IL C [Choline][Cl]
Choline Bitartrate	IL D [Choline][Bitartrate]
1,3-Dimethylimidazolium Methylsulfate	IL E $[DiMIM][MeSO_4]$
Triethylsulfonium Bis(trifluoromethylsulfonyl)imide	IL F $[Me_3S][NTf_2]$
1-butyl-3-methylimidazolium (2-methoxyethyl)sulfate	IL G $[BMIM][MeOEtSO_4]$
N,N-dimethyl-N-octyloctan-1-aminium Methane sulfonate	IL H $[N_{1,1,8,8}][MeSO_3]$

Tetrabutylphosphonium	IL I
Chloride	$[P_{4,4,4,4}][Cl]$
N,N,N-trimethyl-1-Butanaminium	IL J
Dimethylphosphate	$[N_{4,1,1,1}][(MeO)_2(HO)PO]$
Trihexyltetradecylphosphonium	IL K
Bis(trifluoromethylsulfonyl)imide	$[P_{6,6,6,14}][NTf_2]$
Trimethylsulfoxonium	IL L
Iodide	$[Me_3SO][I]$
Trimethylsulfonium	IL M
Iodide	$[Me_3S][I]$

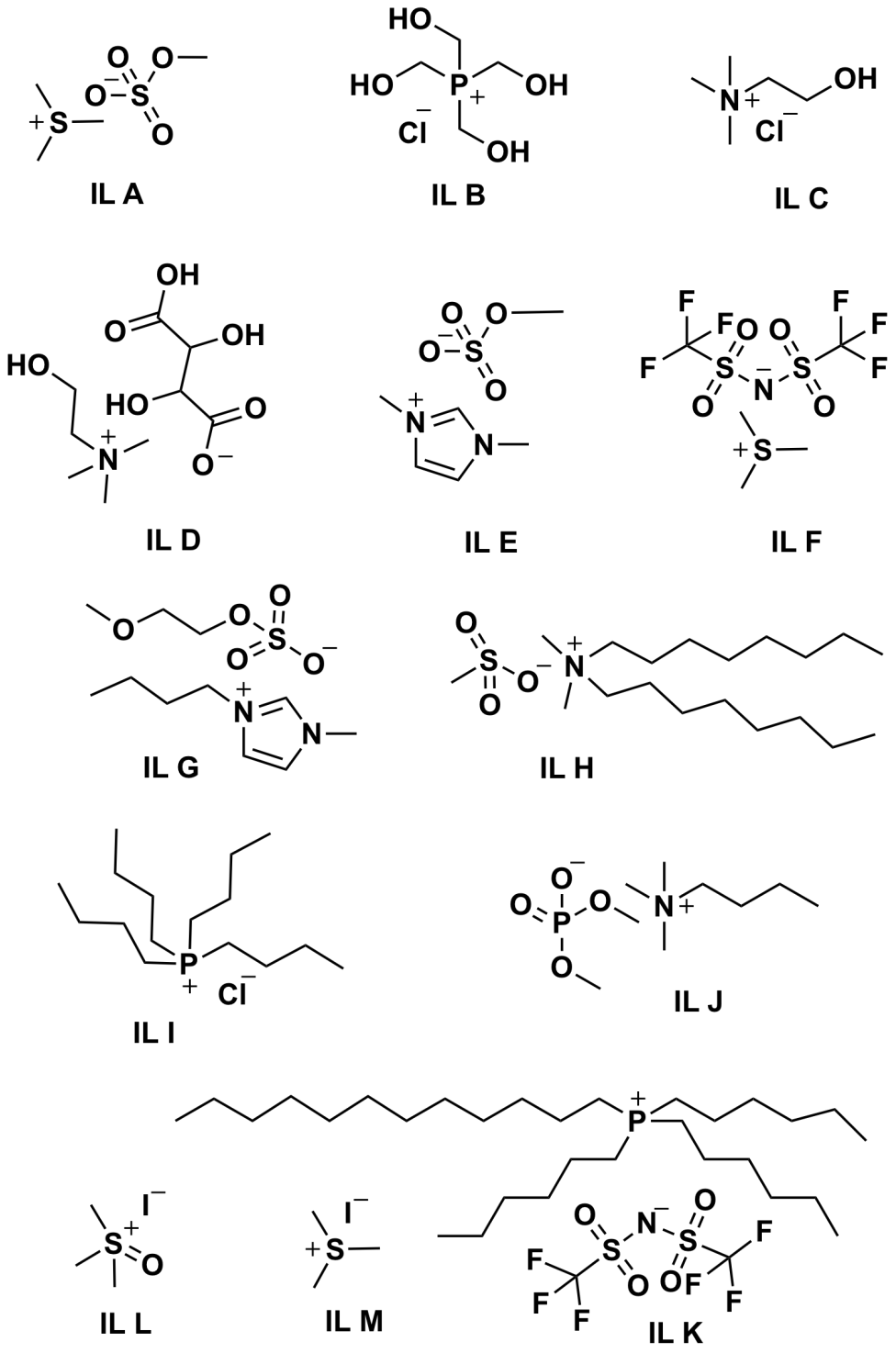


Figure 1: ILA $[Me_3S][OSO_3Me]$, ILB [Tetrakis][Cl], ILC [Choline][Cl], ILD [Choline][Bitartrate], ILE $[EMIM][MeSO_4]$, ILF $[Me_3S][NTf_2]$, ILG $[BMIM][MeOEtSO_4]$, ILH $[N_{1,1,8,8}][MeSO_3]$, ILI $[P_{4,4,4,4}][Cl]$, ILJ $[N_{4,1,1,1}][(MeO)_2(HO)PO]$, ILK $[P_{6,6,6,14}][NTf_2]$, ILL $[Me_3SO][I]$ and ILM $[Me_3S][I]$

Chapter 1

Introduction

1.1 Industrial Enzymatic Synthesis and Reaction Engineering

Global consumption of products such as fuel, food, textiles, cosmetics and pharmaceuticals, entails the expense of large amounts of raw material and energy and generates chemical pollution leading to incalculable adverse impacts on our environment and quality of life. The second Global Chemicals Outlook (held in 2019) estimates a current chemical production capacity of 2.3 billion tonnes and global annual waste generation of an estimated 2.01 billion tonnes in 2016 with a projection to double by 2030 [Alpizar et al., 2019]. However, providing exact numbers through economy-wide material flow accounting (an established method to monitor annual flows of material and energy), is limited, because it omits the integration of in-use material stocks [Wiedenhofer et al., 2019]. Methods such as Life Cycle Assessment (LCA) and Techno Economic Analyses evaluate only separate production processes [Curran, 2014], albeit predicted LCA scenarios of electricity supply confirm global environmental benefits of low carbon technologies [Hertwich et al., 2015].

Conventional chemical processes have been economically viable despite being at points immensely wasteful and associated with poor efficiency and yields. Stricter environmen-

tal regulations and fines for waste and emissions guide the progress of transition towards cleaner industrial production through bioprocesses, such as enzymatic synthesis [Dewan, 2012]. However, unless all process metrics are optimised, up-scaled implementation of bioprocesses does not take place, since conventional syntheses set the bar based on their high economical profitability, i.e. challenging existing routes is difficult, with little to no economic incentive for changing the process [Woodley, 2019b]. Demonstrating superiority over conventional ways of production demands the streamlining and improvement of enzymatic chemical synthesis on every technical scale, including increase in yield, savings of raw material, energy, chemicals and water as well as decreased emissions and waste [Jegannathan and Nielsen, 2013]. This in turn is greatly in compliance with sustainability and has the potential to limit the logistical burden to both the environment and the individual.

Due to simplified production routes omitting protection- and deprotection steps of active groups, and mitigated waste generation through less chemical harsh substances, commercialisation of bioprocesses of technical enzymes in the pharmaceutical sector has been successful. For example the synthesis of a noradrenergic antidepressant against Fibromyalgia by the company Pfizer, where the lipase showed 98 % regioselectivity, entailing a 58 % improvement in yield of commercial product and a reduction in process waste of 1300 MT per year [Hayes et al., 2011]. On the contrary, application of a monoamine oxidase for the production of a bicyclic proline intermediate was deemed economically not feasible, despite a 59 % decrease in raw materials, 32 % decrease in water and 63 % decrease in process waste, due to the necessity of mitigating irreversible inhibition by addition of bisulphate, when compared to the intensive, established process, which requires an excess of metal-based oxidant and 8 reaction steps [Li et al., 2012]. This demonstrates that despite the benefits at hand, only a bioprocess that entails no major disadvantage will succeed an established process.

Five major performance metrics can be identified as mandatory for the economics of implementing sustainable manufacturing: reaction yield (product/reactant), space time yield (product/reactor volume/hour), total product concentration (product/volume), effectiveness of applied biocatalyst (product/biocatalyst), and the ratio of waste to product, or 'E-factor' [Clark, 2009, Woodley, 2019b]. The lack of stability of enzymes at high tem-

peratures and in turbulent flow regimes is the main application barrier for biocatalysis in the chemical industry and challenging to these metrics [Sheldon, 2007]. To help overcome those technical barriers, integrated process development combining protein, reaction and process engineering is paramount and entails the identification of highly selective and intensified operations (biocatalysts), use and recyclability of process streams, few as possible process steps as well as few pH, temperature and solvent changes throughout a given process. However, engineering these steps optimally is associated with elaborate costs and time consuming analysis and hence is dissuasive for implementing them industrially [Woodley, 2019b].

The supremacy of the carbon-based fuel industry, based on the fact that (fossil) fuel is the basis of current societal life, provides a viable chance for pioneering for ways in overcoming these technical barriers for biocatalysis and might pave the way for low-priced chemical production via enzymatic synthesis in general [Chapman et al., 2018]. Since biofuels have a significantly reduced environmental impact owing to less expulsion of carbon monoxide, sulphur oxides, nitrogen oxides and particulate matter, they are an attractive option to lower charged costs for emissions, satisfying economic pressure [Liew et al., 2014]. Controversially, conventional biofuel production relies on the conversion of raw biological materials, necessitating the use of sodium methoxide for plant oil triglyceride conversion to fatty acid methyl esters and hence, resulting in a highly contaminated product containing alkali salt and requiring costly purification [DiCosimo et al., 2013]. This disadvantage and moreover, cheap alternative feedstock of lignocelluloses biomass suitable for enzymatic conversion are major driving forces and stimulate the exploration of enzymatic synthesis for biofuel production significantly. Problematic with inexpensive starting materials such as crop residues, municipal waste, soft- and hard-wood and herbaceous biomass is, however, the dissolution, which is hitherto done through extensive mechanical or chemical pretreatment and subsequent processing through hydrolytic enzymes [Cardona et al., 2010]. Currently ionic liquids are intensively researched and are expected to play a key role for viable release of cellulose from lignocellulose material for industrial applications [Rabideau and Ismail, 2015]. The economical scale-up of cellulase catalysed cellulose conversion to glucose for bioethanol is, however, jeopardised, for reasons of poor biocatalyst recovery, slow enzyme-catalysed reaction rates, and low biocatalyst stability under harsh

industrial operation conditions [Taherzadeh and Karimi, 2007]. These issues are exemplary for nearly all enzymatic syntheses processes and multidisciplinary competences are required to tackle these challenges. Engineering of inherently stable biocatalysts, design of compatible reaction media and engineering of adequate immobilisation strategies for recovery are indispensable components to drive the progress towards industrial biocatalysis [Chapman et al., 2018, Woodley, 2019a].

Addressing the first of these points, the subject of investigation to this thesis is an extremophilic alcohol dehydrogenase from the organism *Haloferax volcanii*. *H.volcanii* is a halophilic archaeon originally isolated from the dead sea. It is easily genetically manipulated and provides a well established platform for the expression of halophilic proteins. A plethora of enzymes with unique abilities for extraordinary stability and even entirely new reactions not observed before, are provided by nature. Such enzymes are found mainly in organisms from extreme environments and exploration of novel functionalities and consequent installation of these in routine enzymes, next to advances in protein design optimisation through molecular dynamics simulation will allow for a wider range of available biocatalysts for commercialisation.

Secondly, water as the sole reaction medium for enzymatic catalysis is widely recognised as problematic. Foremost, many interesting substrates and/or products are poorly water soluble and hence are present at low concentrations in the vicinity of the catalyst. Downstream processing, i.e. the recovery of the product and/or catalyst, have to be technically easy and inexpensive. Water with its high boiling point is particularly expensive to evaporate and makes concentration of products unfeasible in large-scale operations [Woodley, 2019a]. Furthermore, many products are highly similar to the substrates they have been converted from, differing only in a single group and finding properties to differentiate them sufficiently to enable efficient separation is hard. Two phase systems of water and organic solvents exhibit two inherent issues. Firstly, enzyme acting as a surfactant at the liquid-liquid interface is frequently observed at high biocatalyst loadings, and potentially leads to significant loss in subsequent product recovery, and secondly, the current use of highly volatile organic compounds (VOCs) is considerably contributing to greenhouse gas emissions and needs to be replaced urgently [Volkamer et al., 2006].

Ionic liquids (ILs) are highly versatile alternative solvents ('molten salts'), because of an estimated 10^{18} different combinations of anions and cations, making them exceptionally tuneable for the provision of solubility for any substrate [Earle and Seddon, 2000]. Additionally, they are characterised through strong chemical and thermal stabilities ensuring compatibility with harsh operation conditions. However, principles of enzyme destabilization in ILs are not well understood and finding an adequate combination of ions for chemo- and biocatalysis remains challenging [Pramanik et al., 2019]. Investigative studies on the interplay between ionic liquid molecular properties and their effects on enzyme stability and activity is crucial to make these solvents palatable to industry, where the speed of the development of a process is economically decisive. Hence, the second point addressed in this study is the identification and evaluation of impact of different ILs on enzyme structure and activity.

Lastly, the idea to apply immobilisation of enzymes for biocatalysis originated with the intention of up-scaling to facilitate clean downstream processing and to reduce the number of process steps through facile separation of the biocatalyst from the product. It became apparent that immobilised enzymes can be used across a broader range of operating conditions due to improved stability on scaffolds or entrapments, which makes them accessible to continuous processes and/or reusable [Sheldon, 2007, 2011]. At best, they can be immobilised in a preferred conformation and preferred location (e.g. for multi-enzyme cascades) and scaffolds can provide additional functionality. Hence, the third point addressed by this study is the immobilisation by nanocapsules. A range of factors for immobilisation have to be taken into account to prevent loss of enzyme activity through conformation disruption. Other disadvantages are the additional cost of the carrier and preparation materials as well as time-consuming preparation methods, which makes multifold benefits by immobilisation conditional. Further, up-scaling has proven to be difficult and optimal ways of operation include the optimisation of the amount of protein that can be loaded without compromising diffusional exchange and mass transfer [Cao et al., 2016, Thompson et al., 2018].

If further progress is to be achieved, optimizing structure function relationships of enzyme based conjugates, while minimizing production costs of individual components (support and biocatalysts), as well as lowering implementation costs for biocatalytic processes is

imperious. This thesis sets out to contribute research on structure function relationships of extremophilic proteins in ionic liquid mixtures and their immobilisation in micelles.

1.2 Extremozymes

Life in the form of single celled organisms arose around 4 billion years ago. During this time bacteria and archaea managed to explore every conceivable environmental niche by adapting their metabolic pathways and molecular machinery. It took another 2 to 3 billion years until eukaryotes appeared and were ready to explore morphological diversity [Parfrey et al., 2011]. The discovery of organisms in extreme environments has led to the realisation that the two prerequisites for life on Earth (and possibly other planets) are liquid water and the ability to utilise redox chemistry as a source for energy [Ball and Stillinger, 1999]. Life in extremes spans all three domains of life (figure 1.1), requiring stable biomolecules and thermodynamically useful reactions. Different strategies to maintain function by keeping the external environment out, developing protective mechanisms, altering physiology or enhancing repair capabilities have been developed by different organisms and have led to the adaptation of nucleic acids, membrane lipids and proteins. Enzymes from extremophiles dubbed extremozymes (this term was first coined by MacElroy in 1974 [MacElroy, 1974]) (figure 1.2), convey potential in helping us understand evolutionary processes and in exploiting their chemistry for reactions that are not feasible in conventional ways. Understanding the alternatives by comparing biomolecules from extremophiles can provide information on protein structure-function relationships and protein folding, vital insights for the successful application in biocatalysis. The increasing availability of DNA sequences and metagenomes from new extremophilic species provides a continuing resource for the discovery of novel enzymes and enzyme functionalities for commercial exploitation.

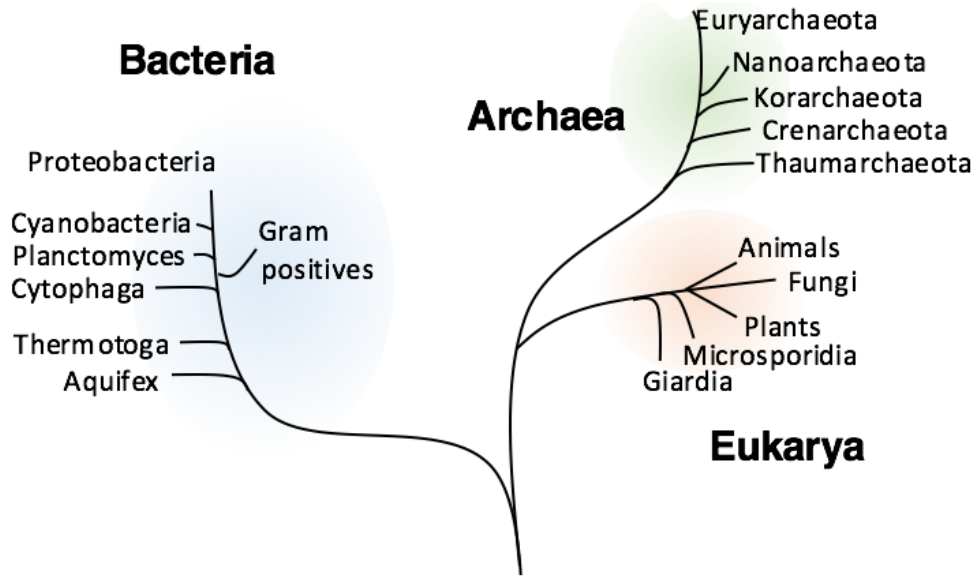


Figure 1.1: Phylogenetic tree of the three domains of life: Bacteria, Archaea and Eukarya, as proposed by [Woese et al., 1990], based on rRNA data.

Foremost, the role of enzyme activity is to maintain a steady flux of products [Cornish-Bowden, 1976] or in other cases, is tightly regulated in conjunction with a metabolic pathway [Fell and Cornish-Bowden, 1997]. The three dimensional fold of an enzyme is only marginally stable with a Gibbs free energy (ΔG) difference of 5 - 15 kcal/mol between the folded and unfolded state [Huyghues-Despointes et al., 2001], and relies on the sum of a small number of molecular interactions, including hydrogen bonds, hydrophobic interactions, ion pairs and Van der Waals interactions. Burial of nonpolar side chains for instance contributes 1.1 $\pm$ 0.5 kcal/mol per buried -CH_2- group upon the removal of the side chain from water. Once inside the interior, London dispersion forces take over for tight packing. The contribution of hydrogen bonds for stabilisation is highly context dependent, but can be averaged to 1.1 $\pm$ 0.8 kcal/mol [Pace et al., 2014], despite the fact that individual H-bonds account to energies of >15 kcal/mol. This comparatively small contribution renders the protein flexible enough to undergo breathing [Rupley and Careri, 1991] and the single atoms to be in constant motion, subject to small temperature dependent fluctuations [Koshland Jr, 1998]. Besides these small breathing movements, conformational changes are essential to convey the complexity of any catalytic activity[Mesecar et al., 1997]. Only a minor proportion of enzymes, depending on

their location and translocation within the cell, undergo major structural rearrangements, ranging from large relative domain movements, loop movements, secondary and tertiary element rearrangements, to utilising disordered regions within the protein, which lack any well-defined structure, for conveying catalytic activity. A much greater proportion of enzymes shows a C-alpha Root Mean Square Deviation (RMSD) of less than 2Å for fulfilling their biological function [Gutteridge and Thornton, 2005]. Moreover, in recent years the importance of structures such as pockets, cavities and tunnels has become increasingly evident. Briefly, these structural elements guide the channelling of substrates and other ligands from the protein surface to active or binding sites. Slight movements of few residues, which essentially serve as gatekeeper, could define an active or inactive state of conformers [Chovancova et al., 2012]. Adversary to stabilising forces are configurational entropic forces, which favour flexibility and together allow motions necessary for catalysis at a rate appropriate to maintain metabolic flux.

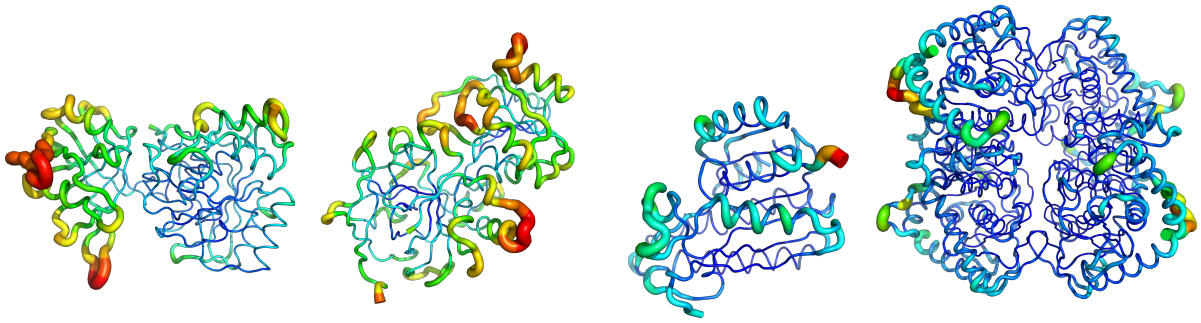


Figure 1.2: from left to right: Visualisation of B-Factors of psychrophilic *Pseudoalteromonas Sp.* protease [1.4 Å], mesophilic horse-liver alcohol dehydrogenase [1.59 Å], thermophilic *Thermococcus Sibiricus* alcohol dehydrogenase [1.55 Å], halophilic *Haloarcula marismortui* glucose dehydrogenase [1.7 Å]; The temperature value (also, temperature factor or B-factor) quantitates the uncertainty for each atom. Atoms with low temperature values are coloured blue, while atoms with high temperature values are coloured red. B-factors (colouring) allows for inferring the rigidity of structures. Values are relative not absolute, hence similar resolution (Å) of structures is required. Enhanced rigidity can be observed in thermophilic and halophilic proteins.

Extremes on the continuum of physical factors, like temperature, pressure, pH, and salinity amongst others, are a strain on cells and biomolecules and lead to inactivation of

non-adapted enzymes. Adaptations of extremozymes correspond to key changes in the amino acid sequence and enzymes typically show a maximum activity at optimal physical and chemical conditions, reflected in the increase of the catalytic constant for an unperturbed enzymatic site [Reed et al., 2013]. A change in this environmental 'hot-spot' conditions leads to the disruption of tuned intermolecular forces and alterations in the flexibility of the protein structure. This observation led to the proposal of a strict link between protein motion, function and stability, which assumes, that activity depends on protein flexibility [Somero, 1978]. A detailed consideration of thermodynamics regarding protein flexibility is given in the subsequent subsections.

It should be mentioned that there is an evolutionary point of view proposing a different explanation for the almost universality of marginal stability. This counter-theory states that because there are many more ways for a protein to fold in a marginally stable fashion than there are for hyperstable conformations, mutational pressure and genetic drift will drive the protein to occupy the most numerous set of states, i.e. the marginal stable ones. This implies that it is theoretically equally possible for a protein to be hyperstable and at the same time highly active as it is for a marginal stable one [Harms and Thornton, 2013], and this was indeed shown to be the case via directed evolution and proposed mathematical models [Elias et al., 2014, Giver et al., 1998]. However, these studies exclude environmental factors in their way of argument and do not take into account the additional flexibility of the protein gained through e.g. thermal energy, which aids the structure in surmounting excess rigidity (see subsection on thermophilic proteins). Hence this theory should be treated not as an explanation for marginal stability, but as a consequence, i.e., that no excess stability is evolved, if there is no requirement for it, because of an intricate balance between entropic and enthalpic forces amendable with the environment. Therefor, the transition between unfolded protein to marginal stability is characterised through a bountiful space of possibilities, whilst the highest reachable stable conformation is unique. To this end, evolution has provided a huge variety in adaptations of extremophilic proteins to similar physical or chemical conditions [Reed et al., 2013]. Combined efforts of biochemical, computational and structural approaches were able to pin point the importance of ion pair formation, hydrogen bonds and hydration, rather than general hydrophobicity for playing a crucial role for the stabilisation of extremozymes. Understanding these

structural properties and complementing the thermodynamic perspective on stabilisation is central to enabling more accurate predictions of the behaviour of natural extremophilic or engineered enzymes under a variety of conditions. The aim is to provide insights on catalytic and structural features auxiliary to the development of novel, efficient and more cost effective bioprocesses.

1.2.1 Thermophilic Proteins

Temperature is a major environmental component that impacts all living organisms. It affects the kinetic energy of biomolecules, their reaction rates and numerous physicochemical properties, that in turn influence the strength of molecular interactions. Adaptation to extreme temperatures has led to enzymes with distinct conformational stabilities, which take advantage of the conformational entropy rise with temperature to gain back mobility. Above a certain temperature (the melting temperature, T_{m}) proteins unfold, meaning that they loose their distinct 3D structure and this results since the unfolded state has a much higher heat capacity, $\Delta \mathrm{C_p}$, than the folded state [Brandts, 1964]. Enzymes with exceptionally high melting temperatures show high conformational stability, $\Delta \mathrm{G_{U\text{-}N}}$, exhibit a higher degree of structural packing [Sterner and Liebl, 2001] and are considered thermophilic (optimal activity $>$ 50 ℃) and hyperthermophilic (optimal activity $>$ 80℃), respectively [Vieille and Zeikus, 2001]. The most extreme example of hitherto discovered thermozymes is an amylopullulanase from a marine hyperthermophilic archaeon with activity at temperatures as high as 142 ℃ [Schuliger et al., 1993].

Amongst acknowledged features contributing to protein stabilisation at high temperatures are an increased number of large hydrophobic residues, facilitating tighter packing of the core. Complementing this, many hyperthermophilic enzymes assume an oligomerised state, when compared to their mesophilic counterparts, due to increased favourable hydrophobic interactions at dimer interfaces, leading to quaternary organisation of tetramers and even octamers [Bräsen et al., 2005, Vieille and Zeikus, 2001]. However, there are also hyperthermophilic enzymes that have implemented the opposite strategy of higher oligomerised states, organising all necessary components within a single subunit and at the same time creating structural rigidity through tighter packing of the hydrophobic core [Park et al., 2013].

Another common strategy of thermophilic proteins is an increase in ion pair content for optimised electrostatic interactions through disulphide bridging and salt bridging [Fukuchi and Nishikawa, 2001]. The desolvation penalty and entropic cost associated with ion pairing found in salt bridges is more easily overcome at higher temperatures. Hence, they become a structurally stabilising element through increasing both short- and long range charge interactions, protecting against thermal denaturation [Lee et al., 2005]. In conjunction with this, asparagine and glutamine are often replaced with their charged versions, aspartic acid and glutamic acid, as they can undergo deamination, which would reduce stability further [Littlechild, 2015]. Whilst the higher content of surface charge residues may prevent aggregation, it can also destabilise the structure if unevenly distributed and implies the necessity of inorganic salt for counterbalancing positively charged surfaces [Liu et al., 2012]. Other observed adaptations contributing to increased protein rigidity are increased helix propensity of residues in α-helices, decreased length of surface loops and anchored C- and N-terminals [Jaenicke, 1996].

The sum of these thermophilic adaptations is reflected in a free energy of stabilisation that is slightly higher than that for mesophilic proteins [Feller, 2010] and which is based on an increase in thermodynamic stability, i.e. an increase in the free energy difference, ΔG, between the folded and the unfolded state [$\Delta G_u = G_u - G_f$]. This is at the same time founded in a decrease of the rate of unfolding, i.e. in an increase in kinetic stability through increasing the free energy difference between the folded state and the transition states of unfolding (figure 1.3). The overall free energy difference between folded and unfolded states is as small as 0.1 kcal/mol per residue [Beker et al., 2017]. Consequently, several mechanisms will act together to shift the melting temperature and can arise from both intra-protein contacts or solvation.

The heat capacity, C_p, upon unfolding, defined as $C_p = -Td^2G/dT^2$, where T is the temperature and G is the enthalpy, is smaller for thermophiles when compared to mesophiles, presumably due to the solvation of hydrophobic interactions contributing positively and the solvation of polar groups contributing negatively to ΔC_p [Prabhu and Sharp, 2005]. Besides this correlation to hydration, the smaller heat capacity suggests further that enthalpic fluctuations of the folded and unfolded states of thermophiles are closer than in mesophiles, since ΔC_p relates to enthalpic fluctuations, $C_p = dH/dT$. Recent MD simula-

tions and visualisation of the free energy landscapes of hyperthermophilic and mesophilic carbonic anhydrase underline this suggestion by showing a less expanded, yet slightly deeper and smoother conformational landscape of the free energy of unfolding for the thermophilic homologues [Parra-Cruz et al., 2018]. The smaller heat capacity of unfolding in some thermophilic proteins can in addition to the stated arguments be traced back to residual native hydrophobic clusters. These residual clusters are hydrophobic groups that are thus not exposed to water even in the unfolded state and the unfolded protein does not assume a fully solvated random coil [Sawle and Ghosh, 2011]. This behaviour is found similarly in halophilic proteins, which are presented in subsection *Halophilic Proteins*.

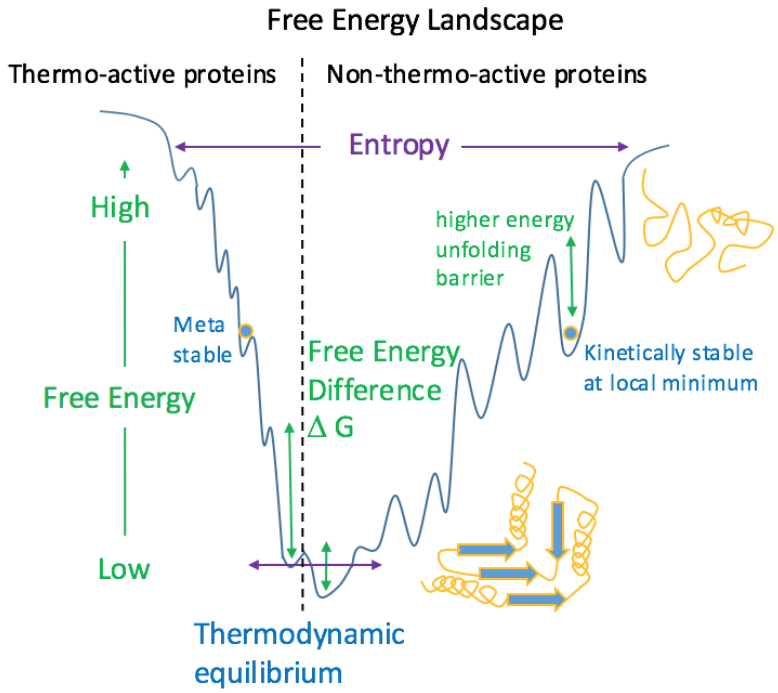


Figure 1.3: Free Energy Landscape of thermo-active proteins (left of the doted line) and mesophilic proteins (right of the dotted line). At the bottom of the trough an optimal thermodynamical equilibrium state is reached. Folds of secondary structural elements explore the conformational landscape towards lower free energy by visiting wells of local minima. These tend to be deeper for mesophilic proteins.

Since ΔC_p entails both enthalpic and entropic contributions, it is difficult to discriminate their impact. The surplus of charged amino acids in thermophilic proteins, for instance, favours the creation of salt bridge and H-bond networks and by this presumably

contributes to lowering the enthalpy of the folded state [Feller, 2010]. On the contrary, introduction of disulphide bridges or X_a-to-proline mutations decreases the entropy of the unfolded state, whereby two possibilities can be imagined: either the energy landscape of the folded state is broad and conformational entropy gets lifted to that of the unfolded random coil, or the unfolded state has a relative lower entropy which is likely caused by residual native interactions reducing the available conformational space [Missimer et al., 2007].

Furthermore, adaptations towards enduring higher temperatures could also be intended to stabilise proteins in a kinetic sense. Since the free energy barrier separating the folded from the unfolded state is higher for thermophilic proteins, they are essentially trapped in the native configuration, even at high temperatures. Localised salt bridges for instance were shown to contribute to kinetic trapping [Cavagnero et al., 1998]. The 'kinetic trapping' theory surmises that proteins in non-aqueous media remain in their native structure, because of an increased amount of hydrogen bonding within the protein resulting in a higher kinetic barrier for structural rearrangements. It is generally assumed that the conformational dynamics of thermophilic proteins are specifically optimised for high temperatures, rendering thermophilic enzymes almost inactive at ambient temperatures as a result of their compactness and excess rigidity [Feller, 2010, Lam et al., 2011]. Basically, they cannot surmount the excess of rigidity, due to a shortage in thermal energy. However, this is a subject of ongoing debate. According to the transition state theory, the catalytic kinetic constant shows an exponential dependence on the free energy barrier for the catalytic reaction ΔG^{++} and a monotonic dependence on temperature: $k_c \propto k_b T\, e^{(-\Delta G^{++}/k_b T)}$. Therefore the lack of activity at moderate temperatures could be solely caused by a higher ΔG^{++}, likely resulting from different arrangements in the catalytic centre, which are 'tuned' to higher temperatures to accommodate for a greater mobility of the substrate [Roca et al., 2007].

1.2.2 Halophilic Proteins

The ability to not only survive, but to thrive in high salt environments earned the halophiles their name by which 'halo' translates to salt and '-philos' means lover to the Greeks. Our definition of extreme, one could argue, is an anthropocentric one, as it im-

plies that it should be hard to thrive in conditions at the extremes of scales. Halophilic proteins, however, not only 'survive' but actually utilise high salt to be able to fold correctly and on the contrary lose their ability to function in low salt. This salt-dependent folding was first experimentally evidenced in 2009 for an esterase from the halophilic archaea *Haloarcula marismortui* [Müller-Santos et al., 2009]. A time-dependent recovery of activity in the enzyme was demonstrated when 0.25 M - 3 M KCl salt was added to pure aqueous conditions. Observations used far-UV circular dichroism, which under aqueous conditions confirmed spectra typical of unfolded protein and a subsequent increase in ellipticity in a salt-dependent manner, pointing to an increase in overall α-helical content as the fold of the protein recovered [Müller-Santos et al., 2009].

Indeed, what we consider extreme environments might have been at the origin of life, and any deviating environment like ours could therefore inversely be called extreme. Abiotic chemical syntheses identify a consensus set of 10 α-amino acids, dubbed 'prebiotic', for which findings strongly suggest the ability to support cooperative folding, such that defined structure is possible with the earliest polypeptides [Lerner et al., 1993, McDonald and Storrie-Lombardi, 2010]. This prebiotic set contains high-propensity amino acids for each basic secondary structure type, including alanine and leucine, which have the highest α-helix propensity and glutamic acid having a high propensity, isoleucine [Pace and Scholtz, 1998], valine and threonine having the highest preference in β-sheets [Street and Mayo, 1999], aspartic acid, aerine, glycine and proline in addition to threonine being crucial in loop positions of beta-hairpins, and again valine, proline, aspartic acid and serine for different types of β-turns [Hutchinson and Thornton, 1994], as well as hydrophobic and hydrophilic amino acids with the potential to support patterning essential for specific secondary structural elements. Notably, this set has a skew for residues associated with halophilicity and lacks residues which are under-represented in halophilic proteins, i.e. it lacks all three basic residues, and the key contributors to extensive van der Waals interactions in hydrophobic cores, the aromatic residues, are also absent.

Arguably, foldability of early polypeptides should be most compatible with the halophile environment. One study for instance demonstrated the construction of a β-trefoil protein from a foldable set containing 79 % prebiotic residues. The protein was foldable in high salt (2 M NaCl) and showed pI values and surface electrostatic charge distribu-

tions typical of halophilic proteins [Longo et al., 2013]. In another study an ancestral thermophilic nucleoside kinase was reconstructed and the amino acid set systematically reduced to 13, where residues inconsistent with the prebiotic set appeared to be catalytically important and not necessarily for structural stability [Shibue et al., 2018]. However, little work has been done on systematically reducing the size of the amino acid set to construct prebiotic proteins, despite the relevance to finding an answer to proteogenesis and possibly abiogenesis (origin of life), if one takes a 'protein-first' view on this matter. The reason for having such few studies at hand is probably the conflicting conclusions derived at what environmental conditions were like at the early stages of life [Kandler, 1998].

Hypersaline environments, as they are found in the Great Salt Lake in Utah or at the lowest point of terrestrial earth, the Dead Sea, harbour abundant moderately halophilic (> 0.5 M), halophilic (> 1.5 M) and extreme halophilic (> 2.0 M) organisms, who defy excess salinity through different adaptation strategies [Mevarech et al., 2000]. Mostly it is halophilic bacteria and eukaryotes that prevent the entry of inorganic salt into the cell and consequently have adapted to synthesise small organic molecules (osmolytes), like ectoine, to balance the extreme osmotic pressure. Halophilic archaea, on the contrary, accumulate high concentrations of inorganic salt intracellularly and have adapted their proteins to remain stable and functional accordingly. While the growth medium contains sodium in several hundredfold excess over potassium, cells concentrate potassium to over 2 M, which is several times higher than the molarity of intracellular sodium. This preference for potassium likely stems from the size difference between these ions, as the bigger size of potassium coordinates (i.e. strips) less water molecules from the protein surface and hence is favoured. Proteins from these organisms are classified as halotolerant (0.2 - 2.0 M), halophilic (2.0 - 3.0 M) and extremely halophilic (3.0 - 4.0M), and encounter water activity as low as 0.75 in near saturated salt solution. Early reviews centring around halophilism started appearing in the 1950s, followed by more reasonable hypotheses on halophilic adaptation in 1974 by Lanyi [Lanyi, 1974] and the field was led by a comprehensive review from Eisenberg in 1992 [Eisenberg et al., 1992]. However, a much more detailed molecular perspective on halophilic adaptation could only be provided in 1995 by Dym, Mevarech and Sussman with the first solved structure of the tetrameric

Haloarcula marismortuii enzyme malate dehydrogenase [Dym et al., 1995], and with the first halophilic genome of *Halobacterium* species NRC-1 sequenced in 2000 the striking negativity of the halophilic proteome became evident [Ng et al., 2000]. In stark contrast to non-halophilic proteomes, which posses a bimodal distribution with acidic and basic proteins, a unimodal distribution of protein isoelectric points (pI) with a mean of 5.0 was observed and subsequent modelling studies, consistent with available structural work asserted an abundance of negatively charged amino acid residues on protein surfaces.

This increase in fraction of the amino acids glutamic acid and aspartic acid is ubiquitous for all known halophilic proteins and hence is the most notable difference to mesophilic enzymes, which on the contrary are stable and functional in ambient conditions. The basic role of the elevated numbers in charges is dual, first, they provide carboxylate groups that maintain the solubility of the protein and second, their electrostatic nature is thought to contribute to repulsion, which offsets the increased stabilisation from the salt's enhancement of the hydrophobic effect, provided that the solvent prefers an interaction with the folded protein over the unfolded one [Mevarech et al., 2000]. The exact mechanisms of how the prominence of carboxylic acid residues inflicts solubility are however still not fully understood. One proposed mechanism constitutes a competition of the acidic amino acids with the ions for water molecules (figure 1.4), since aspartic acid (ASP) and glutamic acid (GLU) are capable of forming hydrogen bonds through their carboxylic side chains, having pKa values of 3.8 and 4.2 at low salt concentrations. These values rise to 4.9 (ASP) and 5.3 (GLU) in high salt and increase the strength of the hydrogen bond [Lenton et al., 2016]. Whilst this hypothesis is supported by crystal struc-

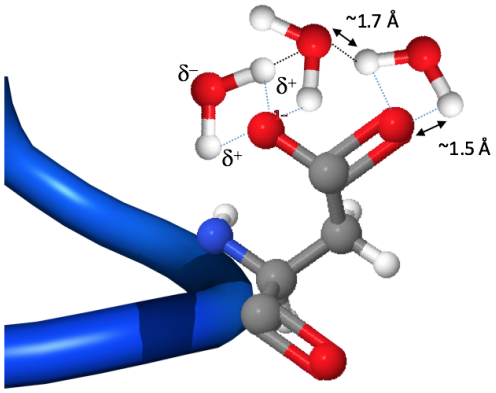


Figure 1.4: Protein backbone with aspartic acid residue interacting with water molecules through hydrogen bonds. On average 2-3 H_{water} and 3-4 O $_{water}$ reside in the first solvation shell of the carboxylic oxygen.

tures that show increased water binding for halophilic proteins with an average of 1.9 water molecules per residue compared to an average of 1.2 in mesophilic proteins [Mevarech et al., 2000], Madern et al pointed out, that under crystallising conditions salt exclusion and improved water binding is expected, suggesting a misrepresentation of native conditions [Madern et al., 2000]. Analysis of the solvent structure and its interaction with the protein surface from crystallised halophilic glucose dehydrogenase shows higher complexity in the first and second solvation shell, which is reflected in a most ordered arrangement of pentagonal rings of water molecules around distinct sites [Britton et al., 2006]. Opposing to this, a study by Qvist et al emphasises that there is no increased water of hydration in halophilic proteins [Qvist et al., 2012]. Using ^{17}O magnetic spin relaxation to monitor water associating with the protein surface and returning to the more mobile bulk solvent, they determined no difference in the amount of water bound to the halophilic over the mesophilic version of their target protein. Moreover, the homology model of a halophilic dihydrofolate reductase shows a similar amount of hydrogen bonding networks when compared to mesophilic counterparts [Kastritis et al., 2007].

An alternative model explaining the large number of negative charges suggests increased binding of hydrated cations to protonated and charged forms of the carboxyl group, which are expected to have distinct affinities for K^+ depending on their pKa values. The potassium ions in turn are thought to maintain a shell of hydration around the protein by coordinating water molecules (figure1.5). Crystal structures showing specific cation-protein-binding are known, but limited, because they require a high resolution (below 2.4Å) to allow discrimination between sodium ions and water, which are isoelectric (10 electrons) [Frolow et al., 1996, Richard et al., 2000]. Therefore, data on coordination geometry is required. When studying aspartic acid molecules in different concentrations of salt it was found that the interaction strength between potassium and the side-chain carboxylate oxygen decreases with increasing salt, however, coordination numbers stay the same, suggesting a higher surface mobility on part of the cations. Underlining this, a study utilising NMR investigated pKa values of carboxylic groups of halophilic proteins and found a significant electrostatic net fold stabilisation, resulting from the cumulative weak interactions between single residues and solute cations [Ortega et al., 2015]. The

same group found preferential ion exclusion and competitive hydration of the unfolded state of halophilic proteins that increases with salt concentration, driving a salt induced destabilisation of the unfolded state in favour of the active conformation of the protein. This in congruence with the low affinity binding of potassium to specific (negative charged) sites on the surface of the folded polypeptide, which will stabilise the folded state.

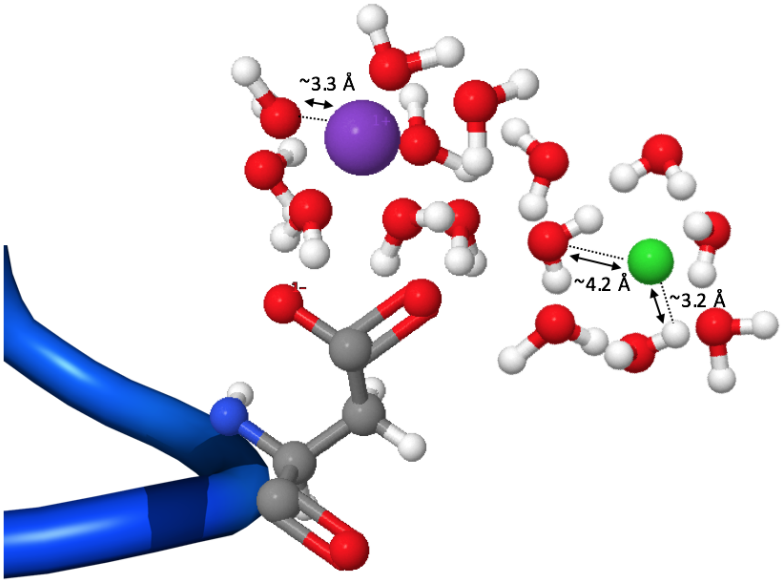


Figure 1.5: Protein backbone with aspartic acid residue, hydrogen bonds are formed to the same extent with carboxylic oxygen, solvation is mainly driven by potassium ions (purple) coordinating on average 16 H_{water} and 5 O_{water} in its first and second solvation shell. Chloride ions(green) coordinate on average 5 H_{water} and 8 O_{water}. Distance between H_{water} - O_{water} in low salt accounts to 1.7 Å, in high salt 1.8 Å. Distance between ASP-ASP residues in low salt accounts to 4 Å on average, in high salt to 9 Å. Radius of potassium ion accounts to approximately 5 Å.

It was demonstrated that anions do not determine the solvation shell, like cations do, through the numerous weak binding sites at the protein surface. However, coordination of anions is a likewise observed feature, most frequently in the inter subunit space, where they are proposed to interact through a limited number of strong binding sites and are closely located at salt bridge clusters in order to stabilise the formation of oligomerised states. Malate dehydrogenase from *Haloarcula marismortui* (hMDH) for instance was shown to be inactive in monomeric form, with dimers and tetramers being equally active and stabilised through chloride coordination [Madern and Ebel, 2007]. This observation

on oligomerised states being a prerequisite is similar to those made on multiple occasions in thermophilic proteins (see subsection *Thermophilic Proteins*).

An early theory (1959) proposed that the elevated surface charge conveyed by the acidic residues may be necessary for electrostatic neutralisation [Baxter, 1959]. However, this theory was already judged as deficient in several aspects a couple of years later (1970) [Lanyi and Stevenson, 1970] and was dismissed as highly unlikely by Mevarech et al in 2000, arguing that already sub-millimolar concentrations of nicotinamide adenine dinucleotide phosphate were able to effectively replace the requirement for molar concentrations of salt in the stabilisation of their protein hMDH, instead the high surface charge is neutralised mainly by tightly bound water dipoles [Mevarech et al., 2000]. Further, overall tetrahedral water structure is more perturbed in high salt, which is reflected in a strengthening of water-water interactions, whilst coordination numbers between water molecules stay the same [Lenton et al., 2016].

The reduction in hydrophobic residues or rather, the less common presence of bulky, aromatic hydrophobic residues is another common property in halophilic proteins Reed et al. [2013]. At salt concentrations above 0.1 M, water is less available to the protein due to the water molecules preferential coordination to the ions. This lower availability of water can cause hydrophobic residues to loose hydration and aggregate, hence strengthening the hydrophobic effect. Comparison of protein sequences between halophilic and mesophilic proteins evidenced a less common presence of tyrosine, phenylalanine and tryptophan in halophilic proteins. The hypothesis suggests that the extent of protein interior hydrophobic surface is reduced in order to lessen the change in solvated hydrophobic area in a high salt environment, i.e. to counterbalance the increased strength of hydrophobic interactions. Confirming this, a study looking at the relative accessible solvent area (ASA) and the apolar contact area (ACA) found a significantly smaller fractional apolar component of ASA for halophilic proteins as well as, that the ACA in the core of the molecules is not different between mesophilic and halophilic proteins, however it is consistently smaller for halophilic ones and even more so, if the core contains only residues with no accessibility to water [Siglioccolo et al., 2011]. This weakening in hydrophobic interaction prevents the hydrophobic core from becoming too rigid, as seen in non-halophiles when under high

salt conditions and hence restores the dynamic motions required for efficient catalysis. Further observations assert that this hydrophobic deficit is accompanied by an unusually large presence of serine in extremely halophilic proteins [Lanyi, 1974]. This borderline apolar residue is thought to be a requirement to compensate the lack in hydrophobic character, through the salting out effect in order to bring about a tight fold and the confirmation required for the functional state of the molecule at high salt concentrations. Opposing to this, bioinformatics analysis of amino acid sequences of mesophilic and halophilic proteins found a tendency towards a lower frequency of serines in halophilic proteins and often those serines are substituted with acidic residues [Zhang and Ge, 2013]. The provided explanation is the exceptional ability of serine to interact with water, which potentially destabilises the protein upon release of bound water molecules, hence is unfavourable at moderately high concentrations of salt. This suggests that extremely halophilic proteins gain advantage of deploying serine, to encourage hydrophobic interactions, whilst moderately halophilic proteins see adverse effects, through increased fluctuations of releasing and binding of water molecules. This would be in congruence with concentration dependant salting out effects (figure 1.6).

In some cases, adaptations to high salt are not found throughout the entire polypeptide chain, but are significantly increased through insertion sequences. These peptide insertions then typically contain large numbers of acidic residues, proposed to have different functions, amongst them, improving the ASA, and may have come about as a way to quickly impart halophilic adaptation to changes in the environment [Marg et al., 2005, Taupin et al., 1997].

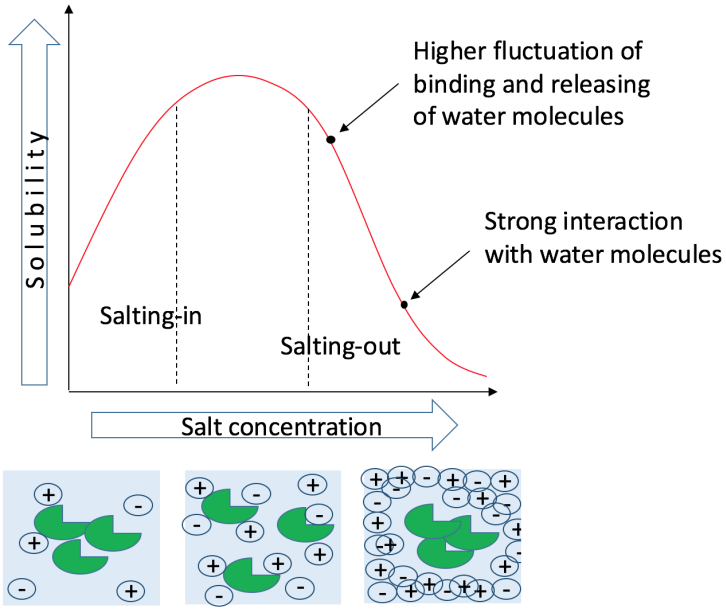


Figure 1.6: Salting-out: The increasing salt concentration draws water molecules from the protein surface to interact with the ions and decreases the number of water molecules available to interact with the charged surface of the protein. To minimise entropy, the protein is forced to merge towards a greater hydrophobic contact area. Hydrophilic amino acids are essential at this point to maintain solvation and dissolution of the protein.

In that sense a crude picture can be drawn where functionality through flexibility in halophilic proteins is asserted through the closely placed elevated surface charges, repelling each other and the lower hydrophobic contact in the core, counterbalancing the hydrophobic effect. However, to enable correct folding and the maintenance of a stable confirmation, borderline polar residues like serine and threonine are either under- or over-represented, depending on the salt concentration the proteins had to adapt to, to prevent destabilisation through faster water dynamics in halophilic proteins and to guide folding and stabilisation in hyperthermophilic proteins, respectively.

1.2.3 Haloalkaliphilic Proteins

As the term 'haloalkali' suggests, high salt environments may also vary in pH with some of these being highly alkaline, e.g. brine pools. Protein adaptations to alkaline pH are subtle, because most haloalkaliphiles have cellular mechanisms deploying a large number of glycosylated proteins to maintain a more neutral pH in their cytoplasm, and hence al-

kaliphilic adaptations are mostly secondary to halophilic adaptations, containing similar to these large proportions of acidic residues (see previous subsection *Halophilic Proteins* [Grant, 2004, Siddaramappa et al., 2012]). Extreme pH conditions cause low availability of metal ions and also reduce levels of ammonium ions. Hence, organisms thriving under these conditions have evolved a multitude of strategies for ammonium and ion uptake, including large sets of transporter genes and genes involved in nitrate reduction and urea conversion clustered next to them. Haloalkaliphilic organisms also show altered ion-specificity for induction of respiratory complexes and ATPases permitting them to cope with inverted proton-gradients [Falb et al., 2005]. These altered pathways may hold enzymes with different substrate scopes, useful for biotechnological applications of which alkaliphilic proteins dominate in numbers over other extremophilic proteins with alkaline cellulases and proteases being implemented for laundry detergents, alkaline cyclomaltodextrin glucanotransferase for the production of cyclodextrin, which is used in large quantities in food, chemical and pharmaceuticals as well as other alkaline enzymes like xylanases used for alkali-treated wood pulp.

1.3 Archaea

Phylogenetically speaking, Archaea have been among us only as recent as 1977 when Carl Woese and George Fox first identified Archaebacteria and distinguished them from other prokaryotic organisms through nucleotide sequence analysis of small subunit ribosomal RNAs (16S rRNA) [Woese and Fox, 1977]. Their monophyletic status became widely acknowledged around 1990 to 1995 following the advent of DNA sequencing, which allowed the analysis of gene markers to establish evolutionary relationships and a new model phylogenetic tree constituting the three major evolutionary lines (domains of life) Bacteria, Eukarya and Archaea was developed [Woese et al., 1990]. Within this early tree topology Archaea comprise of the phyla Euryarchaeota, which include all known methanogens and halophiles (except nanohaloarchaeota) as well as several thermophilic and psychrophilic species, Crenarchaeota, which involve all extreme hyperthermophiles as well as at least

one psychrophile, and Nanoarchaeota and Korarchaeota.

Recent discovery of archaeal superphyla however challenges this rooting of the three domain tree of life once more and instead of placing Eukarya and Archaea as monophyletic sister groups proposes an emergence of eukaryotes from within an archaeal domain [Cox et al., 2008, Lake et al., 1984, Spang et al., 2015]. Since 2010 this two domain tree (figure 1.7) has been changed and challenged several times, upheaved by the discovery of Thaumarchaeota [Brochier-Armanet et al., 2008] and Aigarchaeota [Nunoura et al., 2010], which together with Crenarchaeota and Korarchaetoa form the TACK superphylum [Guy and Ettema, 2011], Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota together forming the superphylum DPANN [Rinke et al., 2013], and the more recent discovery of the Asgard superphylum [Spang et al., 2015] from which or as a sister group of eukaryotes are suggested to have originated. This suggestion is based on the identification of genes encoding eukaryotic signature proteins, including homologues of eukaryotic ribosomal proteins L22e and L28e [Zaremba-Niedzwiedzka et al., 2017] and distant, but still most closest yet discovered homologues of actins [Jékely, 2014] as well as components of the ubiquitin modifier system [Eme et al., 2017] within members of the Asgard superphyla. These eukaryotic signature proteins are thought to be fundamental to eukaryotic cellular complexity and raise a lot of profound phylogenetic questions.

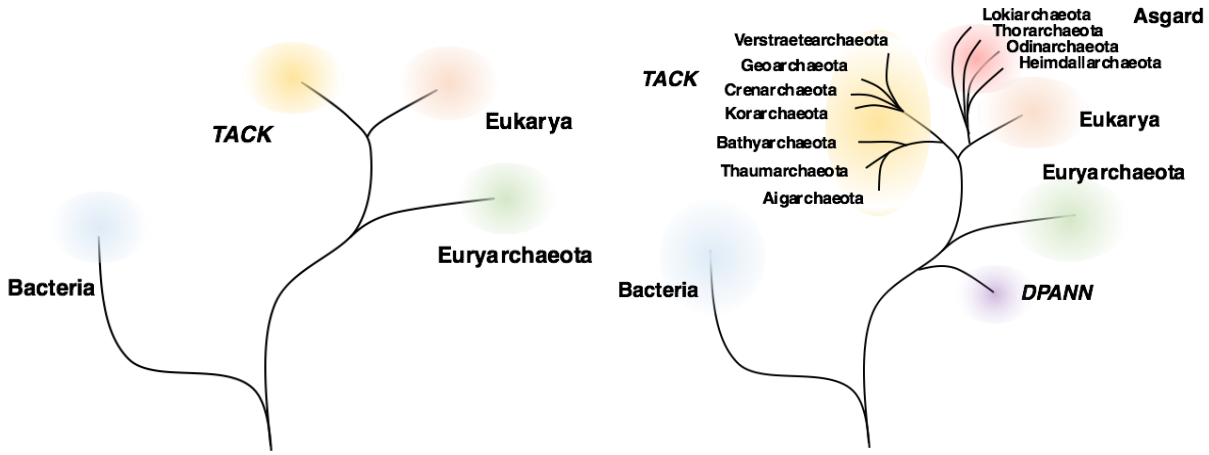


Figure 1.7: Novel two domain tree showing the position of Eukarya originating within Archaea.

These newly discovered species are foremost a treasure trove for phylogenetic insights possibly as far-reaching as Eukaryogenesis [Rossi et al., 2003]. In addition, once a multitude of

genomes is sequenced, they will allow the tracking of the evolution of protein superfamilies within and across these different species and enable predictions of function and structure for uncharacterised proteins [Ovchinnikov et al., 2014, Yang, 2005]. Moreover, elucidation of relationships of specific protein homologs can shed light on common ancestors and their growth environments by analysing orthologs (sequences related to by speciation, and thus presumed to share a common function) and paralogs (sequences related by gene duplication in a common ancestor and potentially divergent in function) across the phylogenetic tree. One such relationship with relevance to hyperthermophily and the debate on the environment of LUCA (last unified common ancestor) seems to have been solved very recently by using this approach. The branching of 376 reverse gyrase sequences, a protein unique to hyperthermophilic organisms and ubiquitously found amongst them, which induces positive supercoiling into DNA to protect DNA from unfolding at very high temperatures, was analysed to reconstruct phylogenetic relationships [Catchpole and Forterre, 2019]. Comparison of the branching of the sequences and the commonly known phylogenetic tree showed a high dissimilarity. This suggested a non-hyperthermophilic origin of LUCA and an early origin of reverse gyrase with likely horizontal gene transfer across domains.

Regardless of this immense archaeal diversity, one main feature, that seems to prevail for these organisms is resilience towards environmental extremes, including salinity, heat, cold, acidity, basicity, pressure and radiation, conditions thought of as similar to those on an early earth, the archaean, which is also why Woese picked their name accordingly. Vice versa, not all Archaea are extremophiles, but were shown to be ubiquitous, some of which have even been identified in the human gut [Gaci et al., 2014].

It is this ubiquity of habitats, ranging from oceans and soils to acidic pools, hydrothermal vents and hypersaline lakes, that poses a major challenge to isolating and cultivating archaea and hence deriving their proteins, let alone expressing them in amounts required for biocatalysis. Nonetheless, quite a few genetic systems have been developed for Halophiles (*Haloferax volcanii*, *Halobacterium salinarum* and *Haloarcula spp*), Methanogens (species from the genera Methanococcus and Methanosarcina), Sulfolobus (*S.solfataricus* and *S.acidocaldarius*) and Thermococcales (mainly *Thermococcus kodakarensis* and *Pyrococcus furiosus*), which have been successfully used for studying archaeal genetics [Allers and

Mevarech, 2005, Atomi et al., 2012].

1.3.1 *Haloferax volcanii*

Haloarchaea belong to the phylum Euryarchaeota, are chemoorganotrophic and require high salt concentrations, usually greater than 1.0 - 1.5 M, for their growth. They possess purple, orange or red pigmentations due to the presence of carotenoids and have obligate salt-dependent enzymes. Some species also possesses a unique proton pump, bacteriorhodopsin, which only depends on sunlight and they can grow at saturated salinity in diverse environments such as the Dead Sea, solar salterns, subterranean salt cores and even cold evaporation ponds, like the Deep Lake in Antarctica[Allers and Mevarech, 2005]. The species used in this work is *Haloferax volcanii*, which was originally isolated from the Dead Sea and is a coccid, red pigmented organism, whose outer cell wall is composed of a glycoprotein S-layer [Mullakhanbhai and Larsen, 1975]. Its lipids are characterised by ether linkages and isoprenoid chains of phytanyl and bis-phtanyl and it possess a polar lipid, archaetidylglycerol methylphosphate, which is the major driver for membrane stability in extreme environments [van de Vossenberg et al., 1998]. *H. volcanii* is an easy-to-culture organism for it is aerobic, has an optimal growth at temperatures of 45 ℃ and NaCl concentrations of 1.7 - 2.5 M and has a generation time of 2.5 - 3 hours. A large genetic tool box has been established (see below) and hence *H. volcanii* serves as a model organism for genetic investigations, such as identifying groups of genes and their particular physiological functions. In addition, biochemical tools have been developed prior to this study that allow the overexpression and purification of histidine and/or Strep-II tagged proteins in *H. volcanii*.

Genetic and biochemical tools in *H. volcanii*

The full genome of *H. volcanii* has been sequenced and the complete genome of the laboratory strain H26 is 4.2 Mb in size with a main chromosome (2.85 Mb) and three megaplasmids pHV1 (86 Kb), pHV3 (442 Kb) and pHV4 (690 Kb) [Charlebois et al., 1991, Hartman et al., 2010], whereby the megaplasmid pHV4 was integrated onto the chromosome in the laboratory strain [Hawkins et al., 2013]. The wild-type isolate strain

DS2 has an additional plasmid pHV2 (6 Kb), which was was removed in the laboratory strain, so as to utilise it for transformations [Wendoloski et al., 2001]. The genome itself has, like most haloarchaeal genomes, a high GC content with 65 %. *H.volcanii* cells are further highly polyploid with up to 20 genome copies present in each cell [Breuert et al., 2006].

Antibiotic selection (Novobiocin and Mevinolin) is possible for *H. volcanii*, however these markers were isolated as mutant alleles of essential genes and therefore homologous recombination with chromosomal alleles might occur, establishing antibiotic resistance in the host strain [Holmes et al., 1991, Lam and Doolittle, 1989]. For this reason, several genes have been deleted in the laboratory strain to render it auxotrophic to selectable markers. Markers relevant to this study are summarised in Table 1.1.

Table 1.1: Auxotrophic selection markers in *H. volcanii* used in this study

Gene	Protein / Biochemistry	Reference
pyrE2	Orotate phosphoribosyl transferase Uracil biosynthesis	[Bitan-Banin et al., 2003]
hdrB	Dihydrofolate reductase Thymidine biosynthesis	[Ortenberg et al., 2000]

Transformation of *H. volcanii* with plasmid DNA is achieved by removing the S-layer using EDTA, which chelates Mg^{2+}, and the addition of polyethylene glycol, which facilitates the uptake of DNA. Transformations with plasmid DNA are made to either facilitate a gene deletion / replacement or the overexpression of genes from episomal plasmids. The gene replacement / knock-out system utilises the *pyrE2* marker (for uracil/5-FOA' conversion) present on the plasmids during pop-in(integration) and pop-out(excision) steps, and the overexpression of genes on episomal plasmids is induced by the addition of tryptophan. The tightly controlled tryptophan inducible promoter *p.tnaA* is placed upstream of the gene to be overexpressed and originates from the *H. volcanii* tryptophanase A gene *tnaA* [Large et al., 2007]. Biochemical tools that allow for the overexpression and purification of histidine- and / or streptavidin- tagged proteins in *H. volcanii* have been developed prior to this study. Genetic modifications of expression strains include the deletion of

the restriction enzyme Mrr, enabling efficient transformation without the need for prior passage of plasmid DNA through *E.coli dam-* strains, as well as modifications of the naturally histidine-rich proteins PitA and Cdc48d, which co-purify alongside the his_6-tagged proteins. Further details on these methods are to be found in *Chapter 2: Materials and Methods*.

Although Haloarchaea (*Halorubrum chaoviator*) even made it into space on board the BIOPAN facility and survived (un)surprisingly well [Leuko et al., 2014], their potential as a source of commercially exploitable biocatalysts remains mostly unrealised for industrial applications [Oren, 2010]. One of the main reasons for the 'under-utilisation' of halophilic proteins in industry is their insolubility and misfolding in low salt conditions, making heterologous expression in mesophilic host systems ambivalent. The frequently required process of refolding is a tedious and low yielding task. Expression in native halophilic hosts on the other hand necessitates the acquisition of such systems and the switching from industrially established hosts, such as *E.coli* has encountered reluctance, despite the well established protocols for *H. volcanii*. The slightly longer generation times and requirements for e.g. high salt protocols of haloarchaeal cells compared to *E.coli* or yeast expression systems appear to be hurdle enough for most to disregard their advantages.
The industrial production of the organic compounds beta-carotene by the green alga Dunaliella [Ben-Amotz and Avron, 1989] and ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid) by halobacteria [Buenger and Driller, 2004] are the notable exceptions of hitherto commercial exploitation of halophiles. Potential applications, where halophilic biocatalysts are superior to their mesophilic homologoues, such as hydrolases and oxidoreductases for the food industry including treatment of saline wastewaters for bioremediation, proteases and dehydrogenases for biocatalysis in non-aqueous media for the production of fine chemicals and cellulases, chitinases and esterases for biofuel production and biopolymers, have been demonstrated, but remain so far industrially unexploited.

1.4 Alcohol Dehydrogenases

Alcohol dehydrogenases (ADHs, Enzyme Commission number 1.1.1.1) belong to the large class of oxidoreductases that catalyse the transfer of electrons between substrate, cofactor and product. More precisely, they catalyse the reversible dehydrogenation of alcohols to aldehydes or ketones, requiring the cofactor nicotinamide adenine dinucleotide or its phosphate ($NAD(P)^+$), which acts as electron acceptor or donor (figure 1.8).

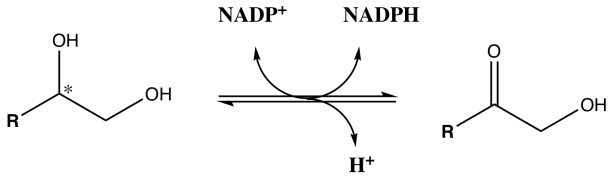


Figure 1.8: Mini-pathway of ADH catalysed $NAD(P)^+$ dependent conversion of chiral alcohol.

There are three subclasses of $NAD(P)^+$-dependent ADHs: short-chain zinc-dependent ADHs, medium-chain zinc-dependent ADHs and iron-dependent ADHs [Reid and Fewson, 1994]. For yeast alcohol dehydrogenase (YADH) and horse liver alcohol dehydrogenase (HLADH), both zinc dependent, the reaction follows an ordered 'bi-bi' mechanism of $NAD(P)^+$ binding first and NAD(P)H dissociating last [Theorell and Chance, 1951], which is in congruence with the principle of proton relay. The mechanistic aspects of their catalysis are similar and can be seen as common for this class of enzymes. To exemplify this reaction mechanism, the oxidation of ethanol to acetaldehyde is illustrated in (figure 1.9). The dehydrogenation starts with the coordination of the ethanol by the zinc ion. The catalytic zinc acts as a Lewis acid and lowers the pKa of the alcohol, removing the proton from the hydroxyl group and releasing it to the solvent. This promotes the transfer of a hydride ion from the carbon atom of the alcohol to the nicotinamide ring. The tetracoordinated positively charged nitrogen of NAD^+ serves as a strong Lewis acid and is essential for enhanced binding of the coenzyme to the alcohol, as this ensures that the hydride is transferred with both electrons in a single transition state, without ever

existing as a free species [Dalziel, 1963, Plapp, 2010].

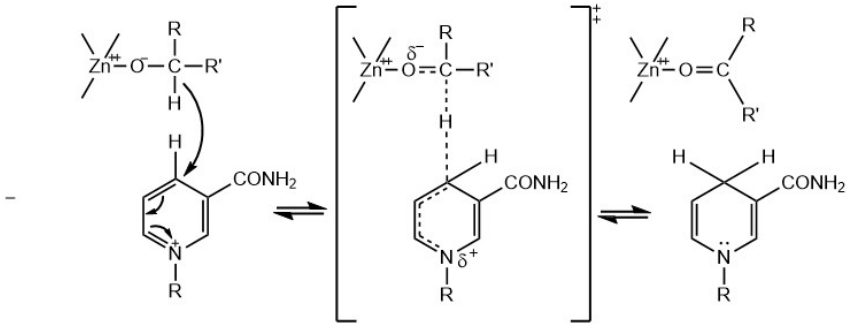


Figure 1.9: Catalytic mechanism of ADH: The hydroxyl group is oxidised by removal of a proton through the catalytic zinc and a hydride is transferred from the alcohol to NAD^+.

Medium-chain ADHs are a structurally well characterised class of enzymes and are composed of two subunits, the substrate binding domain (SBD) and the nucleotide (coenzyme) binding domain (NBD) (figure 1.10).

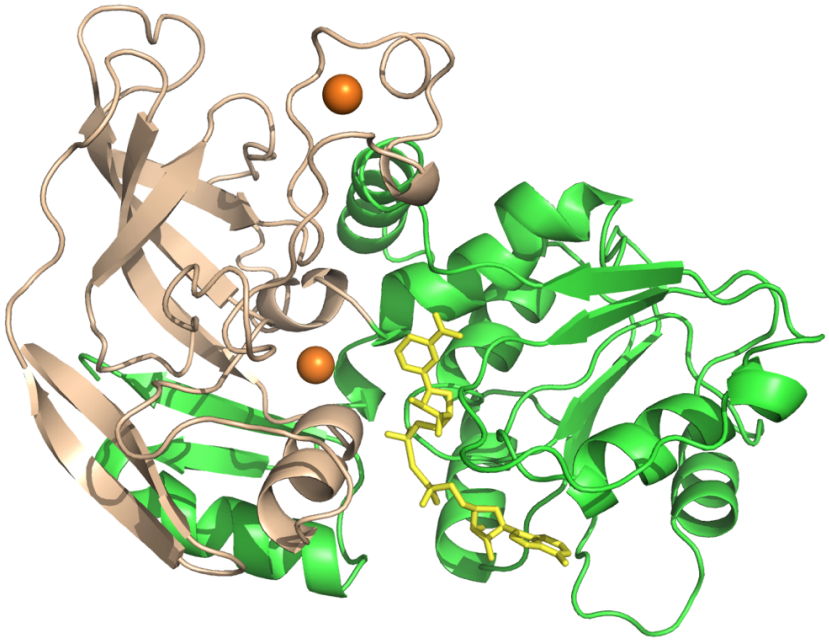


Figure 1.10: Nicotine binding domain (green) with a bound ligand (yellow) and substrate binding domain (wheat) containing the structural and catalytic zinc (orange spheres).

The NBD and the catalytic domain are separated by a deep cleft and upon coenzyme binding, the conformation of the interdomain crevice between the NBD and SBD shifts

from an 'open' to 'closed' position. This interdomain cleft was observed to be significantly larger in the hyperthermophilic *Sulfolobus solfataricus* ADH, when compared to other known ADH structures and loop structures of the conserved NBD and SBD were more flexible. Comparison between crystallographically resolved structures shows high variability between SBDs, albeit a cysteine, a histidine and an aspartic acid residue for the coordination of the catalytic zinc ion are widely conserved. In contrast, the NBD seems well conserved throughout the different species and consists of the well characterised Rossmann fold, a $\beta/\alpha/\beta$-motif [Rao and Rossmann, 1973] (figure 1.11).

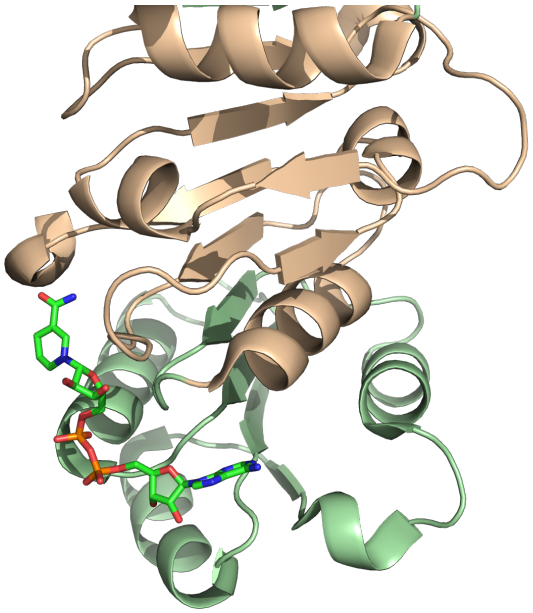


Figure 1.11: Rossmann fold with four α-helices and six β-parallel sheets: nicotinamide binding half (wheat) and adenine binding half (pale green)

Specifically, two residues located in the conserved glycine-rich sequence motif [Gly-x-Gly-x-x-Gly] at the C-terminal end of the second β-strand of the $\beta/\alpha/\beta$-dinucleotide fold are considered to be predictive for coenzyme dependency. If these residues are aspartate or glutamate, NAD^+ will be preferred, whereas if there is a glycine or an arginine, $NADP^+$ specificity is predominant [Pire et al., 2009]. The positively charged residue is predicted to form a salt bridge with the phosphate moiety at the 2'-position of AMP. A noteable halophilic adaptation in a glucose dehydrogenase was reported, where the 2' phosphate of the adenine ribose was additionally coordinated by a nearby cation cluster of two potassium ions, which in turn showed water mediated interactions with the carboxyl group of

an aspartate at the acidic C-terminal tail of the protein [Britton et al., 2006]. Further mechanistic features of the SBD are a hydrophobic pocket formed by certain conserved residues of Leu, Val and Ala, to accommodate the adenine ring of the cofactor, and binding of the nicotinamide moiety is promoted through polar interactions with conserved Val, Ile and Thr residues. Molecular interactions of the pyrophosphate and ribose moieties involve formation of hydrogen bonds.

Besides this, most medium chain ADHs have additional structural zinc bound to four cysteine residues, which maintains the structural stability of the oligomer. However, in *Sulfolobus solfataricus* ADH one of the four cysteine residues is replaced by a glutamic acid residue, which functions to enhance the rigidity of the lobe structure by forming a hydrogen bond network, contributing to increased thermostability of the enzyme.

1.4.1 *Haloferax volcanii* Alcohol Dehydrogenase 2

Alcohol dehydrogenases comprise a group of a multitude of isozymes that are thought to have originated from an formaldehyde dehydrogenase, identical to a class III ADH. Medium chain ADH have multiplied by several gene duplications, forming classes II, I, IV, with the class III/I separation at early vertebrate times. *H. volcanii* has two annotated alcohol dehydrogenases ADH1 and ADH2, both belonging to class I, whereby ADH2 presumably evolved out of a duplicate version of the ADH1 gene. *Hv*ADH2 (37.8 kDa) was first characterised in 2012 as a haloalkaliphilic and thermoactive enzyme for the oxidative reaction that catalyses the reductive reaction at slightly acidic pH [Timpson et al., 2013]. The protein has a tetrameric tertiary (quarternary) structure and is known to form heterodimers /-tetramers with ADH1. It is a zinc dependent ADH with preference for medium chain alcohols, showing highest activity with 1-butanol in 4 M potassium chloride and pH 10. It has a preference for KCl over NaCl and is maximally active between 85 and 90°C. In contrast to *Hv*ADH1 it features a Gly residue at the C-terminal end of the second β-strand and accordingly displays a strong preference for the phosphorylated coenzyme, albeit low activity was also demonstrated with NAD^+. The reason for focusing attention on *Hv*ADH2 over *Hv*ADH1 is its overall higher activity and broader substrate scope with additional acceptance of branched substrates, e.g. isoamyl alcohol and benzyl

alcohol. Binding pockets that are able to accommodate bulky substrates are especially desirable for the synthesis of chiral products with regards to pharmaceuticals [Cassidy et al., 2017]. The enzyme further displayed positive cooperativity in congruence with the aforementioned mechanistic aspects of catalysis of ADHs (figure 1.10 and 1.11) and appeared to follow Michaelis-Menten kinetics for the substrate acetaldehyde.

One particular reason why this specific halophilic protein was chosen for the thesis is its demonstrated remarkable resilience against a range of organic solvents, specifically dimethyl sulfoxide (DMSO) and methanol (MeOH) [Alsafadi and Paradisi, 2013]. Stereo- and regio-selectivity is often essential to create optically pure chiral molecules in the pharmaceutical industry and it is desirable to avoid multiple protection and deprotection steps of molecules. Enzymatic reactions can circumvent these steps, but productivity for a broad range of substrates in traditional aqueous enzymology is limited by sparingly water-soluble ketone substrates, necessitating the application of organic solvents. This and the incompatibility with the solubility of the required redox cofactor NAD(P)H in non-aqueous media, such as DMSO, makes an aqueous organic solvent mixture for biocatalysis highly desirable. Hence, solvent tolerant enzymes that naturally remain stable are highly demanded. Accessibility of hydrophobic substrates plays especially a vital role in asymmetric reduction and *Hv*ADH2 demonstrated absolute enantioselectivity for the synthesis of enantiopure aromatic secondary alcohols [Alsafadi et al., 2017]. According to the enzymatic selectivity for the substrate in the chiral centre, stereospecific ADHs are referred to as either (R)- or (S)-specific [Zhu and Hua, 2010]. The stereoisomer of a chiral molecule is defined according to the Cahn-Ingold-Prelog (CIP) rule [Cahn et al., 1966]. The stereochemical reaction depends by definition on the attack of the pro-R- or pro-S-hydride of NAD(P)H on the *si* or *re* face of the sp^2-hybridised carbon atom in the C=O moiety (figure 1.12). Hence, stereoselective ADHs conform either to Prelog's or anti-Prelog's rule. Docking studies indicate, that Prelog complexes have shorter distances between benzylic carbon atoms of the substrate and the nicotinamide of the coenzyme as well as stronger hydrogen bonds between the nicotinamide and a tyrosine residue, when compared to anti-Prelog. However, enantioselectivity is not only based on the steric confirmation of the binding pocket, it also depends on the chemical structure of the substrate and most dehydrogenases show mixed selectivity. In this regard *Hv*ADH2 is extremely

exceptional as it exclusively follows Prelog's rule, yielding only (S)-alcohols, regardless of the aromatic ketones tested, making it a highly desirable enzyme for enantioselective catalysis.

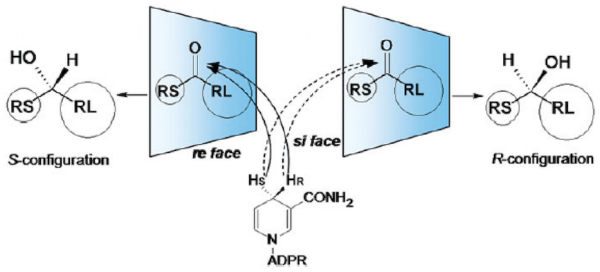


Figure 1.12: Stereochemistry of hydride transfer from NAD(P)H to a carbonyl moiety.

In addition to this, the coding sequence for the enzyme ω-transaminase1 (*gabT1*) is found adjacent to *adh2* on the genome of *H.volcanii* (figure 1.13).

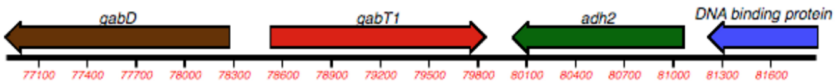


Figure 1.13: Excerpt of *H. volcanii* genome, showing the genes encoding for ADH2 and TA1.

Transaminases convert ketones, the metabolic product of alcohol dehydrogenases, to amines (figure 1.14) and hence these two enzymes may be part of a shared metabolic pathway, as has been shown for other genetically linked proteins [Valencia and Pazos, 2002]. Pathway mapping did not indicate an interaction (*personal communication* with Paradisi group), but the implementation of an enzyme cascade for the production of enantiopure amines is of special interest for the pharmaceutical and agricultural industry, as these molecules serve as builiding blocks for valuable fine chemicals [Nestl et al., 2011].

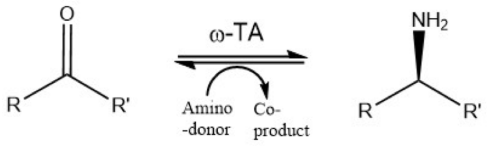


Figure 1.14: Asymmetric synthesis of a chiral amine catalysed by ω-transaminases.

1.5 Ionic Liquids in Green Chemistry

Ionic liquids (ILs) consist of molecular ions and counterions, of which at least either the anion or the cation are irregularly shaped. This irregular molecular shape is the basis for their physical property of having low melting points (below 200 - 100 ℃), since constant fluctuations of H-bonding between molecules and asymmetrical charge distribution within the molecules prevent the formation of a proper crystal lattice, which is found in common salts. As is the case for all classes of solvents, characterisation and prediction of the liquid state through first principle models has remained largely elusive to this day despite decades of research [Eyring and Hirschfelder, 1937, Hildebrand, 1981, Israelachvili and Wennerström, 1996, Lennard-Jones, 1940, Sterpone et al., 2012]. Structure determination at atomic resolution depicts solids adequately, but the fast molecular interactions in liquids evade exact measurement through existing tools. Bulk properties such as relative permittivity, that arise through these interactions, can be measured thermodynamically and reflect the molecular ensemble as a whole, but cannot compensate the lack in understanding of the underlying intermolecular interactions that govern physical properties. Identifying direct links between the network of intermolecular forces based on molecular properties of molecules and the resulting bulk properties of corresponding solvents would facilitate the prediction of structural and thermodynamically driven interactions of the liquid state and hence provide a means of estimating a wide range of physical properties relating to phase transfer equilibria, solvent effects on reactivity and solvation of macromolecules amongst others.

Investigation of the physico-chemical properties of ILs and the design of ILs with desired properties has found increasing interest, due to pressing environmental issues regarding conventional organic solvents and the option to satisfy divers application fields such as electrochemistry, synthesis, catalysis, separation and material science. The divers range of ILs serving the different fields of application can be classified into ion families. Widely

acknowledged classifications are: protic, aprotic and metal-complex ion families, as well as more recently described families of solvate ILs emerging from metal salts and chelating agents, hydrate ILs, which are particularly valuable for biotechnological applications based on their H-bonding interactions and finally salts with high molecular ion structures with importance for medicinal formulations [Kar et al., 2019].

With the exception of the enzyme class of Lipases, proteins rarely show activity in pure ILs. Biocatalysis is hence mostly performed in IL-solvent mixtures [van Rantwijk and Sheldon, 2007]. Different molar ratios between ILs and molecular solvent lead to different structural trends in IL-solvent mixtures and provide an even larger repertoire of tuneable combinations of properties, that can be exploited for different applications and for example can be tailored to proteins. Following an increase of ions over molecular solvent, it is generally distinguished between dilute electrolytes, ion pairs and neutral clusters, charged aggregates, 'diluted' IL and pure IL, (figure 1.15) [MacFarlane et al., 2017].

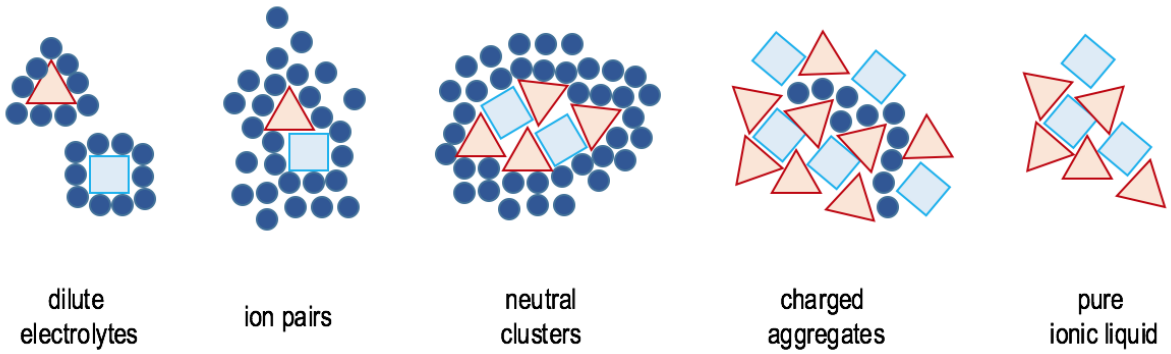


Figure 1.15: Mesoscopic structures following an increase of ions (red and blue) over molecular solvent (dark blue).

Many thermodynamic and transport properties in these mixtures follow approximately simple mixing rules. However, nano level structural effects on the mesoscopic scale are seldomly directly obvious in these bulk properties and require inspection for any given IL, since they might affect for example protein surfaces through direct molecular interactions [Zhang et al., 2017].

1.5.1 Effects of Solvation of Ionic Liquid Ions on Proteins

Aprotic (AILs) and protic ILs (PILs) were distinguished early in the process of classification. While protic ILs (e.g. EAN, PAN) can be easily defined by their ability of proton transfer from an equimolar combination of a Brønsted acid and a Brønsted base [Greaves and Drummond, 2008], creating H-bond donor and acceptor sites on respective ions, such a clear distinction cannot be made for aprotic ILs (e.g. imidazolium, pyridinium, alkylammonium), of which some have the ability to H-bond and some do not [Hayes et al., 2015] which is a major criterion for the interactions impacting protein solubility. In protic ILs the to-be-cation (conjugate acid) becomes the primary H-bond donor (was the H-bond acceptor in form of a Brønsted base) and the H atom is covalently bound to the heavy atom carrying the formal charge, usually N or P. The predecessor to the anion (conjugate base) is a neutral acid (Brønsted acid), which can also form H-bonds with already fully formed anions [Hunt, 2017]. Proton transfer is a chemical equilibrium, which impacts conductivity, vapour pressure, thermal stability, catalytic activity and protein stabilisation. This thermodynamic (i.e. intrinsic and unavoidable) equilibrium process can produce a non-negligible population of neutral species such as ion pairs, impacting the structure of the liquid and leading to an incompleteness of the proton transfer. This incompleteness is an issue with regards to the ensuring and verifying the exact stoichiometry in PILs [MacFarlane and Seddon, 2007]. A small excess of one species over the other can lead to a significant difference in effective proton activity and the obtained liquid might be a mixture of ions and neutral species. In aqueous solution, i.e. in hydrate ILs, a similar neutralisation equilibrium has been observed and has to be taken into account [Mandai et al., 2014, Welton, 2018]. However, the mechanism of proton transfer within the extended hydrogen bond network in PILs and also aqueous mixtures of ILs is not well understood. Thus making predictability of proton movement in the hydration shell of proteins even more challenging. Since proteins themselves posses charged species on their side chains (especially halophilic proteins), conductivity or rather charge transfer is a quality of ionic liquid mixtures, that is of uttermost importance.

Conductivity of ILs cannot be fully understood solely based on the degree of dissociation, rather the mechanism of charge transfer and the size of charge carrying ions is of great importance. Vehicular motion, proton hopping and the Grotthuss mechanism have been

observed in PILs as mechanisms of charge transfer (figure 1.16) [Hayes et al., 2015, Miran et al., 2014, Wojnarowska and Paluch, 2015].

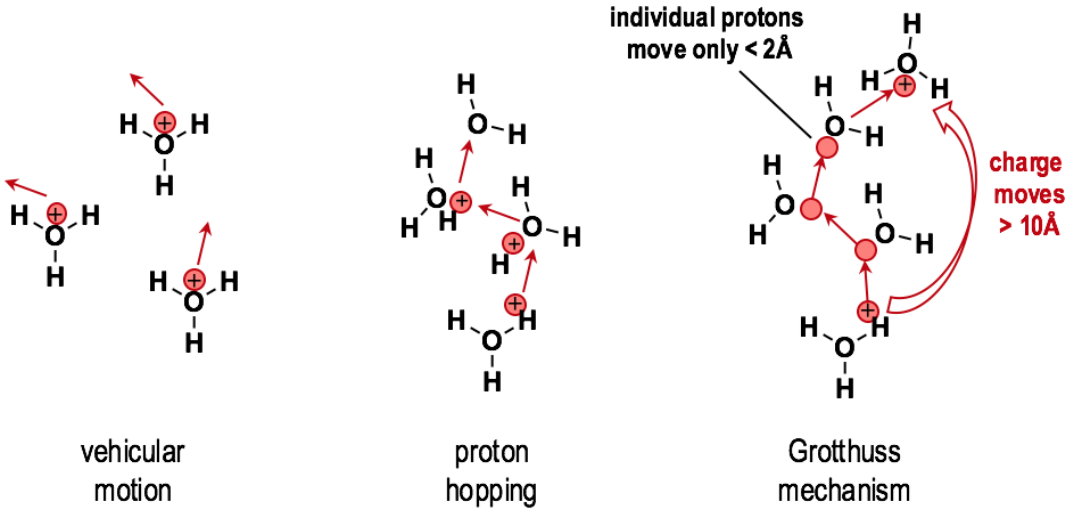


Figure 1.16: Vehicular motion: translational diffusion of carrier ions. Proton hopping: protons jump between neighbouring molecules through the H-bond network. Grotthuss mechanism: only charge moves

Proton migration by translational diffusion of large species (vehicles) like H_3O^+, that can easily slip through the structure via channels is most important with regards to hydrated ILs surrounding protein shells. Consequently, steric arrangements of the structure and resulting viscosities control this kind of charge transfer. The Grotthuss mechanism, in which the protonic defect is moved from one 'electrode end' of the network to another end, despite the protons only being transferred between two neighbouring molecules, on the contrary has implications for pure ILs and aggregate networks of ILs, since the proton transport is decoupled from the ions diffusion and this means that the charge transfer is much faster than the species (protons) controlling the structural relaxation. This makes these kinds of ILs invaluably important for electrochemistry, but will also have implications for IL networks surrounding charged protein surfaces and proton and charge transfer at catalytic sites, which needs to be acknowledged.

Equally important to understanding the mechanistic details of charge transfer, is the possible quantitative analysis of the behaviour of charged species in bulk solution, i.e. the quantification of H-bond donor (HBD) and H-bond acceptor (HBA) strengths of solutes.

The Gibbs free energy change, $\Delta\ G^{\circ}$ = -(α - α_s)(β - β_s) + 6, for the formation of a H-bonded complex in any solvent can be predicted based on HBD (α) and HBA (β) parameters [Cook et al., 2007]. This enables the prediction of the stability of a given complex as well as the solvents polarity [Hiraga et al., 2018]. HBA and HBD scales were first introduced by Abraham through analysis of experimentally derived association constants [Abraham and Platts, 2001] and have been recently expanded by Hunter to explicitly include the influence of the solvent. They were able to obtain self-consistent (i.e. transferable between different solvents) HBA and HBD parameters for a set of anions [Pike et al., 2017] and cations [Pike et al., 2019] commonly used in ILs. Importantly, they found no correlation between the H-bonding properties of the anions and the pKa values of the conjugate acids, which is considered to be a measure of proton transfer from an acid to a base. Based on the HBD-HBA solvent pair mixtures of similar polarities, but different HBD/HBA abilities, solvents can be tuned and compositions will have practical implications. For instance, water has the ability to both donate (alpha = 1.17) and accept protons (beta = 0.47) and if it is mixed with an HBA solvent of low acidity (high basicity) HBD-HBA complexes will form alongside self-associated water structures, having implications for the electronic environment of the IL mixture. Providing a 'mild' H-bonding environment with a relatively low water activity, for instance, was shown to prolong long term stability of DNA dissolved in hydrated choline based ILs at ambient temperatures exceptionally, since the rate of hydrolytic reactions, that would degrade the macromolecule, are considerably slowed down in this kind of media [Vijayaraghavan et al., 2010]. The IL cations were observed via MD and spectroscopic measurements to bind to the groove of the helix through hydrophobic and polar interactions and thereby disrupt the water cage around the DNA, preventing hydrolytic reactions.

Similarly to the diversity of proton transfer mechanisms in ILs, is the unusual nature and wide variety of hydrogen bonds [Hunt et al., 2015]. The commonly recognised 'ionic' H-bond, which is established between a neutral molecule and a charged ion, has to be distinguished from H-bonds in ILs, where two ions interact via a 'doubly ionic' H-bond [Hunt, 2017]. Most H-bonds in ILs are of low strength, making them highly dynamic at ambient temperatures, and unlike molecular solvents, such as water, there are not one donor-/acceptor-type H-bonds, but rather whole networks are formed through bifurcated

or trifurcated and chelating H-bonds. In aprotic ILs the H-bond donor is usually a C-H moiety or else N-H and O-H. Most H-bond acceptors are halogen, oxygen or nitrogen atoms. Cations as well as anions with H-bond acceptor functional groups in their alkyl moieties, such as alcohols, amines and carboxylic acids, can also H-bond together with like charged ions. With the presence of multiple donor and acceptor sites the accessible H-bond network is extensive and it can be highly disordered with vacant H-bonding sites. However, if the number of donor and acceptor sites are almost equal and if these are located in sterically accessible positions across both ions, an associated fluid is established and forms a supramolecular solvent structure [Schröder et al., 2000]. In contrast to molecular liquids, where strong, linear H-bonds lead to a more structured liquid, H-bonds in ILs promote directional interactions at the expense of electrostatic forces, by inducing defects in the Coulomb lattice, which then favours higher order arrangements [Dong and Zhang, 2012, Fumino et al., 2009]. An even wider range of distinctive H-bonding interactions is expected to take place in hydrate ILs, however quantum chemical calculations are needed in conjunction with MD simulations to elucidate this kind of networks and these are computationally expensive for large systems.

Addition of water induces charge flow between oppositely charged ion species through interaction with the ions and by this decreases the Coulomb attraction between them, i.e. screens the electrostatic interaction, without changing the interionic distance between them, since water molecules do not necessarily become interstitial [Konar et al., 2019]. However, dielectric measurements indicate that solvent assisted ion pairs are much more common than direct contact ion pairs or aggregates [Bester-Rogac et al., 2014, Laage et al., 2011]. Small anions of hydrophilic ILs interact strongly with directly surrounding water molecules, however they do not influence further solvation shells [Omta et al., 2003]. They were also shown to interact as mediators between the water molecules and cations, which are to some extent expelled and form clusters similar to micelles [Schröder et al., 2007]. The formation of these structures are driven by entropic forces, since such apolar domains minimise the disturbance of the H-bond network of the water molecules. This surfactant effect gets stronger, the longer the cationic alkyl chain and this can affect the stability and activity of proteins positively through suppression of protein-protein interaction, which prevents aggregation [Tariq et al., 2012b]. However, if coulombic interactions between

the ions are strong and moreover if such ions are substituted with hydrophobic tails, strong dispersion forces are observed. While this does not have immediate implications for proteins as long as these ion cluster are located in the bulk solvent, it is likely, that such ions are expelled from the hydrogen bond network at interfaces and may act as surfactants for solvated proteins.

1.5.2 Alcohol Dehydrogenases in Ionic Liquids

Alcohol dehydrogenases (ADHs) are valuable biocatalysts due to their enantioselective reactions with prochiral ketones and alcohols, as detailed in *Introduction, section 1.4*. However, productivity for a broad range of industrially interesting substrates is limited by the incompatibility of the solubility of the required redox cofactor NAD(P)H in non-aqueous media and opposingly, sparingly water-soluble ketone substrates. Ionic liquids provide a means to establishing compatibility with the biocatalytic reaction.
A range of ionic liquids have been investigated with ADHs. The mechanism of impact of ionic liquids on ADH can be interpreted, such that a classification into: (a) hydrophilic ionic liquids (b) water-immiscible ionic liquids, which drive the partition coefficients of substrates and products in favour of the enzyme without direct interaction and (c) water-miscible, amphiphilic ionic liquids, which provide H-bonding through the hydrophilic part and substrate solubility within their alkyl-tail-pockets, is possible. The classification of hydrophilic ionic liquids can be further divided based on their central ion into either, sulfonic, phosphatidic or halides for anions and sulfonium, phosphonium, ammonium or imidazolium for cations. Amphiphilic ionic liquids can be partioned into hydroxy-functionalised and oligoether-based ILs.

Hydrophilic ionic liquids

The impact of hydrophilic ILs on enzymatic reactions can be interpreted under the perspective of either ion effects or water effects.

Ion effects

The effect of hydrophilic ions, such as Me_4N^+, choline, and imidazolium cations, and Cl^-, Br^-, BF_4^-, PF_6^-, OTf^-, $Me_2PO_4^-$, $Et_2PO_4^-$, $MeSO_3^-$ anions, has been examined for

a number of ADH systems. For the ionic liquids [EMIM][Cl], [BMIM][PF_6], [BMIM][Cl] and [BMIM][Br] an improvement of activity of HlADH was demonstrated at low concentrations between 0.05 and 0.075 M, compared to pure buffer systems [Shi et al., 2006]. Comparability of results in this field is challenging as reported results vary between conversion rates, yields and residual activity. Regardless, conversion rates and half-life of enzymes generally show improvements upon addition of aqueous ionic liquids, relative to buffer solutions at low to moderate concentrations. In particular, improvements have been for [BMIM][OAc] and [EMIM][OAc] at 20 % (v/v) [de Gonzalo et al., 2007] and for [EMIM][$MDEGSO_4$], [EMIM][$MeSO_3$] and [MMIM][Me_2PO_4] at 10 %(v/v) with respect to conversion rates [Kohlmann et al., 2011] and for [EMP][ES], [BMIM][BF_4] and [EMIM][TOS] at 10 % (v/v) with respect to half-life stability [Hussain et al., 2008].
Special care must be taken when deploying imidazole and imidazolium derivatives for alcohol dehydrogenase systems, because of their similarity in structure to parts of the cofactor $NADP^+$,[Dabirmanesh et al., 2012] have been reported as exhibiting competitive inhibition of the active centre of ADHs,[Zhang et al., 2008] and with smaller alkyl chains lengths for these imidazolium-based ionic liquids the inhibitory effect increases[Dabirmanesh et al., 2012]. Further, with regards to deploying anions PF_6^- and BF_4^- it has to be acknowledged, that these ions can be hydrolysed in aqueous solutions. While PF_6^- hydrolysis is mainly promoted only under acidic conditions so that the PF_6^- ion can be suitable for use at moderate temperatures in aqueous solutions, activity in BF_4^- might not be fully attributed to the biocatalyst, as it is not stable in water and will hydrolyse readily.

Chloride-ion based ionic liquids do not seem to act as a challenge to ADH-systems. For HlADH assayed in [BMIM][Cl] a maximum relative activity of 150 % was observed at 0.075M and decreased slowly to 25 % at 0.4 M ionic liquid concentration. Use of [EMIM][Cl] lead to a maximum relative activity of 150 % at 0.05 M, but activity values were ∼50 % less on average at higher concentrations (0.1 - 0.3 M) than for [BMIM][Cl] [Shi et al., 2006]. However, activity ended up being about the same at 0.4 M in both ionic liquids, which was the highest concentration assayed. In comparison to HlADH, YADH activity is not stimulated in [BMIM][Cl] or [MMIM][Cl]. [BMIM][Cl] lead to a slight decrease in activity (∼95 %) at very low concentrations (∼0.01M), but activity

(∼20 % remaining) was maintained up to a concentration of ∼0.6 M ionic liquid, while for [MMIM][Cl] only 80 % remaining activity was observed at 0.01 M and activity had completely declined at ∼0.2M [Dabirmanesh et al., 2011].

Of the other hydrophilic ionic liquid systems that have been assayed, [EMIM][Et_2PO_4] is strikingly detrimental to the enzymatic activity of L. brevis ADH with only 5 % relative activity left at 10 % (v/v) for the reduction of 2-octanone. This contrasts with [MMIM][Me_2PO_4], with 85 % relative activity, and is comparable to the activity of YADH found for [MMIM][Cl] (above), and [MMIM][$MeSO_3$] with 105 % relative activity [Kohlmann et al., 2011].

Simulation studies have been able to shed light on the specific interactions that may be at play in controlling activity. Ionic liquid ions with low charge density (most cations) do not tend to exhibit any preference of vicinity on the protein surface, since coulombic forces for these ions are less dominant due to delocalisation and are competitive with the van der Waals interactions of the alkyl chains. Thus, these ions tend to be more mobile on the protein surface, than those ions with high charge density (most anions) [Lim and Klhn, 2018, Tom et al., 2012] and this gives a rationale for the relatively more significant impact ascribed ionic-liquid anions. At all concentrations, the cations with their alkyl chains are able to contact like-charged-ions, but at low concentrations, since there are only a few about, they will, as single ions or small clusters, be able to closely associate to the protein surface.

The shorter the alkyl chain, the smaller are the polar domains that can interact with other species meaning, [MMIM] or [EMIM] can associate more closely and individually onto the surface of the protein and more easily into the binding pocket, than, for example [BMIM]. The peak in activity at very low concentrations of ionic liquid, and the consequent decrease in activity at increasing concentrations of ionic liquid may be attributed to the different numbers of cations being able to aggregate on the protein surface. As the cations continue to coordinate to the protein surface at increasing concentrations, [Lim and Klhn, 2018] the ions form ever bigger patches, that are still small enough to amass onto the protein surface, but with the increasing patch-size an increased rigidity of the protein is conferred. The discrepancy of ∼50 % less activity between [BMIM] and [EMIM] at medium range concentrations (0.1 - 0.3 M) underlines the different interactions of the

ions with the protein surface, as [EMIM] having shorter chains will be able to associate more closely overall. This close association allows an optimum effect on the flexibility (and activity) of the protein at a slightly lower concentration (0.05 M), than [BMIM] with 0.075 M, and at the same time having a much higher impact on constraining the protein flexibility at medium concentrations. At low ionic strength, the ring and tail of the imidazolium ions are equally often in contact with the protein surface, whereby the tail mainly has a steric interaction (fitting into clefts on the surface) and the ring has a mostly charged interaction, which becomes more pronounced and dominant at higher ionic strengths [Haberler et al., 2011]. In comparison, high-charge density ions, if located on the surface, preferentially reside at positions with oppositely charged amino acids and hardly associate with like-ions. At low ionic liquid concentrations the number of low charge density ions populating the protein surface is far greater than the number of high charge density ions, that prefer the hydration of water molecules in the bulk phase, [Haberler et al., 2011] leading to asymmetric influence at different concentrations on the protein.

Anions with high charge density, being dispersed in bulk water at low concentrations, may strip off a small number of tightly bound, structuring water molecules from the protein surface, leading to an increase in entropy and hence more flexibility and activity of the enzyme, up to the maximum point, upon thereafter the disruption of the secondary structure and subsequent rigidity of the protein (as seen for cellulases) becomes detrimental. This can affect the binding pocket significantly, as was proposed by the study on the enantioselectivity of Thermoanaerobacter ethanolicus ADH [Musa et al., 2008]. In aqueous solution, the binding pocket of T. ethanolicus ADH includes a bound, ordered water molecule, which is displaced when a bulky (R)-substrate binds, yielding a substantial entropy increase [Heiss et al., 2001]. For smaller substrates there is no enantioselective conversion in aqueous solution because they are unable to displace the water molecule. Addition of ∼50 % (v/v) of the hydrolysable ionic liquid [BMIM][BF_4] was able to facilitate the asymmetric reduction of small phenyl-ring containing ketones, with this result suggested as indicating the expulsion of the tightly bound, structured water molecule by the ionic liquid [Musa et al., 2008]. A similar result was obtained using ∼ 50 % (v/v) of the water immiscible [BMIM][NTf_2], where the lower concentration of substrate in the

aqueous phase was used to account for the increase in enantioselectivity, highlighting that a more microscopic view of the reaction may be required to fully interpret the role of the ions.

Water-effects

The change in activity seen for ADHs, spiking around one mole fraction of water ($H_2O \sim 0.9$) highlights the finding of a magic point, reported in a computational solvation study of a small zinc-finger protein in aqueous ionic liquids. [Haberler et al., 2011] This point hints at a change in the organisation of the ions and water molecules and appears to correlate with a transition from dipolar screening (achieved by small water molecules interlacing with protein residues and attenuating residue-residue interactions) to charge screening (by ion-charge interactions) of electrostatic forces. At this transition point the protein is at its highest total energy and lowest level of ordering of the secondary structure [Haberler and Steinhauser, 2011, Haberler et al., 2011]. At comparably high ionic strengths the mean residence times of ions, especially anions, on the protein surface, i.e. their interaction with specific residues, is raised from picoseconds to nanoseconds, [Haberler et al., 2011] and a weakening effect on the binding affinity of the substrate to the enzymes is observed [Dabirmanesh et al., 2011, 2012].

Furthermore, at increasing concentrations of ionic liquids, simulation studies show that the water molecules change their behaviour significantly [Gehrke et al., 2017]. At relatively low mole fractions of water (0.38) water molecules tend to form larger clusters of around 12 molecules, and they are apparently present either as single molecules or on-average as dimers at higher mole fractions of $\sim$0.82 [Gehrke et al., 2017]. Here, the underlying structuring force of the van der Waals-induced, tetrahedral-coordinating water molecules becomes replaced by an ionicity driven network, which removes interstitial water molecules as the number of high-charge density ions increases in the bulk. This is in agreement with the observation of an overall dielectric decrement at reduced mole fractions of water, which is especially consequential in the first hydration shell [Haberler et al., 2012].

Water-immiscible ionic liquids (biphasic systems)

The water-immiscible ionic liquid [BMIM][NTf_2] was one of the first ionic liquids successfully implemented for improving the enantioselective reduction of ketones with ADH, [Eckstein et al., 2004] and since then biocatalysis in multiphasic ionic liquid reaction systems has advanced to show the potential for industrial scale applications of these systems [Meyer et al., 2018]. In comparison with hydrophilic ionic liquids, for which a maximum tolerance of around 20 % (v/v) of [EMIM][OAc] (the best candidate with regards to conversion rates), was found, biphasic systems can provide meaningful application maxima of ∼80 % (v/v) and still retain the enzymes hydration shell [de Gonzalo et al., 2007]. Since the enzyme and the cofactor are in this case immobilised in the aqueous phase, it is the partition coefficient of the substrate and co-substrate in the ionic liquid and aqueous phase that drives the thermodynamics of the reaction. In this regard, these biphasic systems are optimal replacements for conventional organic co-solvents as they can be ion-specifically selected to facilitate different partition coefficients for particular substrates and co-substrates, as has been shown for 1-hexyloxymethyl-3-methyl-imidazolium and 1, 3-dihexyloxymethyl-imidazolium based ionic liquids with alcohols, water, ketones and hydrocarbons [Domańska and Marciniak, 2007]. In particular, programmes such as COSMO-RS have been successfully validated for predicting phase equilibria for such systems containing imidazolium and pyridinium-based ionic liquids[Freire et al., 2007].

Water-miscible, amphiphilic ionic liquids

Hydroxyl-functionalised ionic liquids, including the AmmoEng^{TM} series of task-specific ionic liquids (Figure 1.17), have been shown to enhance the activity of ADHs even at very high concentrations of up to 90 % (v/v) ionic liquid content,[de Gonzalo et al., 2007] with the optimal concentration dependant on the experimental set-up. Up to 90 % (v/v) is tolerated when whole cell-biocatalysts or lysates thereof are applied,[de Gonzalo et al., 2007] whilst for purified enzyme, solution concentrations of 10 % (v/v) AmmoEng 101 were shown to enhance the relative activity of L. brevis ADH by 180 % [Kohlmann et al., 2011].

A major advantage of the use of these amphiphilic ionic liquids is the much higher than usual solubility of difficult substrates, as has been demonstrated for, for example, 4'-Br-

2,2,2-trifluoroacetophenone and 6-bromo-beta-tetralone [Hussain et al., 2008]. Rhodococcus erythropolis ADH and GDH 103 did not show any residual activity for those substrates when used in buffer, but GDH 103 showed 65 % (w/w) residual activity in 10 %(v/v) AmmoEng 102. Both ADH and GDH 103 enzymes showed 50 % (w/w) residual activity in the water-immiscible ionic liquid [C_4C_1Pyr][NTf2] [Hussain et al., 2008]. The amphiphilic oligoether-based ionic liquids AmmoEng 100 and 101, with a 14-carbon coconut oil-derived group, were found to either impart the most activity or increase conversion by 150 % and 180 %, respectively [de Gonzalo et al., 2007, Kohlmann et al., 2011], with the main difference being the anion ($MeSO4^-$ (100) and Cl^- (101)). AmmoEng 102, which has an 18-carbon tallow-derived group, a five-fold longer oligio(ether-) chain than AmmoEng 101 and a greater ability to absorb water due to the higher ether content [Ribot et al., 2012], was found to increase activity of L. brevis ADH by 110 %. In AmmoEng 140 only a 90 % increase in activity was observed.

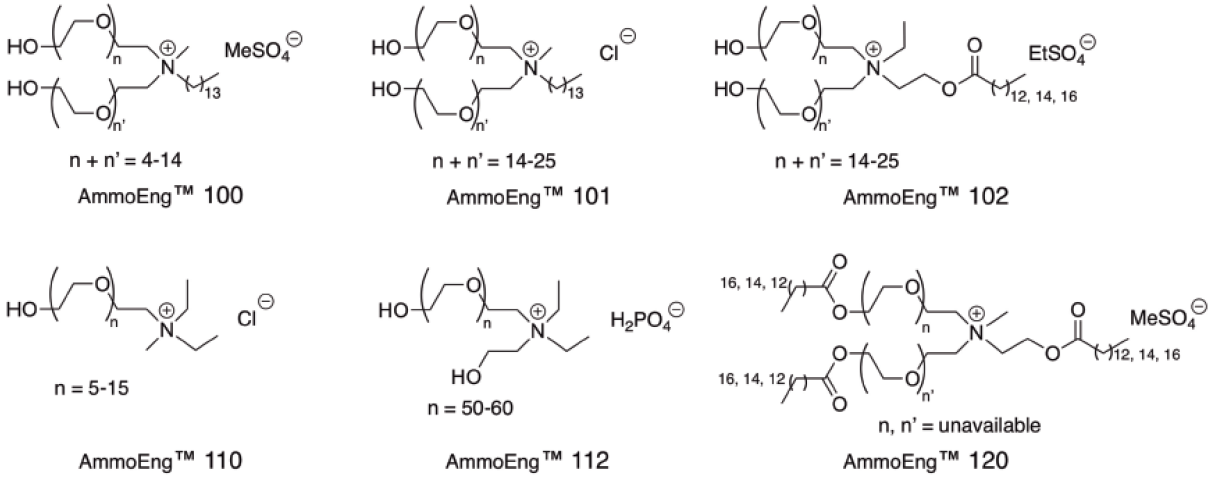


Figure 1.17: AmmoEng^{TM} series of task-specific ionic liquids.

1.5.3 Molecular Dynamics Simulation

A major obstacle for widespread application of ionic liquids for biocatalysis lies in the inability to predict the tolerance of a chosen enzyme to any particular ionic liquid [Kumar and Venkatesu, 2014]. This inability is grounded in a poor understanding of the atomic-scale interactions taking place at the enzyme-ion interface, which are the key drivers for enzyme destabilisation or inactivation [Yang, 2009]. MD simulations were able to broaden

the understanding of fundamental biophysical principles underlying these interactions.

Bacillus subtilis lipase A was recently simulated for 100 ns in [BMIM] in combination with halide anions as well as [TfO^-]. Solvent spatial distribution function analysis revealed a strong binding affinity of the cation towards the enzymes surface based on hydrophobic and π-π stacking interactions [Pramanik et al., 2019]. They observed a lower density for chloride ions at the surface, compared to iodide and bromide. [TfO^-] on the other hand showed a higher tendency to interact with the cation at the enzyme surface. All ILs were found to strip off essential water molecules. They further found for anions a mainly electrostatic driven interaction, whilst for cations Van-der-Waals based interactions with the enzyme surface were prevalent.

Investigations on surface charge modification were done on a 50 ns timescale for *Candida rugosa* lipase and *Bos taurus* α-chymotrypsin in [BMIM][Cl] and [EMIM][C_2SO_4] [Burney et al., 2015]. They found Root mean square deviations (RMSD) were not altered for α-chymotrypsin in ionic liquids compared to water and surface charge modifications towards increasingly negatively charged residues restricted enzyme flexibility only at 75 % of modification in the simulation with [BMIM][Cl]. RMSD values were increased by 0.05 - 0.1 Å for the lipase in ionic liquids regardless of altered surface modifications. In terms of Radial distribution functions (RDF) a slight increase of cation coordination to the surface was observed, whilst the opposite was found for anions for both enzymes. However, anions tended to coordinate closer to the surface compared to cations.

Further, a mesophilic carbonic anhydrase and its engineered halophilic version were simulated for 600 ns in common salts, whereby the sodium cation was paired with Nitrate, Hydrogen sulfate and Chloride [Warden et al., 2015]. Regardless of their pairing with different counterions all sodium ions showed an increased coordination to the enzyme surface for the halophilic version, when compared to the wildtype, whilst again, the opposite was true for the anions, which showed less coordination to the increased negatively charged surface.

More broadly this behaviour of cations coordinating closer can be rationalised by the hydrogen-bonded network of the anions and water in the bulk phase. The cations are not able to contribute to this network and are therefor expelled to the protein surface, where they mainly interact based on their slightly amphiphilic character. Hydrophilic an-

ions tend to prefer the positively charged surface and show strong Coulomb interactions, which result in a longer residence time of anions, when compared to cations. Due to interactions between the ions and the protein surface, water gets displaced, but can have positive and/or negative effects for protein solubility. Hydrophobic ILs on the contrary can form biphasic systems resulting in less depletion of water molecules surrounding the surface and hence increase protein solubility.

1.6 Nanoreactors

Compartmentalisation is one of the key factors essential to all forms of life. It facilitates organisation and metabolism and is involved in movement and communication, since localisation in compartments ensures, that multistep reactions can take place without interference by other compounds and enables the flux in and out of a microenvironment [Lane, 2014]. Enzyme cascades localised in such compartments, for instance, govern cell morphology and allow interactions with the environment through signaling molecules and transport of nutrients and energy [Ivanov et al., 2019].

Mimicry of the properties of these natural constructs through synthetic materials is used to implement nanoreactors for green chemistry [Gaitzsch et al., 2015]. Water as a reaction medium is most attractive from an economic and environmental point of view since it is save and abundant. However, its use is mostly limited to the work-up, due to insolubility of organic compounds [Clark, 1999]. A solution to this problem are nanocompartments, which are able to create a separate microenvironment and enable heterogeneous catalysis, while maintaining the high selectivity of the catalyst. While in homogenous catalysis both the biocatalyst and substrate, are present and molecularly dissolved in the same phase [Tariq et al., 2012a], heterogeneous catalysis refers to biocatalysts being aggregated to a support [Cole-Hamilton, 2003]. Advantageous aspects over non-supported homogenous systems are site isolation of reactive components enabling enzyme cascades, conversion of hydrophobic substrates in water and facile biocatalyst recovery. However, major drawbacks of solid phases are the obstruction of active sites and mass transfer limitations due

to increased reactant diffusion, that arises through formation of temperature gradients along solid supports [Dekker et al., 1995]. Hence, the aim is to simultaneously maintain the homogeneity of the catalyst and substrate and provide their inclusion and confinement. A possible solution is the use of biphasic liquid-liquid systems and compartmentalisation, which enables biocatalyst segregation and shielding. Such forms of catalyst confinement lead to a high local concentration of the substrate at the active site, which results in higher reaction rates and better conversion and also shield the catalysts from their surroundings [Lipshutz et al., 2011].

Extensive literature has been established in this field [Che and van Hest, 2016, Khullar et al., 2013, Thomas et al., 2007, Vriezema et al., 2005], where four major classes of nanoreactors have emerged for application in catalysis. These are micelles, polymersomes, dendrimers and nanogels. Only micelles will be discussed in detail, since it is this class of nanoreactors, that is used in this study.

1.6.1 Micelles

Micelles belong to the class of self-assembled nanoreactors, which are defined as macromolecular architectures, that are non-covalently assembled from their constituent building units. In case of micelles, the constituent parts are amphiphilic molecules, consisting of a hydrophilic head, that can be neutral, anionic or cationic and a hydrophobic head. The term 'head' is used if lipids are the molecules used for micelle formation. However, micelles can also be made up of polymers. In that case hydrophilic an hydrophobic parts are referred to as 'blocks'. Above the critical micellar concentration (CMC) this structure leads to a self organisation into micelles [Dwars et al., 2005], see figure 1.18 (left structure). The driving force for the self-organisation is solvent exclusion and intermolecular interactions such as hydrogen bonds in some cases. Different micellar morphologies can be obtained by varying the packing parameter, which is defined as

$p=v/a_o l_c$,

where v is the volume l_c is the length of the hydrophobic chain and a_o is the optimal area of the head group [Jain and Bates, 2003]. The general rule is, that spherical micelles form, if $p \leq 1/3$ and cylindrical, worm-like micelles form when $1/3 \leq p \leq 1/2$. The typ-

ical micelle has a hydrophobic core, that leads to a concentration effect for hydrophobic substrates and this ensures higher reaction rates than those performed in bulk. Micelles find application in organic synthesis allowing reactions in water [Duplais et al., 2010].

Polymeric vesicles or polymersomes, which derive their names based on the similarity to liposomes, on the other hand, have been successfully deployed for effectively sequestering enzymes in their hydrophilic core, creating a highly concentrated environment for confined catalysis [Nallani et al., 2007]. Similar to liposomes, they are bilayered hollow spheres, that are self-assembled from amphiphilic units, see figure 1.18 (right structure). In case of polymersomes these units are block copolymers, which comprise two or more homopolymer subunits linked by covalent bonds [Discher et al., 1999]. Liposomes in comparison are composed of phospholipids, which can form unilamellar or multilamellar vesicles. These are much less mechanically stable, hence polymersomes are preferred over liposomes for nanoreactor applications [Schoonen and van Hest, 2016].

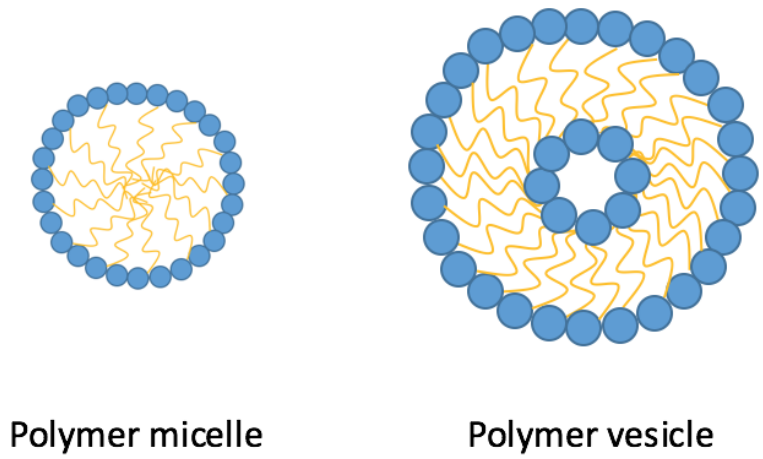


Figure 1.18: Polymer structures that can serve as nanoreactors

1.7 Thesis Outline and Aims

The aim of this study was to investigate the applicability of halophilic enzymes in ionic liquid systems with the intention of developing a lab-scale process for the conversion of alcohols and to further understand the role of molecular interactions, that drive protein stability and activity in aqueous ionic liquid mixtures. Halophilic enzyme-IL systems investigated to date cover halophilic cellulases from *Halorhabdus utahensis* [Zhang et al., 2011], and a row of enzymes engineered towards halotolerance [Burney et al., 2015, Warden et al., 2015]. Most IL-enzyme systems assayed, however, consist of mesophilic enzymes and current reviews are available [Pramanik et al., 2019, Schröder, 2017], demonstrating the diversity of components present in these systems and the great complexity, that follows from this. Investigations in this study were set-up to explore similarities between *Hv*ADH2-IL systems to put forward hypotheses relating halophilic Alcohol Dehydrogenase 2 activity, structure and ionic liquid composition.

Methods for the immobilisation of enzymes applied in bio-processes are numerous [Brena et al., 2013], but entrapping biocatalysts in nanoreactors, such as polymersomes and micelles presents the advantage of creating vessels that can be specifically defined in size, catalyst loading and enable precisely controlled enzyme cascades [Nordwald, 2015, Peters et al., 2014]. It was previously demonstrated, that microfluidics can be used to create multicompartmentalised polymersomes from double emulsions [Perro et al., 2010] and that ionic liquids can be successfully encapsulated to form IL-filled micelles [Guarnido et al., 2019]. So far, no system combining ionic liquid and enzyme in one vesicle has been proposed.

In order to achieve these goals, an existing lab strain of *H. volcanii*, used for the overexpression of proteins, was modified for the heterologous overexpression of Alcohol Dehydrogenase 2 and ω-transaminase 1, see *Chapter 2*. Expression and purifications conditions of ADH2 were optimised in the following, see *Chapter3*. The activity of the enzyme was then experimentally assayed in aqueous ionic liquid mixtures, which were based on thir-

teen different ion pairings, see *Chapter 4*. Following the experimental characterisation, *in-silico* studies were carried out to investigate structural changes of halophilic ADH2, ranging from quaternary rearrangements to residue-residue interactions. These investigations were intended to give insights into the mechanistic role of prevalent acidic amino acid surface residues and reveal details on preferential binding sites of anions and cations to the protein surface and their role in disrupting or enhancing solvating interactions, see *Chapter 5*. To give a first proof of principle for an industrially attractive end application, the system, showing the highest enzyme activity was selected to be encapsulated in polystyrene based micelles for immobilisation with the aim of exploring novel routes for continuous chemical synthesis, see *Chapter 6*.

Chapter 2

Plasmid and Strain Construction

2.1 Background

Haloferax volcanii laboratory strain H26 had been previously genetically modified in multiple steps to yield strain H1895, which is optimised for the overexpression of proteins [Strillinger et al., 2016]. Stepwise genetic modifications are described in detail in section *Strain Construction* below. The laboratory overexpression strain H1895 was used in this thesis to construct an optimised masterstrain for the overexpression of the enzymes alcohol dehydrogenase 2 (ADH2) and ω-transaminase 1 (ω-TA1) and was further optimised for tryptophan-inducible protein expression from episomal plasmids.

ADH2 had been previously shown to be most active in its tetrameric form and to form homo-tetramers as well as hetero-tetramers with ADH1 [Timpson et al., 2013]. ω-TA1 had been shown to form dimers. Genetic deletions of homologous alcohol dehydrogenase 1, alcohol dehydrogenase 2 and ω-transaminase 1 from the genome were therefore required to ensure purification of recombination constructs of tagged-protein, without copurification of native protein. Hexa-histidine and Streptavidin tags are short amino acid sequences enabling the chromatographic purification of proteins. *H.volcanii* strain 1895 had been modified to eliminate co-purification of histidine-rich proteins PitA and Cdc48d, so that metal-affinity chromatography yields sufficiently pure fusion proteins. Early termination

can occur resulting in N-terminal his6x-tagged premature peptides. A second purification step based on the strong non-covalent interaction of streptavidin with biotin can additionally be introduced by fusing a strepII-tag to the C-terminal end to eliminate immature protein contaminants. Besides this, purification of his6x-tagged proteins requires imidazole for elution. Imidazolium had been previously shown to decrease the activity of alcohol dehydrogenases significantly, due to removal of catalytic zinc from the enzyme [Dahl and McKinley-McKee, 1981]. Three strategies were explored in this study to eliminate ADH2 inactivation. First, immediate buffer exchange of imidazolium-eluted proteins had been previously shown to restore ADH2 activity. Second, the chelating agent EDTA can be substituted for imidazolium to facilitate elution, but requires immediate buffer exchange also. Third, providing a strepII-tag as an alternative tag to the his6x-tag in order to circumvent the use of zinc-removing agents altogether. All three possibilities were explored in this thesis and required the construction of differently tagged plasmid constructs. However, issues with C-terminally His-tagged protein and N-terminally StrepII-tagged protein had been previously observed in the Allers lab-group for other proteins. Hence, plasmids for the expression of N-terminal his6x-tagged or strepII-tagged as well as N-terminal his6x-tagged and C-terminal strepII-tagged proteins were constructed. Details and results regarding the experimental investigation of these strategies are given in *Chapter 3: Optimisation of Enzyme Expression and Purification Conditions.*

Simultaneous to this, strain H1895 was further optimised for tryptophan-inducible protein expression through tryptophanase A (*tnaA*) deletion from the genome. Tryptophan is used as an inducer for protein expression under the control of the *p.tnaA* promotor on episomal plasmids [Large et al., 2007] and it is expected, that deletion of tryptophanase, an enzyme which breaks down its inducer tryptophan, will allow for either, the use of less inducer for equivalent protein expression or higher levels of protein expression, which is especially important for cultivation in up-scaled tank reactors. To this end, plasmids for the deletion of *tnaA* were constructed and $\Delta tnaA$ strains were transformed with plasmids for the expression of halophilic β-galactosidase (BGHA) [Holmes et al., 1997] under the control of the *tnaA* promotor.

The *Results* section of this chapter details how plasmids and strains used in this thesis were derived through cloning and transformation procedures mentioned in the *Materials*

and Methods section of this chapter.

2.2 Materials and Methods

2.2.1 Materials

Oligonucleotides

All primers were designed with MacVector Version 14.5.2 (MacVector, Inc.) and synthesised by Eurofins MWG, Germany. Database software FileMakerPro 14.0.2 (FileMaker, Inc.) is used to store information about existing oligonucleotides, plasmids, strains and genomes.

Table 2.1: PCR Primers used in this study

Number	Sequence (5'-3')	Use
o159	GTAAAACGACGGCCAGT	Forward primer for pBluescript
o364	TTAAGTTGGGTAACGCCAGGG	Reverse primer for pBluescript
o627	GCCTGCCGATTACTTCA CATTCGC	Forward primer internal to *p.tnaA* promoter
o1415	CAAGAACGTGCGCTTCCAGC	Forward primer for *adh2*FR
o1422	ACGAGGACCTGCTGGTGTTGGG	Reverse primer for *tnaA*
o1429	CAGGCtGaTCAGGCCA GGTCGTTCAGCGCGG	Reverse primer for *gabT1* for pTA1392 (6xHis expression)
o1430	GGAACacaTGTTCATGA GAGAACGGACGACG	Reverse primer for *gabT1* for pTA1392 (6xHis expression)

o1485	ACGGTCTTCGAGTTCG CGCAGGTCACG	Reverse primer for *adh2*FR
o1504	GTCGGGGTAcCCCGCCCGCTCG AAGACGG	Forward primer to amplify *gabT1* for pTA131 (deletion construct)
o1505	GGCGTCTAgAAGCTCACCGTCG AACTGGC	Reverse primer to amplify *gabT1* and *adh2* for pTA131 (deletion construct)
o1512	GTGTGctagcCTCGTCGACTTTC ACCATGACC	Reverse primer to amplify *adh2* for pTA1392 (6xHis-StrepII expression)
o1513	ACCGgctagcGACTCGTTGCGTTT CGGCCG	Reverse PCR primer to amplify *gabT5* for pTA1392 (6xHis-StrepII expression)
o1514	TAGTGctagcGGCCAGGTCGTTC AGCGCG	Reverse primer to amplify *gabT1* (6xHis-StrepII expression)
o1562	TTATCAACATATGGGACGACA GT	Reverse primer for Δ*gabT1* for pTA131 (deletion construct)
o1573	AGGATCCAAACTGAATCCTAAT AAATT	Reverse primer for Δ*gabT1*

*capital letters represent perfect matches to the genome sequences and small letters represent missmatches

Plasmids

Plasmid construction is detailed in *Chapter 3, Plasmid Construction*. All plasmids were confirmed by restriction digest and sequencing.

Table 2.2: Plasmids used in this study

Plasmid	**Origin**	**Use**	**restriction sites**
pTA131	pGB70 pBluescript II SK+	Deletions in Δ*pyrE2* background	BamHI/XbaI
pTA1229	pTA131	*adh1* flanking regions for deletion	BamHI/ ClaI/XbaI

pTA1230	pTA131	*adh2* flanking regions for deletion	BamHI/ KpnI/XbaI
pTA1743	pTA131	*adh2* and *gabT1* with flanking regions	KpnI/XbaI
pTA1754	pTA1743	*adh2* and *gabT1* flanking regions for deletion	BamHI
pTA1755	pTA131 pTA875	*tnaA* with flanking regions	BamHI/XbaI
pTA1756	pTA1755	*tnaA* flanking regions for deletion	AgeI
pTA1503	pTA875	15 bp flanking region of *tnaA*	AgeI
pTA1508	pTA1503	15 bp flanking region of *tnaA*	BamHI/XbaI
pTA1185	pTA9	*hdrB* open reading frame	NcoI/SphI
pTA1730	pTA1717	$tnaA\Delta EcoN$	BamHI/XbaI
pTA1762	pTA1756 pTA1185	$\Delta tnaA::hdrB^{+}$	AgeI BamHI-HF
pTA875	pIL11	*tnaA* with flanking regions	BspHI/NcoI
pTA963	pTA230	For tryptophan inducible gene expression of N-terminal 6xHis-tag protein	KpnI/NdeI
pTA1205	pTA963	Expression of 6xHis-*adh2*	PciI/BamHI
pTA1543	pTA941 pTA1228	Expression of *bgaHa* reporter	NdeI/NotI
pTA1544	pTA941 pTA1228	Expression of *bgaHa* reporter	NdeI/NotI
pTA1392	pTA1228	Template for N-terminal 6xHis-tag and StrepII tag expression	PciI/EcoRI
pTA1670	pTA1392	Expression of 6xHis-*gabT5*	PciI/EcoRI
pTA1703	pTA1392	Expression of 6xHis-*gabT1*	NcI/EcoRI
pTA1759	pTA1392	Expression of 6xHis-*gabT1*-StrepII	NdeI/NheI
pTA1760	pTA1392	Expression of 6xHis-*gabT5*-StrepII	NdeI/NheI
pTA1761	pTA1392	Expression of 6xHis-*adh2*-StrepII	NdeI/NheI
pTA1863	pTA1392	Expression of StrepII-*adh2*	PciI/EcoRI

TA = Thorsten Allers

***Haloferax volcanii* strains**

Table 2.3: Strains used in this study

Strain	Genotype	Notes
H26	*pyrE2+::[ΔpyrE2 NovR], ΔpyrE2*	Constructed by TA [Allers et al., 2004]
Parent: H1895	*ΔpyrE2, Nph-pitA, Δmrr, ΔhdrB,* textitCdc48d-Ct, ΔpilB3C3	Constructed by TA [Strillinger et al., 2016]
H1969	{*p.tnaA::bgaHa*}	overexpression pTA1543
H1970	{*p.tnaA::bgaHa*}	overexpression pTA1544
H2466	{*p.tnaA::6xhis-gabT5*}	overexpression pTA1670
H2467	*adh1+::[Δadh1]*	Integration pTA1229
H2468	*adh2+::[Δadh2]*	Integration pTA1230
H2684	*tnaA+::[tnaAΔEcoN]*	Integration pTA1730
H2685	*tnaA+::[ΔtnaA]*	Integration pTA1756
H2686	*tnaA+::[tnaAΔEcoN]*	Integration pTA1762
H2599	*tnaA+::[tnaAΔEcoN]*	Integration pTA1730
H2600	*tnaA+::[tnaAΔEcoN]*	Integration pTA1730
Parent: H2467	H1895 + *adh1+::[Δadh1]*	
H2538	*Δadh1*	pop-out pTA1229
Parent: H2538	H2467 + *Δadh1*	
H2571	*adh2+::[Δadh2]*	Integration pTA1230
Parent: H2571	H2538 + *adh2+::[Δadh2]*	
H2638	*Δadh2*	pop-out pTA1230
Parent: H2638	H2571 + *Δadh2*	
H2687	*tnaA+::[tnaAΔEcoN]*	Integration pTA1730
H2688	*tnaA+::[ΔtnaA]*	Integration pTA1756
H2766	*gabT1+::[Δadh2::ΔgabT1]*	Integration pTA1754

[ ] indicates integrated plasmid, { } indicates episomal plasmid, TA = Thorsten Allers

Strain	Genotype	Notes
Parent: H2688	H2638 + *tnaA+::[ΔtnaA]*	
H2764	*ΔtnaA*	pop-out pTA1756
Parent: H2764	H2688 + *ΔtnaA*	
H2812	*gabT1+::[Δadh2::ΔgabT1]*	Integration pTA1754
H2836	{*p.tnaA::6xhis-2xstrep-gabT1+*}	overexpression pTA1759
H2837	{*p.tnaA::6xhis-2xstrep-gabT5+*}	overexpression pTA1760
H2838	{*p.tnaA::6xhis-2xstrep-adh2+*}	overexpression pTA1761
Parent: H2812	H2764 + *gabT1+::[Δadh2::ΔgabT1]*	
H2974	*ΔgabT1*	pop-out pTA1754
Parent: H2974	*ΔpyrE2, Nph-pitA, Δmrr, ΔhdrB, Cdc48d-Ct, ΔpilB3C3*, *Δadh1*, *Δadh2*, *ΔtnaA*, *ΔgabT1*	
H3094	{*p.tnaA::6xhis-adh2+*}	overexpression pTA1205
H3095	{*p.tnaA::6xhis-gabT1+*}	overexpression pTA1703
H3096	{*p.tnaA::6xhis-2xstrep-gabT1+*}	overexpression pTA1759
H3097	{*p.tnaA::6xhis-2xstrep-adh2+*}	overexpression pTA1761
H3098	{*p.tnaA::2xstrep-adh2+*}	overexpression pTA1863
Parent: H2685	H1895 + *tnaA+::[ΔtnaA]*	
H2765	*ΔtnaA*	pop-out pTA1756
Parent: H2599	H1895 +*tnaA+::[tnaAΔEcoN]*	
H2682	*tnaAΔEcoN*	pop-out pTA1730
Parent: H2682	H2599 + *tnaAΔEcoN*	
H2832	{*p.tnaA::bgaHa*}	overexpression pTA1543
H2833	{*p.tnaA::bgaHa*}	overexpression pTA1544
Parent: H2600	H1895 + *tnaA+::[tnaAΔEcoN]*	
H2683	*tnaAΔEcoN*	pop-out pTA1730

[ ] indicates integrated plasmid, { } indicates episomal plasmid, TA = Thorsten Allers

Strain	Genotype	Notes
Parent: H2683	H2600 + *tnaA*Δ*EcoN*	
H2834	{*p.tnaA::bgaHa*}	overexpression pTA1543
H2835	{*p.tnaA::bgaHa*}	overexpression pTA1544
Parent: H1325	Δ*pyrE2, Nph-pitA,* Δ*mrr,* Δ*hdrB,* Δ*adh1,* Δ*adh2*	
H1332	{*p.tnaA::6xhis-adh2+*}	overexpression pTA1205

[ ] indicates integrated plasmid, { } indicates episomal plasmid, TA = Thorsten Allers

Escherichia coli **strains**

Table 2.4: *E.coli* strains used in this study

Strain	Genotype	Notes
XL1-Blue MRF'	*endA1, gyrA96 (NalR), lac [F' proAB lacIqZM15 Tn10 (TetR)],* Δ*(mcrA)183,* Δ*(mcrCB-hsdSMR-mrr)173, recA1, relA1, supE44, thi-1*	Standard cloning strain for blue/white selection using pBluescript-based plasmids Tetracycline resistant, Restriction endonuclease and recombination deficient, *dam+*. From Stratagene.

Chemicals and Enzymes

All chemicals were purchased from Sigma and all enzymes from New England Biolabs (NEB) unless stated otherwise. Specific buffers and solutions are mentioned with the appropriate method.

Haloferax volcanii **Media**

All media is sterilised by autoclaving (1 min, 121 ℃), solutions are sterilised by filtration through a 0.2 μm filter. All liquid media is stored in the dark at room temperature (RT).

Plates are stored in sealed bags at 4 ℃ and dried at 30 ℃ for approximately 20 minutes before use.

30 % salt water

240 g NaCl, 30 g $MgCl_2.6H_2O$, 35 g $MgSO_4.7H_2O$, 7 g KCl, 20 ml 1 M Tris.HCl, pH 7.5, dH_2O to 1 L.

18 % salt water

200 ml 30 % SW, 100 ml dH_2O, 3 mM $CaCl_2$ is added after autoclaving.

10 X Ca

5 % Casamino acids, 17.6 mM KOH, (Not autoclaved, used immediately).

Trace elements

1.82 mM $MnCl_2.4H_2O$, 1.53 mM $ZnSO_4.7H_2O$, 8.3 mM $FeSO_4.7H_2O$, 200 μM $CuSO_4.5H_2O$. Filter sterilised and stored at 4 ℃.

Hv-Min salts

0.4 M NH_4Cl, 0.25 M $CaCl_2$, 8 % v/v of trace element solution. Stored at 4 ℃.

Hv-Min carbon source: 10 % DL-lactic acid Na_2 salt, 8 % succinic acid Na_2 salt.$6H_2O$, 2 % glycerol, pH to 7.0 with NaOH. Filter sterilised.

Hv-Ca Salts

362 mM $CaCl_2$, 8.3 % v/v of trace elements, 615 μg/ml thiamine, 77 μg/ml biotin.

KPO_4 Buffer

308 mM K_2HPO_4, 192 mM KH_2PO_4 pH 7.0.

10 X YPC

5 % yeast extract (Difco), 1 % peptone (Oxoid), 1 % casamino acids, 17.6 mM KOH.

(Not autoclaved, used immediately).

Hv-YPC agar

1.6 % Agar (Bacto), 18 % SW, 1 X YPC, 3 mM $CaCl_2$. Microwaved without 10 X YPC to dissolve agar. 10 X YPC added, then autoclaved. $CaCl_2$ added prior to pouring.

Hv-Ca agar

1.6 % Agar (Bacto), 18 % SW, 1 X Ca, 0.84 % v/v of Hv-Ca Salts, 0.002 % v/v of KPO_4 buffer (pH 7.0). Microwaved without 10 X Ca to dissolve agar. 10 X Ca added, then autoclaved. Hv-Ca Salts and KPO_4 Buffer added prior to pouring.

Hv-YPC broth

18 % SW, 1 X YPC, 3 mM $CaCl_2$. $CaCl_2$ added after autoclaving.

Hv-Ca+ broth

18 % SW, 30 mM Tris.HCl pH 7.0, 1 X Ca, 2.5 % v/v of Hv-Min carbon source, 1.2 % v/v of Hv-Min Salts, 0.002 % v/v of KPO_4 buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine; 30 % SW, dH_2O and Tris-HCl pH 7.0 autoclaved. Everything else added when cooled.

Hv-Min agar

1.6 % Agar (Bacto), 18 % SW, 30 mM Tris.HCl pH 7.5, 2.5 % Hv-Min carbon source, 1.2 % Hv-Min Salts, 0.002 % v/v of KPO_4 buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine; Microwaved to dissolve agar; Tris.HCl pH 7.5 added, then autoclaved. Hv-Min carbon source, Hv-Min Salts, KPO_4 buffer, biotin and thiamine added prior to pouring.

Supplements

Supplements are used for growth of *Haloferax volcanii* auxotrophic mutants as shown in the table below. Where a supplement is named, its addition is required for growth.

Table 2.5: Media supplements used with *Haloferax volcanii* and mutant growth

Supplement	**final**	**Genotype**	**Hv-YPC**	**Hv-Ca**
Uracil (Ura)	50 μg/ml	Δ*pyrE2*	+	Ura-
Thymidine (Thy)	50 μg/ml	Δ*hdrB*	Thy-	Thy-*
5'-Fluoroorotic acid (5'-FOA)**	50 μg/ml	Δ*pyrE2*	counter selection	

*In addition to thymidine, Δ*hdrB* strain cultures are supplemented with hypoxanthine in Hv-Ca. ** 5'-FOA media is additionally supplemented with 10 μg/ml uracil.

Buffered Spheroplasting Solution

1 M NaCl, 27 mM KCl, 50 mM Tris.HCl, pH 8.5, 15 % sucrose.

Unbuffered Spheroplasting Solution

1 M NaCl, 27 mM KCl, 15 % sucrose, pH 7.5.

Spheroplast Dilution Solution

23 % SW, 15 % sucrose, 37.5 mM $CaCl_2$.

Regeneration Solution

18 % SW, 1 X YPC, 15 % sucrose, 30 mM $CaCl_2$.

Transformation Dilution Solution

18 % SW, 15 % sucrose, 30 mM $CaCl_2$.

***Escherichia coli* Media**

All media is supplied by Sigma and sterilised by autoclaving, unless otherwise stated. Plates are stored in sealed bags at 4 ℃.

Lysogeny Broth (LB)

10 g tryptone (Bacto), 5 g yeast extract (Difco), 10 g NaCl, 0.002 M NaOH, pH 7.0, addition of 10 g agar (Bacto) for LB agar plates, dH_2O to 1 L

SOC

20 g tryptone (Bacto), 5 g yeast extract (Difco), 0.58 g NaCl, 0.186 g KCl, 2.03 g $MgCl_2$, 2.46 g $MgSO_4$, dH_2O to 1 L

Supplements

Ampicillin at a final concentration of 50 μg/ml

Buffers and Solutions

Solutions for Colony Hybridisation and Southern Blot

Blotting Solutions

25 ml 10 M HCl added to 975 ml dH_2O

87.66 g NaCl and 20 g NaOH, dH_2O to 1 L (Denaturing solution)

50 ml 20 X SSPE, 450 ml dH_2O (2 X SSPE)

20 X SSPE

3 M NaCl, 230 mM NaH_2PO_4, 32 mM EDTA, pH 7.4

Neutralising Solution

1.5 M NaCl, 0.5 M Tris.HCl, 1 mM EDTA

Prehybridisation Solution

6 X SSPE, 1 % SDS, 5 x Denhardts, 200 μg/ml salmon sperm DNA (Roche, boiled for 5 minutes prior to addition)

Hybridisation Solution

6 X SSPE, 1 % SDS, 5 % dextran sulphate

100 X Denhardts Solution

2 % Ficoll 400, 2 % PVP (polyvinyl pyrrolidone) 360, 2 % BSA (bovine serum albumin, Fraction V).

Low Stringency Wash Solution

2 X SSPE, 0.5 % SDS

High Stringency Wash Solution

0.2 X SSPE, 0.5 % SDS

Buffers and Solutions for Genomic DNA Spooling

ST Buffer

5 M NaCl, 1 M Tris.HCl, pH 7.5, autoclaved and stored RT

Lysis Solution

0.5 M EDTA, pH 8.0, 10 % SDS, autoclaved and stored at RT

2.2.2 Methods

DNA and Plasmid preparation

PCR components

Template DNA was either taken from genomic DNA (strain H26) or plasmid DNA. DNA polymerases for amplification were Phusion (Finnzymes) and Q5 Hot Start High-Fidelity (NEB). Only Q5 Hotstart was subsequently used for amplification of DNA for plasmids in this thesis. The following reaction conditions were applied:

Total Volume [50 μl]

200 μM dNTPs

1 μM Primer forward

1 μM Primer reverse

1 ng - 1 μg genomic and 1 pg - 1 ng plasmid template DNA

1 X Reaction Buffer Q5

1 X Q5 High GC Enhancer

0.02 U/μl HS Q5 Polymerase

PCR conditions

Step	Temperature / Time
Initial Denaturation	98 ℃, 30 s
Denaturation	98 ℃, 10 s
Annealing	Tm_e ℃, 20–30 seconds/kb
Annealing*	Tm_s ℃, 20–30 seconds/kb *(raise 1 ℃/ cycle)
Polymerisation	72 ℃, 25 s
Final Polymerisation	72 ℃, 5 min
Hold	4 ℃

Annealing* applies for Touch Down PCRs *(see below).*

Annealing temperatures (Tm ℃) are calculated using equation:

$$81.5 + (16.6 \times log_{10}[Na^+]) + (0.41 \times \%GC) - (100 - \%homology) - (\frac{600}{length(bp)}) \quad (2.1)$$

%GC is the percentage of guanine and cytosine present in the primer, %homology is the percentage of sequence homology between primer and template and length refers to the primer length.

Touch Down PCR was used, if the sequence homology of the oligos to template had not been 100 %. Two annealing temperatures were calculated based on the actual homology (Tm_s) and on 100 % homology (Tm_e). Starting from Tm_s the temperature was raised by 1 ℃ per cycle until Tm_e is reached. A further 20–30 cycles are performed based on Tm_e.

Agarose Gel Electrophoresis

Agarose gels were prepared using Agarose and Tris-borate EDTA (TBE) buffer. DNA samples were applied to the gel in a 1/6 mixture with 5x-gel loading dye. Gels (10 cm) were run at 110 V for 50 minutes in TBE buffer. TAE buffer was used when high quality resolution was required and TAE gels (25 cm) were run overnight (16 hours) with buffer circulation at 50 V. Molecular markers run alongside samples were either a 1 kb ladder (NEB) or a 100 bp ladder (NEB). Ethidium bromide was either directly added to the gel to a final concentration of 0.5 μg/ml or was added in a 1:10000 dilution in water to excised

gels. Appropriate bands were excised from the gel, whilst other sample lanes were protected using foil. DNA was visualised by UV light using a UV transilluminator (UVP inc.).

Nucleic Acid Purification

Purification of nucleic acid after PCR or agarose gel extraction was carried out using Macherey-Nagel purification kits according to user manual. Columns used in the kit are silica-gel based membranes and adsorb DNA in salt and pH dependent buffers. Small oligonucleotides or proteins will not bind and are separated by ethanolic washing. DNA was eluted in 30 μl.

Restriction Digest

All restriction enzymes were purchased from New England Biolabs (NEB). Digests were performed according to manufacturers instructions, using provided buffers (NEB) cutsmart, buffer 3.1 or buffer 1.1. 5 U enzyme were used per μg DNA. Incubation periods varied between 1 h and overnight. Buffers for double digests were selected appropriately with at least 75 % enzyme activity remaining.

Ligation of DNA

Ligation reactions were performed using 5 U T4 DNA Ligase (NEB), 1 X T4 DNA Ligation Buffer per μg DNA and incubated in a 16 ℃ waterbath overnight. Molar ratios of 3:1 insert:vector DNA were used.

Ethanol precipitation of ligased DNA

Ligation reactions were acidified with 0.1 volumes of 3 M sodium acetate buffer (pH 5.3) and an excess of 100 % ethanol was added, before ligations were incubated at –20 ℃ for one hour. Precipitated DNA was centrifuged at 4 ℃, 14.000 x g and restored in 4 μl H_2O.

Sequencing of plasmid DNA

All sequencing reactions were analysed by Deep Seq, University of Nottingham and use dideoxy chain termination method [Sanger et al., 1977].

Nucleic acid quantification

Concentration and purity of plasmid DNA samples was assessed with spectrophotometric measurements (Beckman Coulter DU 530) at 260 nm and the 260:280 nm absorbance ratio, respectively.

Microbiological Methods for *Escherichia coli*

Growth and storage of *E. coli*

Cultures on solid media and liquid cultures of 1-10 ml were grown overnight in a static incubator (LEEC) at 37 ℃. Liquid cultures were rotated at 8 rpm. Liquid Cultures in flasks (300 ml) were grown in an Innova 4330 floor-standing shaking incubator (New Brunswick Scientific) at 110 rpm. Solid media plates and liquid cultures were stored at 4℃ for short term storage. Plasmid transformed cells are stored long term in 80 % glycerol (25 % (v/v)) after flash freezing at –80 ℃.

Preparation of electrocompetent cells

Liquid cultures (5 ml) were grown overnight and diluted 1/100 in LB broth supplemented with appropriate antibiotics at 37 ℃ to A_{650} = 0.5 - 0.8. Cells were pelleted at 6000 *x g* for 12 minutes at 4 ℃ and the supernatant discarded. The pellet was resuspended in an equal volume of ice-cold sterile 1 mM Hepes (pH 7.5) buffer. These steps were repeated with 0.5 volumes 1 mM Hepes, 0.25 volumes 1 mM Hepes plus 10 % glycerol and 0.1 volumes 1 mM Hepes plus 10 % glycerol. Cells were snap frozen in liquid nitrogen and stored in 100 μl aliquots at –80 ℃.

Transformation of *E. coli* by electroporation

50 μl electro-competent cells, thawed on ice, were inserted into a pre-chilled cuvette (1 mm FlowGen, Bioscience) and 1 μl plasmid DNA ($\sim$ 1 μg) was added. The cuvette was placed into an *E. coli* gene pulser (BioRad) and pulsed with 1.8 kV. 1 ml SOC media was added and the mixture was transferred to a 4 ml round bottomed tube and incubated at 37 ℃ for 1 hour. Cells were diluted 1:10 or 1:100 and plated onto LB plates supplemented with ampicillin and incubated at 37 ℃ overnight.

Mini scale purifications of plasmid DNA from *E.coli*

Purification of plasmid DNA was carried out with Macherey-Nagel Plasmid DNA purification kits according to user manual. Cells were lysed in alkaline buffer and DNA was bound to an anion-exchange resin column in high salt. DNA was eluted in 30 μl and stored at –20 ℃.

Maxi scale purification of *E.coli*

300 ml overnight cultures of transformed XL1-Blue cells were pelleted and purification of plasmid DNA in quantitative amounts was carried out according to the protocol of Macherey-Nagel Plasmid DNA purification kits with provided solvents. Plasmid DNA was stored in 200 μl TE at –20 ℃.

Microbiological Methods for *Haloferax volcanii*

Growth and storage of *Hv. volcanii*

Cells were either grown on solid media (plates) or in liquid culture (5 ml, 50 ml or 333 ml) on a rotor (8 rpm) in a 45 ℃ static incubator (LEEC) or at 110 rpm in a Innova 4330 floor-standing shaking incubator (New Brunswick Scientific). Plates were stored short term at room temperature. Cultures were stored long term in 20 % (v/v) glycerol (80 % plus 6 % salt water) after flash freezing at -80 ℃.

Transformation of *Hv. volcanii*

PEG600 was used to transform *Hv. volcanii* according to [Cline et al., 1989]. *Hv. volcanii* encodes a restriction endonuclease (Mrr) that targets methylated DNA and necessitates the passage of plasmid DNA through a *dam- E. coli* host prior to transformation [Holmes et al., 1991]. All strains used in this study for transformation possess a Mrr deletion (Δmrr) and can be transformed directly with *dam+* plasmid DNA.

A 5 - 10 ml Hv-YPC overnight culture was pelleted at 3600 *x g* for 8 minutes after an OD_{600} of 0.8 was reached. Pellets were gently resuspended in 2 ml buffered spheroplasting solution and transferred to a 2 ml round-bottomed tube and pelleted at 3600 *x g* for 8 minutes. Cells were gently resuspended in 600 μl buffered spheroplasting solution. 200μl resuspended cells per transformation were transferred to a clean 2 ml round bottomed

tube and a 20 μl drop of 0.5 M EDTA was placed on the side of the horizontally held tube, which was then inverted to mix and incubated for 10 minutes at RT. Plasmid DNA samples were prepared in unbuffered spheroplasting solution and 0.5 M EDTA. 1- 2 μg DNA was added to spheroplasts in the same manner as ETDA and incubated for 5 minutes at RT. 250 μl of 60 % PEG was added to each transformation in the same manner as EDTA, but the tube was shaken horizontally and left at RT for 30 minutes. 1.5 ml spheroplast dilution solution was added and left at RT for 2 minutes. Spheroplasts were pelleted at 3600 x g for 8 minutes and resuspended in 1 ml regeneration solution (+40μg/ml thymidine for $\Delta hdrB$ strains). Whole pellets were transferred to a 4 ml sterile tube and left undisturbed at 45 ℃ for 1.5 - 2 hours to regenerate. Pellets were resuspended by tapping on the side of the tube and further incubated at 45 ℃ for 3 - 4 hours on a rotator. Cells were transferred to a 2 ml round-bottomed tube and pelleted at 3600 x g for 8 minutes. Cells were gently resuspended in 1 ml transformant dilution solution and plated on Hv-Cas media in appropriate dilutions. After 5 days incubation at 45 ℃ single colonies were picked and plated on fresh Hv-Cas to promote the pop-in of exogenous DNA. After a growth period of 5 days at 45 ℃ colonies were picked and incubated in YPC-Broth in three consecutive overnight cultures to prepare cells for the pop-out on Cas-5-FOA plates. After 5 days of incubation at 45 ℃ colonies were patched (2x 40) onto YPC-plates. The following colony lift screens for homozygous deletion strains and was afterwards confirmed by southern blot.

Colony Lift

Candidate and control colonies were patched onto Hv-YPC (+Thy if required) and incubated at 45 ℃ for 3 days. Colonies were transferred onto Zeta-Probe GT positively charged membranes (Bio-Rad) for radioactive labelling. The membrane was rolled onto the surface of the plate and incubated until it appeared wet. The membrane was removed and placed, colonies facing upwards, onto Whatman paper soaked in 10 % SDS to lyse cells. The membrane was removed again, when colonies were glossy (approx. 10 min.) and placed onto Whatman paper soaked in denaturing solution. When colonies appeared slimy (after approx. 15 min.) the membrane was placed onto blotting paper soaked in neutralising solution for 5 minutes. The membrane was placed onto fresh neutralising

solution soaked blotting paper for further 5 minutes and was then put on fresh blotting paper for drying. When membranes were completely dry, the DNA was UV cross-linked to the membrane with 120 mJ/cm2.

Genomic DNA extraction by DNA spooling

A 5 ml overnight culture was grown at 45 ℃. 1 ml culture was pelleted at 3600 *x g* for 5 minutes in a round-bottomed 2 ml Eppendorf tube and resuspended in 200 μl ST buffer and lysed by the addition of 200 μl lysis solution. The aqueous layer was overlaid with 1ml EtOH (100 %) and DNA was spooled onto a gel-loading tip and washed twice in fresh 1ml EtOH (100 %). Spooled DNA was resuspended in 500 μl TE and precipitated with 50 μl 3 M sodium acetate (pH 5.2) and 400 μl isopropanol and centrifuged at 3600 *x g* for 5 minutes. The pellet was washed with EtOH (70 %) and allowed to dry thoroughly. DNA was resuspended in 100 μl TE at 45 ℃ for over an hour and stored at 4 ℃.

Southern Blot Vacuum Transfer

Purified genomic DNA was digested overnight with selected enzymes cutting between 500 and 1000 bp upstream and downstream of the gene of interest (to be deleted), and was run on a 200 ml 0.75 % agarose gel in 1x TAE for 16 hours at 50 V connected to a peristaltic pump for buffer circulation. After staining with 0.5 μg/ml of ethidium bromide for 30 minutes for visibility of bands the gel was acid-nicked for 20 minutes in 0.25 M HCl, then washed in dH_2O for 10 minutes and denatured in denaturing solution for 45 minutes. A 15 x 25 cm Zeta-Probe GT membrane (Bio-Rad) was activated in dH_2O for 5 minutes and equilibrated in denaturing solution for 2 minutes. The vacuum transfer was performed with a Vacugene XL gel blotter and a Vacugene Pump (Pharmacia Biotech) for 1 hour at 40 mBar. After transfer the membrane was very briefly neutralised in 2 x SSPE and left to dry before crosslinking DNA at 120 mJ/ cm^2 UV.

Colony Hybridisation and Southern Blot

Membranes from colony lifts or blotted gels were pre-hybridised with 40 ml pre-hyb solution containing denatured salmon sperm DNA (10 mg/ml) for 3 hours at 65 ℃. DNA-probes were denatured at 100 ℃ for 5 minutes and then kept on ice. Radiolabelled DNA

was prepared using $\sim$ 50 ng DNA and 0.74 Mbq of [α-^{32}P] dCTP (Perkin Elmer) by adding the radioisotope and HiPrime random priming mix (Roche) and incubating for 20 minutes at 37 ℃. Probes for Southern blots contained a 1 kb ladder (1 μg/ml). The probe was then purified from unincorporated isotopes by using a P-30 column (Bio-Rad), mixed with salmon sperm DNA (10 mg/ml) and incubated at 100 ℃ for 5 minutes, before being put on ice. The pre-hyb solution was discarded and 40 ml hybridisation solution and the DNA probe were added to membranes. Membranes were incubated overnight at 65 ℃ and then washed twice with 50 ml low stringency wash solution for 10 minutes and 30 minutes and then washed twice with 50 ml high stringency wash solution for 30 minutes each. Membranes were dried before being wrapped in saran wrap and exposed to a phosphorimager screen (Fujifilm BAS Casette 2325) for over 24 hours. The screen was scanned by a Molecular Dynamics STORM 840 scanner.

Methodology for Gene Deletion in *Haloferax volcanii*

Generating a Deletion Construct

A deletion construct for creating a genomic clone was generated by inserting the upstream and downstream flanking regions of the gene of interest into pTA131. This can either be done by restriction digest, if the gene of interest has appropriate restriction sites within the coding sequence at both (5' and 3') ends, or by PCR, if no appropriate sites for restriction are present. In the latter case the gene of interest was PCR amplified with additional 1 kb upstream and downstream flanking regions, introducing restriction sites for insertion, schematic shown in figure 2.1.
The gene deletion construct was then derived through PCR amplification of the whole plasmid apart from the gene (pTA131+Flanking regions) using outward facing primers. The forward primer has to bind exactly just downstream of the stop codon and incorporates a novel *Bam*HI, *Bgl*II or *Bcl*I site and the reverse primer binds at the start codon, incorporating a novel *Nde*I site. In the following the PCR product was restricted with appropriate enzymes and gel purified. The backbone of the vector, including the flanking regions of the gene of interest, was then self-ligated to obtain the $\Delta gene$ construct. Plasmids were transformed into XL-1 blue and plated onto LB+Amp. Empty vector was

digested, re-ligated and transformed into XL-1 blue as a negative control.

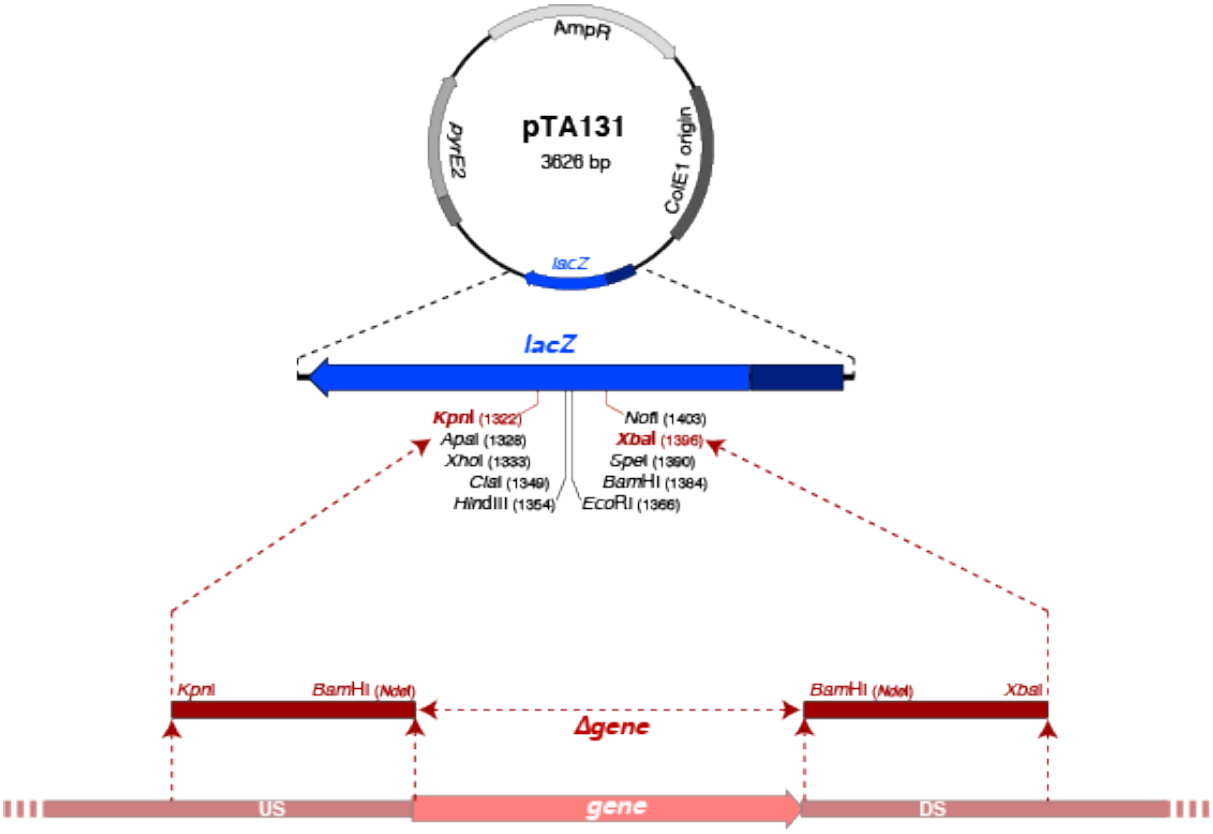


Figure 2.1: The construct pTA131 is derived from pBluescript and carries the *pyrE2* marker (uracil biosynthesis in *H.volcanii*) for selection through the pop-in/ pop-out gene knock out system. The gene of interest for deletion is cloned into pTA131 including upstream and downstream flanking regions and disrupts the *lacZ* gene (*β-galactosidase* gene used for blue/white selection in *E.coli*. Consequently, from the gene outward facing primers are designed to amplify the plasmid backbone with the flanking regions of the gene and incorporating at either end a BamHI/KpnI site for re-ligation to obtain a Δ*gene* deletion construct. *colE1 origin* (*E.coli* origin of replication).

Pop-in / Pop-out Method

Gene deletion constructs were integrated at the gene locus of interest by transforming appropriate Δ*pyrE2 H.volcanii* strains with Δ*gene* derivatives of pTA131 containing the *pyrE2* marker, which encodes for orotate phosphoribosyl transferase, an enzyme involved in uracil biosynthesis Allers et al. [2004]. Transformed strains were plated onto Hv-

Cas+Thy plates to select for integrated *pyrE2*-marked plasmid pop-ins, enabling cells to synthesize uracil. A pop-in colony was picked and grown non-selectively overnight in a 5 ml Hv-YPC+Thy culture until A_{650}=1.0. This culture was diluted 1:500 into fresh Hv-YPC+Thy and growth and dilution repeated once. The relief from the selection for uracil by providing it with Hv-YPC media leads to a loss of the integrated plasmid and the native gene is replaced only with its flanking regions on the chromosome. Screening for successful pop-outs was performed using Hv-Ca+5FOA+Thy plates as a final culture step to recover either wild-type or deletion mutants, depending on the location of recombination. Replacement of the gene with a selectable marker such as *trpA+* enables direct selection for deletion mutants.

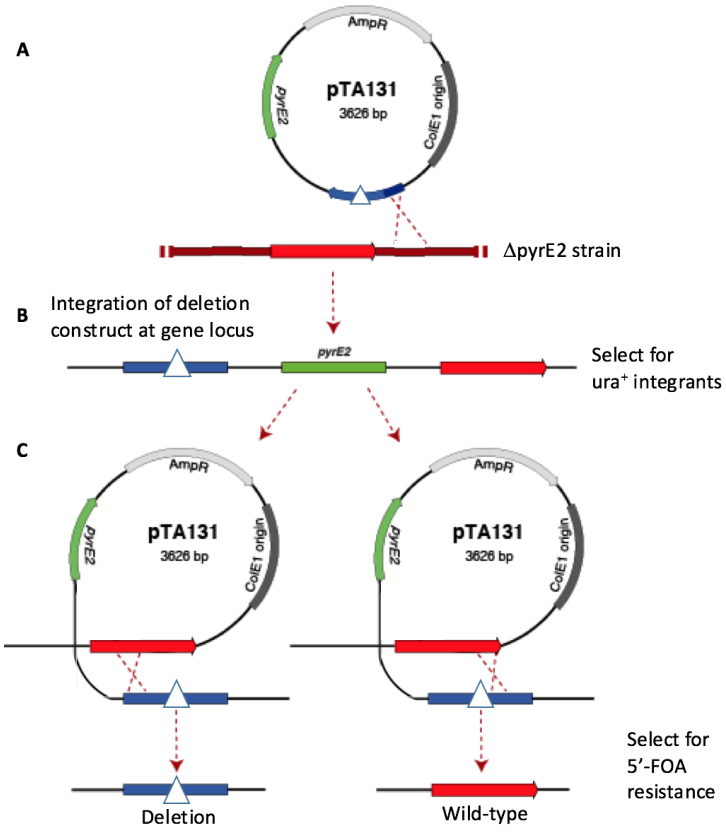


Figure 2.2: (A) Δ*pyrE2* strains are transformed with a *pyrE2+* gene deletion construct and (B) integration (pop-in) of the deletion construct is selected for by the ability to grow on media lacking uracil. (C) Eliminating uracil selection pressure allows for the excision (pop-out) of the construct through crossing-over. Recombination between the regions of homology can be either upstream (left) or downstream (right). Pop-outs are selected for by plating cells on 5-FOA, resulting either in a deletion or wildtype.

Since *H.volcanii* is polyploid and transformed strains hence can be merodiploid, i.e. pop-out colonies could have a mixture of deletion and wildtype alleles on different chromosome copies, a selection for strains that are homozygous is required. Colony hybridisation and Southern blotting of genomic DNA must be used to confirm the presence of deletions.

Methodology for Gene Overexpression in *Haloferax volcanii*

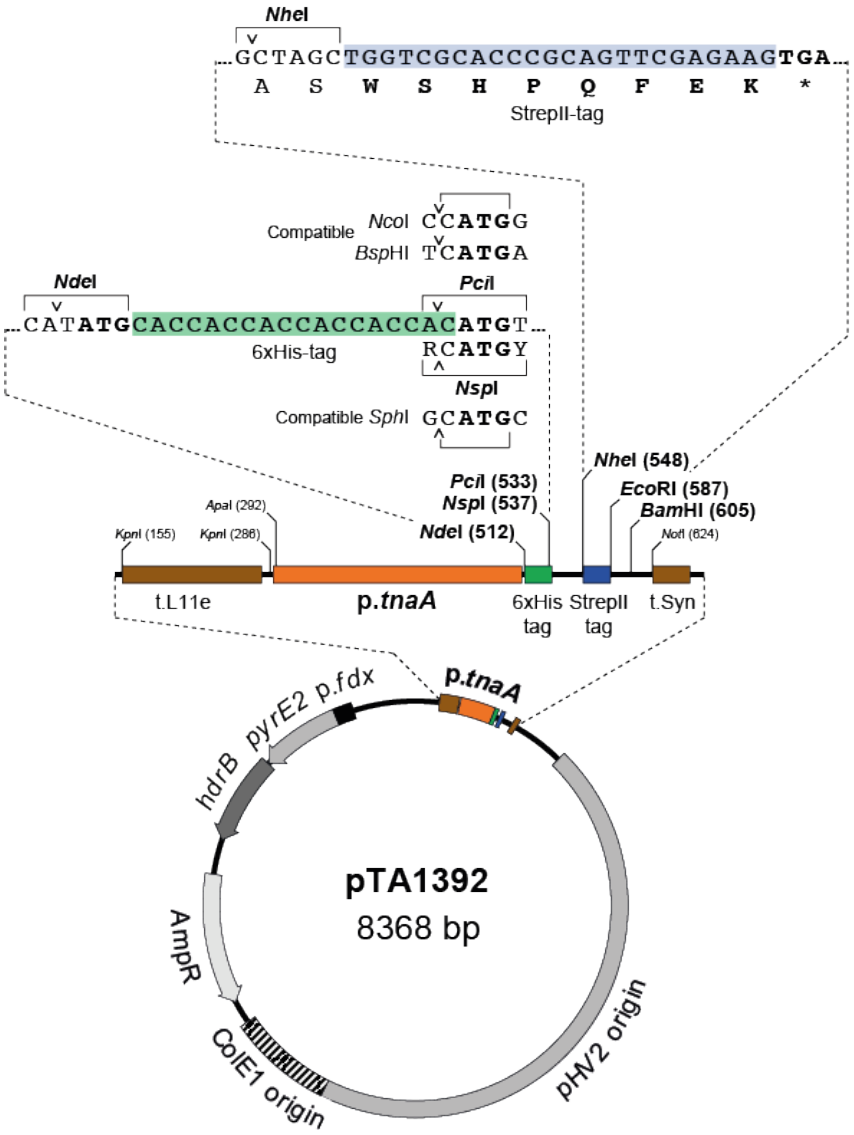


Figure 2.3: pTA1392 for conditional overexpression of proteins under the control of the tryptophan inducible promotor *p.tnaA*, containing *pyrE2* and *hdrB* selection markers. Restriction sites allow for the construction of N-terminally his6x- and/or C-terminally strepII-tagged protein.

Episomal constructs for the expression of his6x- and/or strepII-tagged proteins were generated by inserting the gene of interest into pTA1392. Constructs including the his6x-tag were derived through replacement of the start codon of the gene with *Nco*I, *Bsp*HI, *Pci*I or *Sph*I sites, while *Nde*I was used for constructs excluding the his6x-tag. Constructs including the strepII-tag were derived through replacement of the stop codon of the gene with *Nhe*I compatible sites (*Avr*II, *Spe*I or *Xba*I), while *Eco*RI or *Bam*HI sites after the stop codon were used to exclude the strepII-tag.

2.3 Results

2.3.1 Plasmid Construction

This section details the construction of plasmid vectors used in this study. All plasmids were confirmed by restriction digest and sequencing. The following abbreviations apply for all figures in this section: *AmpR* (ampicillin resistance gene, *E.coli*), *colE1 ori* (*E.coli* origin of replication), *f1(+) ori* (*E.coli* origin of replication), *hdrB* (thymidine biosynthesis, *H.volcanii*, *hisx6-tag* (hexahistidine tag), *lacZ* (*β-galactosidase* gene used for blue/white selection in *E.coli*, MCS (multiple cloning site), *p.fdx* (ferredoxin promotor), *pHV2* (*H.volcanii* origin of replication), *p.tnaA* (tryptophan inducible promotor), *p.lac* (promoter for *lacZ*), *pyrE2* (uracil biosynthesis, *H.volcanii*, *strepII-tag* (streptavidin tag), *t.L11e* (terminator), *t.Syn* (terminator).

Genomic Clones and Deletion Constructs

Genomic clones and derived deletion constructs were constructed as described in the *Material and Methods* section of this chapter using pTA131. The vector is based on the standard *E.coli* cloning vector, pBluescript II SK+ and the vector pGB70 (figure 2.4). The MCS is located within the *lacZ* gene. Expression of *lacZ* results in blue colonies when plated onto media containing 5-bromo-4-chloro-3-indolyl--D-galactopyranoside (X-gal), whereas, if *lacZ* is disrupted upon insertion of a sequence in the MCS, the colonies

are white.

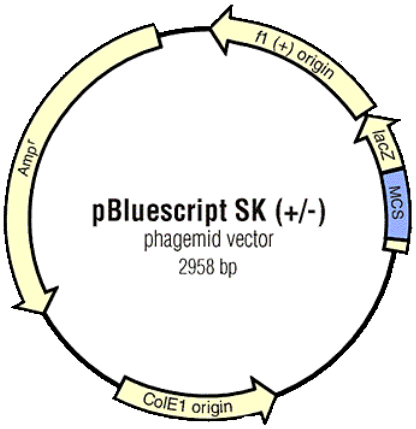


Figure 2.4: (A) pBluescript II SK+. Standard *E.coli* cloning vector, (B) pGB70; Figure adapted from Stratagene.

pTA1229 and pTA1230, *adh1*(HVO_2428) and *adh2*(HVO_B0071) deletion constructs

The deletion constructs pTA1229 and pTA1230 containing upstream and downstream flanking regions of *adh1* and *adh2*, respectively, were previously constructed by Thorsten Allers in 2010.

pTA1743 and pTA1754, *gabT1+adh2* (HVO_B0070+HVO_B0071) genomic clone and deletion construct

Genomic DNA encoding for the adjacent genes *adh2 and gabT1* was amplified from wild-type *H.volcanii* strain (H26) via PCR using the oligonucleotides o1504 and o1505 to incorporate a *KpnI*HF and a *Xba*I site, respectively, and resulted in a 3910 bp fragment including flanking regions. Amplified DNA fragments were purified, digested and ligated into pTA131, which was digested with the same enzymes and gel purified, resulting in a 3300 bp backbone fragment. *E.coli* transformants were grown on LB+Amp and mini-preps were digested with *Mlu*I and *Sfo*I to verify the insertion and the genomic clone pTA1743 was isolated from maxi-prep.

The backbone and flanking regions of *adh2+gabT1* were PCR amplified from the genomic clone pTA1743 using the oligonucleotides o797 and o1573, whereby o1573 incorporates a novel *Bam* site in place of *Nde*I, because the downstream flanking region possesses

a native *Nde*I site. The PCR fragment was gel purified and digested with *Bam*HI for self-ligation to create the *adh2+gabT1* deletion construct pTA1754. The construct was verified by digest with *Nde*I and *Sca*I (figure 2.5).

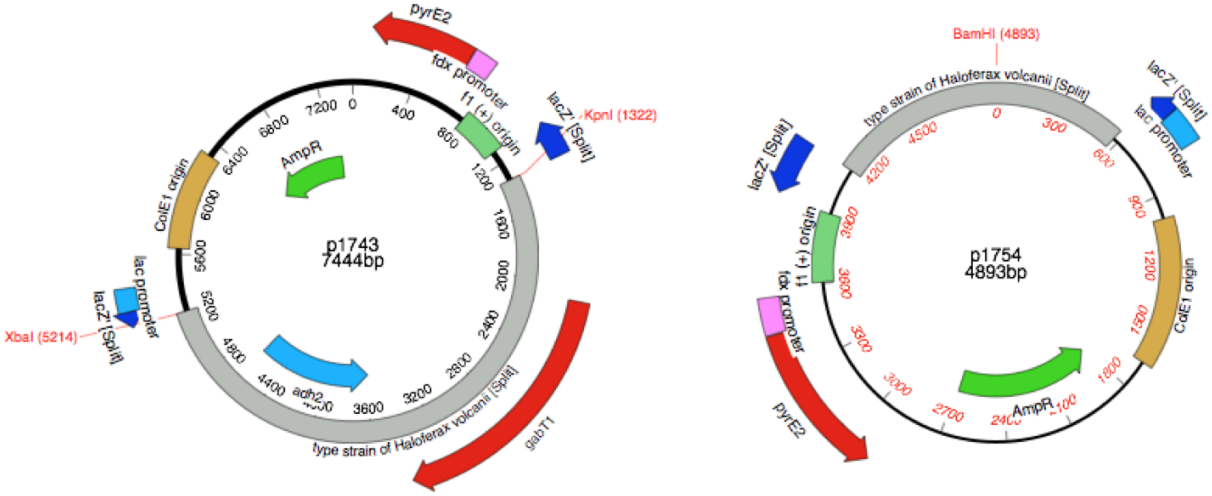


Figure 2.5: pTA1743 - construct containing flanking regions and genes *adh2+gabT1* (left) and pTA1754 - deletion construct containing only flanking regions of *adh2+gabT1* (right).

pTA1730, *tnaA*Δ*EcoNI* (HVO_0009) deletion construct

The deletion construct pTA1730 containing a disrupted version of the tryptophanase A *tnaA* gene was constructed by Thorsten Allers in 2015.

pTA1508, *tnaA* (HVO_0009) deletion construct

The deletion construct pTA1508 containing *tnaA* flanking regions was constructed by Rebecca Milner in 2014.

pTA1755 and pTA1756 *tnaA* (HVO_0009) genomic clone and deletion construct

The genomic clone p875 containing *tnaA*+flanking regions was digested using restriction enzymes *Bam*HI and *Xba*I. The resulting 3803 bp fragment was ligated into pTA131 using the same enzymes for restriction. The resulting construct was verified by digest with *Not*I.

The deletion construct pTA1756 was derived by restriction of pTA1755 with *Age*I and self-ligation of the backbone containing only flanking regions of *tnaA*.

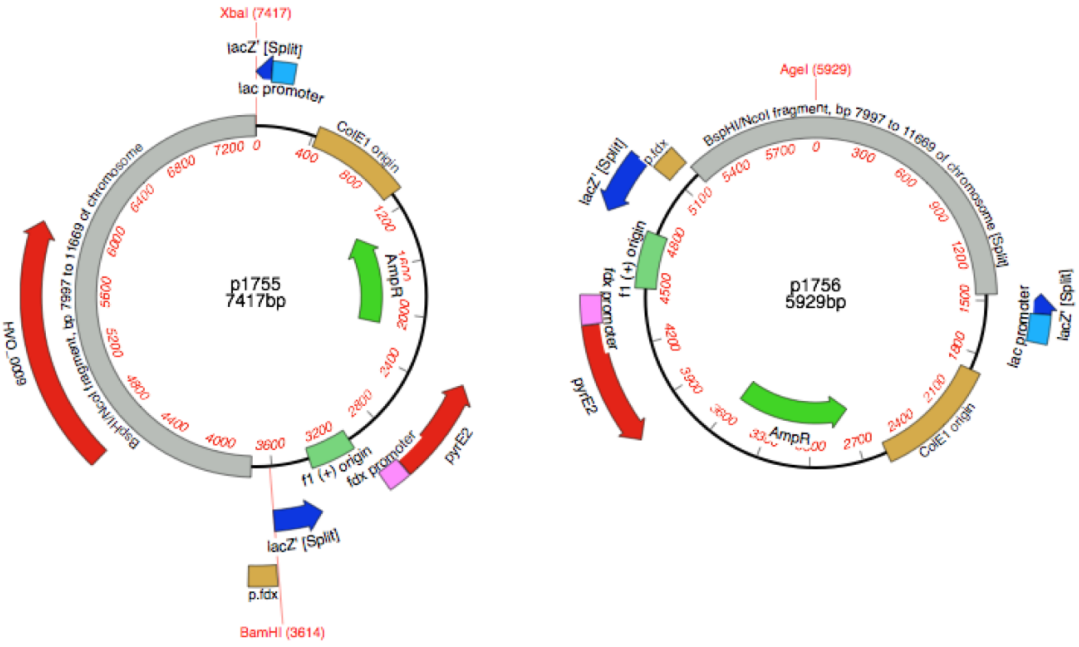


Figure 2.6: pTA1755 - construct containing flanking regions and gene *tnaA* (left) and pTA1756 - deletion construct containing only flanking regions of *tnaA* (right).

pTA1762, Δ*tnaA::hdrB*$^{+}$ (HVO_0009) deletion construct

The deletion construct pTA1762 was constructed by insertion of a 727 bp fragment of *hdrB* restricted from p1185 with *Bam*HI, into *Age*I restricted pTA1756. However, this plasmid could not be verified.

Episomal Plasmids for Overexpression of Tagged Proteins

The construct pTA1392 was used as template vector for conditional overexpression of his6x- and/or strepII-tagged *H.volcanii* proteins as described in the *Material and Methods* section of this chapter.

pTA1670 for overexpression of N-terminally his6x-tagged omega transaminase 5

This plasmid was previously constructed by Emma Sayers in 2015 for the conditional overexpression of N-terminally his6x-tagged *gabT5*.

pTA1703 for overexpression of N-terminally his6x-tagged ω-transaminase 1

Genomic DNA of *gabT1* encoding for ω-transaminase 1 was amplified from wild-type

H.volcanii strain (H26) via PCR using the forward and reverse primers o1429 and o1430 to incorporate a *Bcl*I and a *Pci*I site, respectively, and resulted in a 1332 bp fragment. The PCR product was digested with *Pci*I and *Bcl*I and ligated into the expression plasmid pTA1392 at *Pci*I and *Bam*HI sites downstream of a his6x-tag, restricting the strepII-tag from the construct. Digest for verification was performed using *Xba*I.

pTA1759 for overexpression of N-terminally his6x-tagged and C-terminally strepII-tagged ω-transaminase 1

gabT1 was amplified from the vector pTA1703 using the primers o1514 and o627 resulting in a 1316 bp fragment, which was inserted into pTA1392 at *Nde*I and *Nhe*I restriction sites downstream of a his6x-tag and upstream of a strepII-tag. Digest for verification was performed using *Mlu*I and *Sca*I.

pTA1760 for overexpression of N-terminally his6x-tagged and C-terminally strepII-tagged ω-transaminase 5

gabT5 was amplified from the vector pTA1670 using the primers o1513 and o627 resulting in a 1379 bp fragment, which was inserted into pTA1392 at *Nde*I and *Nhe*I restriction sites downstream of a his6x-tag and upstream of a strepII-tag. Digest for verification was performed using *Sca*I.

pTA1761 for overexpression of N-terminally his6x-tagged and C-terminally strepII-tagged alcohol dehydrogenase 2

adh2 was amplified from the vector pTA1205 using the primers o1512 and o627 resulting in a 1048 bp fragment, which was inserted into pTA1392 at *Nde*I and *Nhe*I restriction sites downstream of a his6x-tag and upstream of a strepII-tag. Digest for verification was performed using *Sfo*I and *Sca*I.

pTA1863 for overexpression of N-terminally strepII-tagged alcohol dehydrogenase 2

adh2 was amplified from the vector pTA1205 using the primers o1694 and o1695 resulting in a 1071 bp fragment, which was inserted into pTA1852 at *Pci*I and *Eco*RI restriction sites downstream of a his6x-tag and upstream of a strepII-tag. Digest for verification was

performed using *Sty*I and *Ale*I.

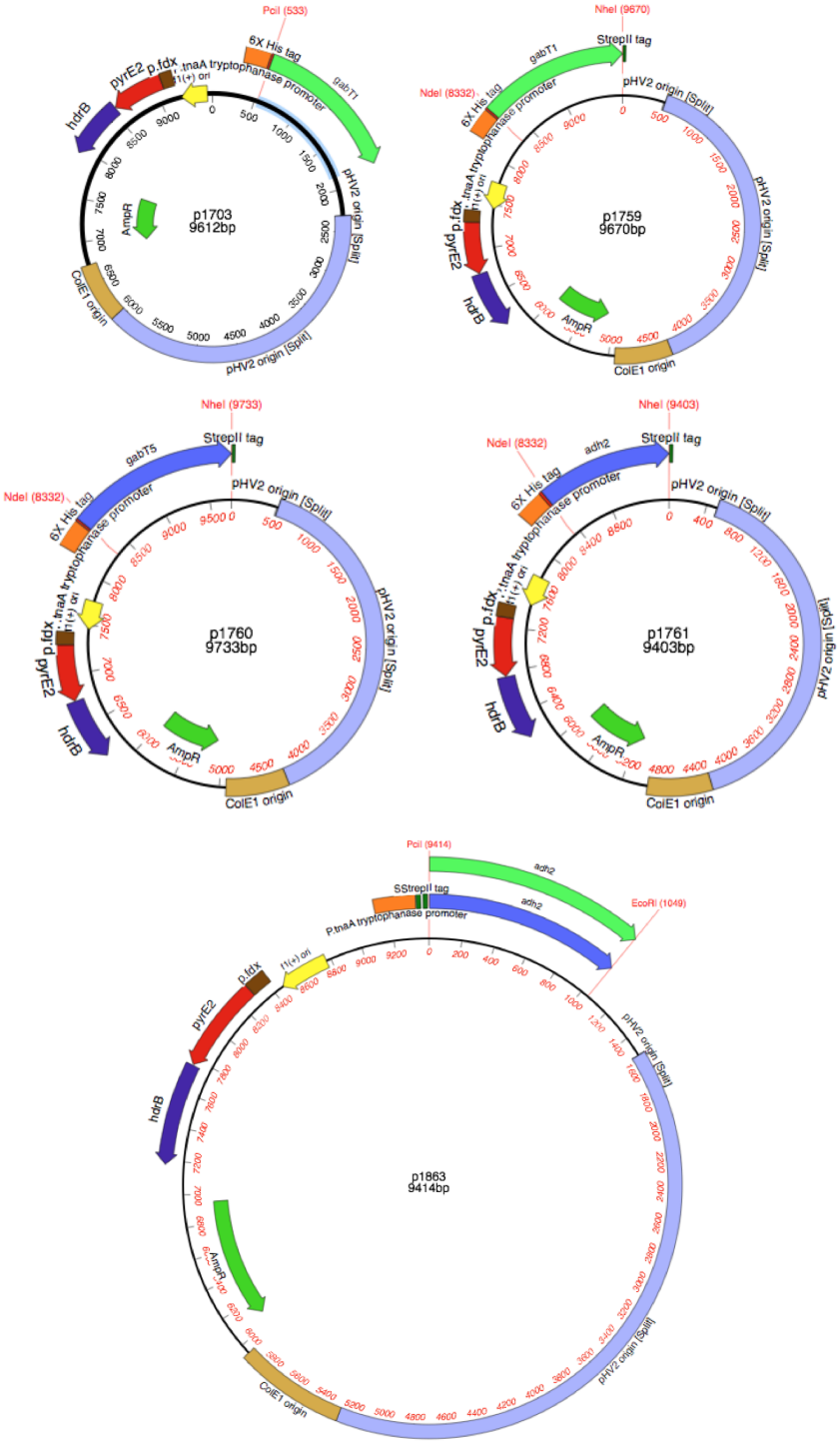


Figure 2.7: pTA1703 - episomal construct for *gabT1* expression; restriction site *Bcl*I was deleted during construction, pTA1759 - episomal construct for *gabT1* expression, pTA1760 - episomal construct for *gabT5* expression, pTA1761 - episomal construct for *adh2* expression, pTA1863 - episomal construct for *adh2* expression.

Episomal Plasmids for the Expression of β-galactosidase

pTA1543 and pTA1544, *β-galactosidase* expression

The genomic clones pTA1543 and pTA1544 were previously constructed by Stefan Groetzinger in 2014 for the expression of the *bgaHA* reporter gene under the control of the *p.tnaA* tryptophan-inducible promoter.

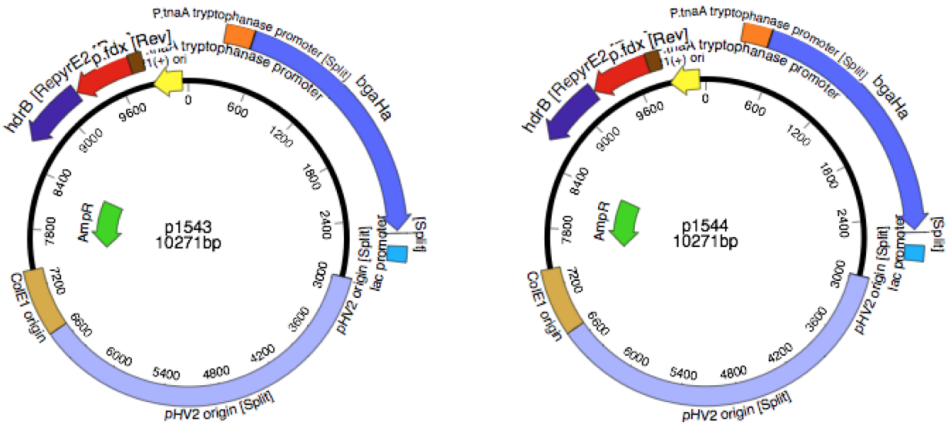


Figure 2.8: pTA1543 (left) and pTA1544 (right) - episomal constructs for *bghHa* expression.

2.3.2 Strain Construction

The main parental strain of *H. volcanii* used in this study was H1895, which was previously derived by Strillinger et al [Strillinger et al., 2016]. Strain H1895 was herein used to derive the masterstrain H2974 devoid of the genes encoding for co-purifying homologous proteins ADH1, ADH2 and ω-TA1, and it was improved for tryptophan-inducible expression through deletion of the gene *tnaA* encoding for tryptophanase A, an enzyme capable of breaking down the inducer, resulting in the strain H2974. Additionally, strain H1895 was solely deleted for *tnaA*, resulting in a new strain H2765 for overexpression. The wildtype strain H26 was used in this study solely for amplification of DNA by PCR. The strain H1332 was used as a reference strain for Δ*tnaA* strains for the overexpression of N-terminal his6x-tagged ADH2.

In the following, a description of previously derived strains is given and in the following

subsections, strains derived in this study are described.

H26 (Δ*pyrE2*)

H26 is the standard wild-type laboratory strain derived from the wild-type isolate DS2 by curing of the plasmid pHV2 and deletion of *pyrE2* (for uracil biosynthesis) [Allers et al., 2004, Wendoloski et al., 2001]. The origin of pHV2 therefore can be used in episomal vectors.

H1895 (Δ*pyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB, ΔpilB3C3*)

H1895 was derived from H1424 for improved protein overexpression and purification through the deletion of the genes *pilC3* and *pilB3* [Allers, 2010, Strillinger et al., 2016, Stroud et al., 2012]. These genes encode an ATPase and an integral membrane protein of the pilus assembly system, which enables biofilm formation, this is undesired for cultivation in stirred tank reactors. Cell motility stays intact. *mrr* encodes a restriction endonuclease that cleaves foreign 5-GATC methylated DNA, which would require the passaging of genomic clones through a *dam*$^-$ *E. coli* strain and was consequently deleted in strain H1209. This strain was further developed to reduce the co-purification of histidine rich proteins with his6x-tagged proteins by replacing the *pitA* gene of *H. volcanii*, which contains a histidine-rich linker region, with a non-histidine rich orthologue from *Natronomonas pharaonis*. Strain H1424 was derived by additionally truncating the *cdc48d* gene by its histidine rich C-terminus [Stroud et al., 2012].

H2682 and H2683 (Δ*pyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB, ΔpilB3C3, tnaAΔEcoN*)

Strains H2682 and H2683 were derived from H2599 and H2600, respectively, by Laura Mitchell in 2016 through pop-out of deletion construct pTA1730 (*tnaAΔEcoN*).

H1969 and H1970 (Δ*pyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB, ΔpilB3C3, p.tnaA bgaHA)*

Strains H1969 and H1970 were derived from H1895 by Stefan Groetzinger in 2014 through transformation with pTA1543 and pTA1544 (both *p.tnaA bgaHA*), respectively, for expression of the reporter β-galactosidase.

H1332 (Δ*pyrE2, Nph-pitA, Δmrr, ΔhdrB, Δadh1, Δadh2, p.tnaA::his6x-adh2*

Strain H1332 was derived from H1352 through transformation with pTA1205 for the over-expression of N-terminal his6x-tagged ADH2.

Masterstrain for Protein Overexpression

Gene Deletions

Strain H2974 (Δ*pyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB, ΔpilB3C3, Δadh1, Δadh2, ΔtnaA, ΔgabT1*) was derived from H1895 in four deletion steps, resulting in the deletions for *adh1, adh2, tnaA and gabT1.*

Δ*adh1*

HVO_2428(*adh1* was deleted from strain H1895 using pTA1229 to generate pop-in strain H2467 and pop-out strain H2538. Deletion of *adh1* was confirmed by colony hybridisation and by Southern blot using a 2300 bp radiolabelled probe of *adh1* flanking regions restricted from pTA1229 using *Pvu*II (figure).

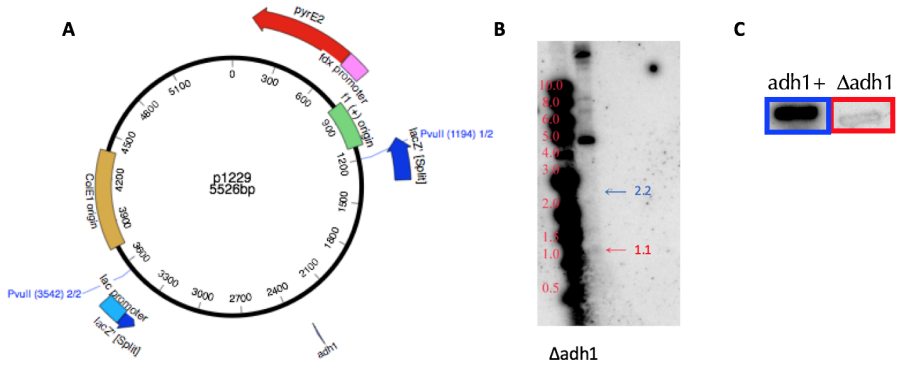


Figure 2.9: (A) pTA1229: deletion construct containing flanking regions of *adh1* used for probing Southern blots (B) probe bound to deletion of *adh1* (C) Colony Hybridisation.

Δ*adh2*

HVO_B0071(*adh2* was deleted from strains H1895 and H2538 using pTA1230 to generate pop-in strains H2468 and H2571, respectively and pop-out strain H2638 from H2538. Deletion of *adh2* was confirmed by colony hybridisation and by Southern blot for H2468

and H2638 using a 2600 bp radiolabelled probe of *adh2* flanking regions restricted from pTA1230 using *Pvu*II (figure).

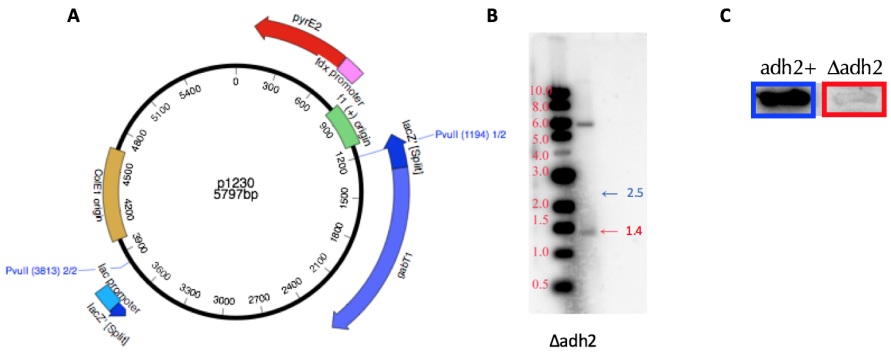


Figure 2.10: (A) pTA1230: deletion construct containing flanking regions of *adh2* used for probing Southern blots (B) probe bound to deletion of *adh2* (C) Colony Hybridisation.

Δ*tnaA*

HVO_0009(*tnaA* was deleted from strain H2638 using pTA1756 (Δ*tnaA*) and pTA1730 (*tnaA*Δ*EcoN*) to generate pop-in strains H2688 and H2687, respectively and pop-out strains H2764 from H2688. Deletion of *tnaA* through pTA1756 was confirmed by colony hybridisation for H2688 and H2687 using a 1488 bp radiolabelled probe of *tnaA* restricted from pTA875 with *Age*I (figure). Deletion of *tnaA* with pTA1730 could not be confirmed via colony hybridisation (figure). Deletion of *tnaA* was also confirmed by Southern blot for H2688(figure) using a 1826 bp radiolabelled probe of *tnaA* flanking regions restricted from pTA1756 using *Bam*HI and *Stu*I (figure).

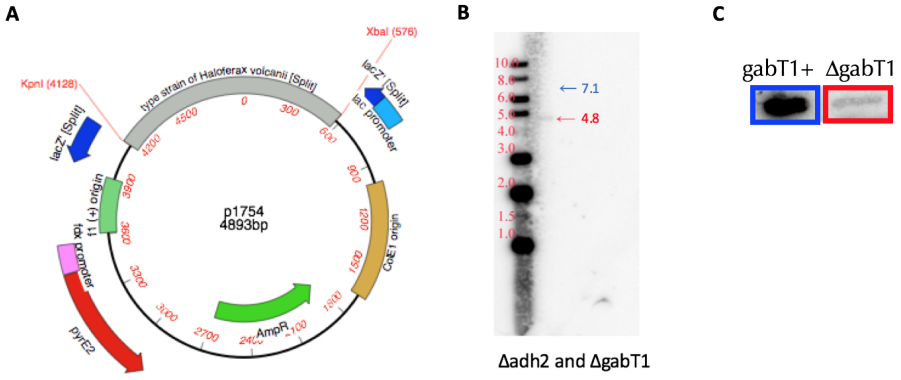


Figure 2.11: (A) pTA1756: deletion construct containing flanking regions of *tnaA* used for probing Southern blots (B) probe bound to deletion of *tnaA* (C) Colony Hybridisation.

$\Delta gabT1$

HVO_B0070(*gabT1* was deleted from strain H2764 and H2638 using pTA1754 ($\Delta adh2{+}gabT1$) to generate pop-in strains H2812 and H2766 respectively, and pop-out strain H2974. Deletion of *gabT1* was confirmed by colony hybridisation using a 2514 bp radiolabelled probe of *gabT1* restricted from pTA1743 with *Nde*I and *Aat*II (figure). Deletion of *gabT1* was also confirmed by Southern blot using a 1341 bp radiolabelled probe of *tnaA* flanking regions restricted from pTA1754 using *Kpn*I and *Xba*I (figure).

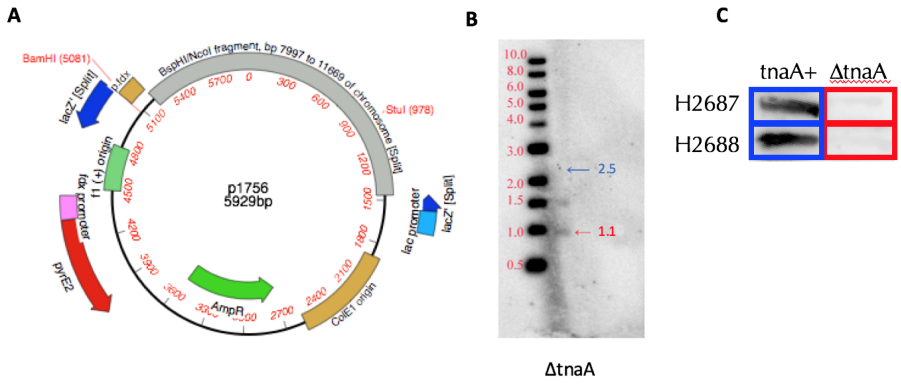


Figure 2.12: (A) pTA1754: deletion construct containing flanking regions of *adh2+gabT1* used for probing Southern blots (B) probe bound to deletion of *adh2+gabT1* (C) Colony Hybridisation.

Strains with Episomal Plasmids for Overexpression

Results from purifications of constructed strains and plasmids are presented in *Chapter 4: Optimisation of Enzyme Expression and Purification Conditions*.

Episomal plasmids for expression of ω-TA5 (*gabT5*)

Strain H1895 was transformed with pTA1670 (*p.tnaA::his6x-gabT5*) for overexpression of N-terminal his6x-tagged ω-TA5 under the control of *p.tnaA* and resulted in the strain H2466.

Strain H2764 was transformed with pTA1760 (*p.tnaA::his6x-gabT5-strepII*) for overexpression of N-terminal his6x-tagged and C-terminal strepII-tagged ω-TA5 under the control of *p.tnaA* and resulted in the strain H2837.

Episomal plasmids for expression of ω-TA1 (*gabT1*)

Strains H2764 and H2974 were transformed with pTA1759 (*p.tnaA::his6x-gabT5-strepII*) for overexpression of N-terminal his6x-tagged and C-terminal strepII-tagged ω-TA1 under the control of *p.tnaA* and resulted in the strains H2836 and H3096, respectively.
Strain H2974 was transformed with pTA1703 (*p.tnaA::his6x-gabT1*) for overexpression of N-terminal his6x-tagged ω-TA1 under the control of *p.tnaA* and resulted in the strain H3095.

Episomal plasmids for expression of ADH2 (*adh2*)
Strain H2974 was transformed with pTA1205 (*p.tnaA::his6x-adh2*) for overexpression of N-terminal his6x-tagged ADH2 under the control of *p.tnaA* and resulted in the strain H3094. Strains H2764 and H2974 were transformed with pTA1761 (*p.tnaA::his6x-adh2-strepII*) for overexpression of N-terminal his6x-tagged and C-terminal strepII-tagged ADH2 under the control of *p.tnaA* and resulted in the strains H2838 and H3097, respectively.
Strain H2974 was transformed with pTA1863 (*p.tnaA::strepII-adh2*) for overexpression of N-terminal strepII-tagged ADH2 under the control of *p.tnaA* and resulted in the strain H3098.

tnaA-deleted Strains

Gene Deletions

Strain H2765 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*hdrB,* Δ*pilB3C3,* Δ*tnaA*) was derived with the novel deletion *tnaA* for episomal overexpression of proteins.

Δ*tnaA*
HVO_0009(*tnaA* was deleted from strain H1895 using pTA1756 (Δ*tnaA*) to generate pop-in strain H2685 and pop-out strain H2765. Deletion of *tnaA* was confirmed by colony hybridisation using a 1488 bp radiolabelled probe of *tnaA* restricted from pTA875 with *Age*I (figure 2.13).

*tnaA*Δ*EcoN*

HVO_0009(*tnaA* was deleted from strain H1895 using pTA1730 (*tnaA*Δ*EcoN*) to generate pop-in strain H2684. Colonies were screened for deleted genoptypes, but deletions could not be confirmed by colony hybridisation using a 1488 bp radiolabelled probe of *tnaA* restricted from pTA875 with *Age*I (figure).

Δ*tnaA::hdrB*$^+$

HVO_0009(*tnaA* was deleted from strain H1895 using pTA1762 (Δ*tnaA::hdrB*$^+$) to generate pop-in strain H2686. Colonies were screened for deleted genoptypes, but deletions could not be confirmed by colony hybridisation using a 1488 bp radiolabelled probe of *tnaA* restricted from pTA875 with *Age*I (figure).

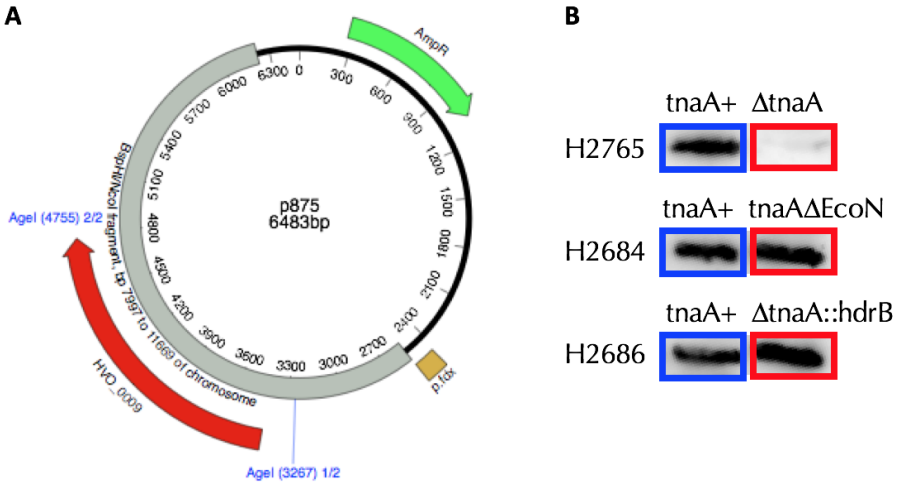


Figure 2.13: (A) pTA875: construct containing *tnaA* used for probing colony hybridisations (B) Colony Hybridisation.

Strains with Episomal Plasmids for Expression of β-galactosidase

Results regarding the expression of protein from constructed strains and plasmids are presented in *Chapter 4: Optimisation of Expression and Purification Conditions*.

Episomal plasmids for expression of β-galactosidase (*bgaHA*)

Strains H2682 and H2683 were each transformed with pTA1543 and pTA1544 (both: *p.tnaA::bgaHA*) for expression of β-galactosidase under the control of *p.tnaA* and resulted in the strains H2832, H2833, H2834 and H2835, respectively.

2.4 Discussion

Summarising the results, the deletions of the genes *adh1, adh2, tnaA* and *gabT1* were successfully confirmed by colony hybridisation and Southern blot and yielded the masterstrain H2974, which was transformed with episomal plasmids for overexpression. This also confirmed the successful construction of deletion constructs. With this, strains for experimental work for the next chapters were obtained.

Of the three deletion plasmids constructed for *tnaA* deletion/replacement/disruption (pTA1756 $\Delta tnaA$, pTA1762 $\Delta tnaA::hdrB$, pTA1730 $tnaA\Delta EcoN$), only construct pTA1756 resulted in successfully transformed strains. Construction of pTA1762 was attempted three times, but sequencing could not confirm the correct construct. Construction of this plasmid required blunt-end ligation, which is more difficult to obtain and is a likely reason for unsuccessful cloning. Disruption of *tnaA* with the previously to this study obtained plasmid pTA1730 could not be confirmed by colony hybridisation and did not yield successfully transformed strains.

Since the two strategies, (1) obtaining the masterstrain and (2) obtaining strains for assessing *tnaA* deletion, were pursued in parallel, additional *tnaA* deleted strains, that had been derived previously to this study, were transformed with plasmids for the expression of the reporter protein β-galactosidase. These strains were used to analyse expression levels of *tnaA* deleted against non-deleted strains; strains with novel gene deletions derived in this study were not used for $\Delta tnaA$ analysis.

2.5 Conclusion and Future Perspectives

A new template strain H2765 for optimised expression of episomal plasmids, which includes the novel deletion of the *tnaA* gene, was derived from the previous template strain H1895. However, deletion was only confirmed by colony hybridisation. Future work re-

quires the confirmation of the deletion by Southern blotting.

The cloning strategy for the construct pTA1762 $\Delta tnaA::hdrB$, and possibly pTA1730 $tnaA\Delta EcoN$, needs to be revised and new vectors need to be constructed.

Chapter 3

Optimisation of Expression and Purification Conditions

3.1 Background

Tryptophan is the inducer of the *p.tnaA* promoter on episomal plasmids and, as mentioned in *Chapter 2*, is degraded by tryptophanase A, an enzyme that is natively expressed under the control of the *p.tnaA* promoter to break down its own inducer. Deletion of *tnaA* will eliminate the degradation of tryptophan and hence, all tryptophan added is expected to be used for protein expression. Tryptophan is energetically costly to cells, and genes involved in tryptophan synthesis and degradation are tightly regulated. The *p.tnaA* promoter has effectively zero basal activity, and is rapidly and strongly induced in the presence of $\geq$ 1 mM tryptophan, leading to 100-fold induction [Large et al., 2007]. Further, the enzyme tryptophanase in *E.coli* converts tryptophan to indole [Li and Young, 2013], a signalling molecule involved in various cell processes, amongst them quiescence, whereby cell division is blocked in the presence of indole [Chen et al., 2015]. In order to analyse the expected difference in protein expression, Δ*tnaA* and *tnaA+* strains with episomal plasmids for *bgaHa* gene expression under the *p.tnaA* promoter were assayed by blue-white screening (figure 3.1) and ortho-nitrophenyl-para-galactoside (ONPG)-assay with the aim

of comparing expression levels between strains. It was previously demonstrated that the *bgaHa* gene is a suitable marker for the analysis of heterologous gene expression levels [Dyall-Smith, 2008, Strillinger et al., 2016], since the expression of active BGHA shows a linear correlation with measured rates for the catalytic conversion of ortho-nitrophenyl-para-galactoside (ONPG) to ONP monitored at 405 nm (figure 3.1).

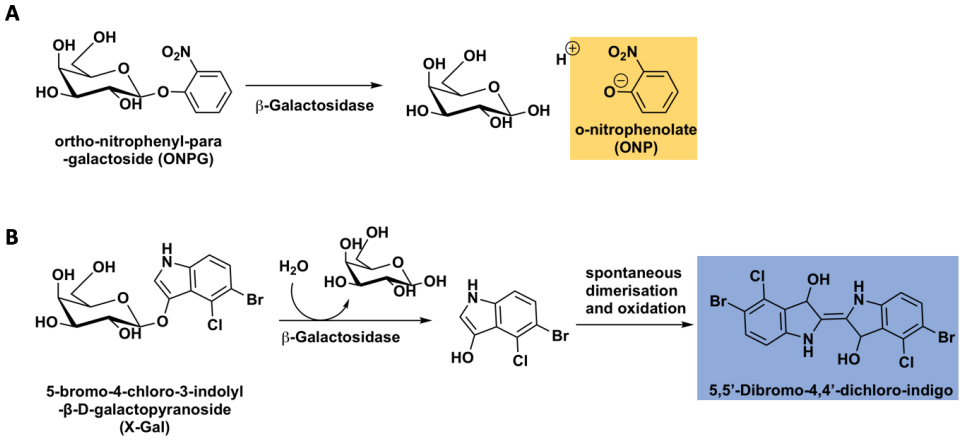


Figure 3.1: (A) Conversion of ONPG by β-galactosidase to ONP, which appears yellow. (B) Conversion of X-Gal by β-galactosidase resulting in a blue compound.

A third approach towards comparing expression levels was taken after the protocol for expression and purification of ADH2 with äkta Prime plus had been optimised, whereby expression levels of his6x-tagged ADH2 from a $\Delta tnaA$ strain and a $tnaA+$ strain were compared based on their elution profiles.

Secondly, strains derived from masterstrain H2974, which is deleted for genes *adh1, adh2, tnaA* and *gabT1* (see *Chapter 2, Results & Discussion*), with the aim of expressing his6x-tagged and/or strepII-tagged ADH2, ω-TA1 and ω-TA5, were in the first instance evaluated for purity of expressed protein through metal affinity and Strep-Tactin chromatography (figure 3.2). In a second step, enzyme activity was analysed through UV-Spectrophotometry, based on the conversion of the cofactor $NADP^+$ to NADPH. To this end, the purification of his6x-tagged and strepII-tagged proteins on gravity columns (see *Materials and Methods of Chapter 3* was used for initial characterisation of ADH2, ω-TA1 and ω-TA5 and consequently expanded to purify protein with äkta Prime plus. The protocol for the äkta Prime plus was particularly optimised to yield ADH2 in its

highly active tetrameric form. Development of an optimised purification protocol for ω-TA1 was not pursued, after collaboration partners (Jennifer Cassidy under supervision of Francesca Paradisi, Nottingham University, Chemistry Department, 2017) working on the identification of suitable substrates for these particular enzymes did not detect sufficient activity for either of the transaminases. Hence, pursuit of this enzyme was dropped and only work on ADH2 continued.

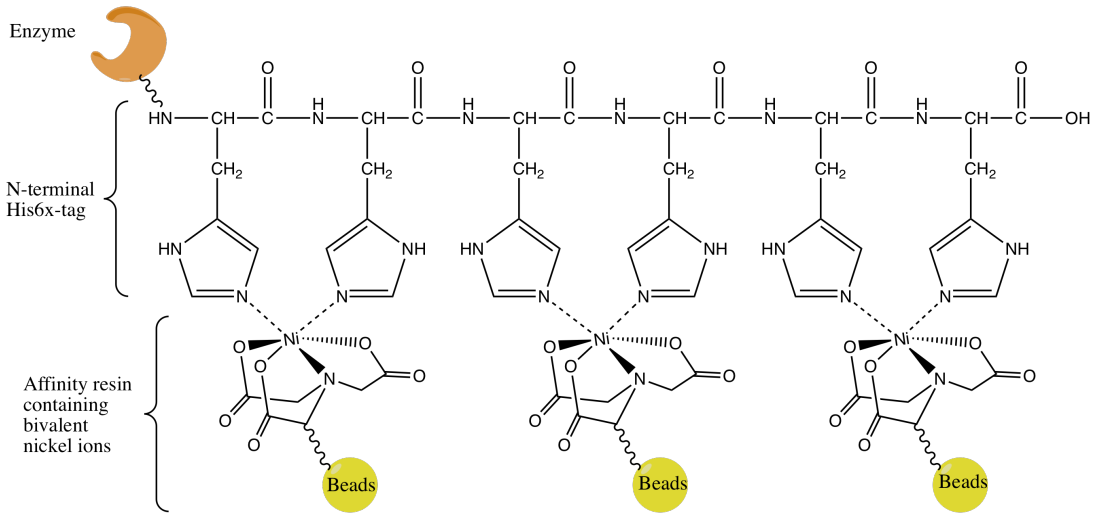


Figure 3.2: (A) The enzyme, which is N-terminally modified with a stretch of 6 polyhistidine, binds to the affinity resin, whereby two neighbouring histidine form a complex with one Ni^{2+} with a binding affinity of K_d = 10 µM. The tagged proteins are then eluted by an addition of imidazole to the solution, which outcompetes the histidine bound to the metal, thus releasing the proteins from the matrix [Bornhorst and Falke, 2000]. (B) The enzyme, which is N- or C-terminally modified with a short peptide recognition sequence (WSHPQFEK), is immobilised to streptactin-coated surfaces with a K_d = 1 µM. Upon addition of desthiobiotin, the tag is displaced and enzymes are eluted [Skerra and Schmidt, 2000].

3.2 Materials and Methods

3.2.1 Materials

***Haloferax volcanii* strains**

Table 3.1: Strains used in this chapter

Strain	Genotype	Notes
Parent: H1895	Δ*pyrE2, Nph-pitA,* Δ*mrr,* Δ*hdrB,* textitCdc48d-Ct, ΔpilB3C3	Constructed by TA [Strillinger et al., 2016]
H1969	{*p.tnaA::bgaHa*}	overexpression pTA1543
H1970	{*p.tnaA::bgaHa*}	overexpression pTA1544
H2466	{*p.tnaA::6xhis-gabT5*}	overexpression pTA1670
Parent: H2764	Δ*pyrE2, Nph-pitA,* Δ*mrr,* Δ*hdrB,* Δ*adh1,* Δ*adh2,* Δ*tnaA*	
H2836	{*p.tnaA::6xhis-2xstrep-gabT1+*}	overexpression p1759
H2837	{*p.tnaA::6xhis-2xstrep-gabT5+*}	overexpression p1760
H2838	{*p.tnaA::6xhis-2xstrep-adh2+*}	overexpression p1761
Parent: H2974	Δ*pyrE2, Nph-pitA,* Δ*mrr,* Δ*hdrB, Cdc48d-Ct,* Δ*pilB3C3*, Δ*adh1*, Δ*adh2*, Δ*tnaA*, Δ*gabT1*	
H3094	{*p.tnaA::6xhis-adh2+*}	overexpression pTA1205
H3095	{*p.tnaA::6xhis-gabT1+*}	overexpression pTA1703
H3096	{*p.tnaA::6xhis-2xstrep-gabT1+*}	overexpression pTA1759
H3097	{*p.tnaA::6xhis-2xstrep-adh2+*}	overexpression pTA1761
H3098	{*p.tnaA::2xstrep-adh2+*}	overexpression pTA1863

Parent: H2682		
H2832	{*xp.tnaA::bgaHa*}	overexpression p1543
H2833	{*xp.tnaA::bgaHa*}	overexpression p1544
Parent: H2683		
H2834	{*xp.tnaA::bgaHa*}	overexpression p1543
H2835	{*xp.tnaA::bgaHa*}	overexpression p1544
Parent: H1325	Δ*pyrE2, Nph-pitA,* Δ*mrr,* Δ*hdrB,* Δ*adh1,* Δ*adh2*	
H1332	{*xp.tnaA::6xhis-adh2+*}	overexpression p1205

[ ] indicates integrated plasmid, { } indicates episomal plasmid, TA = Thorsten Allers

Chemicals, Buffers and Media

H.volcanii media has been described in detail in *Chapter 2, Materials and Methods Section.*

Buffers and Chemicals for Protein Expression and Purification by Gravity Column

Tryptophan

0.2 g (∼ 3 mM) powdered tryptophan for induction, step 1

0.1 g (∼ 4.5 mM final concentration) powdered tryptophan for induction, step 2

Buffer A for gravity columns

1 ml 20 mM HEPES, pH 7.5, 20 ml 2 M NaCl, 500 μl 1 mM PMSF, dH_2O to 50 ml, add 5 ml protease inhibitor as required

Elution buffer A for gravity columns

Buffer A with 100 mM, 200 mM and 500 mM imidazole

Buffer B for gravity columns

1 ml 20 mM HEPES, pH 7.5, 20 ml 2 M NaCl, 500 μl 1 mM PMSF, dH_2O to 50 ml

Elution buffer B for gravity columns

0.016 g D-desthiobiotin dissolved in 15 ml buffer B (5 mM)

Buffer C for gravity columns

1 ml 20 mM HEPES, pH 7.5, 20 ml 2 M NaCl, 500 μl 1 mM PMSF, dH_2O to 50 ml

Elution Buffer C for gravity columns

Buffer C with 5 mM or 10 mM EDTA, unless stated otherwise

3.0 % SDS-PAGE gel (stacking)

0.77 ml 30 % acryl/bias, 0.2 ml 1.5 M Tris pH 6.8 (0.4 % SDS), 25 μl 10 % AMPS (ammonium persulfate, Sigma), 2.5 μl TEMED (tetramethyleethylenediamine, Sigma), 0.77 ml H_2O

12.5 % SDS-PAGE gel (resolving)

3.13 ml 30 % acrylamide/bis-acrylamide (Severn Biotech Ltd), 1.88 ml 1.5 M Tris pH 8.8 (0.4 % SDS), 37.5 μl 10 % AMPS (ammonium persulfate, Sigma), 3.75 μl TEMED (tetramethyleethylenediamine, Sigma), 2.46 ml H_2O

SDS-PAGE running buffer

0.25 M Tris, 1.92 M glycine, 1 % SDS

3.0 % Native-PAGE gel (stacking)

0.77 ml 30 % acrylamide/bis-acrylamide, 0.2 ml 1.5 M Tris pH 6.8, 25 μl 10 % AMPS (ammonium persulfate, Sigma), 2.5 μl TEMED (tetramethyleethylenediamine, Sigma), 0.77 ml H_2O

12.5 % Native-PAGE gel (resolving)

3.13 ml 30 % acrylamide/bis acrylamide (Severn Biotech Ltd), 1.88 ml 1.5 M Tris pH 8.8, 37.5 μl 10 % AMPS (ammonium persulfate, Sigma), 3.75 μl TEMED (tetram-

ethyleethylenediamine, Sigma), 2.46 ml H_2O

Native-PAGE running buffer

0.25 M Tris, 1.92 M glycine

Gel-loading Dye for native-Page

5 ml H_2O, 4.375 ml 1 M Tris.HCl pH 6.8, 25 mg bromphenol blue, 0.625 ml 80 % glycerol

Buffers and Chemicals for Protein Expression and Purification by AeKTA prime plus

20 % Ethanol

SDW

0.5 M NaOH

Loading buffer

20 mM Hepes, pH 7.5, 2 M NaCl, 20 mM imidazole

Wash buffer

20 mM Hepes, pH 7.5, 2 M NaCl

Elution buffer

20 mM Hepes, pH 7.5, 2 M NaCl, 50 mM EDTA

Buffers for BCA Protein Assay

BSA Buffer

20 mM HEPES, pH 7.5, 0.15 M NaCl

Sample Buffer

20 mM HEPES, pH 7.5, 2 M NaCl

ADH2 Storage and Relative Activity Assay Buffers and Solutions

100 mM $NADP^+$ prepared in H_2O

2 M ethanol

Glycine-KOH Buffer

50 mM glycine-KOH, pH 10 or pH 8, 4 M KCl, 3 M KCl or 2 M KCl respectively

Glycine-KOH Buffer + $ZnCl_2$

50 mM glycine-KOH, pH 10 or pH 8, 3 M KCl, 20 μM $ZnCl_2$

50 mM Glycine-KOH

3.75 g glycine in 0.995 L, adding 5.2 ml 5 M KOH

ONPG-Assay

BGAH Buffer

2.75 M NaCl, 0.5 M Tris, 10 μM $MnCl_2$ (store at 4 ℃), before use add 0.1 % 2-mercaptoethanol

ONPG solution

6 mg/ml ortho-nitro-phenyl-beta-D-galactoside dissolved in bgaH buffer

Other

TE

10 mM Tris.HCl, pH 8.0, 1 mM EDTA

TBE

89 mM Tris.HCl, 89 mM boric acid, 2 mM EDTA

Gel Loading Dye (5x)

50 mM Tris.HCl, 100 mM EDTA, 15 % Ficoll 400 (w/v), 0.25 % bromophenol blue (w/v), 0.25 % Xylene Cyanol FF (w/v)

3.2.2 Methods

Protein expression and purification

Overexpression of proteins in *Hv. volcanii*

A starter culture of 5 ml Hv-Ca+ was inoculated with 1-2 colonies and grown overnight at 45 ℃ on a rotator. After an $OD_{600} \sim 0.1$ was reached cells were diluted (1:250, 1:100, 1:50 and 1:25) into fresh 5 ml Hv-Ca+ and grown at 45 ℃ until $OD_{600} \sim 0.1$ was reached. Cultures were used to inoculate 50 ml Hv-Ca+ broth in 1:100, 1:200 and 1:500 dilutions and incubated for 24 h in a 45 ℃ waterbath shaking at 150 rpm until an $OD_{600} \sim 0.5$ was reached. Cultures were diluted (1:120 or 1:60) into 333 ml YPC broth and incubated overnight in a 45 ℃ shaking incubator at 150 rpm until an $OD_{600} \sim 0.5$ was reached. When $OD_{600} \sim 0.5$ was reached, expression was induced upon addition of 0.2 g tryptophan (3 mM final concentration), followed by a second induction step with 0.1 g tryptophan (4.5 mM final conc.) after 1 hour. Expression was performed at 45 ℃ for two hours total in a shaking incubator at 150 rpm.

Optimised overexpression of ADH2 in *Hv. volcanii*

A starter culture of 5 ml Hv-Ca+ was inoculated with 1-2 colonies and grown overnight at 45 ℃ on a rotator. After an $OD_{600} \sim 0.1$ was reached cells were diluted (1:250, 1:100, 1:50 and 1:25) into fresh 5 ml Hv-Ca+ and grown at 45 ℃ until $OD_{600} \sim 0.1$ was reached. Cultures were used to inoculate 50 ml Hv-Ca+ broth in 1:100, 1:200 and 1:500 dilutions and incubated for 24 h in a 45 ℃ waterbath shaking at 150 rpm until an $OD_{600} \sim 0.5$ was reached. Cultures were diluted (1:120 or 1:60) into 333 ml YPC broth and induced with 0.1 g tryptophan and incubated in a 45 ℃ shaking incubator at 150 rpm until an OD_{600} of $\sim$ 1.2 - 1.5 was reached (after 24 - 36 h). When $OD_{600} \sim 1.2$ was reached, cells were spun down and were either prepared for purification (see *Protein Purification*), or frozen at -20 ℃ for later purification.

Protein Purification by Gravity Column

Column residue IMAC Sepharose 6 Fast Flow (GE Healthcare) bead suspension was used per sample and washed twice with an excess of dH_2O before being charged with 1 ml 0.2M $NiSO_4$ solution for 30 minutes at 4 ℃. Beads were washed twice with an excess of dH_2O and further washed once with 1 ml buffer A (500 mM Imidazole) and three times with 5 ml buffer A (20 mM Imidazole). Beads were resuspended in $\sim$ 1 ml buffer A (20mM Imidazole). Cells were pelleted at 3600 x g for 10 minutes at 4 ℃ and resuspended in 7 ml buffer A (20 mM containing 1 X Sigma Fast protease inhibitor). Cells were then lysed by sonication on ice (3x 30 s at 8 μ) until the lysate became clear. Lysed cells were centrifuged at 48000 x g for 30 min at 4 ℃. Supernatant was subsequently syringe filtered using to 0.8 μm, 0.45 μm and 0.2 μm filters and a sample of 37.5 μl was taken (to be run on gel). The debris pellet was resuspended in 200 μl buffer A and an aliquot sample was taken. Equilibrated beads were added to the cell lysate and rotated gently for 1 hour at 4 ℃. The mixture was loaded onto a Poly-Prep gravity column (Bio-Rad) and the flowthrough collected and applied a second time onto the beads. The column was washed with 5 ml buffer A (20 mM imidazole) and a second time with 10 ml buffer A. The bound proteins were eluted in two to three fractions with 2x 1 ml buffer A (100mM imidazole), 2x 1 ml buffer A (200 mM imidazole) and optional 2x 1 ml buffer A (500 mM imidazole). Fractions were pooled and either dialysed against buffer B and stored (overnight) or diluted with buffer B and immediately applied to a Strep column. 0.5 ml Strep-Tactin Sepharose (IBA) per sample was aliquoted into a Poly-Prep gravity column and equilibrated twice with binding buffer B. The protein sample was applied to the column three times. The column was washed five times with buffer B before elution with elution-buffer B (2.5 mM D-destiobiotin) in three steps of 0.4 ml, 0.7 ml and 0.4 ml. Aliquot samples of 37.5 μl each were collected after each step and analysed by SDS-PAGE.

Protein Purification by äkta prime plus (Amersham Biosciences)

All solutions applied to the äkta purification system were syringe filtered through a 0.55μm filter and degased for 5 minutes in an ultrasonic bath. The äkta prime plus was stored in 20 % ethanol and was purged with dH_2O before use for purification. HisTrap HP (GE Healthcare) 1 ml (IMAC) columns were regenerated with 0.5 M NaOH after each

use and recharged with 0.2 M $NiSO_4$ solution before being stored at 4 ℃ or RT in 20% ethanol. The attached column was consequently washed with dH_2O, binding buffer and wash buffer, 5 ml each, 1 ml / min, before the protein sample was loaded onto the column via the tubing inlet (not injection valve) at 0.3 - 0.5 ml / min. Cells of expression cultures were pelleted at 3600 *x g* for 10 minutes at 4 ℃ and resuspended in 5 ml wash buffer. Cells were then lysed by sonication on ice (3x 30 s at 8 μ) until the lysate became clear. Lysed cells were centrifuged at 48000 *x g* for 30 min at 4 ℃. Supernatant was subsequently subjected to a 0.45 μm filter before being applied to the column. The column was then washed with wash buffer until the baseline, monitored via UV-absorbance, was attained and the system was autozeroed. A gradient of 100 % elution buffer over 50 ml was set and His-tagged proteins eluted between 11 % to 20 % elution buffer. Captured fractions were immediately stored on ice and screened to find the most active ones. Selected fractions were pooled and dialysed via Viva Spin columns (Sartorius) using a 3 M glycine-KOH buffer, pH 8 with or without Zinc.

Protein Characterisation Methods

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Gels were prepared in a disposable cassette (Invitrogen) and consisted of a stacking gel and a running gel. The running gel (12.5 % SDS-PAGE) was poured first into the cassette and covered with isopropanol until settled. After thorough removal of the isopropanol the stacking gel (3 %) was poured and a comb inserted. Protein samples were applied onto gel in a 1/5 mixture with 4x gel loading dye. Gels were run at 200 V, 36 mA for 1 hour and 20 minutes in 1 X SDS-Page running buffer. Gels were stained with coomassie blue and destained with water overnight.

Native - Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Gels were prepared in a disposable cassette (Invitrogen) and consisted of a stacking gel and a running gel. The running gel (12.5 % Native-PAGE) was poured first into the cassette and covered with isopropanol until settled. After thorough removal of the isopropanol the stacking gel (3 %) was poured and a comb inserted. Protein samples were applied onto gel in a 1/5 mixture with 4x gel loading dye. Gels were run at 25 V, 36 mA

for 3 hour in 1 X Native-Page running buffer. Gels were stained with coomassie blue and destained with water overnight.

BCA Protein Assay

The Pierce Bradford Concentration Assay was used to assess the protein concentration of purified protein samples according to the provided protocol. A standard curve using Bovine Serum Albumin (BSA) was measured based on 2-3 replicates per standard and sample. Working reagents (WR) A and B were added to the prepared standards and samples and plates were incubated at 37 ℃ shaking in an Epoch 2 Microplate Spectrophotometer for 30 minutes before being measured at 562 nm (figure 3.3).

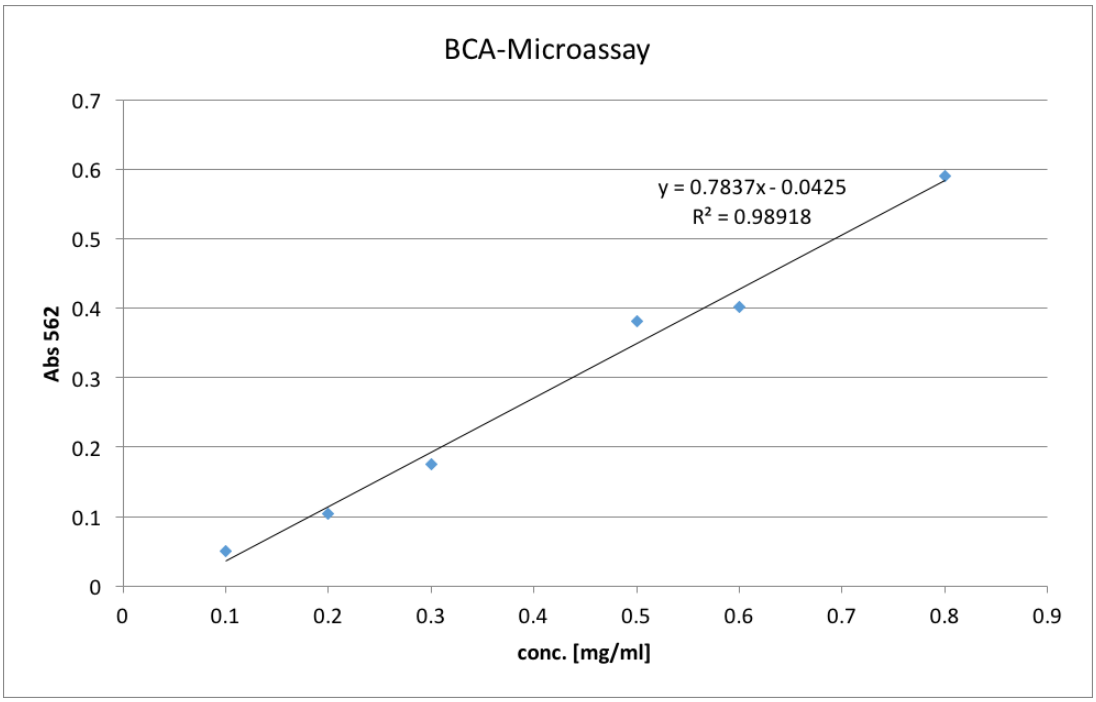


Figure 3.3: Bradford Concentration Assay using known concentrations of Bovine Serum Albumin.

ONPG Assay

The lactose analogue o-nitrophenyl beta-D-galactopyranoside (ONPG) was used to assess the activity of halophilic β-galactosidase (BGHA). Cells were grown overnight to OD_{600} $\sim$ 0.4 and induced with Tryptophan concentrations ranging from 0.0 mM to 5.0 mM for 2 hours. 200 μl induced cells were mixed with 700 μl bgaH buffer and 2 drops of 0.1 % SDS in Eppendorf tubes. The reaction was started by addition of 150 μl ONPG solution and the time was noted. Reactions were stopped upon addition of Na_2CO_3. Reactions were transferred to 1 ml cuvettes for measurement on a UV/Vis-Spectrophotometer. Upon hydrolysis of ONPG to ONP the solution turns yellow. By measuring the rate at which the colour intensity increases the activity of the enzyme can be calculated according to equation.

$$\beta - GalUnit(Miller'sUnit) = \frac{(OD_{420} \times 1000)}{(OD_{600} \times cellvolume \times t_e - t_0)} \tag{3.1}$$

Equation: The intensity of ONP is measured at OD_{420}. The unit for the cell volume is calculated in ml and time is measured in minutes. t_0 is the time at the start of the reaction and t_e is the time at the end of the reaction. A strain with no active bgaH gene is used as the negative control.

Blue-White Screening

Cas+trp plates were used to screen for the expression of β-galactosidase. Tryptophan was added to the plates in different concentrations up to 5.0 mM. After a 24 h incubation time at 45 ℃, colonies were sprayed with 5-bromo-4-chloro-3-indolyl--d-galactopyranoside (X-Gal). X-gal, when cleaved by -galactosidase, yields galactose and 5-bromo-4-chloro-3-hydroxyindole. The latter dimerizes and is oxidized into 5,5'-dibromo-4,4'-dichloro-indigo, which appears blue in visible light. X-gal itself is colorless, hence the presence of blue coloured product may therefore be used as a test for the presence of active β-galactosidase.

Spectrophotometric Measurement (Relative Specific Activity)

An Epoch 2 Microplate Spectrophotometer (BioTek) was used in connection with 96-well UV transparent plates to monitor the absorption at 340 nm of NADPH. The extinction coefficient and pathlength specific for this system for NADPH is 3.3 and was supplied by BioTek. The relative specific activity is calculated with the following equation:

$$U/mg = \frac{AverageSlope}{3.3} * \frac{1}{V_a} * \frac{1}{c_a} * V_f \tag{3.2}$$

whereby V_a is the volume of enzyme added to the reaction, c_a is the concentration of enzyme added and V_f is the total volume of the reaction. The average slope was calculated from the obtained measurement values over time:

$$AverageSlope = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} \tag{3.3}$$

Data was plotted using Excel (Microsoft). Assays were performed at 50 ℃.

3.3 Results

3.3.1 Analysis of Expression Constructs

Purifications by Gravity Column

Initially, purifications of enzymes ADH2, ω-TA5 and ω-TA1, were performed using gravity columns, since in this way multiple samples can be purified simultaneously and this accelerated the initial screen for protein expression (figures 3.4, 3.5 and 3.6). Use of Ni^{2+}-NTA columns bears the disadvantage of non-specific binding of natively histidine-rich protein, whilst StrepII-tags are highly specific peptide sequences for obtaining purely recombinant protein.

ω-Transaminase 5

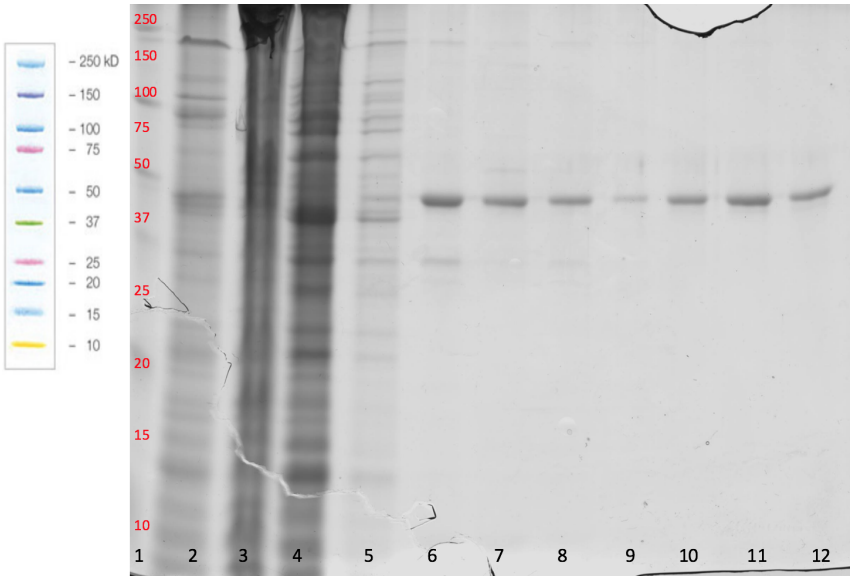


Figure 3.4: SDS-Page of purified fractions from expression of ω-transaminase 5 (49 kDa) in strain H2837; lane 1: Marker, lane 2-5: flowthrough and washes, lane 6-8: elutions with imidazole, lane 9: wash, lane 10-12: elutions with desthiobiotin.

ω-Transaminase 1

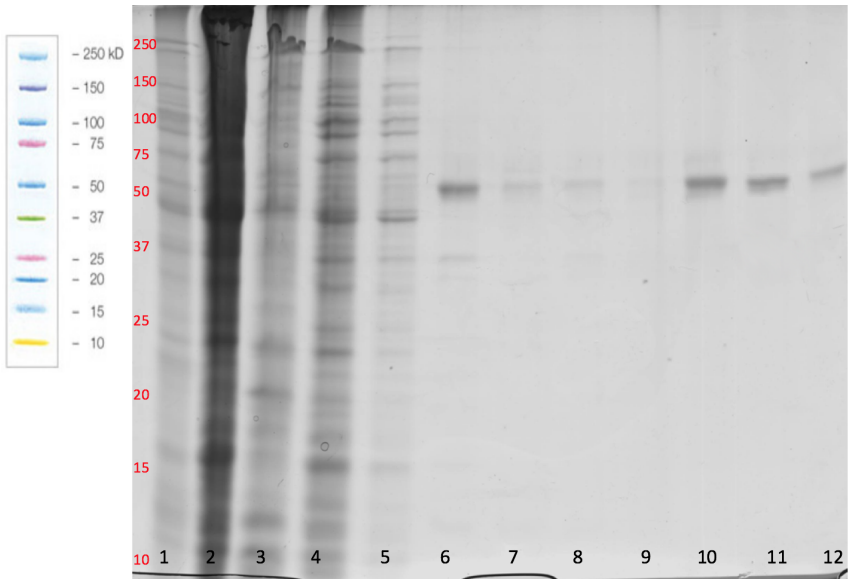


Figure 3.5: SDS-Page of purified fractions from expression of ω-transaminase 1 (46 kDa) in strain H3096; lane 1: Marker, lane 2-5: flowthrough and washes, lane 6-8: elutions with imidazole, lane 9: wash, lane 10-12: elutions with desthiobiotin.

Alcohol Dehydrogenase 2

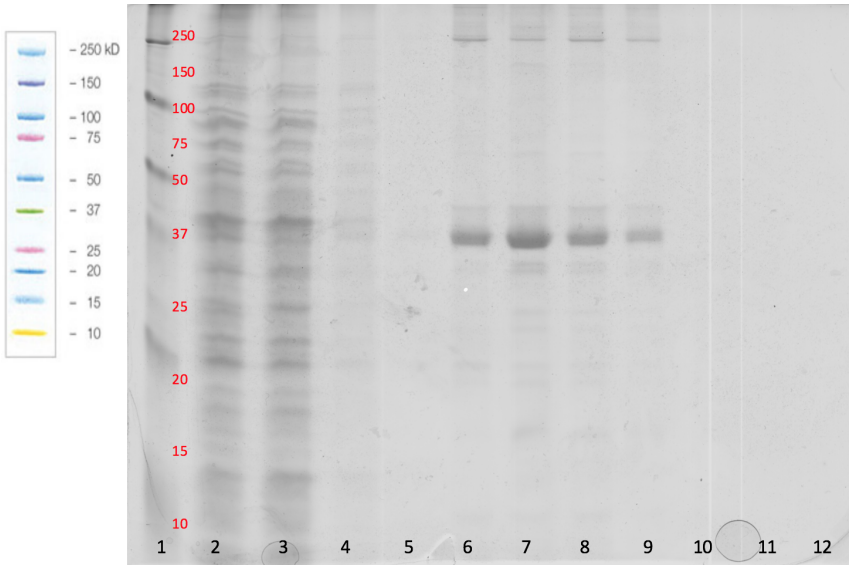


Figure 3.6: SDS-Page of purified fractions from expression of ADH2 (37 kDa) in strain H3097; lane 1: Marker, lane 2-5: flowthrough and washes, lane 6-8: elutions with EDTA, lane 9: wash, lane 10-12: elutions with desthiobiotin.

Purification of TA5 showed co-purification of a histidine-rich contaminant, especially in the first fraction. StrepII-tag purification resulted in fractions containing only recombinant protein. Purification of TA1 resulted as well in the co-purification of a histidine-rich contaminant, but less then for TA5. Almost all the protein was eluted in the first fraction (50 mM imidazole), hinting at weaker binding of TA1 to the resin, when compared to TA5. This might be down to sterical hindrance at the N-terminus or might have been simply correlated to less successfully charged resin, explaining also less co-purification. The StrepII-tag elutions of TA1 resulted in a highly purified sample. Co-purifying bands for both TAs could result from early termination products, since the bands run lower than the expected protein.

ADH2 (figure 3.6) was eluted with EDTA instead of imidazole, since activity assays performed after imidazole elution showed no activity. Elution with EDTA on gravity columns resulted in high amounts of recovered protein, however, co-purification of histidine-rich contaminants was observed. The imidazole elution of ADH2 showed a similar profile to TA1, whereby most protein eluted in the first fraction from the Ni^{2+}-NTA column and desthiobiotin elutions from Streptactin columns showed equally pure protein. In comparison, EDTA elution did not seem to be compatible with a consequent Strep-purification step, since no protein was recovered from Streptactin fractions. EDTA is a chelating agent and might interfere with the resin of the Streptactin column. If this strategy was to be pursued, a buffer exchange step would be required between the two purification systems.

Purifications by äkta Prime Plus

Eluted fractions from gravity columns were assayed for activity after dialysis. Activity of ADH2 was none after elution with EDTA, imidazole or desthiobiotin. Therefore, a different purification approach by äkta Prime plus was investigated for His- and Strep-tagged protein using HisTrap and StrepTrap HP purification columns. His-tagged protein was eluted using EDTA or imidazole (*data for imidazole not shown*). Figure 3.7 shows the chromatogram of EDTA eluted proteins, which were expressed according to the standard protocol for protein expression. Figure 3.8 shows the chromatogram of EDTA eluted proteins following expression based on the optimised protein expression protocol (both expression protocols see *Methods, Protein expression and purification*). For comparison between the two protocols a native-Page was prepared (figure 3.9). A chromatogram of Strep-tagged protein which was eluted using desthiobiotin, is shown in figure 3.10. Relative specific activity units provided in the tables below are averaged over multiple purification runs (table 3.2 and table 3.3).

EDTA elution from HisTrap following standard expression protocol

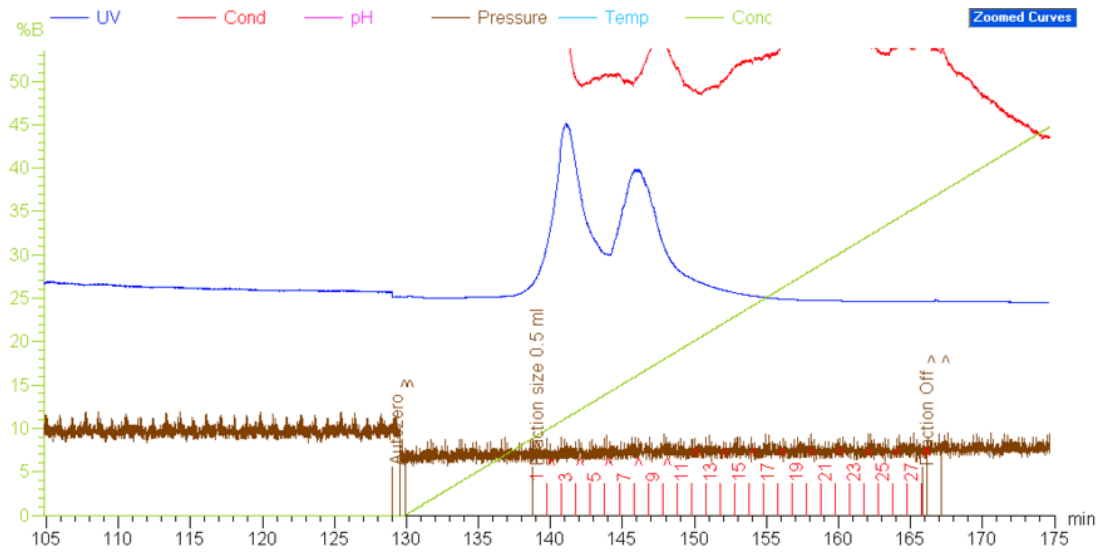


Figure 3.7: Chromatographic elution profile of His6x-tagged ADH2 expressed in H3094 using the shorter induction period (2 hrs) of the standard protocol. Multiple purifications were performed and resulted consistently in two eluting peaks, whereby only the tail fractions (11-15) of the second peak showed higher activity (> 2 U/mg).

EDTA elution from HisTrap following optimised expression protocol

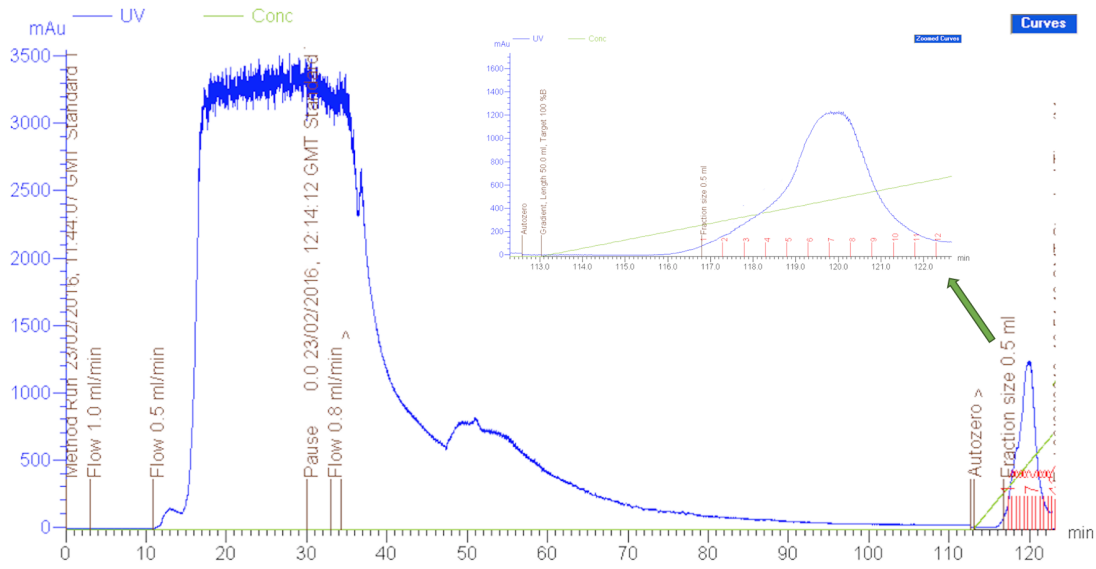


Figure 3.8: Chromatographic elution profile of His6x-tagged ADH2 expressed in H3094 using the longer induction period (24 -36 hrs) of the optimised protocol. Multiple purifications were performed and resulted consistently in one eluting peak, whereby all fractions showed high activity (> 2 U/mg). Complete chromatogram is shown for clarity as well as zoomed curve of eluted peak (smaller figure).

Native Page of ADH2 from standard vs optimised expression protocol

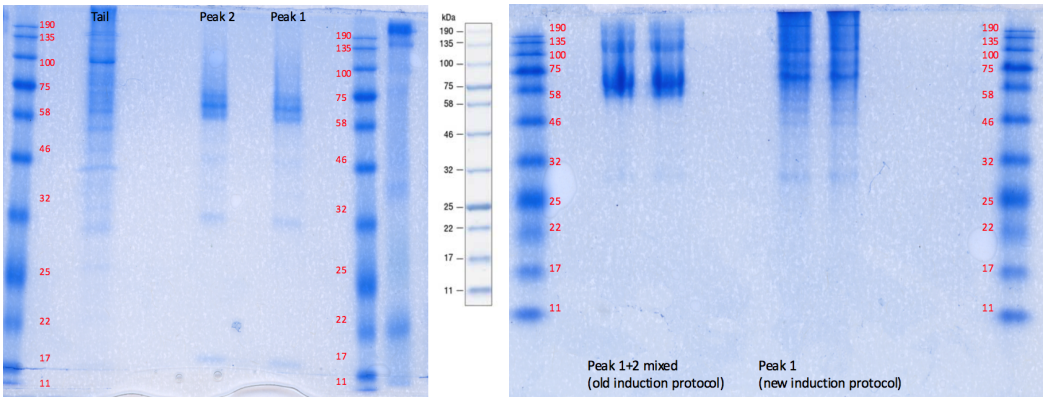


Figure 3.9: Left - Native Page of his6x-tagged ADH2 following expression of short induction (standard protocol): Tail shows a broader band at higher molecular weight, than Peak 2 and 1. Right - Native Page of his6x-tagged ADH2 following standard (old) induction protocol (peaks 1, 2 and tail mixed, seen on the lefthandside of the gel) and his6x-tagged ADH2 following optimised (new) induction protocol (peak 1, seen on the righthandside of the gel).

Table 3.2: Results from BCA and activity assays of his6x-tagged ADH2 (H3094) following expression according to standard protocol vs. optimised protocol

	H3094 standard	**H3094 optimised**
Protein concentration	1.3 mg/ml	1.7 mg/ml
Rel. Spec. Activity	2.3 U/mg	9.6 U/mg

StrepTrap elution

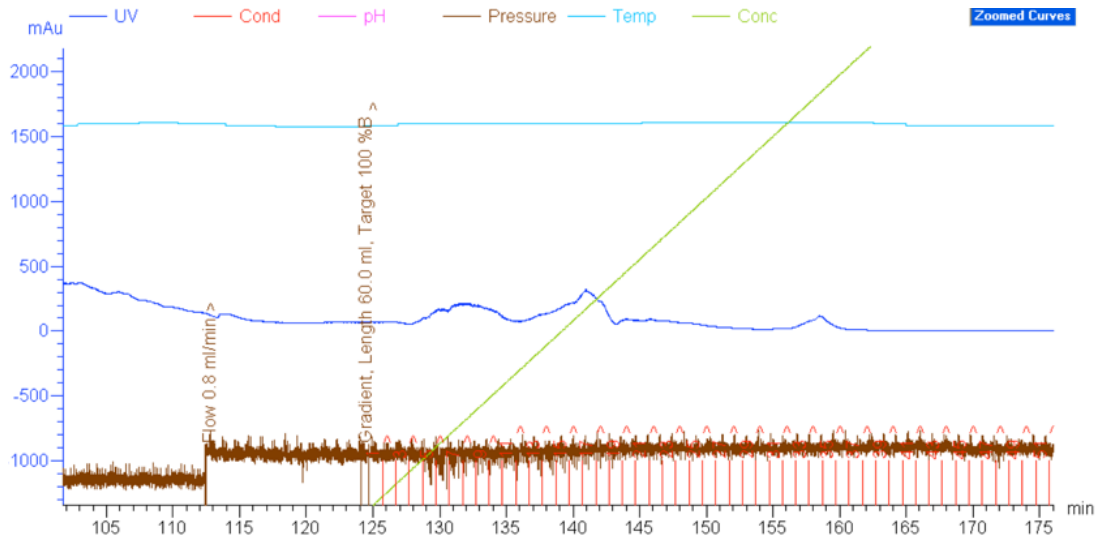


Figure 3.10: Chromatographic elution profile of StrepII-tagged ADH2 expressed in H3098 using the shorter induction period (2 hrs) of the standard protocol.

Comparison of elutions using Imidazolium, EDTA and Desthiobiotin

Table 3.3: Results from BCA and Activity Assays of his6x-tagged ADH2 (H3094) and strepII-tagged ADH2 (H3098) following expression according to standard protocol

	H3094 imidazolium	**H3094 EDTA**	**H3098 Desthiobiotin**
Protein concentration	1.8 mg/ml	1.5 mg/ml	0.8 mg/ml
Rel. Spec. Activity	0.6 U/mg	2.3 U/mg	0.3 U/mg

3.3.2 Analysis of Tryptophanase A Deletion

Blue-white Screening

The initial assay for analysing the *tnaA* deletion used simple Blue-white screening. Different tryptophan concentrations were added to Cas-plates and strains H1969, H1970 (functional tryptophanase A) and strains H2832, 2833, 2834 and 2835 ($\Delta tnaA$) were cultivated on the plates. All strains contained episomal plasmids for *bgaHa* expression. According to the gradient of tryptophan it is expected to see different shades of blue and less blue for tryptophanase A functional strains, since expression of β-galactosidase should be higher in $\Delta tnaA$ strains. The first concentration range assayed was 0, 0.5, 1.0, 2.5 and 5.0 mM tryptophan and cells were plated in a 1:10 dilution to a final OD_{600nm} of 0.1. No difference between the plates containing tryptophan nor the different strains was observed, whereby all colonies stained equally blue and cells were too densely plated. The follow-up experiment assayed a gradient consisting of lower concentrations of 0, 0.2, 0.25 (standard concentration), 0.4, 0.6, 0.8 and 1.0 mM tryptophan and cells were plated in a 10^{-6} dilution. Again, no difference between the plates containing tryptophan, nor the different strains was observed, whereby all colonies stained equally blue (figure 3.11).

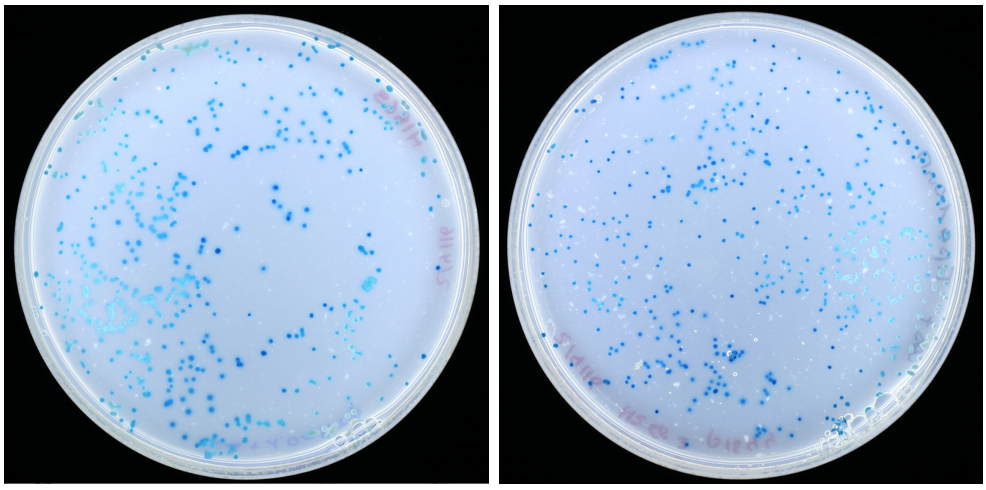


Figure 3.11: Blue-white screen of H1969 (left) and H2834 (right) on 0.2 mM tryptophan Cas-plates

Tryptophan-inducible promoters of other species are known to be strongly induced once stationary phase is reached [Kajitani and Ishihama, 1983]. Stationary phase in *H.volcanii* is on average reached after 12-24 h, and this seems to be the most plausible explanation for not observing a difference between concentrations.

ONPG-Assay

Subsequently, β-galactosidase activity was assayed by ONPG-assay, which is performed in liquid media. *H.volcanii* cells reach stationary phase, depending on starting dilution, at an OD of 0.7-0.8. Consequently, assays were carried out after cultures reached an OD_{600nm} of 0.4, before stationary phase. Additionally, the negative control strain H2838, containing no plasmid for *bgaHa* expression, was assayed. Assays were performed multiple times, data is only shown for two representatives (figure 3.12 and figure 3.13). Strains were cultivated in duplicates and measurements were taken in triplicates.

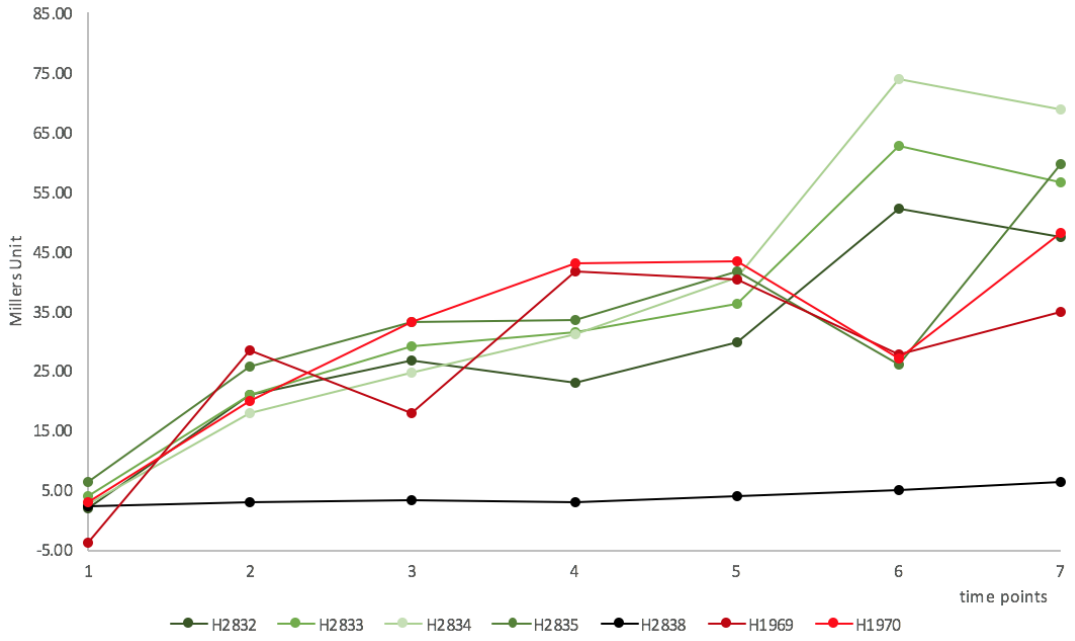


Figure 3.12: ONPG-Assay (1.5 hrs): time points represent 15 min steps at start at 0 minutes. Strain H2838 is the negative control, without an episomal plasmid for β-gal expression (black line). Strains H1969 and H1970 are *tnaA+* (red lines). Strains H2832, H2833, H2834 and H2835 are Δ*tnaA* (green lines).

It was theorised, that due to the variability between stopping the reaction and transferring

more than 42 samples for measurement, artefacts, such as observed for time point 6 in figure 3.12 may arise and lead to greater inaccuracy. It was therefore decided to reduce the number of strains for consequent measurements.

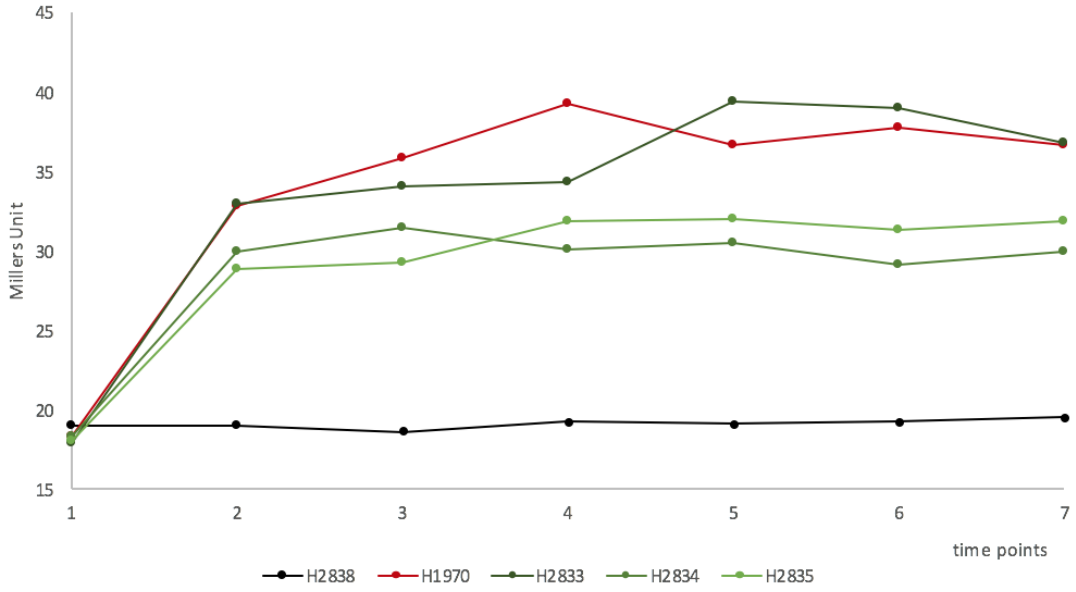


Figure 3.13: ONPG-Assay (1 hr): time points represent 15 min steps. Strain H2838 is the negative control, without an episomal plasmid for β-gal expression (black line). Strains H1969 and H1970 are *tnaA+* (red lines). Strains H2832, H2833, H2834 and H2835 are Δ*tnaA* (green lines).

Comparison of Elution-profiles

After establishing an optimised purification protocol for ADH2 expression, the strain H3094 (Δ*pyrE2, Nph-pitA,* Δ*mrr,* Δ*hdrB, Cdc48d-Ct,* Δ*pilB3C3,* Δ*adh1,* Δ*adh2,* Δ*tnaA,* Δ*gabT1,* {*p.tnaA::6xhis-adh2+*}), which was subsequently used for all activity assays in ionic liquids, was compared in terms of expression to the *tnaA+* strain H1332 (Δ*pyrE2, Nph-pitA,* Δ*mrr,* Δ*hdrB,* Δ*adh1,* Δ*adh2,* {*xp.tnaA::6xhis-adh2+*}). It was theorised that upon using the prolonged induction protocol, a difference between the *tnaA* deleted (H3094) and non-deleted strain (H1332) would be detected in form of less peak intensity and consequently less protein concentration for H1332 (figures 3.14 and 3.15 and table 3.5).

Both strains were cultivated simultaneously according to protocol and both cultures were

frozen and purified the next day. In both cases 10 active fractions were collected and consequently dialysed and concentrated to the same volume (400 µl).

H1332

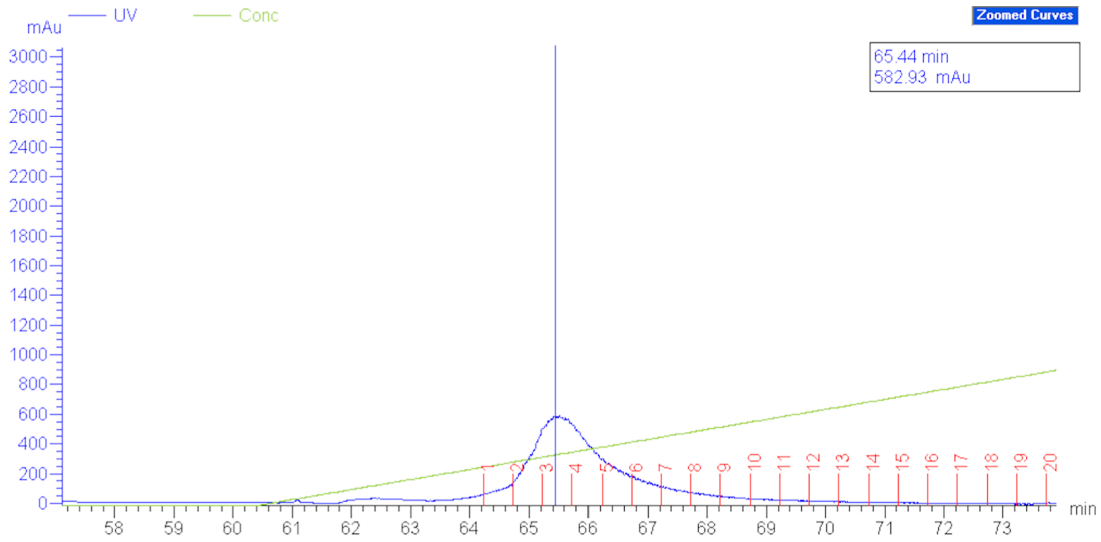


Figure 3.14: his6x-tagged ADH2 expressed in H1332 (*tnaA+*), maximal mAU: 582.93.

H3094

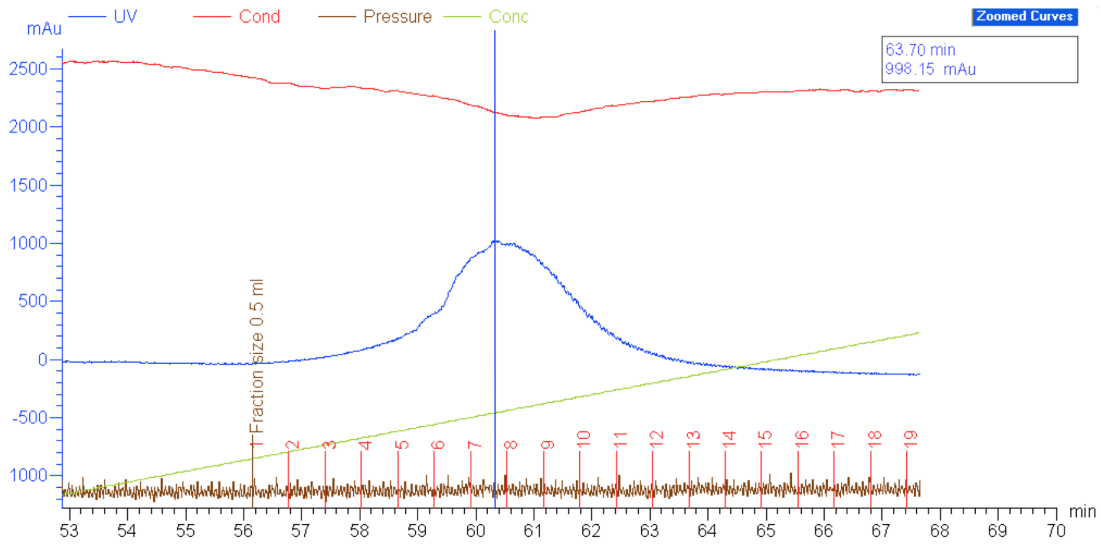


Figure 3.15: his6x-tagged ADH2 expressed in H3094 (Δ*tnaA*), maximal mAU: 998.15.

Table 3.4: Results from BCA and Activity Assays of his6x-tagged ADH2 from strains H3094 ($\Delta tnaA$) and H1332 ($tnaA+$) following expression according to optimised protocol

Strain	**H3094**	**H1332**
Protein concentration	1.31 mg/ml	0.85 mg/ml
Rel. Spec. Activity	5.3 U/mg	4.6 U/mg

3.4 Discussion

It has previously been demonstrated that halophilic proteins bind less SDS than their non-halophilic counterparts, resulting in reduced mobility and overestimation of molecular weights [Monstadt and Holldorf, 1991]. This lack of SDS binding arises through repulsion of the unusually high amount of negatively charged amino acids occupying on average 25 % of the total MW of halophilic proteins, but, which in turn are suggested to aid migration. However, SDS denatures the protein and interacts to some degree through its hydrophobic hydrocarbonated portion with hydrophobic regions of the protein, contributing to a consistent protein shape. Thus, migration through the gel is on size only, and is not influenced by innate secondary, tertiary conformations, or side chain charges. Confounding characteristics of the protein are eliminated by SDS, hence, size is the prevailing migration factor. But this is not the case if SDS is repelled or not used at all. Slightly higher migration was observed for ADH2 (figure 3.6), but was not observed for either of the transaminases (figure 3.4 and 3.5). Regarding the native-page, the influence of SDS is lacking and the influence of charge and sterical conformation of ADH2 monomers, dimers and tetramers are clearly observable. All conformational states of ADH2 run about 5 - 10 kDa lower than expected and this congruent with the argument, that the negative charges contribute to a faster migration.

The obtained results from evaluating chromatographic elution peaks through activity assays and native-page demonstrate, that ADH2 is highly active in its tetrameric form.

Formation of homotetramers proceeds through two different pathways, either through monomer-dimer-tetramer (MDT) or monomer-dimer-trimer-tetramer (MDTT) formation [Jaenicke and Lilie, 2000]. The latter is generally less favoured and associated with disease, since the protein can be trapped in intermediate oligomeric states and improper oligomers could be harmful to an organism [Selkoe, 1998]. A faster rate of association for MDT is promoted, favouring this pathway, based on the optimisation of electrostatic interaction and desolvation forces of only one of the interfaces, not both. This promotes a fast association for dimerisation, but will presumably lead to a slower association for tetramerisation. It was noted that a steady state can be achieved once a majority of homodimers have been formed, trapping a substantial amount of protein in the homodimer and homotrimer state and tetramerisation does not reach completion in large portions of evaluated simulation space [Powers and Powers, 2003]. This could be an explanation for the observed predominant presence of dimers of ADH2 when the short induction protocol was used and the predominant formation of tetrameric and highly active protein upon using the prolonged induction protocol. A longer induction period for overexpression of protein leads to an overall higher protein content in the cell and probably moves the steady state towards tetramerisation. Protein turnover, the balance between synthesis and degradation, is one of many forms of regulation that are coordinated to achieve a unified cellular response [Hinkson and Elias, 2011]. It is possible, that the native expression of alcohol dehydrogenases, i.e. without overexpression, is an additional way to control ADH activity through monomer-dimer-tetramer abundance.

Recovery of enzyme activity from imidazolium elutions was very low in comparison to EDTA elutions. This is in alliance with the observations from literature for other alcohol dehydrogenase [Dabirmanesh et al., 2012]. Our collaborators from the Paradisi group were using a protocol that involves imidazolium elution and were able to recover sufficient activity after buffer exchange for the very same enzyme. Differences in the purification protocol itself might be the reason, i.e. faster elution or buffer exchange, could lead to a timely recovery. The advantage of specific binding of StrepTrap purification is significant as can be seen on the SDS-Page gels after desthiobiotin elution, where no co-purification at all is observed. However, activity was very poor for both constructs, regardless of the presence of the tag at the N- or C-terminal end.

Tryptophan-inducible promoters of other species are known to be strongly induced once stationary phase is reached [Kajitani and Ishihama, 1983]. Stationary phase in *H.volcanii* is on average reached after 12-24 h and colonies grown on plates for blue-white screening most likely exceeded this time frame. However, the results from the ONPG-assay suggest no difference between *tnaA* deleted and non-deleted strains. The strains, which are presumably *tnaA* deleted, were constructed as mentioned in *Chapter2* with a plasmid containing a truncated version of *tnaA* (pTA1730). The same plasmid and plasmids pTA1756 and pTA1762 were used to construct tryptophanase A deleted strains, of which however, only the strain transformed with plasmid pT1756 was confirmed for deletion. The two other constructs failed to produce a $\Delta tnaA$ clone. The assay itself was apparently working, since the negative control strain did not lead to signal detection.
Comparison of elution profiles is not a sufficiently valid method to make a meaningful statement. Despite the similarity of strains and cultivation and expression and purification conditions, small differences in growth rate can already lead to a skewed result and altered protein concentrations even for the very same strain. However, the difference between elution profiles of *tnaA* deleted and non-deleted strain is almost double in detected mAU and might hint at a beneficial effect for overexpression. Unfortunately, no software for peak integration was available. Protein concentration after dialysis, which poses again a risk for false interpretation, was however in congruence with elution profiles.

3.5 Conclusion and Future Perspectives

An optimised induction/expression protocol for the tetrameric protein ADH2 was established and highly active protein expressed in H3094 was used for activity assays in ionic liquids (*Chapter4*). Expression according to the optimised protocol was also performed for strain H1332 and shows that the protocol is not specific for H3094. Although both strains express ADH2, it is likely that the optimised protocol would also be applicable for other tetrameric proteins following the MDT pathway for quaternary structure formation, for obtaining fully formed protein.

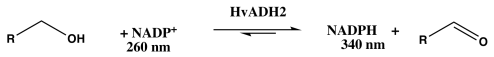
No definite conclusion can be drawn as to whether the deletion of tryptophanase A enhances protein expression or not. ONPG-assays were not able to confirm a benefit. Future work should include the verification of *tnaA* deletion of strains used in the ONPG-assay. However, it might be that only a prolonged induction period yields an observable difference in expression levels and hence, if cultivation and expression in flow reactors would be realised in the future, it would be advisable to revise the assessment of *tnaA* deleted strains.

Chapter 4

Assays of ADH2 in Ionic Liquids

4.1 Background

Following the successful establishment of a purification procedure, the enzyme *Hv*ADH2 was assayed in thirteen different ionic liquids with the aim of assessing enzymatic activity. Spectrophotometric assays were carried out to assess the relative specific activity of *Hv*ADH2 in aqueous ionic liquid mixtures of various concentrations, and activities were compared against a reference buffer system for the following reaction:



The reference buffer system consisted of a 4M KCl 50mM Glycine-KOH buffer pH10, since the enzyme shows highest catalytic activity for the oxidative reaction under haloalkaliphilic conditions and the tetrameric structure is optimally maintained at 4M KCl [Timpson et al., 2013]. Glycine was further shown to generally improve protein stability [Platts and Falconer, 2015] and maintains pH10 during freezing-thawing cycles [Pikal-Cleland et al., 2002].

The aim of the study detailed in this chapter was to provide experimentally obtained values on the change of *Hv*ADH2 activity in ionic liquids that can be correlated to investigations of *in-silico* methods used to characterise the structural changes of *Hv*ADH2 in

ionic liquids in the following chapter (*Chapter 5*), and to identify a system, that retains ADH2 activity best for the subsequent encapsulation (*Chapter 6*).

Ionic liquids and their classification in this thesis (figure 4.1) were chosen in concordance to literature, reviewed in *Chapter 1.5.1*, so that the results can be discussed in context, despite the lack of a uniform boundary of knowledge concerning ionic liquid impact on proteins in general or more specifically, their impact on alcohol dehydrogenases or halophilic proteins. As clarified in *Chapter 1.5.1*, ionic liquid ions and their impact on ADHs can be divided into three classes (hydrophilic, water-immiscible and amphiphilic [Schindl et al., 2019]), whereby these can be subdivided according to either the presence of functional groups, such as hydroxy or ether groups, or by a central atom, that either carries the charge (mostly in the case of cations) or is central to the charge (mostly anions).
In this sense, ionic liquid ions used in this study, (see figure 4.2), were chosen in order to cover a broad range of classes defined.

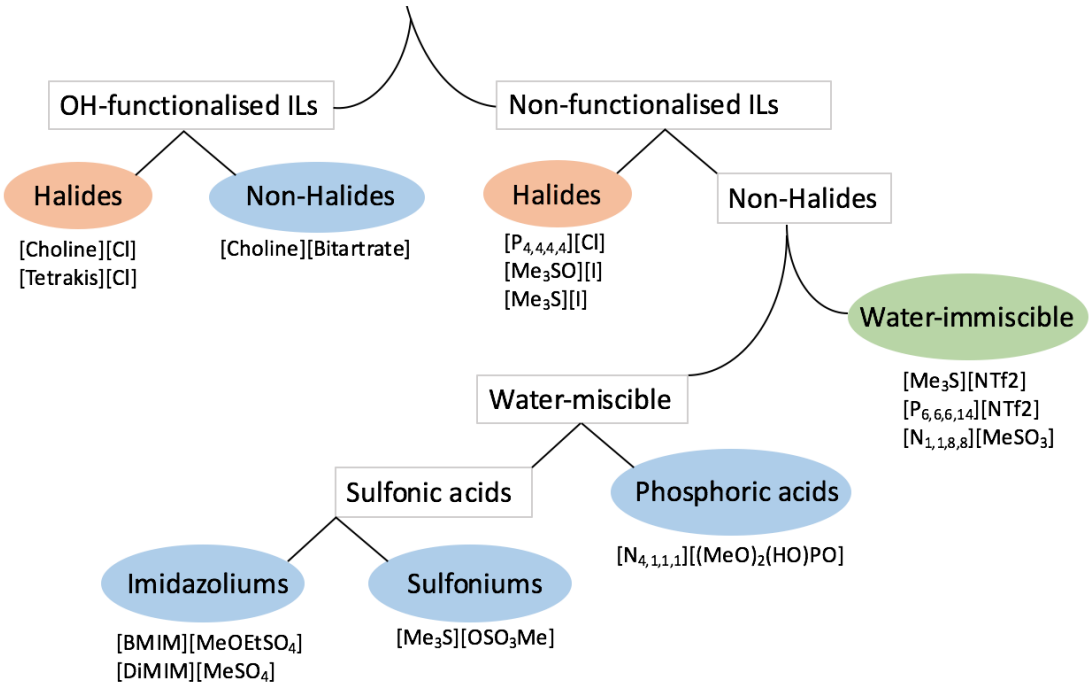


Figure 4.1: Targeted classification of ILs in this study: OH-functionalised vs non-functionalised, halides vs non-halides and water-miscible vs water-immiscible.

Anion groups chosen for this study covered halides (chloride, iodide), bistriflimides (NTf_2),

sulfonic acids (Me_3SO_3, OSO_3Me and $MeOEtSO_4$), phosphatic acids ($(MeO)_2OPO$) and carbonic acids (Bitartrate). Cation groups covered phosphonium ($P_{6,6,6,14}$, $P_{4,4,4,4}$ and Tetrakis), ammonium ($N_{1,1,8,8}$, $N_{4,4,4,4}$, Choline), sulfonium (Me_3S and Me_3SO) and imidazolium (DiMIM and BMIM).

The following combinations were chosen and classified according to the functional groups or central ions, as summarised in the 'classification-tree' in figure 4.2.

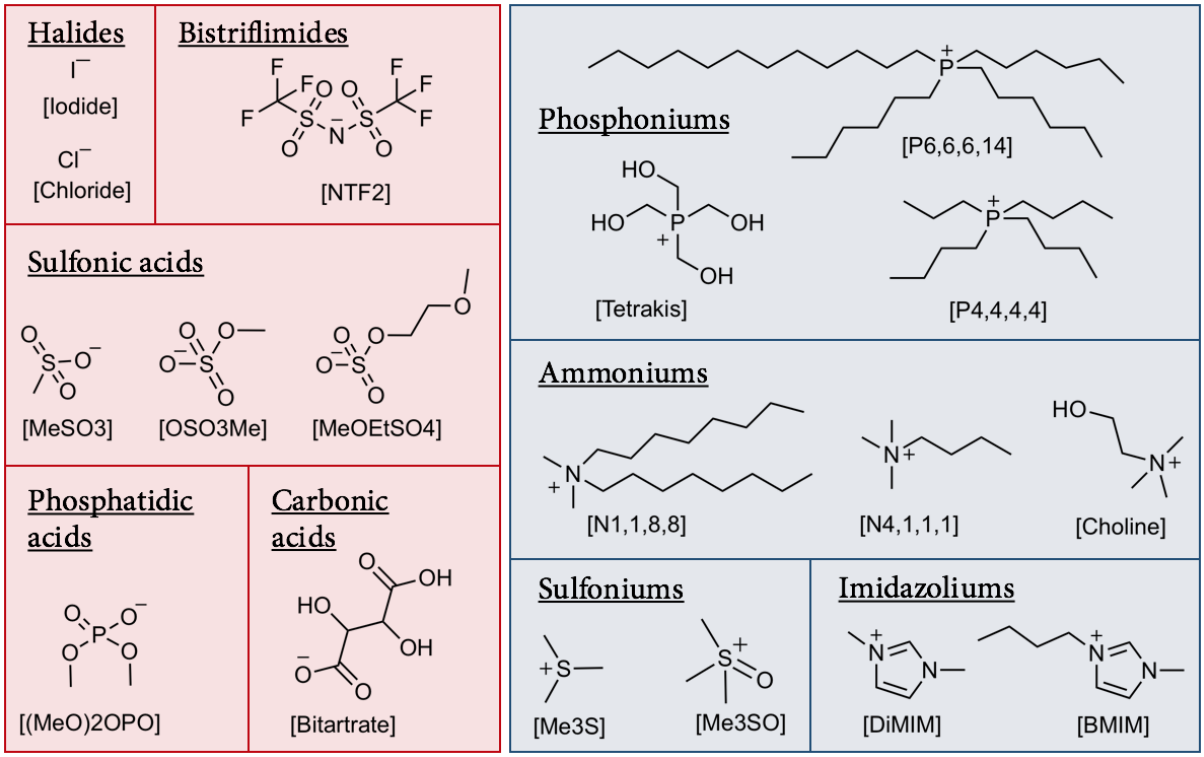


Figure 4.2: Anions (red) and Cations (blue) used in this study.

'Hydroxy-functionalised' was chosen as a classification, since the OH-group is present in many amphiphilic ionic liquids and represents a distinct molecular property. This distinct molecular property is capable of mimicking a protic environment similar to water by forming OH-bridges, and such an environment is prefered by enzymes. The beneficial effect of short-chain hydroxy-functionalised IL $[MTEOA][MeSO_4]$ on *Rhodococcus ruber* ADH-A for the asymmetric reduction of ketones was demonstrated previously [de Gonzalo et al., 2007]. 'Halides' as classification was chosen, since these ions have been found to dissolve cellulose very effectively, the substrate to biofuel production in which the final step is catalysed by ADH [Lo et al., 2015]. However, these types of ILs tend to denature

enzymes, hampering biocatalytic processing of the dissolved cellulose [Toral et al., 2007], whilst halophilic cellulases have shown increased activity in halide-ILs such as [EMIM][Cl], [BMIM][Cl] or [AMIM][Cl] [Zhang et al., 2011]. And finally, hydrophilic against water-immiscible ILs were chosen to be distinguished, since biphasic IL-systems were one of the first explored successfully for ADHs [Eckstein et al., 2004]. Hydrophilic ILs can be further subdivided according to their central ions, phosphor, sulphur or nitrogen, which exhibit different electronic structures and give rise to complex, delocalised systems [Santos et al., 2018], impacting enzyme activity and structure in multifaceted ways.

It is therefore anticipated, that by maintaining consistency of specific ions and counter ions or at least similarity of these, different angles on the classifications can be evaluated by matching the nature of cations and anions and their combinations to enzyme activity in context to enzyme structure.

For instance, a range of sulphur based ions were coupled with either hydrophilic or water-immiscible counterions and can be analysed in the counterions context. By this, it is hoped that patterns may be inferred for protein stabilising or destabilising properties. More specifically for example, **ILA** and **ILE**, (see *Register for of Ionic Liquids used for this thesis*), share the same anion ($[OSO_3Me]^-$), which is paired with another sulfonium based cation (**ILA**) or an imidazolium ion (**ILE**), respectively. Comparing activities of ADH2 in **ILA** and **ILE** in this chapter and structural diversions of ADH2 in the next chapter might allow for conclusions about the influence of the not-commonly-shared cations (sulfonium or imidazolium) on $[OSO_3Me]^-$ and their combinatory influence on the protein.

4.2 Material and Methods

4.2.1 Material

Ionic Liquids

The salts were either purchased from distributors (see table 4.1) or were taken from in-house synthesised IL-stocks (*Engineering, Sustainable Process Technologies (SPT) group*).

Table 4.1: Ionic Liquids used in this Study

Name	**abridged**	**Supplier**
Trimethylsulfonium Methylsulfate	IL A $[Me_3S][OSO_3Me]$	ACROS
Tetrakis(hydroxymethyl)phosphonium Chloride	IL B [Tetrakis][Cl]	ACROS
Choline Chloride	IL C [Choline][Cl]	ACROS
Choline Bitartrate	IL D [Choline][Bitartrate]	ACROS
1,3-Dimethylimidazolium Methylsulfate	IL E $[DiMIM][MeSO_4]$	Fluka
Triethylsulfonium Bis(trifluoromethylsulfonyl)imide	IL F $[Me_3S][NTf_2]$	Solvent Innovation
1-butyl-3-methylimidazolium (2-methoxyethyl)sulfate	IL G $[BMIM][MeOEtSO_4]$	SPT
N,N-dimethyl-N-octyloctan-1-aminium Methane sulfonate	IL H $[N_{1,1,8,8}][MeSO_3]$	SPT
Tetrabutylphosphonium Chloride	IL I $[P_{4,4,4,4}][Cl]$	Quill
N,N,N-trimethyl-1-Butanaminium Dimethylphosphate	IL J $[N_{4,1,1,1}][(MeO)_2(HO)PO]$	SPT
Trihexyltetradecylphosphonium Bis(trifluoromethylsulfonyl)imide	IL K $[P_{6,6,6,14}][NTf_2]$	Fluka
Trimethylsulfoxonium Iodide	IL L $[Me_3SO][I]$	Sigma Aldrich
Trimethylsulfonium Iodide	IL M $[Me_3S][I]$	Sigma Aldrich

Preparation of Ionic Liquid Mixtures

A unified overview of IL-mixture preparations is given in figure 4.3. A more detailed explanation follows below.

IL/Conc.	0.025 M	0.15 M	0.75 M
A	dissolved	pH change	pH change
B	pH change	pH change	pH change
C	dissolved	dissolved	2 M KCl
D	0.15 M KCl and	pH change	not soluble
E	dissolved	dissolved	dissolved
F	emulsion	emulsion	emulsion
G	dissolved	pH change	not soluble
H	emulsion and	pH change	not soluble
I	dissolved	not soluble	not soluble
J	dissolved	dissolved	1 M KCl
K	emulsion	emulsion	emulsion
L	dissolved	dissolved	not soluble
M	dissolved	dissolved	dissolved

Figure 4.3: Overview on IL-mixture preparation. dissolved - IL dissolved readily in the standard 4 M KCl buffer, pH change - pH changed upon addition of IL, but was adjusted accordingly with KOH. emulsion - an emulsion was formed and the mixture was used as such for assays. M KCl - a lower salt buffer was used. not soluble - this mixture could not be obtained and measured.

ILA dissolved readily in 4 M KCl buffer at all concentrations. At molarities of 0.15 M, 0.3 M 0.75 M, a pH change was observed to 4.5. pH was adjusted accordingly. ILB dissolved readily, but pH changed even at the lowest concentration (0.025 M) to 5 and to 4.5 at higher concentrations. pH was adjusted accordingly. ILC dissolved readily at concentrations 0.025 M and 0.15 M. ILC added at a concentration of 0.75 M crashed out and was therefore dissolved in glycine-KOH buffer containing only 2 M KCl. No pH change was observed. ILD crashed out at all three concentrations, when dissolved in 4 M KCl buffer and only concentrations 0.025 M and 0.15 M of IL could be dissolved in 0.15 M KCl buffer. Hence, only two concentrations were measured in the following experiments for ILD. pH change was observed to 5 and was adjusted with KOH. ILE dissolved readily

at all concentrations and no change in pH was observed. ILF formed an emulsion at all concentrations. No change in pH was observed. ILG dissolved at a concentration of 0.025 M, where no change in pH was observed. At a concentration of 0.15 M, pH changed to 3 and was adjusted. A mixture with 0.75 M ILG could not be produced, as it was not found to be soluble in any buffer. ILH formed an emulsion and at 0.75 M crashed out in any buffer. pH changed to 3 and was adjusted. ILI dissolved readily at 0.025 M, but crashed out at higher concentrations, when pH was adjusted from 4 to 10. Since pH was meant to be consistent for all measurements, only one concentration (0.025 M) was measured for ILI. ILJ dissolved readily in 4 M KCl buffer for low concentrations. 1 M KCl buffer was used for dissolution of 0.75 M ILJ. pH was not affected. ILK formed an emulsion and was used as such. No change in pH was observed. ILL dissolved at a concentration of 0.025 M readily in 4 M KCl buffer. 0.15 M ILL dissolved in 4 M KCl upon heating to 50 ℃. 0.75 ILL could not be dissolved in any buffer. No change in pH was observed. ILM dissolved readily at all concentrations, whereby 0.75 M had to be heated to 50 ℃ for dissolution. No change in pH occurred.

ADH2 Storage and Measurement Buffers and Solutions

100 mM $NADP^+$ prepared in SDW
2 M ethanol

Glycine-KOH Buffer
50 mM glycine-KOH, pH 10 or pH 8, 4 M KCl, 3 M KCl or 2 M KCl respectively

Glycine-KOH Buffer + $ZnCl_2$
50 mM glycine-KOH, pH 10 or pH 8, 3 M KCl, 20 μM $ZnCl_2$

50 mM Glycine-KOH
3.75 g glycine in 0.995 L, adding 5.2 ml 5 M KOH

4.2.2 Methods

Preparation of Aqueous Ionic Liquid Mixtures

Ionic liquids for assays were prepared as mixtures with 4 M glycine-KOH buffer, pH 10, in concentrations ranging from 0.025 M to 4 M. Ionic Liquids were weighed in Eppendorf tubes (m[g] = M*V[l]*MW), to which three quarters of the total volume of buffer was added, before the pH was checked. If pH differed from 10, it was adjusted with μl volumes of 5 M KOH solution and buffer was added to the final volume. If ionic liquids at higher concentrations were found to be insoluble in the 4 M KCl buffer, glycine-KOH buffers with lower potassium chloride concentrations (0.15 M, 1M, 2M or 3 M) were used. Findings on dissolutions and pH adjustments are detailed in the *Results section* of this chapter.

Spectrophotometric Measurement and Relative Specific Activity Calculation

An Epoch 2 Microplate Spectrophotometer (BioTek) was used with 96-well UV transparent plates to monitor the increase in absorption (slope) at 340 nm of NADPH. The extinction coefficient and pathlength specific for this system for NADPH is 3.3 and was supplied by BioTek. The relative specific activity was calculated with the following equation:

$$U/mg = \frac{AverageSlope}{3.3} * \frac{1}{V_a} * \frac{1}{c_a} * V_f \tag{4.1}$$

where V_a is the volume of enzyme added to the reaction, c_a is the concentration of enzyme added and V_f is the total volume of the reaction. The average slope was calculated from the obtained measurement values over time:

$$AverageSlope = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \tag{4.2}$$

Data is plotted using Excel (Microsoft) using the following formulas for standard deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_{i1}-x_{i2})^2}{n-1}} \tag{4.3}$$

for triplicate results and

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_{i1}-x_{i2})^2}{2n}} \tag{4.4}$$

for pairs of duplicate results (x_{i1}, x_{i2}), where n is the number of differences. Since the population mean in the latter case is exactly known, the sum of squares is divided by the

number of the differences n (not by *n-1*) and because the obtained variance (squared SD) of the differences is twice that of the individual results, the sum must also be divided by 2 [Doerffel and Eckschlager, 1981, Inczedy et al., 1998, Synek, 2008].
Blanks for individual IL-mixture concentrations were measured on the EPOCH (Spectrophotometer) alongside samples and were substracted prior to calculations. Assays were performed at 50 ℃.

Material for YADH assay

Freeze dryed Yeast alcohol dehydrogenase, 15 KU (Sigma)

10 mM Phosphate buffer, pH 7.6

100 mM NAD^+ prepared in SDW

4.3 Results

4.3.1 Screening Activity of ADH2 in Ionic Liquid Mixtures

The relative specific activity of *Hv*ADH2 was used to characterise the enzyme in terms of performance under optimised standard conditions (glycine-KOH buffer, pH 10, 4M KCl) for comparison to ionic liquid mixtures. The initial assays screened all ionic liquids (IL A-M) to obtain trends, see figure 4.4. Based on these trends, 3 ILs (**A**, **K** and **M**) were selected for investigation at additional concentration ranges (figure 4.7). In the following charts (figure 4.4 and 4.7), the standard glycine-KOH Buffer, 4M KCl, pH 10 served as a reference and is shown in blue. Concentrations of IL-buffer mixtures are shown in orange, whereby decreasing colour intensity reflects increasing concentration (legends given). The first boxplot shows a summary of ADH2 relative specific activities in different IL mixtures. Charts showing measured absorption curves following the increase in NADPH concentration were moved to the appendix. Initial screening assays were

carried out in duplicates and two separate sets of experiments. Assays covering additional concentration ranges were carried out in triplicates of two separate experiments.

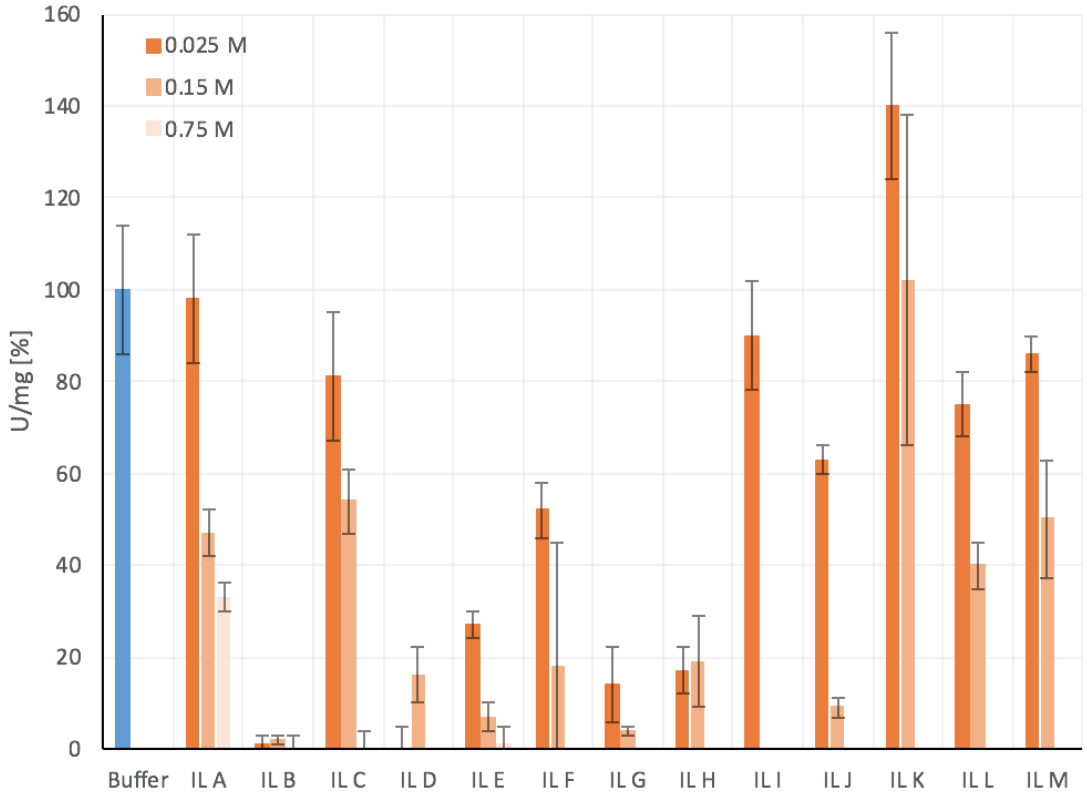


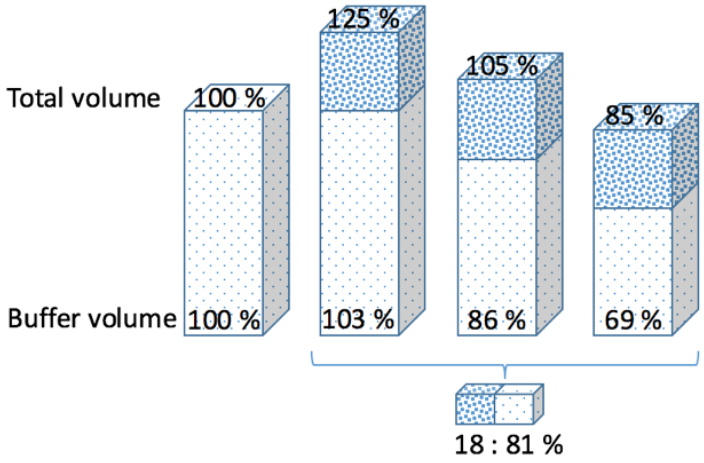
Figure 4.4: Comparative overview of ADH2 activity in buffer and IL mixtures. Relative specific activity [U/mg] was calculated from slopes of measured NADPH increase and is given in percentage relative to the reference buffer system (100 %), which serves as the standard for comparison. Experiments were triplicates are reflected in the error bars.

4.3.2 Detailed Investigation of IL K

Investigation of Solubility of NADPH

The activity of ADH2 in **ILK** mixtures was extraordinarily high, compared to the reference buffer. In order to rule out the possibility that this higher activity can be ascribed to an exclusion of NADPH (or the enzyme) from emulsion patches of **ILK** ions, leading to a lower 'active' volume and hence an apparent higher concentration, which would be detected in a higher signal, an experiment was conducted, in which the stochiometrics of the enzyme are constant, while the total volumes and the 'active' volumes are varied and

NADPH increase is measured. The following set-up was created:



Experimental expectation								
NADPH	upon **exclusion** from IL volume				upon **inclusion** from IL volume			
	Buffer	**IL-mixtures**			**Buffer**	**IL-mixtures**		
U/mg expected	Reference	103%	86%	69%	Reference	**125%**	**105%**	**85%**

RESULTS					
Solvent		**Buffer**	**IL K mixture [0.23 M]**		
Total Volume		200 µl	250 µl	210 µl	170 µl
Buffer Volume		200 µl	203 µl	171 µl	138 µl
U/mg measured		**0.02 % Enzyme**			
Detected slope		0.02	0.03	0.03	0.03
	U/mg	**1.1** (100 %)	**2.06** (187 %)	**1.73** (157 %)	**1.43** (130 %)
Relative slope		0.03	0.03	0.03	0.03
	U/mg	1.65 **(100 %)**	2.06 **(124 %)**	1.73 **(104 %)**	1.43 **(86 %)**

Figure 4.5: Experimental set-up and expectation for the investigation of solubility of NADPH in ILK.

Samples containing **ILK** and buffer (4 M KCl, glycine-KOH, pH 10) were prepared in one pot, before being split, to ensure exact emulsion ratios. Enzyme and substrate (ethanol) were added according to the final volumes to maintain the same end concentration. The reaction was started upon addition of $NADP^+$.

If NADPH is excluded from emulsion patches of **ILK**, the 'active' volume, i.e. the volume in which NADPH can be detected, should be equal for the sample with 25 % more total volume and the buffer. Apparent enzyme activity should be 25 % higher in the direct comparison of measurements, since there ought to be 25 % more enzyme in the same amount of 'active' volume, since the emulsion is presumed to be 'excluded' from

the measurement. On the other hand, if NADPH is not excluded from the emulsion a much higher activity increase is expected. The latter was observed, in addition a greater slope was observed for **ILK** samples. Calculation of a proportionate slope reveals that exactly 25 % more enzyme was present. To observe consistency of this set-up, two further samples with smaller total volumes, but the same emulsion-buffer ratio, were measured and demonstrated consistency.

Investigation of absorption curve

When assaying **ILK** it was noticed that for some measurements a specific 'kink' in the absorption is registered between minute 5 and 10 (figure 4.6). This kink was attributed to a change in phase behaviour, since **ILK** is immiscible with the buffer. To investigate whether its appearance can be circumvented by applying a different mode of shaking, the originally used orbital-mode was compared to double-orbital-mode.

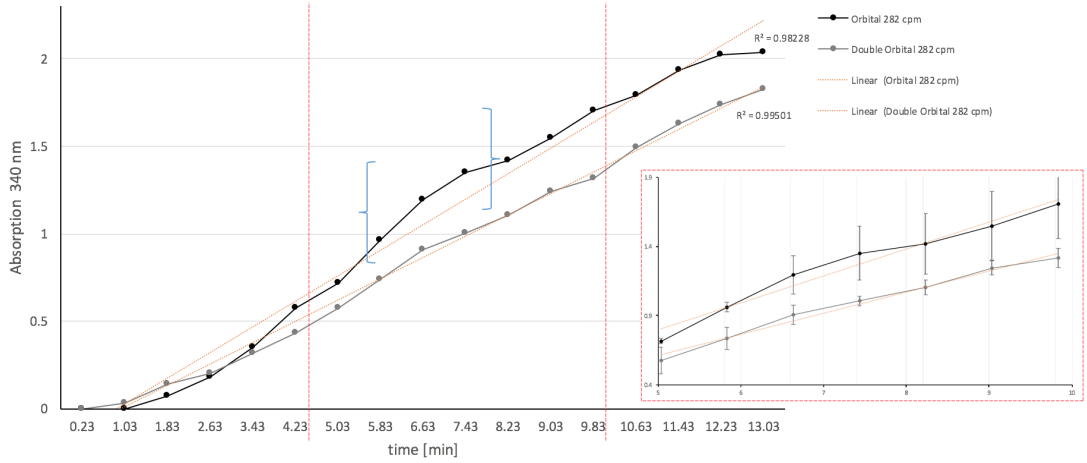


Figure 4.6: Investigation of a 'kink' (indicated by blue brackets) in absorption for *Hv*ADH2 activity in ILK.

4.3.3 Screening Activity of ADH2 in IL A, K and M

Ionic liquids **A**, **K** and **M**, which maintained ADH2 activity best in the initial screening assay, were selected for investigation in a broader range of IL concentrations (0.075 M, 0.3 M and 0.6 M) to the end of identifying a system of high activity. Overall, trends for concentrations measured in the initial assays (0.025 M and 0.15 M) were reproduced,

with the exception of 0.025 M for **ILM**, which was higher than in the preliminary assay (figure 4.7).

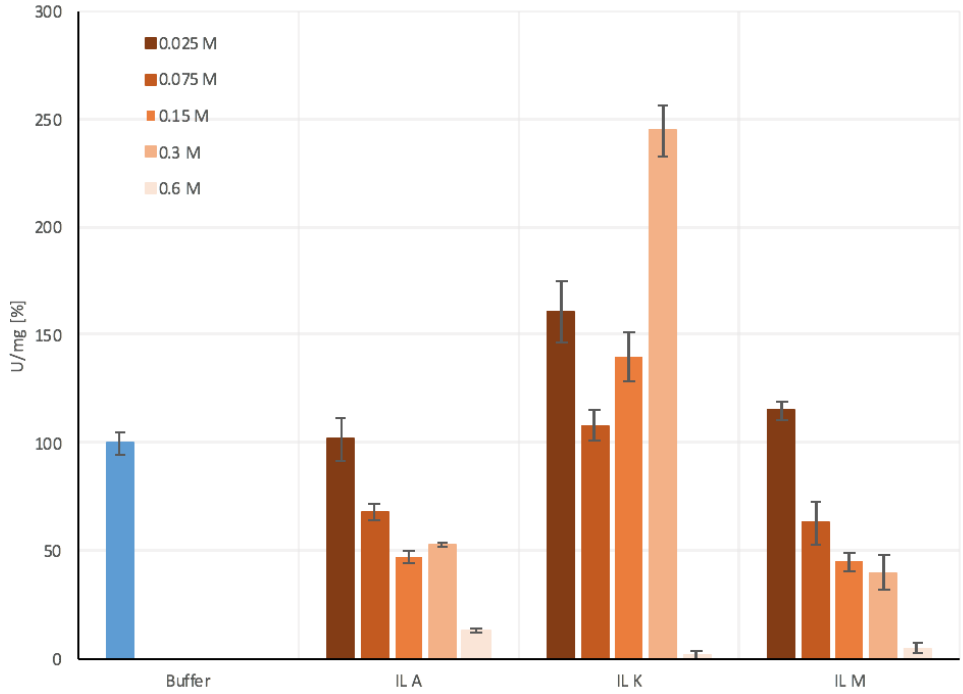


Figure 4.7: Comparative overview of ADH2 activity in Buffer and IL mixtures. Relative specific activity was calculated from slopes of measured NADPH increase and is given in percentage relative to the reference buffer system (100 %), which serves as the standard for comparison.

4.3.4 Long-term stability of *Hv*ADH2

The relative specific activity of HvADH2 was used to characterise the long-term stability of the enzyme under storage conditions (- 20 °C) as well as the residual activity of the enzyme after prolonged incubation at room temperature in ionic liquid mixtures and reference buffer conditions.

Long-term stability of ADH2 under storage conditions

Table 4.2 shows different batches of purification of HvADH2 and indicates the relative specific activity measured after storage at - 20 °C from one to up to 30 days.

Batch	Days at - 20 °C	[U/mg]
17-03-06	1	2.65
	24	1.78
	28	1.28
17-03-21	1	2.24
	9	2.17
	15	2.11
17-05-04	1	6.35
	15	5.27
	30	5.08

Table 4.2: Overview on relative specific activity of HvADH2 and its long term stability.

Long-term stability of ADH2 in IL A, K and M at RT

In order to assess long-term stability of *Hv*ADH2 in **ILA**, **ILK** and **ILM** (0.15 M), activity was measured after up to 4.5 hrs after incubation of the enzyme in the ionic liquid at room temperature.

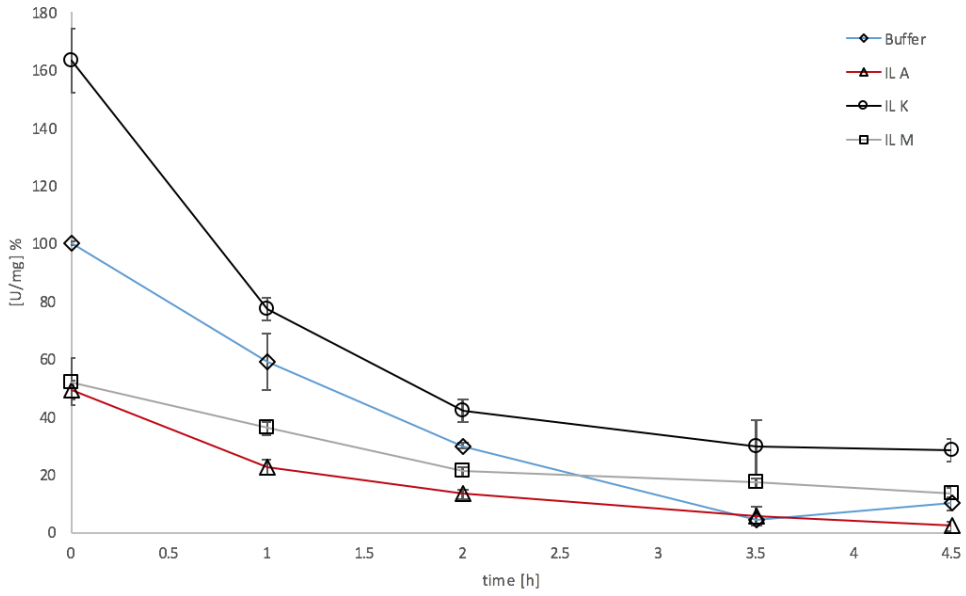


Figure 4.8: Long-term stability of ADH2 in IL A, K and M

A slight recovery of enzyme activity past 3.5 hrs is observed for the buffer system,

while all ionic liquids lead to a decrease in activity over time (figure 4.8). However, the remaining activity for **ILK** is around 40 % and 20 % for **ILM**.

Influence of Zinc and Molarity on HvADH2

Assays containing μmolar concentrations of Zn^{2+} were carried out to verify that Zinc stays bound and is not partially removed from the enzyme during elution with EDTA. Zinc ions are supplied in form of $ZnCl_2$. In the standard glycine-KOH buffer with 4 M KCl addition of 20 μM Zn^{2+} led to a decrease in activity with only 3.6 U/mg compared to 6.4 U/mg without Zn^{2+}. Conversely, for experiments conducted in glycine-KOH buffer with 2 M KCl, a gain in activity is noted with 1.5 U/mg compared to 0.8 U/mg, upon addition of Zn^{2+}. Supplying additional Zn^{2+} is therefore not necessary under standard conditions and opposing to this has a detrimental effects on the activity.

4.3.5 Yeast alcohol dehydrogenase in IL K

Yeast alcohol dehydrogenase (YADH) was assayed in ionic liquid $[P_{6,6,6,14}][NTf_2]$ (**ILK**) to assess whether **ILK** activity for the halophilic protein was due to specific molecular interactions, or whether biphasicity was the major reason for improved activity. Measurements were carried out in 0.25 M **ILK** for both enzymes. Reference buffer for YADH was a physiological phosphate buffer.

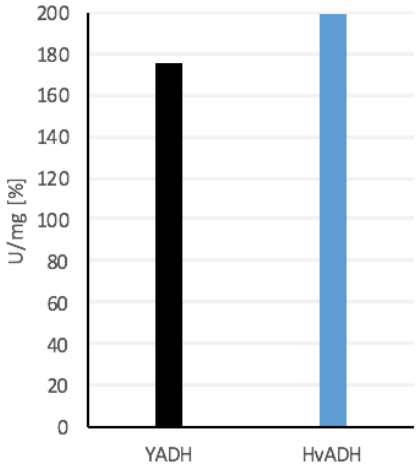


Figure 4.9: Comparison of relative specific activity of YADH and *Hv*ADH2 in 0.25 M IL K mixture.

4.4 Discussion

A summary of results of ADH2 activity at 0.025 M IL is shown in figure 4.10, whereby anion-cation combinations were defined based on the central ion and/or functionalised property. At this concentration six out of thirteen ILs (45 %) were able to maintain enzymatic activity above 80 % and eight out of thirteen were able to maintain activity above 50 %.

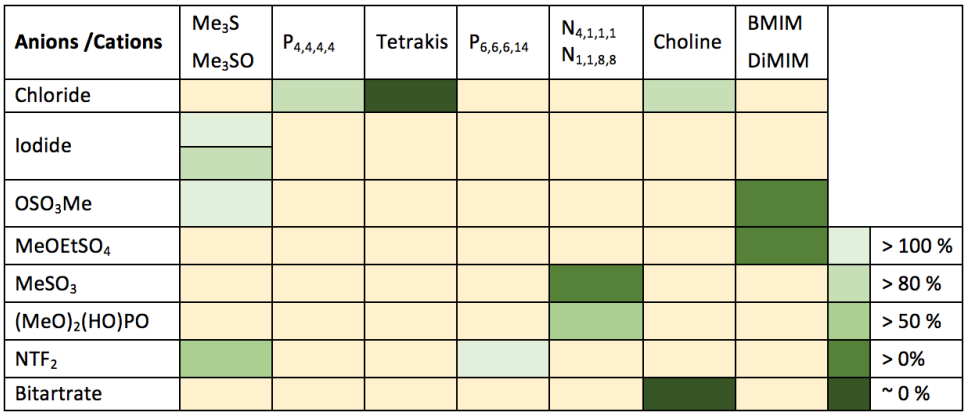


Figure 4.10: Cation-anion correlated effects on enzyme activity at IL concentrations of 0.025 M. Percentages indicate enzyme activity.

Sulfonium-based cations [Me_3S] and [Me_3SO] demonstrate above average activity, while sulfur based acidic anions [$MeSO_3$], [OSO_3Me] and [$MeOEtSO_4$] are slightly below average, except in combination with sulfonium-based cations. Literature reports no sulfonium-based ILs in conjunction with alcohol dehydrogenases or any other type of enzyme. Sulfur based acidic anions demonstrate great variability of impact on enzymatic ADH activity. For instance, while *R.ruber* ADH-A activity was diminished in 20 % v/v [EMIM][$MeSO_3$] [de Gonzalo et al., 2007], *L.brevis* ADH activity was improved to 105 % in 10 % v/v of the same IL-system [Kohlmann et al., 2011].

Phosphorus-based ions in this study showed a mixed picture for ADH2 activity. While

alkyl-tail phosphonium-based ILs showed improved enzyme activity, hydroxy-functionalised phosphonium ions were strikingly detrimental. There are no examples in the literature for ADH activity in phosphonium-based ILs. Similarly converse effects on activity have been observed for phosphatic acid based ILs for *L.brevis* ADH activity [Kohlmann et al., 2011] and *Rhodococcus erythropolis* ADH activity [Hussain et al., 2008]. Most striking example is the activity difference of *L.brevis* ADH in 10 % v/v [BMIM][Et_2PO_4] with 5 % activity compared to in 10 % v/v [BMIM][Me_2PO_4] with 80 % activity.

In general hydroxy-functionalised ILs diminished activity rapidly, especially [Tetrakis][Cl] and [Choline][Bitartrate]. Hydroxy-functionalised ILs for mesophilic ADH-systems reviewed in literature demonstrated preservation or improvement of activity [de Gonzalo et al., 2007, Hussain et al., 2008, Kohlmann et al., 2011]. Many of these ILs have comparable long alkyl tails and are amphiphilic, whereas the hydroxy-functionalised ILs used in this study do not share these attributes. However, the best improvement of enzyme activity was reported for [MTEOA][$MeSO_4$] [de Gonzalo et al., 2007] which is similar to the [Tetrakis] ion, except for the central ion being a nitrogen and one less OH-group. It is most probable, that the presence of the hydroxy-groups on the cation or anion lead to a strong interaction with the carboxy-group of the acidic residues on protein surfaces. The competition for hydrogens in the solvation shell of the protein and the increased ability of OH-groups to donate hydrogen bonds directly to acidic residues will have a strong impact on halophilic proteins when compared to mesophilic proteins.

ADH2 showed below average activity in both imidazolium-based ILs (**ILE** and **ILG**) which both have sulfonic acid counterions. In comparison to literature, where imidazolium ions with halides as their counterions were able to enhance HlADH activity [Shi et al., 2006] or at least moderately maintain it for YADH at comparable concentrations [Dabirmanesh et al., 2011, 2012], the pairing with sulfonic acid based counter ions appears to be less stabilising. Both imidazolium ions ([EMIM] and [BMIM]) are paired with a counter ion of similar tail length in this study and unlike reports in literature, the shorter alkyl chain imidazolium [EMIM] maintains activity better by around 10 % compared to the slightly longer alkyl tail of [BMIM]. Possible reasons for impaired activity are the

similar tail lengths of the counter ions, since the ions might be more homogenously coordinated to each other, when compared to a pairing with a halide as reported in literature where the chloride is found above or beneath the ring as a competitor to stacking interactions, able to jump from one side of the imidazole ring to the other [Das et al., 2018, Shukla et al., 2010]. The arrangements achieved with halides would allow more space for water molecules coordinating to amino acid surface residues, while a more compact ordering of anions and cations achieved through ring stacking interactions would exclude such. As such, all halide containing ILs (4/5), with exception of the one paired with a hydroxy-functionalised cation, where amongst the top 45 %.
Sulphur based ions paired with a hydrophobic (water-immiscible) counter ion, in both instances as cations and anions, maintain activity moderately or poorly for *Hv*ADH2. No such system has been reported in literature for alcohol dehydrogenases, however, the low activity hints at an interaction between the ions and the protein surface and not the establishment of a purely biphasic system, as reported for systems such as [BMIM][NTF_2] [Eckstein et al., 2004]. In comparison, the water immiscible **ILK** [$P_{6,6,6,14}$][NTf_2], where both ions exhibit hydrophobic character established a biphasic system, reflected in the preservation of the enzymatic activity.

The hydrophobic ionic liquid K [$P_{6,6,6,14}$][NTf_2] enhanced enzymatic activity at low to moderate concentrations considerably. The highest activity achieved was twice the activity observed in buffer. The possibility of attributing the high activity to an artefact of measurement based on the immobilisation of the cofactor in the aqueous phase leading to an apparent higher activity, was investigated by an experimental set-up, in which the enzyme concentration was related to the total volume for $NADP^+$/NADPH retention. Conversly, enzymatic activity for this system does not follow a linear trend with increasing IL concentration, since a higher activity was observed for both concentrations of 0.025 M and 0.3 M, in comparison to 0.075 M and also 0.15 M. It is likely, that the enzyme resides mainly in the aqueous buffer phase, but accumulates at the biphasic systems interface through hydrophobic interaction or is exposed to long range molecular interactions through structuring of water, which is modulated by the ionic liquid and depending on the concentration [Bruce et al., 2017, Coles et al., 2017]. Both factors might affect the

stability (maintance of rigid core structure) or flexibility (fluctuations from rigid core structure, to which it reverts back without loss of integrity) of the enzyme. Hence, these factors might explain the 'dip' in activity between low to moderate concentrations.

Characterisation of long-term stability of the enzyme showed promising results for its application. Activity in buffer showed a decrease, but started to increase again after 4 hours of incubation, which is in congruence with findings of collaborators from the Paradisi group (University of Nottingham, Chemistry Department).

4.5 Conclusion and Outlook

Ionic liquid $[P_{6,6,6,14}][NTf_2]$ preserved enzymatic activity best. Although this is most likely because of the formation of a biphasic system, it was demonstrated, that the ionic liquid has an influence on the enzyme, reflected in a non-linear concentration dependency. Investigation of solubility coefficients of the cofactors as well as the substrate and product could be done in the future by fine tuning the length of alkyl chains for providing a system, which provides the best yield. Further, more **ILK** concentrations could be explored. So far, a concentration of around 0.23 M was found to yield the highest ADH2 activity. Sulfonium based ionic liquids containing the ion $[Me_3S]^+$ and $[Me_3SO]^+$ showed very good results and investigation of other ionic liquids containing this class of ions could be done to obtain a more complete picture of synergistic molecular interactions of sulfonium-based compositions.

Chapter 5

Molecular Dynamics Simulations

5.1 Background

The focus of investigations in this chapter were short range electrostatic interactions between ionic liquid ions and acidic protein surface residues. In light of the hypothesis, that ion effects are mainly driven by short range rather than long range electrostatic interactions, atomistic molecular dynamics (MD) simulations, which treat solvent explicitly, have become increasingly important for investigating such effects, since these interactions take place at the nanoscopic level, which can still hardly be explored by experimental techniques [Greco et al., 2014]. Illuminating these ion specific interactions was expected to resolve trends for the distributions of the divers ion species in protein hydration shells and their consequential impact on protein structure. Protein stability and solubility depends on its interface with the solvent [Zangi, 2009] and will be different between mesophilic and halophilic proteins. In addition to particular interactions of single cations and anions with distinct amino acids at the protein surface, which will be more influential in the case of halophilic proteins, the ionic impact on protein solubility is mainly attributed to modifying the water structure and hence the protein hydration [Batchelor et al., 2004]. Albeit the influence of a central ion on surrounding water structures is restricted to its first hydration shell [Omta et al., 2003], the degree of hydration or hydrogen bonding may

have consequences for the solvent properties. This becomes already clear from the fact, that halophilic archaea accumulate potassium, rather than the not-so-different-sodium, given the multitude of ions there are, in their cytoplasm. Analyses in this chapter are expected to shed more light on the complexity of the systems at hand.

To correlate experimentally found activities of *Hv*ADH2 found in *Chapter4* to the presence of ionic liquid ions, a homology model of the tetramer was built and trajectories derived from molecular modelling were analysed for atomicscale interactions. A number of molecular dynamics analysis methods, such as Radial Distribution Functions (RDFs) [Meeron, 1958], Root Mean Square Deviations (RMSDs) [Maiorov and Crippen, 1994], Radius of gyration (Rgyr) [Lobanov et al., 2008] and Gibbs Free Energies [Marcus, 1991, Privalov and Makhatadze, 1993] were applied to trajectories and used to build Free energy landscapes (FELs) [Papaleo et al., 2009]. FELs represent a mapping of all conformations, that the protein underwent during simulation against their corresponding energy levels in form of the Gibbs free energy. During folding proteins seek the lowest energy state by exploring a complex landscape of peaks and valleys. In this studies case, it is not the free energy landscape of folding, that is explored, but the influence of molecular interactions of small charged molecules on the fold of the protein. During simulations (and experimental as well as native conditions) the protein is subjected to thermal motion and hydrophobic, hydrophilic and electrostatic interactions. In response to these the protein seeks a lower energy state by starting off at a local peak and gradually or rapidly sliding down into a basin. This basins can either be the native basin of attraction or competing ones, in which the protein-fold can become trapped. Over time, thermal motion and Van-der-Waals and electrostatic forces will batter the fold hard enough to cause the protein to overcome the energy barrier and rise to the top of the well, where upon it can slide down the next well and eventually, in most cases, end up in the native basin of attraction. The variables in the energy funnel equal vast numbers, since every atom in a protein requires six degrees of freedom ([x,y,z] for centering (translational movement) and [x',y',z'] for orientation (rotational movement)), albeit atoms in proteins are not independent in orientation, which cuts down on the degrees of freedom. However, it poses an enormous computational challenge.

From these free energy landscapes, a range of conformations corresponding to interesting

energy levels, can be 'pulled from' the trajectories, visualised and used for docking studies to investigate changes in the active center. Ionic liquids used in this chapter were the same as in *Chapter4*, see figure 5.1.

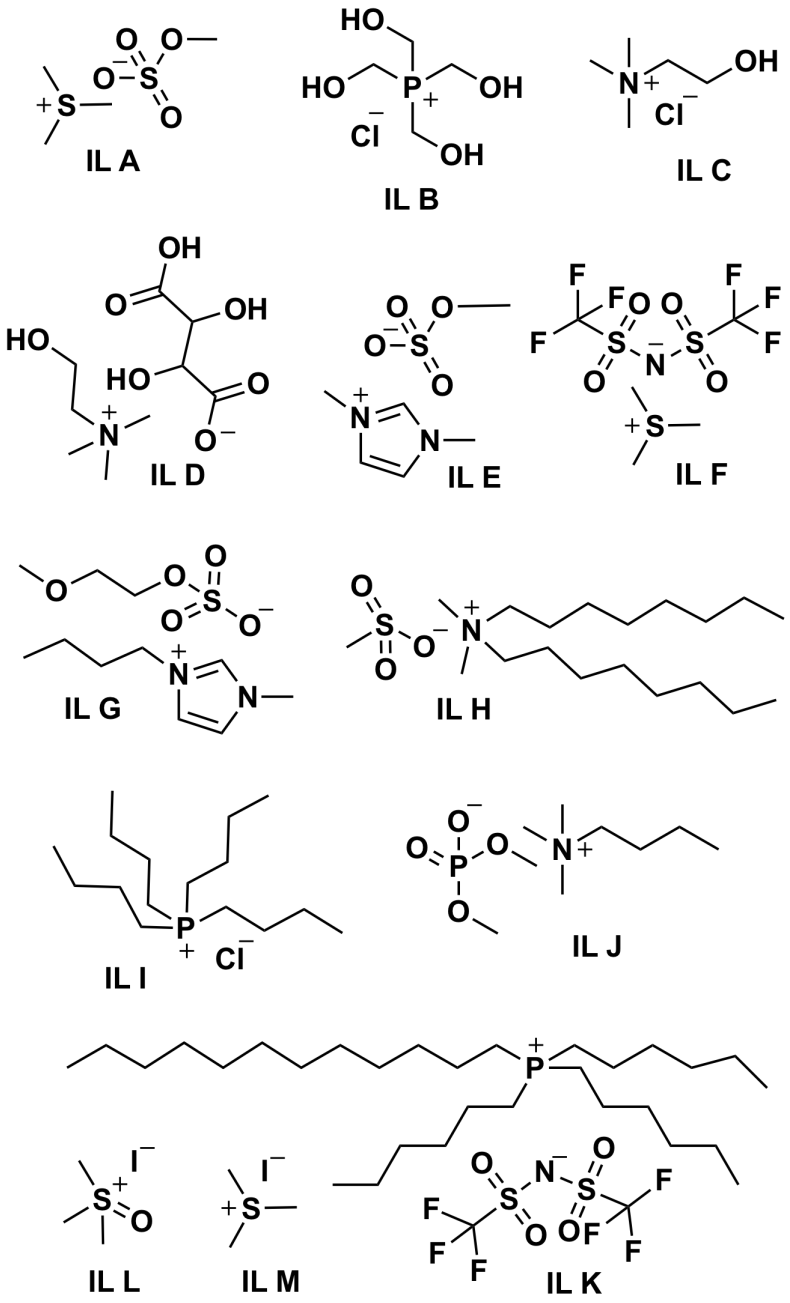


Figure 5.1: ILA $[Me_3S][OSO_3Me]$, ILB [Tetrakis][Cl], ILC [Choline][Cl], ILD [Choline][Bitartrate], ILE $[EMIM][MeSO_4]$, ILF $[Me_3S][NTf_2]$, ILG $[BMIM][MeOEtSO_4]$, ILH $[N_{1,1,8,8}][MeSO_3]$, ILI $[P_{4,4,4,4}][Cl]$, ILJ $[N_{4,1,1,1}][(MeO)_2(HO)PO]$, ILK $[P_{6,6,6,14}][NTf_2]$, ILL $[Me_3SO][I]$ and ILM $[Me_3S][I]$

Focus of investigation was concentration 0.15 M and systems **ILA** [Me_3S][OSO_3Me], **ILB** [Tetrakis][Cl], **ILK** [$P_{6,6,6,14}$][NTf_2] and **ILM** [Me_3S][I] for multiple concentrations.

5.2 Theoretical Background

5.2.1 Molecular Dynamics Simulations

At their core, MD simulations are time dependent descriptions of the classical equations of motion, based on the mechanical description of the potential energy of the systems components, so-called 'force fields'. The principal of force-fields are described in detail in *Chapter 5, Subsection 5.7.*

Integration of Newton's equation of motion leads to a numerical description of the positions and velocities of all particles in the system at each time step and results in a trajectory of coordinates, that depicts the dynamic behaviour of the system over a given time span.

The definition of a reasonable integration time step Δt is crucial to exploit the physical systems (Cores) capability optimally. This time step is limited by the fastest motion of the system and this is most often the vibration of a hydrogen bond. Hence, the longest time step is in this case limited to 1 fs in order to be 1 magnitude slower than this simulation. This in turn limits the absolute length of the simulation and detailed views into biomolecular systems require large amounts of CPU time. On average, several hundred nanoseconds are computed, a time frame in which protein domain motions and protein backbone motions can be resolved sufficiently. However, processes such as protein denaturation or folding cannot be treated, since these take seconds to hours.

5.2.2 Newton's Equation of Motion

Starting an MD simulation requires the definition of the box size of the system as well as the coordinates and velocities of all atoms in the system. Initial coordinates are either retrieved from X-ray or crystallographic investigations or necessitate the building and optimisation of a homology model. Homology models are described in detail in *subsection 9*

of this Chapter. Coordinates are represented by the systems topology, which is generated by a force field and which stays constant throughout the simulation run, meaning, that e.g. no bond breaking events take place. Only the coordinates and velocities of the atoms of a molecule are updated during the simulation run. The initial velocities are usually unknown and are therefore generated with a *Maxwell-Boltzmann distribution*:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}} e^{\frac{m_i v_i^2}{dx}}, \tag{5.1}$$

where k is the *Boltzmann's constant* and T the given absolute temperature. The simulation starts by calculating the acceleration a for every atom i according to *Newton's law*:

$$F_i = m_i a_i, \tag{5.2}$$

where F_i is the force on particle i and m_i is its mass. The force can then be further expressed as the gradient of the potential energy derived from the force field and the acceleration as the second derivative of the positions:

$$\frac{\partial V}{\partial x_i} = F_i = m_i a_i = m_i \frac{\partial^2 x_i}{\partial t^2}, \tag{5.3}$$

where V is the potential energy function, x_i the Cartesian coordinate for i=1, ..., $3N$ for N atoms, and t the time. The total energy E of the system is defined as the sum of all kinetic energy, E_{kin} and potential energy, E_{pot} contributions:

$$E = E_{kin} + E_{pot} = \frac{1}{2} m v^2 + E_{pot} = \frac{1}{2} m (\frac{\partial x}{\partial t})^2 + V(x). \tag{5.4}$$

Since it is not possible to solve the potential energy function analytically, a numerical integration algorithm is used to calculate the positions $x_i(t+\Delta t)$ at the next time step $(t+\Delta t)$. The algorithm, which is most commonly implemented for this, is the *leap-frog* [Hockney et al., 1974] algorithm, which is based on the *Verlet* algorithm [Verlet, 1967]. The *sander* and *pmemd* (AMBER) integrators are closer to *velocity Verlet* than to *leap-frog*, as the way the kinetic energy is estimated comes from *velocity Verlet*. In terms of rounding errors, *sander* and *pmemd* are neither "pure" leapfrog nor "pure" *velocity Verlet*.

Integration algorithms assume the positions, velocities and accelerations by a Taylor series

expansion:

$$\begin{aligned} x(t+\Delta t) &= x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + ... \\ v(t+\Delta t) &= v(t) + a(t)\Delta t + \frac{1}{2}b(t)\Delta t^2 + ... \\ a(t+\Delta t) &= a(t) + ... \end{aligned} \tag{5.5}$$

For the *Verlet* algorithm each next time step is calculated with the knowledge of the atom positions of the last and current time step only. *Verlet* therefor uses the acceleration and positions from time *t-dt* and t to calculate new positions at *t+dt*:

$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \tag{5.6}$$

$$x(t-\Delta t) = x(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \tag{5.7}$$

Summing up equations 5.6 and 5.7 gives equation 2.8, which does not explicitly use velocity information.

$$x(t-\Delta t) = 2x(t) - x(t-\Delta t) + a(t)\Delta t^2 \tag{5.8}$$

For the *leap-frog* algorithm the velocities v are calculated at time *t-Δt/2*. The velocities together with the positions x at time t are then used to calculate the positions at time *t+dt*. Thereafter, the coordinates and velocities are updated using the acceleration a at time t:

$$v(t+\frac{1}{2}\Delta t) = v(t-\frac{1}{2}\Delta t) + a(t)\Delta t \tag{5.9}$$

$$x(t+\Delta t) = x(t) + v(t+\frac{1}{2}\Delta t)\Delta t \tag{5.10}$$

In that sense, the velocities *leap* over the positions, then the positions *leap* over the velocities. This leads to an explicit calculation of the velocities, however not at the same time as the positions. Thus, the velocities at time t must be approximated.

5.2.3 Constant Temperature and Constant Pressure Dynamics

The simulations produce trajectories, which, depending on the factors held constant during the simulation, constitute different statistical ensembles. A microcanonical ensemble is defined as a mechanical system, which is assumed to have an exactly specified total energy, that does not change over time nor is there any exchange of energy or particles

with the environment. In the microcanonical ensemble NVE, the number of particles N, the volume V and the energy E are held constant for the whole simulation and the temperature is calculated. However, in many cases it is also desirable to hold the temperature T constant. Since the temperature corresponds to the kinetic energy of the system, it is controllable by adding or removing kinetic energy from the system. This is achieved by coupling the system to an external heat bath. In the Berendsen's weak coupling method [Berendsen et al., 1984] the velocities of all atoms are scaled at each step, so that the rate of change of temperature is proportional to the temperature difference between the heat bath and the system:

$$\frac{\Delta T}{\Delta t} = \frac{1}{\tau}[T_0 - T], \tag{5.11}$$

where T_0 is the target temperature, T the total temperature of the system and τ the coupling parameter or rather the macroscopic relaxation time. The velocities are scaled by the factor λ to approach the desired temperature:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}(\frac{T_0}{T} - 1)}. \tag{5.12}$$

The magnitude of τ controls how tightly the system is bound to the bath. A large τ couples weakly, a small τ couples strongly.

Further, many experimental investigations are made under constant pressure conditions and make MD simulations, which use an isothermal isobaric ensemble NPT, desirable. This is achieved through weak coupling to an external bath, whereby the pressure P is controlled by scaling the coordinates:

$$\frac{\Delta P}{\Delta t} = \frac{1}{\tau}[P_0 - P], \tag{5.13}$$

whereby τ is again the coupling constant. The scaling factor μ is used to rescale the box size and the atomic coordinates at each integration time step Δt:

$$\mu = [1 - \frac{\Delta t}{\tau}(P_0 - P)]^{\frac{1}{3}} \tag{5.14}$$

Pressure scaling can be done isotropically, whereby the scaling factor is the same for all three spatital directions, or anisotropically, whereby the box dimensions change independently.

5.2.4 Particle-Mesh Ewald Electrostatics

Calculation of nonbonding interactions necessitates the computation of all atom pairs in the molecular system, which is computationally very expensive. The electrostatic interaction energy between molecules is calculated as the sum of pairs of point charges, using *Coloumb's law*:

$$E_{coul} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}), \tag{5.15}$$

where E_{coul} is the total electrostatic interaction energy between all atoms i ... N, q_i and q_j the atomic point charges, ε_0 the permittivity or dielectric constant of free space and r_{ij} the distance of the point charges. The double sum implies the computationally expensive scaling of N^2, which is practically impossible for larger systems. In order to facilitate better computability, electrostatic interactions are split into short range and long range parts, whereby the cutoff is usually between 8 and 12 Å. The short range part is calculated with particle-particle interactions and the long-range part is calculated on a grid with periodic boundary conditions (*see section 5.5*, that uses *Fast Fourier transformation* (FFT). The grid is calculated using the *Particle Mesh Ewald* method [Darden et al., 1993], which is based on the *Ewald summation* method Ewald [1921]. For more information see Cooley and Tukey or Darden et al [Cooley and Tukey, 1965, Darden et al., 1993].

5.2.5 Solvent Treatment

Simulation of biologically relevant molecules, such as proteins, necessitates an appropriate treatment of the surrounding solvent, which is water in most cases. Water can be treated by implicit or explicit molecular models. The model of choice depends on the effects subjected to investigation, whereby implicit models are able to mimic basic electrostatic effects of water on the solute, but cannot reproduce direct, site specifc interactions like hydrogen bonding between water molecules or water molecules and solvent. Hence, the most common way to tackle solvent treatment of bulk properties correctly is the use of explicit solvent models. However, this is computationally very expensive. In the following, implicit solvent models will be outlined only briefly for sake of completeness, and then explicit solvent models will be explained in more detail, since these are the ones relevant to this sudy.

Implicit Solvent

With implicit solvent models, dielectric properties of the solvent can be considered without using explicit atomistic solvent models. The fastest, but crudest method is to rescale the dielectric permittivity of free space ε_0 by a factor D in *Coulomb's law* (Eq. 5.15) to account for the damping effect of solvent on electrostatic interactions with $\varepsilon = \mathrm{D}\varepsilon_0$. Another possibility is a distance dependent dielectric model based on the separation of point charges, whereby the relative permittivity is proportional to the distance in the way $\varepsilon = \varepsilon_0 r_{ij}$, so that the interaction energy between the two charges q_i and q_j according to *Coulomb's law* becomes:

$$E_{coul} = \frac{1}{4\pi\varepsilon}\frac{q_i q_j}{r_{ij}^2} \tag{5.16}$$

Other implicit models treat the solvent as a continuous medium, like for instance the *generalized Born* (GB) method [Gasteiger and Engel, 2006].

Explicit Solvent

Explicit solvent models are intended to be kept as simple as possible, whilst still being able to reproduce experimentally known thermodynamic and structural properties of bulk water. A number of established models for water are available, since it is the solvent of most biological interest. The most straightforward method is to use 'simple' atomistic solvent models. Decrease of the computational cost, that comes with including huge numbers of water molecules, is achieved by holding the water geometry rigid. Bond stretching and angle bending modes are neglected, so that only nonbonding interactions are considered. The models represent the water molecules by a number of interaction sites based on their rigid geometry. The two most common families of water models are the *Transferable Intermolecular Potential* (TIP) and the *Simple Point Charge* (SPC) models. TIP3P [Jorgensen et al., 1983] and SPC [Berendsen et al., 1981] represent the simplest and most commonly used three-point water models. The electrostatic interactions are represented by positive partial charges at the hydrogen atom positions, which are exactly balanced by a negative charge at the position of the oxygen atom. The *van der Waals* interactions (*see section Classical FF*) are only mimicked by one interaction site at the oxygen atom. Further frequently used water models are TIP4P [Jorgensen and Madura, 1985] and SPC/E [Berendsen et al., 1987]. Only TIP3P and TIP4P are relevant to this

study.

TIP4P is a four site water model, whereby the point charge of the oxygen is shifted into the area between the hydrogen atoms and the oxygen atom. When used in conjunction with *Ewald summation* methods a reparametrisation TIP4-EW [Horn et al., 2004] is necessary. Besides these planar water models, tetrahedral five point models are also available, such as TIP5P [Mahoney and Jorgensen, 2000] or ST2 [Stillinger and Rahman, 1974]. All of these models represent bulk properties of water well, however, representation of properties such as dielectric constant, density, vaporisation enthalpy or diffusion coefficient varies significantly between the different models. For literature on the critical comparison between models see [Vega et al., 2009].

5.2.6 Periodic Boundary Conditions

Simulating a box in vaccuum creates a boundary, which is unphysical, meaning that water molecules will 'evaporate' at the solvent/vaccuum interface. A solution to eliminate such effects is the use of *Periodic Boundary Conditions* (PBC). Different box shapes, e.g. cubic, rhombic, dodecahedron, truncated octahedron, etc. are possible to inflict PBC. The box containing the solute and solvent is surrounded by its periodic images. A cubic box is therefore surrounded by 26 neighbour boxes, which are all translated unitcells. A molecule, that leaves the box on one side enters it at the same time on the opposite side (figure 5.3).

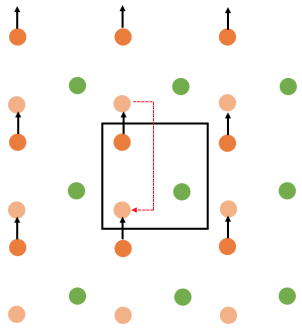


Figure 5.2: Periodic Boundary Conditions in 2D - in case of a cube it would be surrounded by 8 of its images for all three axes in the 3 dimensions, since simple alignment to the side of the cubes would obscur the directions and would pose an issue for molecules escaping at the corners, which are not planar, but inflict all three dimensions.

An atom of a real molecule (not image) must not interact with another real atom and its image at the same time. This requirement poses a problem for long range nonbonding interactions, like electrostatic interactions, and therefore a spherical cutoff must be defined (see section 5.4), that is smaller than half the smallest box dimension.

5.2.7 Classical Force Fields

MD simulations are based on a purely mechanical description of the potential energy of the system, a so called force field. This force field, defines molecular mechanics parameters (in contrast to quantuum mechanics) and ignores the electronic motions by calculating the energy of the system as a function of the nuclear positions only [Leach]. Several assumptions must be made to do this. The most important one is the *Born-Oppenheimer approximation*:

$$\Psi_{tot}(R,r) = \Psi_n(R)\Psi_{el}(R,r), \tag{5.17}$$

which is integrated to the *Schrödinger equation*:

$$H\Psi = E\Psi = i\hbar\frac{\partial\Psi}{\partial t}, \tag{5.18}$$

where R and r are the positions and momenta of the nuclei and electrons, and H is the Hamiltonian (the sum of potential and kinetic energy), Ψ the wavefunction and $\hbar = h/2\pi$, where h is the *Planck's constant*. The wavefunction is a function of the coordinates and momenta of the nuclei and electrons to describe all particles of the system. The *Born Oppenheimer approximation* separates the total wavefunction Ψ_{tot} into the nuclear Ψ_n and the electronic wavefunction Ψ_{el}, where the much heavier nuclei are fixed on the time scale of electronic vibrations. Because of this, the *Schrödinger equation* can be separated into a time dependent part for the motion of the nuclei and a time independent part for the motion of the electrons. Thus, the *Born-Oppenheimer approximation* first made it possible to write the energy as a function of the nuclear positions only.

Despite classical force fields being rather simple functions, their accuracy can in some cases compete with high level quantuum mechanical calculations. Of course, they cannot provide a description of properties that depend on the distribution of electrons in the system, such as either chemical reactions, bond cleaving or bond formation events.

Contributing components to most of the force fields can be classified into intra-molecular

(bonding) or inter-molecular (non-bonding) forces. The bonding contributions are, bond stretching, angle bending and dihedrals or torsions. The non-bonding forces are described by *van der Waals* (vdW) interactions and electrostatic interactions. Hence, In its easiest form, the potential energy function V looks like this:

$$V = V_{bonds} + V_{angles} + V_{dihedrals} + V_{non-bonding}. \tag{5.19}$$

The bonding parameters are calculated as simple harmonic potentials with equilibrium distance (r^0), angle (θ^0) and a torsion term for the rotation around chemical bonds with a penalty for deviation from their equilibrium values. Non-bonding parameters are, in contrast, computationally expensive to calculate, since they must be calculated as pair potential of each atom with each other atom. The electrostatic interactions are calculated according to the previously mentioned *Coulomb's law*, whereas the *van der Waals* interactions are represented by a *Lennard-Jones* potential:

$$V_{bonds} = \sum_{bonds} \frac{1}{2} k_r (r_i - r_i^0)^2 \tag{5.20}$$

$$V_{angles} = \sum_{angles} \frac{1}{2} k_\theta (\theta_i - \theta_i^0)^2 \tag{5.21}$$

$$V_{dihedrals} = \sum_{dihedrals} \frac{1}{2} V_n (1 + cos(n\omega - \gamma))^2 \tag{5.22}$$

$$V_{non-bonding} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} (4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^1 2 - (\frac{\sigma_{ij}}{r_{ij}})^6 + \frac{q_i q_j}{4\pi\varepsilon_0}]) \tag{5.23}$$

In the development of the force fields, the parameters are adjusted to match experimental outcomes, since a good parametrisation is crucial for a good representation of the potential energy surface and since the force fields are based on such a simplicistic model. High resolution crystal structures or thermodynamic averages, as well as high level *ab initio* calculations serve for fitting of parameters. The parameters are then summarised by the definition of different atom types. These atom types have to be defined in the input of the calculations and contain the important information to calculate all bonding and non-bonding interactions.

State-of-the-art force fields for biomolecular simulations are: AMBER [Cornell et al., 1995, Duan et al., 2003], GROMOS [Oostenbrink et al., 2004, van Gunsteren et al., 1996], CHARMM [MacKerell Jr et al., 1998] and OPLS Jorgensen and Tirado-Rives [1988]. All

of these use similar functional terms following equations 5.20 - 5.23, but differ in their paramters. This can lead to a quite extensive definition of atom types. However, the more atom types defined, the broader the variety of molecules treatable with the assigned force field.

Atom types of force fields for biomolecular simulations are parametrised for the simulation of proteins and hence, represent the 20 proteinogenic amino acids. Force fields that are designed to model specific classes of molecules, have a more tailored range of atom types, than general purpose ones, and will also include atom types for commonly present complexing ions, such as zinc. In the AMBER force field, 41 different atoms are defined, among them, for example, two different atom types for an aromatic carbon atom at the junction between a six and a five membered ring (as present in tryptophan) and a carbon atom in an isolated five membered ring (as in histidine). However, these parameters will only poorly describe non-standard amino acids or organic molecules, such as ligands or ionic liquid molecules. Thus, it is necessary to create and incorporate new parameters that also fit to the force field. The AMBER package [Case et al., 2008] contains *antechamber* [Wang et al., 2006], a tool for generating force field parameters, which uses the *general AMBER force field* (gaff) [Wang et al., 2004]. It is an expansion of the AMBER force field with the intention of ease for rational drug design with parameters for almost all organic molecules made of C, O, N, H, S, P, F, Cl, Br and I. Further, it is also possible to assign completely new atom types or parameters using these tools.

5.3 Applied Methods

All simulations were performed on the High Performance Cluster (*Augusta*) at the University of Nottingham, which comprises of a mixture of classical x86 hardware and GPGPU (General-purpose computing on graphics processing units) accelerator hardware, a total of 4,720 cores, 31 TB of RAM (Random-access memory), 6 GPGPUs and 750 TB of storage.

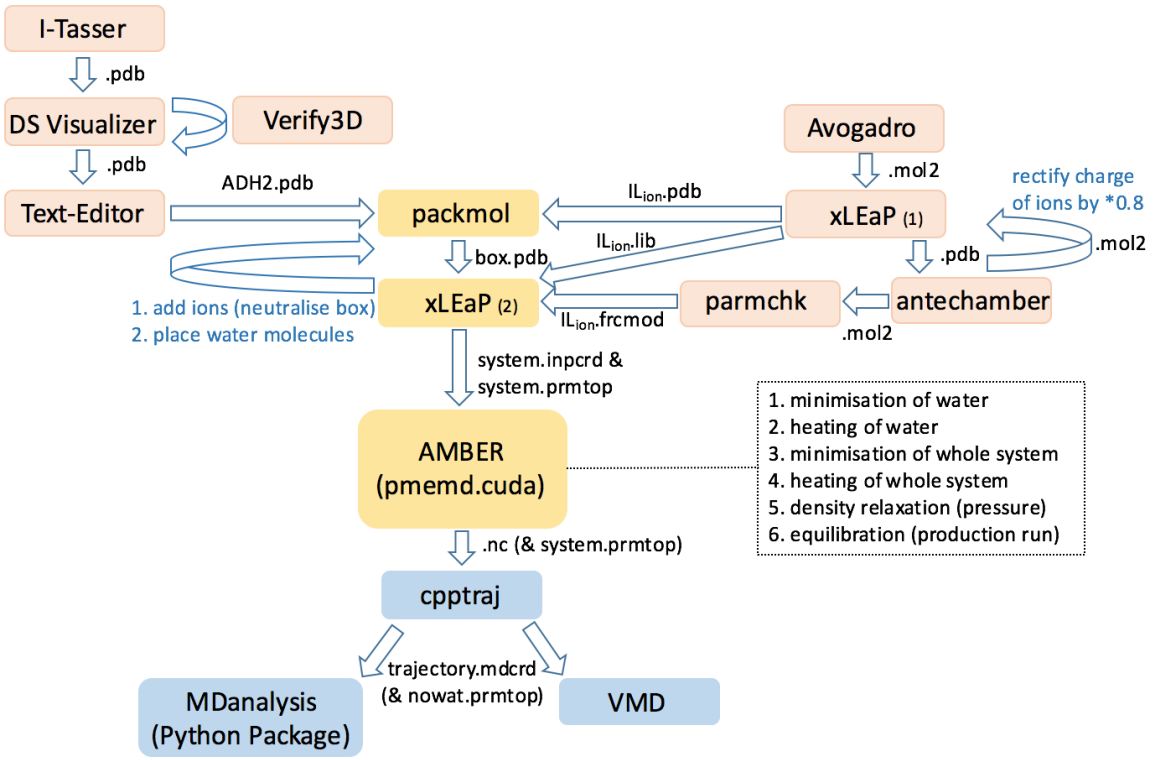


Figure 5.3: Scheme for setting up a molecular dynamics simulation.

A scheme for setting up a molecular dynamics simulation is presented in figure 5.2. Components are described as follows: I-Tasser - to generate a homology model, DS Visualizer - to build a homology model and optimise geometry, Verify3D - to assess quality of homology model, Text-Editor - to adjust atom names and chains in .pdb file; Avogadro - to build molecular structures of ions and optimise geometry, xLEaP (1) - to transform coordinates and yield library files (specify atom names) for ions, antechamber - to calculate charges based on the general amber forcefield (gaff) for ions, parmchk - to assign parameters, such as bonded or non bonded, for ions; packmol - to build the box containing protein and ions, xLEaP (2) - to define force fields (topology) and coordinates, adding single atom ions (K^+) and water molecules (model: TIP3PBOX), pmemd.cuda - to run minimisation, heating and equilibration steps, cpptraj - to convert trajectory file, VMD - to visualise, MDanalysis - to analyse trajectories.

Homology models from I-Tasser [Roy et al., 2010], Phyre2 [Kelley et al., 2015] and Swiss-Model [Schwede et al., 2003] were compared through visualisation with PyMOL [DeLano,

2002] with regards to conserved residues and overall structural conformation. The chosen homology model was used to build the homo-tetrameric form of *Hv*ADH2 with Discovery Studios Visualizer [Biovia, 2017]. Quality of the geometry optimised homology model was assessed with PROCHECK from the online software ERRAT [Colovos and Yeates, 1993]. A text editor was used to prepare .pdb files for input to Packmol [Martínez et al., 2009] and xLEaP [Case et al., 2008]. Based on input .pdb files for the protein and ionic liquid structures, the software package Packmol is used for building the box containing the later simulation system, which serves as input .pdb file for the program xLEaP. It creates an initial point for MD simulations by packing molecules in defined regions of space, whereby, the packing guarantees that short range repulsive interactions do not disrupt the simulations. The software suite used for running simulations, is AMBER18 (*Assisted Model Building with Energy Refinement*) [Lee et al., 2018], whereby xLEaP, antechamber and parmchk [Case et al., 2008] are used as the preparatory programs before simulations, pmemd.cuda 9.2 [Lee et al., 2018] is used as the simulation program and cpptraj [Case et al., 2008] is used to transform obtained trajectories for visualisation after the simulation runs. Simplified, xLEaP reads in force field, topology and coordinate information, taken from predetermined coordinate files such as .pdb, and generates topology and restart files. *sander* is part of AmberTools [Case et al., 2008], and is the basic energy minimizer and molecular dynamics program. This program relaxes the structure by iteratively moving the atoms down the energy gradient until a sufficiently low average gradient is obtained. The molecular dynamics portion generates configurations of the system by integrating Newtonian equations of motion (for detailed explanation see *subsection 5.2. pmemd* (Particle Mesh Ewald Molecular Dynamics) is a version of sander that is optimized for speed and for parallel scaling and the pmemd.cuda variant runs on GPUs. The force fields employed in this study are leaprc.ff14SB for non-coordinating protein residues and the Zinc Amber force field [Peters et al., 2010] for the Zn^{2+} coordinating environment. For this residues for coordinating the catalytic and structural zinc were renamed to match force field specifications. Chosen parameters apply for catalytic residues 'CY4' and 'HD2' instead of CYS-38 and HIS-59 and structural residue 'CY1' replaced CYS-89, -92, -95 and -103. Ionic liquid structures were built with Avogadro [Hanwell et al., 2012] and structures were optimised with xLEaP using the force field leaprc.gaff. The obtained

.pdb files serve as input for antechamber, which calculates RESP charges based on the bcc charges code. The total charge of each ion has to be reduced by applying a factor of 0.8 to all the atomic charges of the individual ions according to evidence suggested in literature [Sprenger et al., 2015]. After rectifying the charges the new .mol2 file serves as novel input to antechamber from which the library file (.lib) is obtained. While the programme parmchk, part of the antechamber suite, is used to create a .frcmod file with the missing parameters. In a first step, a box was built with packmol, containing only the protein, which was neutralised with potassium and chloride ions using xLEaP. In a subsequent step, the .pdb file containing the neutralised protein and ions was used to build the final box, containing different amounts of ions, which corresponded to molarities used in the experimental set-up, yielding .prmtop (topology) and .inpcrd (coordinate) files. In the following, an energy minimisation run with water restraints within 3 Åby applying a 10 kcal/mol.A**2 restraint force was performed, followed by a heating step with water restraints (3 Å) of the system to 323.15 K, followed by a second minimisation which included the water molecules, followed by a heating step which restraint the protein (3 Å), but not the water, followed by an equilibration step with NPT conditions (isothermalisobaric ensemble) for 0.5 ns which restraint the protein (3 Å). Minimisation runs are performed before a final MD simulation to allow the structure to cross over small potential energy barriers. Production runs were between 150 and 200 ns. Trajectories were saved every 5000 steps during the simulation for later analysis, and Gibbs free energy calculations using thermodynamic integration were subsequently performed. The temperature corresponded to the 50 ℃ celsius used in the experimental set-up and was 323.15 K. The exact *in-silico* conditions care given in the *Appendix*.

5.3.1 Molecule number calculations for Packmol

In order to mimic experimental conditions, molarities and compositions of system components, including all ions were calculated according to

$$N_{ions} = N_A * c_{ion}(mol/l) * V(l), \tag{5.24}$$

whereby N_{ions} is the number of ions, N_A is the avogadro number 6.022*10^{23}, c_{ion} the concentration of the ion and V the volume. Exact molecule numbers per system are tabelled

in figure 5.4. Potassium and chloride ions vary, since the system was neutralised, before a defined amount of ions was added. A box size of 100x100x100 was chosen, leading to a volume of 100000 Å(=10^{-21} L) , and the protein was placed at the centre of mass. Since the number of water molecules decreases with increasing numbers of IL_{ions} following an increase in $IL_{concentration}$, the number of potassium and chloride ions needs to be adjusted accordingly.

IL concentration	IL ions/box	K Cl ions/box
0.025 M	14	~1980
0.075 M	40	~1900
0.15M	80	~1800
0.3M	163	~1600
0.6M	326	~1400
0.75M	408	~1200
0 M	N/A	~2000

Figure 5.4: Molecules per box.

5.3.2 Comparative Modelling

The software used for comparative modelling in this study was Phyre2 (*Protein Homology/AnalogY Recognition Engine*) [Kelley et al., 2015], Swiss-model [Schwede et al., 2003] and I-TASSER (*Iterative Threading ASSEmbly Refinement*) [Yang and Zhang, 2015]. Software for visualisation included PyMOL and Discovery Studio Visualiser (DSV) [Biovia, 2017], whereby DSV was also used to build the final tetrameric model using the *superimpose module*. Software used for sequence alignment was BLAST (basic local alignment search tool) and software used for homology model evaluation was Verify3D.
Comparative modelling methods build 3D models for proteins of unknown structure (target) on the basis of sequence similarity to proteins of known structure (templates) (figure 5.5). This generation of credible structures is possible, since small divergence in the amino acid sequence will often only lead to small changes in the overall 3D structure of the folded protein [Chothia and Lesk, 1986]. *de novo* prediction is an alternative, but delivers still far less accurate results, e.g. ROSETTA [Rohl et al., 2004]. The accuracy obtained with comparative modelling can vary substantially depending on alignment between target and

templates. Methods specifically designed for sequence structure alignment searches, such as PSI-BLAST [Altschul and Koonin, 1998], are useful to investigate conserved regions of the target protein, before consulting methods for predicting the 3D structure from known protein structures (comparative modelling methods). Later obtained models of the target, can then be more easily evaluated for their quality to preserve conserved residues and hence readily identified with regards to posed research questions.

Comparative modelling approaches consist usually of four steps. First, known structures related to the target sequence must be found (sequence alignment searches), so that in a second step, target and template structures can be aligned. Based on the template structure, the structure for the target can be built in the next step. This is done, either based on one template structure (*Swiss model*) or by 'piecing-together' through reassembling structural fragments from threading (fold recognition) templates (*I-TASSER, Phyre2*). As a final step the model quality is assessed or scored, and steps need to be repeated until a satisfactory model is obtained.

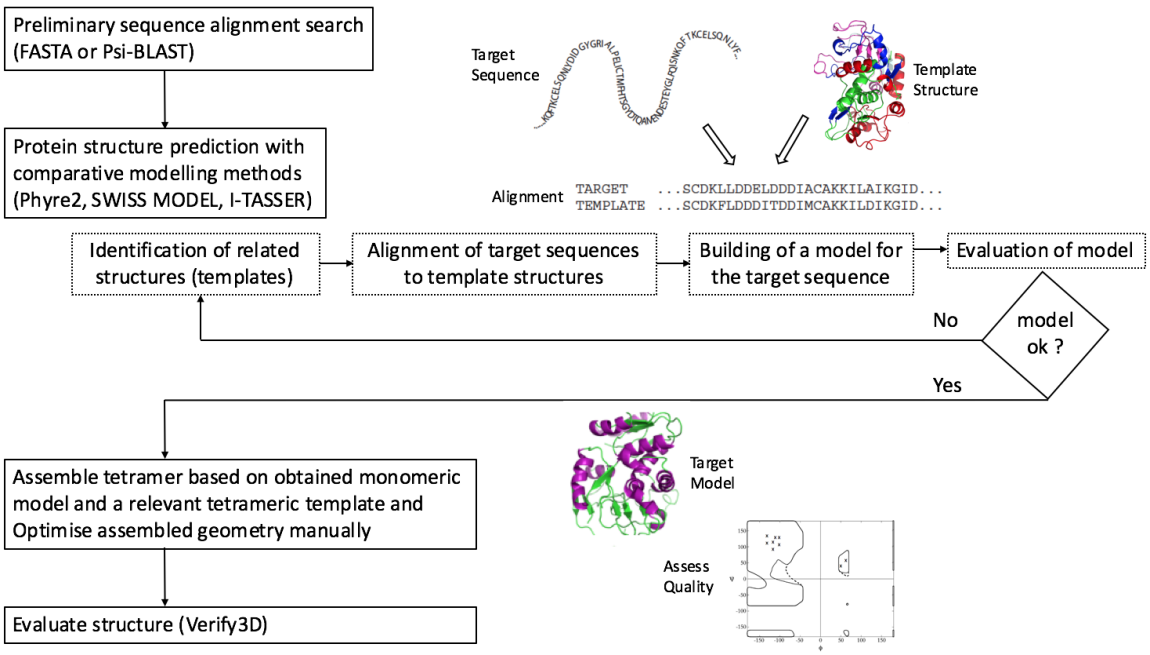


Figure 5.5: Scheme for building a homology model.

Depending on the template structure and on the method used for building the homology model, the model will have the quaternary structure of the template, e.g. in case of

Swiss-model a dimeric template structure will result in a dimer for the model. In case of I-TASSER, however, the derived model is monomeric and hence necessitates further steps to obtain the final model, if the native structure is a tetramer. For this purpose, a structure based design software, such as Discovery Studios Visualizer (unlicensed version) can be used. In this instance, the 3D geometry needs to be optimised after assembly of the tetrameric structure and the quality of the final model needs to be evaluated.

Sequence Alignment

The identification of protein structures related to the target sequence (homologs) and the selection of a suitable template structure is based on the comparison of the protein sequences. This comparison can be performed by either simple pairwise comparison or can include evolutionary information on the target and/or template and results in a list of which amino acids in the two sequences are equivalent and which are not. Methods, that use evolutionary information, align and search through multiple sequences and are superior, since conserved regions can be better described. Most commonly used for this is PSI-BLAST. It includes evolutionary information on the target sequence by using initial database hits to refine its search in an iterative way. The first database hits for a given sequence are used to build a multiple sequence alignment, a position specific scoring matrix is then constructed based on the alignment and is then used to search the database for new homologs. This process is repeated until no new homologs are detected. In addition to this approach, most comparative modelling software uses a consecutive step, which uses evolutionary information to construct a profile. This profile is a set of vectors, where each vector contains the frequency of each type of amino acid in a particular position of the multiple sequence alignment. In this way, parts of the sequence that are conserved can be detected. Usually, higher overall sequence similarity between the target and the template guarantees a better model, but factors such as ligands, quaternary interactions and resolution of template crystal structures need to be taken into account in the selection process. Modelling specifically protein-ligand interactions, for example, will prioritise conserved active center residues over tertiary structure preservation and will require a high resolution template.

Building the Model

A number of methods have been developed for building homology models. Earlier methods can be classified into rigid-body assembly, e.g. Swiss-model; segment matching, e.g. SEGMOD [Levitt, 1992]; and modelling by satisfaction of spatial restraints, e.g. MODELLER [Šali and Blundell, 1993]. Newer methods rely on meta-prediction, e.g. I-Tasser, Phyre2. In the following, only rigid-body assembly and meta-prediction will be briefly elaborated and corresponding methods, Swiss-model, I-Tasser and Phyre2, which were used in this study, will be explained.

Modelling by rigid-body assembly

Modelling by rigid-body assembly is the oldest method, but is still widely applied. Here, the model is assembled from a small number of rigid bodies obtained from the core of aligned regions. Assembly starts by building a framework from the coordinates of aligned C_α atoms. The assembly involves the fitting of the rigid bodies onto the framework and rebuilding the non-conserved parts, i.e. loops and side chains. The main difference between the programs lies in how these non-conserved parts are built. They can either be changed stepwise, one residue at a time or through mean-field minimisation methods.

Swiss-model follows the steps depicted in figure 5.6. It creates a template structure library based on structures deposited in the Protein Database (PDB), which are preliminary assessed using empirical force field energy [van Gunsteren et al., 1996] and ANOLEA (Atomic Non-Local Environment Assessment) mean force potential scores [Melo and Feytmans, 1998, Sippl, 1990]. ANOLEA uses a mean force potential at the atomic level, which is based on a particular definition of atom types. Forty different atom types are defined for all the heavy atoms of the 20 existent amino acids. In a strict physico-chemical point of view, all the atoms with different environments, connectivity and chemical nature, would be different. The mean force potential is calculated to evaluate the non-local environment (7 Åradius and more distant than 11 residues) of a given residue by compiling each pairwise contact symmetrically, using the following expression:

$$\Delta E^i j(l) = RTln[Mij\sigma] - RTln[Mij\sigma]\frac{f^{ij}(l)}{f^{xx}(l)}, \tag{5.25}$$

where M_{ij} is the total number of observations for the atomic pair ij and corresponds to:

$$Mij = \sum_{l=1}^{n} f(i,j,l), \tag{5.26}$$

and sigma represents the weight given to each observation. $f^{ij}(l)$ is the relative frequency of occurence of the atomic pair ij in the class of distance l and corresponds to:

$$f^{ij}(l) = \frac{f(i,j,l)}{Mij}, \tag{5.27}$$

and $f^{xx}(l)$ is the relative frequency of ocurrence of all atomic pairs in the class of distance l and can be expressed as:

$$f^{xx}(l) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} f(i,j,l)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} f(i,j,l)}. \tag{5.28}$$

Local pair-wise alignment is calculated based on an iterative least square algorithm (for more information see [Huang and Miller, 1991]). The model itself is built by generating the core through averaging C_α backbone atom positions of the templates, weighted by sequence similarity. Gaps or overlaps in the alignment are generated based on an ensemble of fragments using constraint space programming. In the next step, side chains are modelled starting with conserved residues and picking conformers from a backbone dependent rotamer library (for more information see [Lovell et al., 2000]).The last step is a minimisation process regularising the joined rigid fragments using the GROMOS96 force field [van Gunsteren et al., 1996].

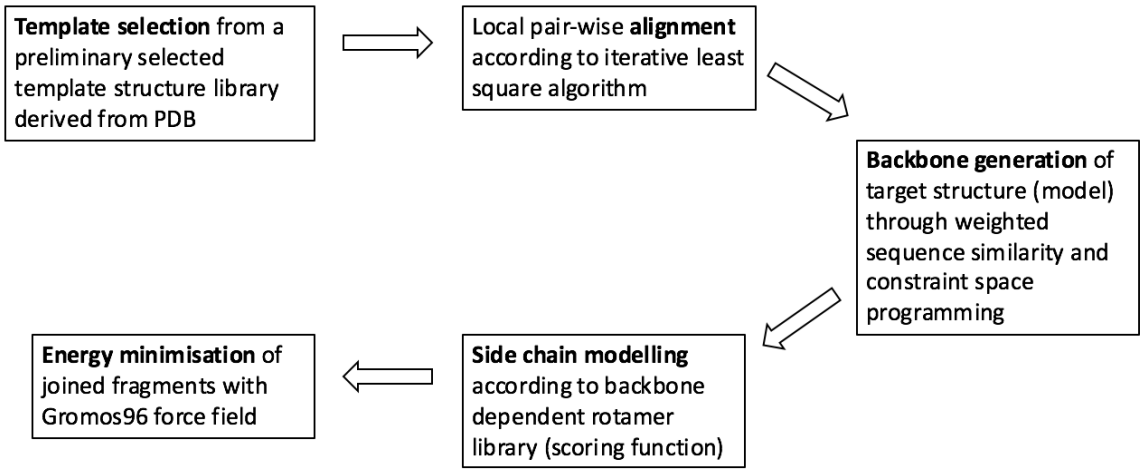


Figure 5.6: Algorithmic Pipeline of Swiss-model comparative modelling

The above stated does not claim to be a complete description of the Swiss-model workspace, nor will the two model below be described in full detail. The description should suffice as an overview on the difference between the methods, to help give an insight on why the homology models built in this study vary to a certain extent (see *Results* section). Hence, for more information on the Swiss-model workspace, see [Bordoli and Schwede, 2011] and [Schwede et al., 2003].

Meta-prediction

In meta-prediction multiple individual threading or other prediction programs are consulted and their information is combined in order to increase performance. They are either based on the idea that if several independent methods produce similar models it is a strong indicator that the models are correct, e.g. 3D-SHOTGUN [Fischer, 2003]. Or else, the top template hits are selected for further consideration and serve to excise continuous fragments from template structures, e.g. Phyre2 (figure 5.7) and I-TASSER (figure 5.8).

Phyre2 follows the steps depicted in figure 5.7. HHblits, a hidden markov model based tool representing query and database sequences, is used to perform an iterative sequence search and is with this, a faster alternative tool to PSI-BLAST [Remmert et al., 2012]. Hidden Markov Models (HMM) are statistical Markov models in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states [Baum and Petrie, 1966], whereby a Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event [Gagniuc, 2017]. The resulting sequence alignment is used to predict the secondary structural elements of the target with PSIPRED [McGuffin et al., 2000]. PSIPRED is based on PSI-BLAST (PSI-BLAST based secondary structure prediction) and is another comparative modelling program. The obtained sequence alignment and secondary structures are combined into a query HMM, which is then scanned against a fold library database of HMMs of proteins with known structures using the alignment algorithm HHsearch [Söding, 2004]. Top scoring alignments are used for the construction of backbone-only models, which, however, often contain insertions and deletions (indels).

Indels are processed using a library of fragments with 2 - 15 amino acids in length, which are derived from a complete fragmentation of the structure database. Geometry of flanking regions and distances between endpoints are taken into account and fitted via cyclic coordinate descent, which is a robotic arm algorithm, that attempts to fit the ends of the fragment to the backbone while minimising changes in the dihedral angles (for more details see [Canutescu and Dunbrack Jr, 2003]). Finally, fitted fragments are ranked using a combination of empirical energy terms and the top scoring model is selected for fitting side chains to the backbone using the R3 protocol (for more information see: [Xie and Sahinidis, 2005]).

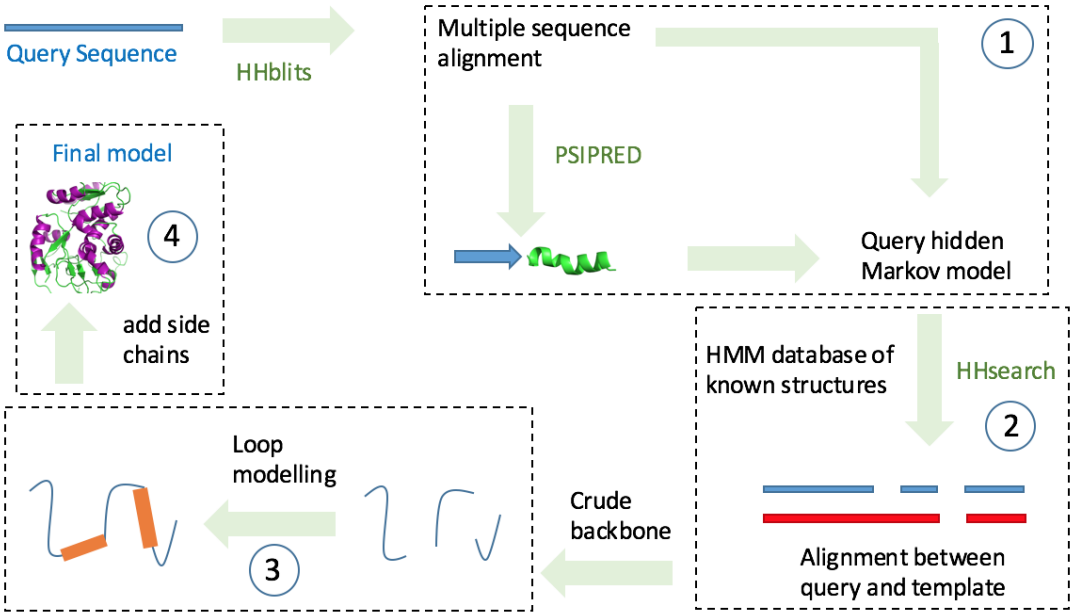


Figure 5.7: Algorithmic pipeline of Phyre2 comparative modelling: (1) Gathering homologous sequences by scanning the target sequence against a protein sequence database, which contains sequences with sequence identity lower than 20 % with HHblits. Resulting sequence alignment is used to predict secondary structures with PSIPRED and both, alignment and secondary structural elements, are combined in a query hidden Markov model. (2) Scanning of query HMM against database of HMMs of proteins of known structure. Top scoring alignments are used to construct backbone-only models. (3) Indels are then corrected by loop modelling. (4) Side chain placement. Source: modified from [Kelley et al., 2015]

I-Tasser follows the steps depicted in figure 5.8. I-Tasser is similar to Phyre2 in that sense,

as it starts by identifying template proteins from solved structure databases and matching them against the query sequence by PSI-BLAST. However, not only whole structures are matched, but templates containing similar structural motifs are matched also. Deep multiple sequence alignments (MSAs) are generated by iterative sequence homology searches through multiple sequence databases. 11 programs are combined to identify structural templates from the PDB library. As with Phyre2, a sequence profile is derived from the multiple alignment and PSIPRED is used to predict secondary structures. In the following, the query sequence, consisting of sequence profile and predicted secondary structure, is then threaded through a representative PDB structure library using LOMETS (LOcal MEta-Threading-Server) [Wu and Zhang, 2007], a meta-threading server, that combines 7 threading programs (FUGUE [Shi et al., 2001], HHSEARCH [Söding, 2004], MUSTER [Wu and Zhang, 2008], PROSPECT [Xu and Xu, 2000], PPA [Wu et al., 2007], SP3 [Zhou and Zhou, 2005] and SPARKS [Zhou and Zhou, 2004]).

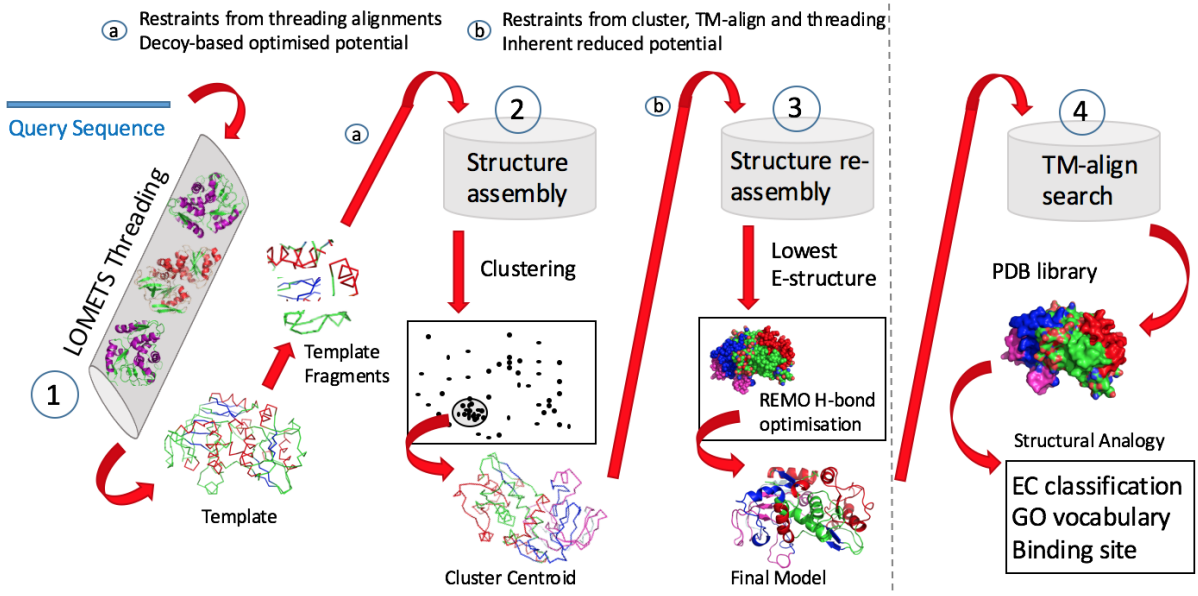


Figure 5.8: Algorithmic pipeline of I-TASSER comparative modelling: (1) Threading (alignment of sequence to structure) using a bundle of 7 different programs, (2) Continous fragments serve as template structural conformations, non-aligned segments are inserted by *ab-initio* modelling. Conformations are clustered and best cluster centroids selected. (3) Re-iteration of previous step using only structural information from the cluster. Final model is built based on best centroids derived from the second step. (4) Function classification. Source: modified from [Roy et al., 2010]

The capital difference between homology modelling and threading is, that homology modelling only aligns sequence to sequence, while threading aligns sequence to structural templates. Highly important in the latter case, is that the quality of the template alignment is evaluated based on the statistical significance of the best threading alignment. The Z-score which is defined as the energy score in standard deviation units relative to the statistical mean of all alignments is used to assess the difficulty of modelling the target. Continuous fragments are then excised and serve to assemble structural conformations of the sections, that aligned well. The unaligned regions are built by *ab initio* modelling [Wu et al., 2007], whereby a lattice system is introduced to reduce the entropy of conformational search. The fragment assembly is performed using a modified replica-exchange Monte Carlo simulation (repeated random sampling) technique [Zhang et al., 2002] and is guided by a composite knowledge-based force field. Conformations are clustered by the algorithm SPICKER [Zhang and Skolnick, 2004] according to their free energy states, whereby lowest are ranked highest. Cluster centroids are then obtained by averaging all clustered decoys.

The next step is basically an iteration of the previous, only, that the assembly simulation is now started only from the selected cluster centroids, to refine the global topology. Decoys of the second run are clustered again and the lowest energy structures serve as the input for REMO [Li and Zhang, 2009], a program that builds the final model from C_α and optimises hydrogen bonding networks.

In the final step, the function of the query protein is inferred by structurally matching predicted 3D models against proteins of known structure and function in the PDB, relying on Enzyme Commission (EC) numbers [Barrett, 1997], Gene Ontology (GO) terms [Ashburner et al., 2000] and binding sites repositories. Further, the quality of the prediction is assessed by a confidence score, C-score, which is defined based on the quality of the threading alignments and the convergence of I-TASSERs structural assembly refinement simulations:

$$C - score = ln(\frac{M/M_{tot}}{< RMSD >} * \frac{1}{N} \sum_{i=1}^{N} \frac{Z(i)}{Z_0(i)}), \tag{5.29}$$

where M is the number of structure decoys in the cluster, M_{tot} is the total number of decoys generated during the simulation run, $RMSD$ is the average root-mean-squared deviation of the decoys to the cluster centroid. $Z(i)$ is the Z-score of the best template generated by

ith threading in the seven LOMETS programs and $Z_0(i)$ is a program-specified Z-score cutoff for distinguishing between good and bad templates.

Model Evaluation

The quality of information extracted from any model is determined by the quality of the model itself. Hence, any model must be evaluated before it can be put to use, since a number of pitfalls during the building process can occur, such as, distorted geometries of side chains or too dense packing, and alignment errors in unaligned regions, as severe as false fold assignment (especially in loop regions). The model can be evaluated as a whole or in individual regions. Any comparative modelling method will evaluate its model with a score, which is an initial indicator and the established chart (figure 5.9) provides a guideline.

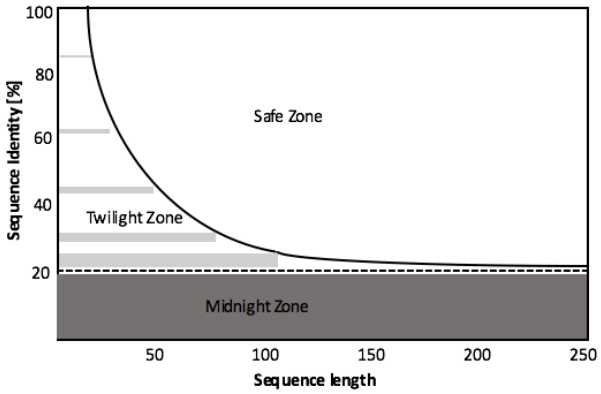


Figure 5.9: Three zones of protein sequence alignments: Two protein sequences can be regarded as homologous, if the percentage sequence identity falls in the *safe zone*. Sequence identity values below the zone boundary, but above 20 % are considered to be the twilight zone, where homologous relationships are less certain. The region below 20 % is the midnight zone, where homologous relationships cannot be reliably determined. (*Source: modified from [Rost, 1999]*

Basic requirement for a model is its correct stereochemistry, which is usually satisfied, due to the use of force fields in most modelling programs. Further, distribution of structural features from known proteins and the deviation thereof can be used as an indicator for errors in the model. Such features include packing, formation of hydrophobic cores, residue and atomic solvent accessibility, spatial distribution of charged groups and atom-atom

contacts, and main chain hydrogen bonding. Methods for checking structures are available, most commonly used are PROCHECK and VERIFY3D. Verify3D uses 3D profiles derived from the local environment around each residue and matches them to reference structures to analyse the compatibility with its own amino acid sequence (1D) based on a scoring system for each of the 20 amino acids [Eisenberg et al., 1997]. PROCHECK [Laskowski et al., 1993] checks the stereochemical quality of a protein structure, analysing its overall and residue-by-residue geometry and resulting in a number of PostScript plots, e.g. Ramachandran plots [Ramachandran and Sasisekharan, 1968].

5.3.3 Analytical Methods

Analyses were performed using the *MDAnalysis* Package from *Python* [Gowers et al., 2016]. Root mean square deviations (RMSD) were calculated from Cα-backbone atoms for the whole length of the trajectory. Inspection of the RMSD plots for the protein backbone atoms indicated that the simulations had typically stabilized after 60 ns, hence the first 60 ns of each trajectory were disregarded for Radial distribution function (RDF) analysis. Radius of gyration and Gibbs free energy were calculated using GROMACS [Lee et al., 2015]. Scripts for processing gromacs files are presented in the appendix. Free Energy Landscapes (FELs) were plotted using Matlab. AutodockVina [Trott and Olson, 2010] and AutodockTools was used for docking [Morris et al., 2009].

RMSD

The RMSD $\rho(t)$ as a function of time is calculated as

$$\rho(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w_i(x_i(t) - x_i^{ref})^2}. \tag{5.30}$$

It is the Euclidean distance in configuration space of the current configuration (possibly after optimal translation and rotation) from a reference configuration divided by $1/\sqrt{\text{N}}$, where N is the number of coordinates.

The weights w_i are calculated from input weights *weights* w'_i as relative to the mean:

$$w_i = \frac{w'_i}{< w' >}. \tag{5.31}$$

RMSF

The root mean square fluctuations of an atom i is calculated as the time average

$$\rho = \sqrt{< (x_i - < x_i >)^2 >}. \tag{5.32}$$

Radial Distribution Functions

The RDF *$g_{ab}(r)$* between types of particles a and b is defined as

$$g_{ab}(r) = (N_a N_b)^{-1} \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} < \delta(|r_i - r_j| - r) >, \tag{5.33}$$

which is normalised so that the RDF becomes 1 for large separations in a homogenous system. The RDF effectively counts the average number of b neighbours in a shell at distance r around a particle a and respresents it as a density.

The radial cumulative distribution function is

$$G_{ab}(r) = \int_0^r dr' 4\pi r'^2 g_{ab}(r') \tag{5.34}$$

and the average number of b particles within radius r

$$N_{ab}(r) = \rho G_{ab}(r) \tag{5.35}$$

where ρ is the appropriate density. This latter function can be used to compute coordination numbers such as the number of neighbours in the first solvation shell *$N(r_1)$*, where r_1 is the position of the first minimum in *$g(r)$*.

Free Energy Landscapes

FELs represent a mapping of all possible conformations a molecule adopted during a simulation together with their corresponding energy typically reported as the Gibbs Free Energy. Classical MD simulations using all-atoms force fields and explicit solvent are able to capture the structure and motion of biomolecules in detail [van Gunsteren et al., 2006]. However, by generating a huge amount of data, these simulations pose a problem of analyzing obtained data sets. A key concept to interpret and illustrate biomolecular processes is provided by the molecules free energy landscape

$$\Delta G(r) = -k_B T ln P(r), \tag{5.36}$$

where P is the probability distribution of the molecular system along some (in general multi- dimensional) coordinate r. Common choices for the coordinate r include the fraction of native contacts, the radius of gyration, and the root mean square deviation of the molecule with respect to the native state. Characterized by its minima (which represent the metastable conformational states of the systems) and its barriers (which connect these states), the energy landscape allows to account for the pathways and their kinetics occurring in a biomolecular process [Onuchic, 1997, Wales, 2003].

A suitable representation of the energy landscape should reproduce the correct number, energy and location of the metastable states and barriers. Unfortunately, these crucial quantities often get lost when the energy landscape is projected on a low-dimensional subspace. It could happen that less minima are obtained in lower dimensions because several minima may overlap along the reduced coordinate *r1*. A loss of information about connectivity of barriers between minima might occur as well. This demonstrates an oversimplification of the free energy landscape of biomolecules. Hence, systematic dimensionality reduction methods were developed to overcome this limitation [Ichiye and Karplus, 1991, Noé et al., 2007, Rao, 2010].

In many cases it may be important to use internal coordinates such as the backbone dihedral angles ϕ, ψ in a PCA rather than the usual Cartesian coordinates. The basic idea of the dihedral angle principal component analysis (dPCA) is to perform a PCA on sin- and cos-transformed dihedral angles [Mu et al., 2005]

$$\begin{aligned} q_{2n-1} &= cos\varphi_n, \\ q_{2n} &= sin\varphi_n, \end{aligned} \tag{5.37}$$

where n = 1,...,N and N is the total number of peptide backbone and side-chain dihedral angles used in the analysis. Equation (4) represents a transformation from the space of dihedral angles $\{\phi_{\mathrm{n}}\}$ to a linear metric coordinate space (i.e., a vector space with the usual Euclidean distance), thus avoiding problems arising from the circularity of angular variables [Altis et al., 2007, Glimm, 1996].

It was concluded, that since a cartesian PCA can lead to artifacts due to the mixing of internal and overall motion, the use of internal coordinates such as the backbone dihedral angles in the dPCA would be necessary to obtain reliable results. However, time

constraints on the study did not allow a full exploration of this method. Hence, because this chapter does not claim to provide an in-depth study of the transitioning between conformational states, but rather looks at single conformations at local minima or maxima, it was seen as sufficient to build the free energy landscapes based upon the two reaction coordinates RMSD and Rgyr, while being aware of the limitations and possible incompleteness. Specifically, if structures close to each other in one apparent local minimum were to be pulled from trajectories, they might stem from two or multiple separate minima. For this reason, if a structure was pulled from a local minimum, only the uttermost energy minimum point was selected and by this assumed to represent the structure at the ultimate minimum of the whole trajectory, neglecting structures from neighbouring (possible hidden) minimas.

Computation of a representation of the free energy landscape (FEL) sampled by trajectories has also the purpose of selecting representatives to be used in docking calculations. To this end, Amber trajectories (.mdcrd and .prmtop) were concatenated using the Python package *MDAnalysis.Universe* to obtain gromacs input files (.gro and .xtc) for calulcating FELs.

The calculation was done using the GROMACS utility g_sham. g_gyrate, g_rmsf, g_rms were used to obtain .xvg files, which were then processed with the perl script sham.pl (*see Appendix*) to generate the proper input file with multiple columns, each representing a different coordinate for g_sham. The Gibbs Free Energy is calculated from a probability distribution together with a specific temperature. The correct temperature for the system, 323.15 K, was provided.

The g_sham output file, .xpm, was converted with the python script, xpm2txt.py (*see Appendix*), to a 3-column .txt file which was used to plot the FEL in Matlab. The FEL was inspected for Free Energy minima and maxima and coordinates were noted. The python script, get_timestamp.py (*see Appendix*), was used to obtain the corresponding timestamps of the coordinates from the g_sham .xvg input file by providing the coordinates. Those values were then used to extract representative structures from the concatenated trajectory using trjconv (gromacs).

In the next step, these structures were then used for docking using Autodock Vina.

AutoDock Vina

Molecular docking is a computational procedure that attempts to predict noncovalent binding of macromolecules or, more frequently, of a macromolecule (receptor) and a small molecule (ligand) efficiently, starting with their unbound structures, structures obtained from MD simulations, or homology modeling. The goal is to predict the bound conformations and their corresponding binding affinities. Docking programs generally use a scoring function, that can be seen as an attempt to approximate the standard chemical potentials of the system. The scoring function from AutoDock Vina [Trott and Olson, 2010] in its general functional form for the conformation-dependent part is defined as

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij}), \tag{5.38}$$

where the summation is over all of the pairs of atoms that can move relative to each other, normally excluding 1 - 4 interactions, i.e. atoms separated by 3 consecutive covalent bonds. Here, each atom i is assigned a type t_i, and a symmetric set of interaction functions f_{titj} of the interatomic distance r_{ij} should be defined.

This value can be seen as a sum of intermolecular and intramolecular contributions

$$c = c_{inter} + c_{intra}. \tag{5.39}$$

The derivation of this scoring function combines certain advantages of knowledge-based potentials and empirical scoring functions: it extracts empirical information from both the conformational preferences of the receptor-ligand complexes and the experimental affinity measurements.

The conformation-independent function g is defined as

$$g(c_{inter}) = \frac{c_{inter}}{1 + wN_{rot}} \tag{5.40}$$

where N_{rot} is the number of active rotatable bonds between heavy atoms in the ligand, and w is the associated weight.

AutoDock Vina [Trott and Olson, 2010] was used for molecular docking and virtual screening of ethanol into the binding pocket of representative structures of *Hv*ADH2 from FELs. AutoDockTools-1.5.6 was used to define the sampling space, which varied between structures. Grid dimensions varied between 10 and 30 number of points in x-, y,- z- dimension. Three docking runs were performed for each structure with a default exhaustiveness of 8

which determines how exhaustive the algorithm looks for the global minimum in all the binding modes. Output structures were visualised with PyMOL and visually inspected for their positioning and Grid dimensions were adjusted where appropriate. Structures showing binding modes with the lowest values for affinity (kcal/mol) were selected and visualised in PyMOL.

5.4 Results

5.4.1 Obtaining the Homology Model

Homology modelling was performed using Swiss-Model, Phyre2 and I-Tasser. Models were evaluated based on the preservation of conserved residues, to be specific, that is the catalytic triad coordinating the catalytic zinc, CYS-38, HIS-59 and ASP-155, as well as the coordination of the structural zinc by four cysteine residues, 89, 92, 95, 103. Evaluations were in a preliminary step only based on the sequence alignments obtained via the different modelling servers and overall conformation was assessed by visualising and comparing structures directly with PyMOL.

Swiss-Model

The model obtained with Swiss-Model is based on the template structure formaldehyde dismutase from *Pseudomonas putida* deposited in the Protein Database (PDB) as 2dph. Target and template structure show 27% sequence identity. The model preserved the catalytic triad, however, multiple regions showing holes and cavities on the surface are observed (figure 5.10), which are absent in the other two models. Given the observed tendency of halophilic proteins for a highly compact structure, that allows less flexibility, these regions are deemed as not favourable in a reliable homology model.

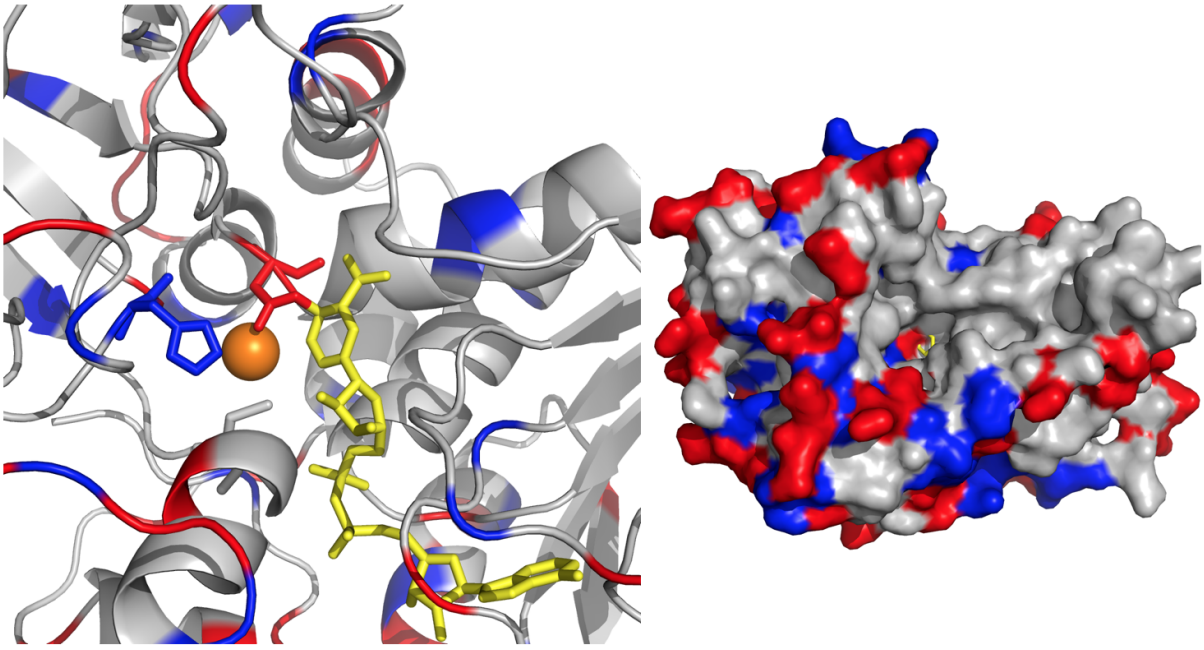


Figure 5.10: Homology model of *Hv*ADH2 obtained with Swiss-model. Left: backbone of ADH2 binding pocket, showing the active center containing the catalytic zinc as well as the catalytic triad, residues ASP-155 (red), HIS-59 (blue), SER-40 (silver), and the cofactor NADPH (yellow). Rright: surface representation of *Hv*ADH2, showing negatively charged residues in red and positively charged residues in blue. Two holes can be observed to the left and right of the structure.

Phyre2

The model obtained with Phyre2 is based on multiple templates, amongst them 2dph (formaldehyde dismutase from *Pseudomonas putida*) and 4cpd (alcohol dehydrogenase from *Thermus sp. ATN1*). The first structure proposed to build ensemble models on was 4gi2, a Crotonyl-CoA Carboxylase/Reductase from Methylobacterium extorquens. The model preserved the catalytic triad very poorly, since the ASP slips behind the cofactor. The overall conformation was better, than the one obtained with Swiss-Model, showing a more equal distribution of acidic resiudes across the surface and no holes were observed (figure 5.11).

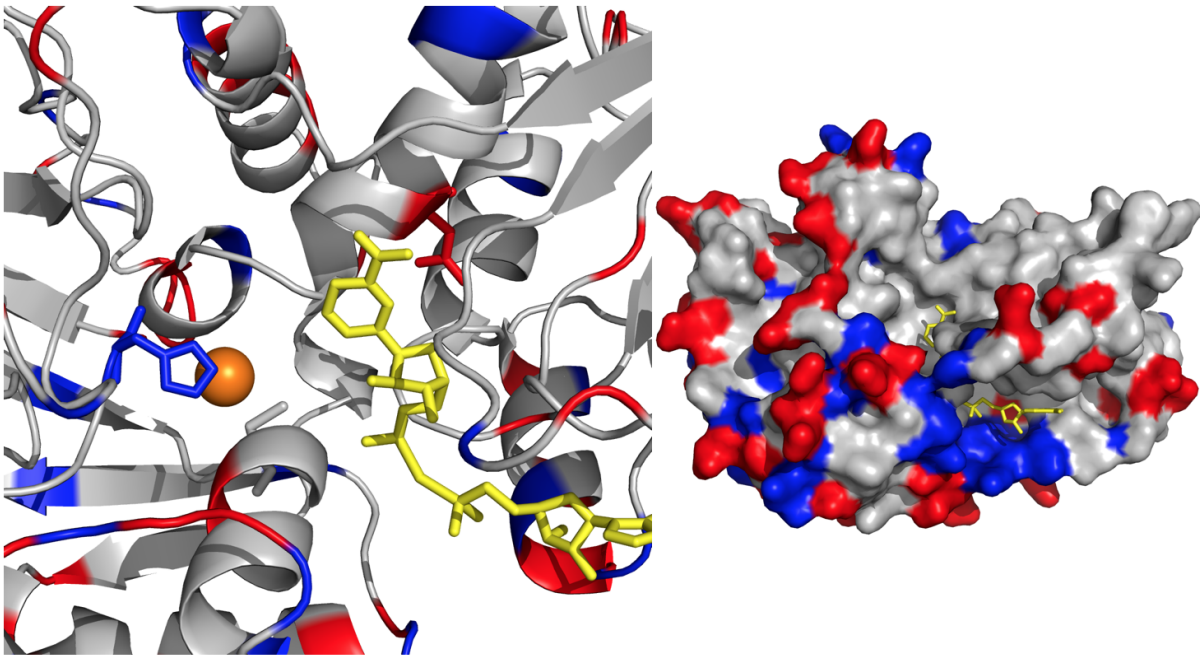


Figure 5.11: Homology model of *Hv*ADH2 obtained with Phyre2. Left: backbone of ADH2, showing the catalytic residues ASP-155 (red), HIS-59 (blue), SER-40 (silver) and the cofactor NADPH (yellow). Right: surface representation of *Hv*ADH2, showing negatively charged residues in red and positively charged residues in blue.

I-Tasser

The model obtained with I-Tasser is again based on multiple templates, however, most of them use different versions of an alcohol dehydrogenase from *Thermus sp. ATN1* (PDB: 4cpd) as template for threading. The model with highest C-score (1.62) was chosen. The catalytic triad is well preserved. The overall conformation shows a much smoother surface and more compact structure, which is to be expected from a halophilic protein (figure 5.12). In addition to this, the automatically calculated B-factor and the visualisation thereof confirmed the preservation of halophilic traits in the model.

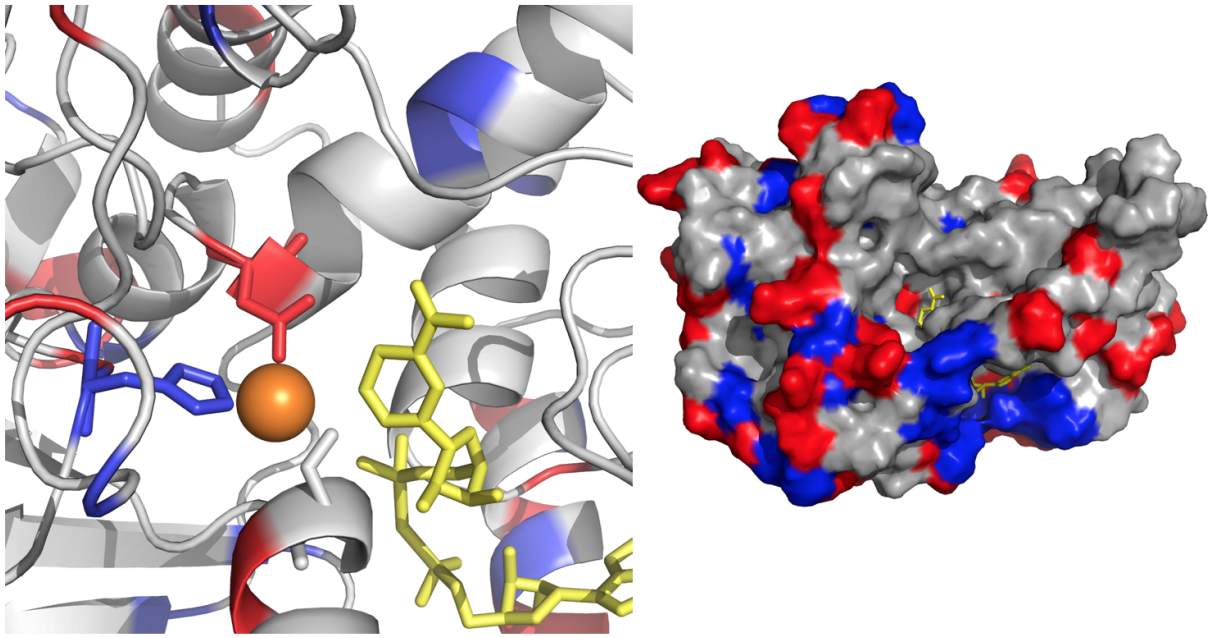


Figure 5.12: Homology model of *Hv*ADH2 obtained with I-TASSER. Top-left: backbone of ADH2, showing the catalytic residues ASP-155 (red), HIS-59 (blue), SER-40 (silver) and the cofactor NADPH (yellow). Top-right: surface representation of *Hv*ADH2, showing negatively charged residues in red and positively charged residues in blue.

Geometry Optimisation of the Tetramer

Following this preliminary evaluation, the model from I-Tasser was chosen to build the tetramer using Discovery Studios Visualiser (figure 5.14). Tetrameric 4cpd was used as a template to superimpose the monomeric model built by I-Tasser to obtain coordinate positions of the four homo-tetramer units (figure 5.13).

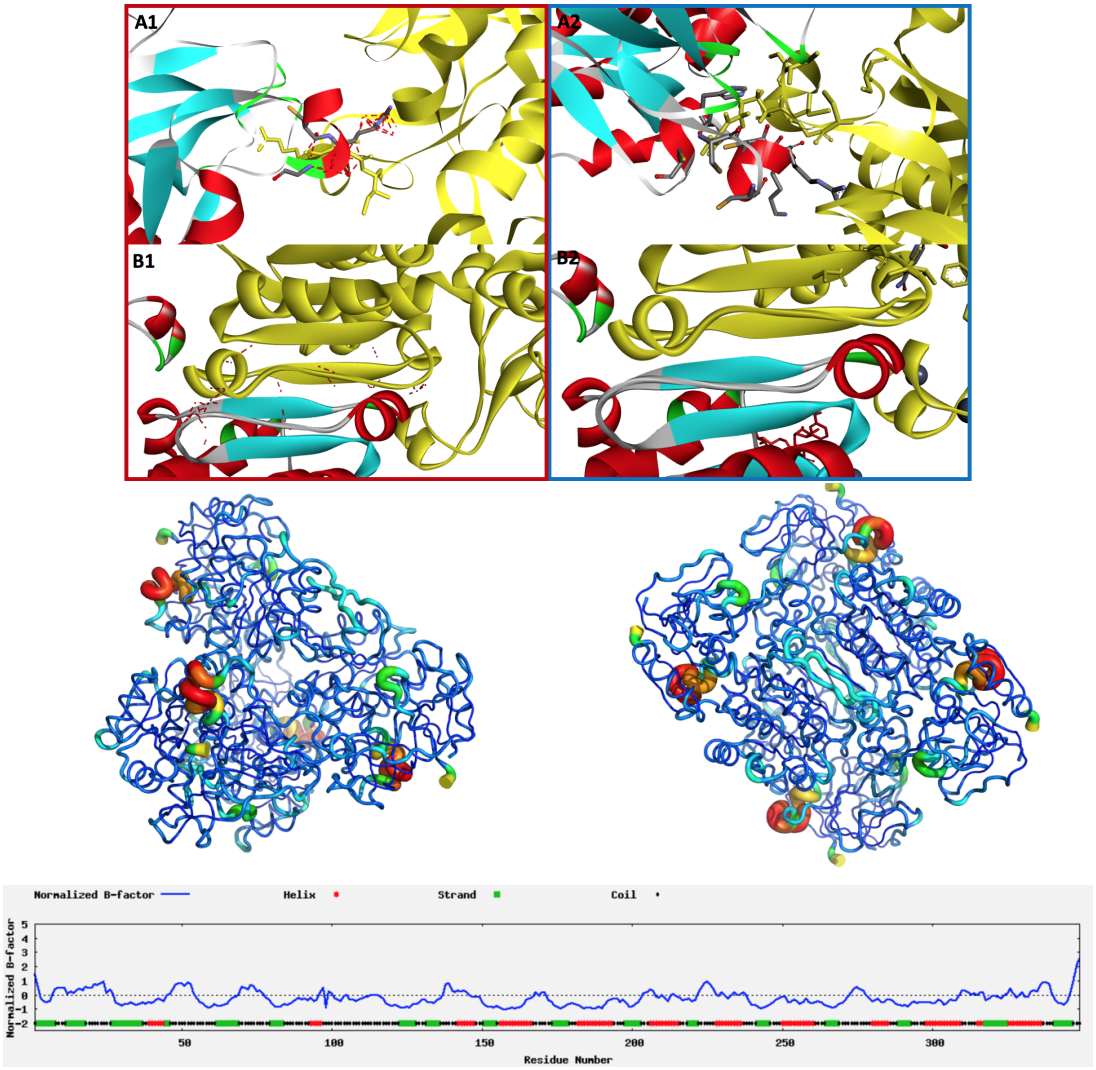


Figure 5.13: Top: Clashes (red dashed-lines) arise between model homo-tetramer units from superimposing the model onto the template structure. Most clashes are found at interfaces of the units, as seen in A1 (structural zinc) and B1. A manual and automated clean up of residues resolves the clashes and disentangles interfaces (A2 and B2). Middle: B-factor visualised after building of the tetramer. Bottom: B-factor distribution across residues as calculated from I-Tasser (bottom).

Verification of the Homology Model of the Tetramer

The optimised structure was then validate by PROCHECK. The Ramachandran plot produced with PROCHECK reported the dihedral angles to be at 74.7% in the most favoured region, 20.1% in the allowed region, 2.8% in the generously allowed region and 2.4% in the disallowed region and confirmed it as a reliable model (figure 5.15). The model was then used to create necessary topology files and was processed with Amber, as described, to run MD simulations. From 200 ns MD time the root mean sqaure deviation (RMSD) was calculated for backbone atoms to validate the model and its stability during the simulation.



Figure 5.14: Secondary structure of final model of the *Hv*ADH2 homo-tetramer shown in cartoon representation.

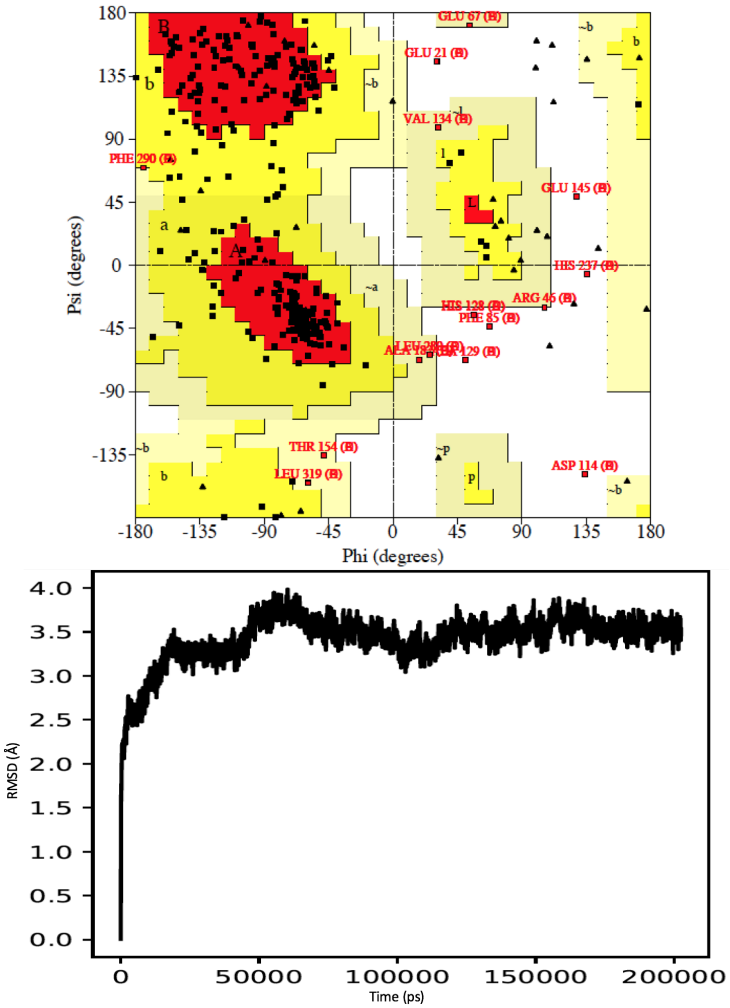


Figure 5.15: Top: Ramachandran plot produced with PROCHECK, dihedral angles are found at 74.7% in the most favoured region, 20.1% in the allowed region, 2.8% in the generously allowed region and 2.4%. Bottom: RMSD of backbone atoms of a 200 ns trajectory, showing stability commencing past 60 ns.

5.4.2 Coordination of ions

RDFs are a standard metric employed in simulation studies to characterize the organization of one species with respect to another. The values of the RDFs presented are ratios of the ion density compared to the bulk ion density within increasing radii of spherical shells about the protein.

Coordination of K^+ and Cl^- ions to the protein surface

In the first instance RDFs between negatively charged surface residues (ASP, GLU) and potassium or chloride were calculated and plotted. Four distinct association distances were apparent across all simulations for K^+ at 2.7, 3.5, 4.4-5.8 and 6-8 Å, with the first and third being the major peaks corresponding to directly coordinated K^+ and hydrogen bonded fully hydrated K^+, also known as solvent separated ion pairs, respectively (figure 5.16). Comparison between different ILs showed for all of them a clear increase in association of K^+ to acidic residues, when compared to the reference without IL. Analysis of the RDFs for Cl^- showed significantly lower populations of chloride anions associating to hydration-shells for all IL systems.

The strongest shift was observed for hydroxy, hydrophobic and sulfo ILs, while imidazolium ILs showed the least effect. Phospho based ILs induced moderate shifts, except for **ILI** ($[P_{4,4,4,4}][Cl]$), which induced a much greater shift towards decreasing association of chloride ions.

Figure 5.17 presents the results of hydroxy-functionalised ILs and demonstrates the trend for all IL systems. RDFs of all other IL-systems were moved to the appendix.

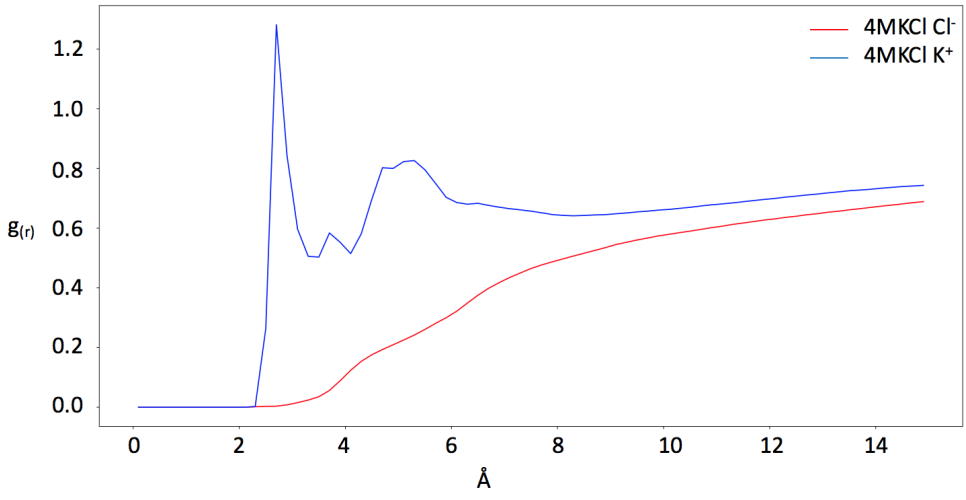


Figure 5.16: RDF of potassium (K^+) (blue) and chloride (Cl^-) (red). First peak (blue) corresponds to directly coordinated K^+, third peak correspondes to hydrogen bonded first fully hydrated K^+. This simulated system contained no ionic liquid ions.

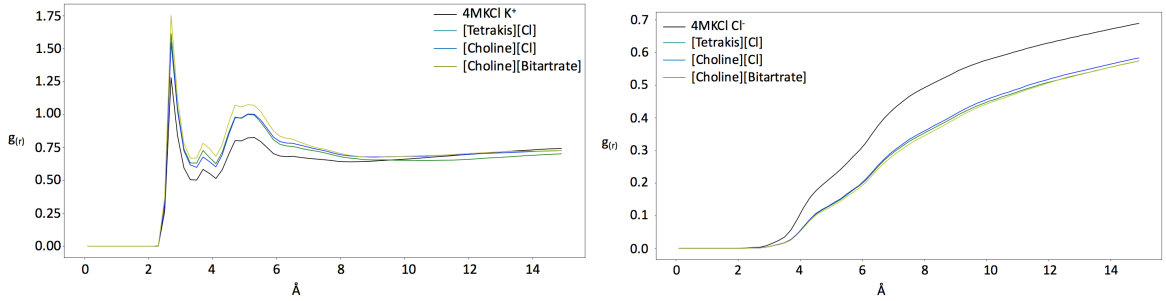


Figure 5.17: RDF of K^+ and Cl^- ions in 4 M KCl in the presence of 0.15 M Hydroxy-functionalised ionic liquids ILB[Tetrakis][Chloride], ILC[Choline][Chloride] and ILD[Choline][Bitartrate].

Coordination of ionic liquid ions to the protein surface

In the second instance RDFs between negatively charged surface residues (ASP, GLU) and IL cations or anions were calculated and plotted. As a generel trend, IL anions tend to associate closer and in greater numbers to the protein surface, than cations do. However, this depends on the combination of cation and anion.

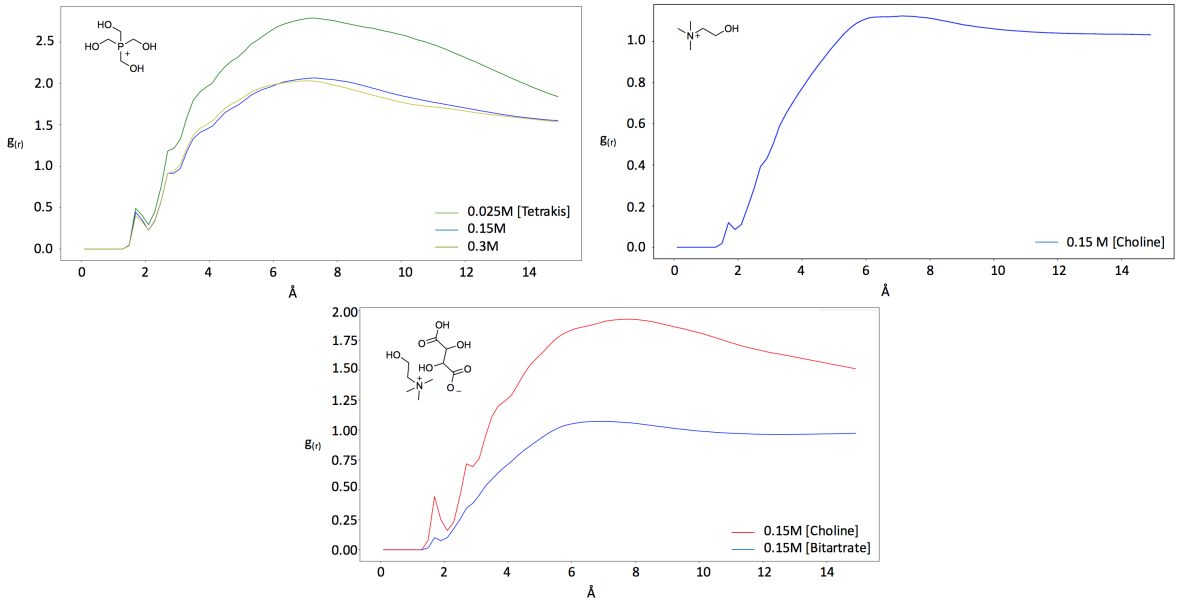


Figure 5.18: RDFs of anions and cations of 0.15 M Hydroxy-functionalised ionic liquids ILB[Tetrakis][Chloride], ILC[Choline][Chloride] and ILD[Choline][Bitartrate] in 4 M KCl.

Figure 5.18 shows the RDFs of hydroxy-functionalised ILs. Regardless of the position

of the hydroxy group on both the cation or anion, a direct interaction with the acidic residues was observed that is similar to potassium, albeit much less frequent, indicated by a g(r) of 0.5 for [Tetrakis] with 4 OH-groups, 0.4 for [Bitartrate] with 3 OH-groups and 0.1 for [Choline] with 1 OH-group. Increase in intensity of *g(r)* corresponds to the increasing numbers of OH-groups on the ions. The association distance indicates a much closer coordination of the hydroxy groups to acidic residues with 1.8 Å when compared to potassium with 2.7 Å. In comparison to potassium, however, there is no defined second peak, but a continuing increase peaking between 6 and 8 Å.

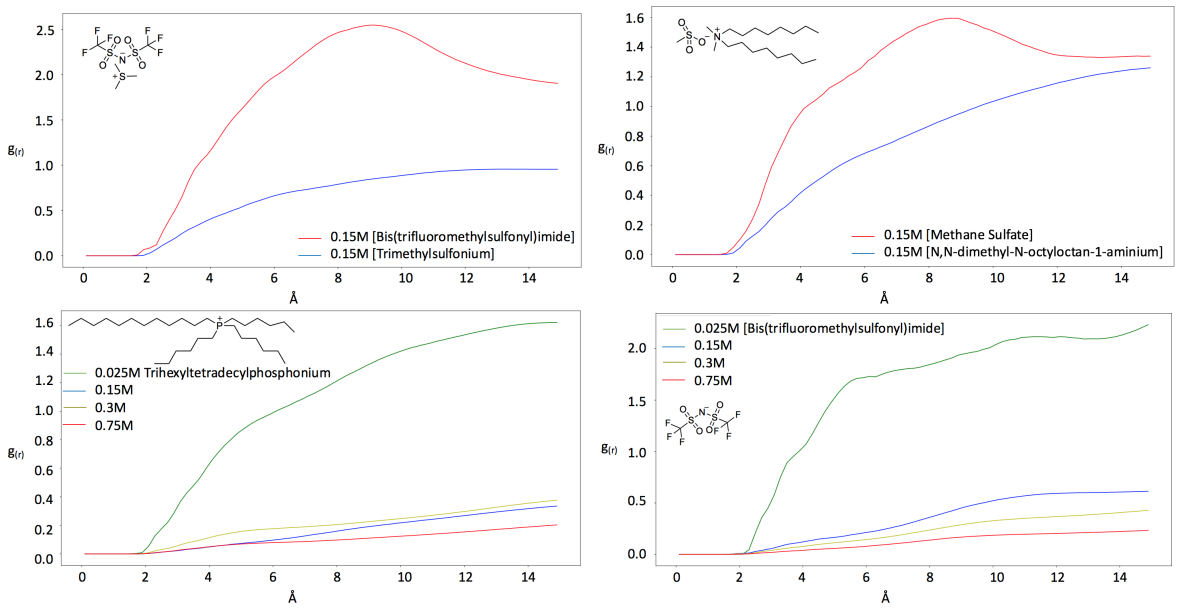


Figure 5.19: RDFs of anions and cations of 0.15 M Hydrophobic ionic liquids F[Me_3S][NTF2], H[$N_{1,1,8,8}$][$MeSO_3$] and K[$P_{6,6,6,14}$][NTF2] in 4 M KCl.

Figure 5.19 shows the RDFs of hydrophobic ILs. [Me3S] in **ILF** shows the same behaviour as in **ILA** (paired with [OSO_3Me]) and **ILM** ([I]) (figure 5.20), associating loosely starting at 2 Å. This indicates a direct interaction with very few ions, and a slow increase to a maximum *g(r)* of 0.8, indicating a gradient with higher density farther away from the surface. NTf_2 shows a similar behaviour for **ILF** and **ILK** (0.025 M) in the first hydration shell of the protein with an indicated peak at around 4 Å with a *g(r)* of 1. The second hydration layer is indicated by a second peak at around 6 Å with a *g(r)* of 2.5 for **ILF** and 1.75 for **ILK**. In the outer hydration layers, NTf_2 ions paired with Me_3S ions decreases, while NTf_2 paired with [$P_{6,6,6,14}$] continuous to increase. This behaviour of **ILK** is only

observed at a concentration of 0.025 M. At all other concentrations, both, anion and cation of **ILK**, show a very low *g(r)*, that slowly increases with distance from the protein surface.

The anion [$MeSO_3$] of **ILH** shows a similar behaviour to the anion [$MeOSO_3$] of **ILA** with two indicated peaks at around 4 Å and 8 Å, with a maximum in approximately the third hydration layer, which was slightly higher for **ILH**. The hydrophobic cation of **ILH** [$N_{1,1,8,8}$] shows an exponential increase with increasing distance from the surface, indicating a gradient.

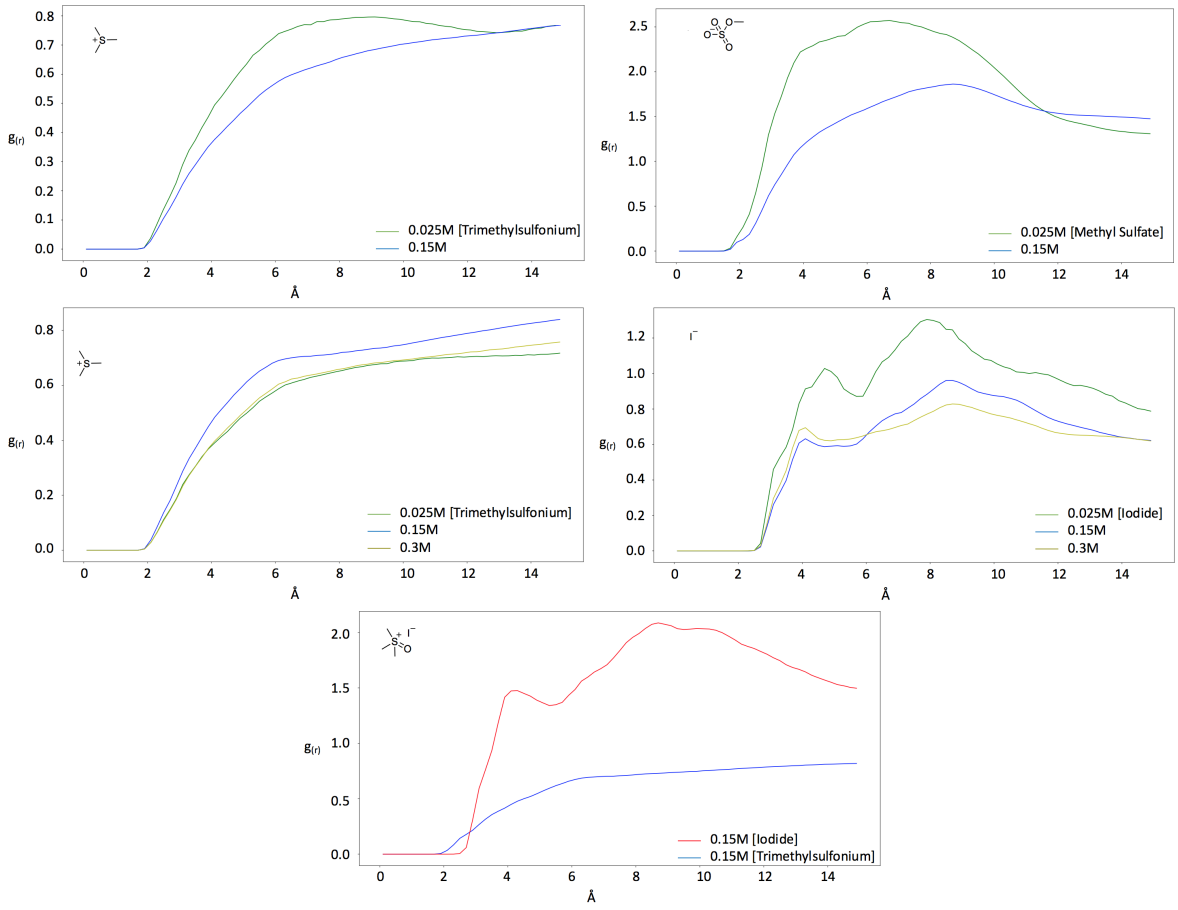


Figure 5.20: RDFs of anions and cations of 0.15 M Sulfo ionic liquids A[Me_3S][OSO_3Me], L[Me_3SO][I] and M[Me_3S][I] in 4 M KCl.

Figure 5.20 shows the RDFs of sulfo-based ILs. The anionic iodide ion shows similar trends in both, **ILL** and **ILM**, with multiple peaks. The trends look similar to the behaviour of potassium. However, a less direct coordination (above 2 Å) and multiple H-bonded hydration shell interactions are observable at 4 Å, 6.2 Å, 8 Å, 9 Åand 11 Å.

However, peak intensity is greater for **ILM** with a *g(r)* of 3.5, where the iodide was paired with [Me_3S], in comparison to ILL (2.0), where the iodide was paired with [Me_3SO]. This hints at a more compact, rigid cage structure around the surface, that is less subject to fluctuations in case of **ILM**, hence, ions possibly linger and lead to a higher intensity. Behaviour of [Me_3S] in **ILA** and **ILM** is similar, however, there is a concentration dependent anion influenced difference. While [Iodide] in **ILM** leads to a less frequent presence of [Me_3S] at the protein surface at 0.025 M compared to 0.15 M, in **ILA** [$MeOSO_3$] leads to an increase in association of [Me_3S] to acidic residues with increasing concentration. The anion [$MeOSO_3$] (**ILA**) is more frequently found in protein surface proximity compared to [Iodide] in **ILL** ([Me_3SO]), but less frequent in comparison to [Iodide] in **ILM** ([Me_3S]).

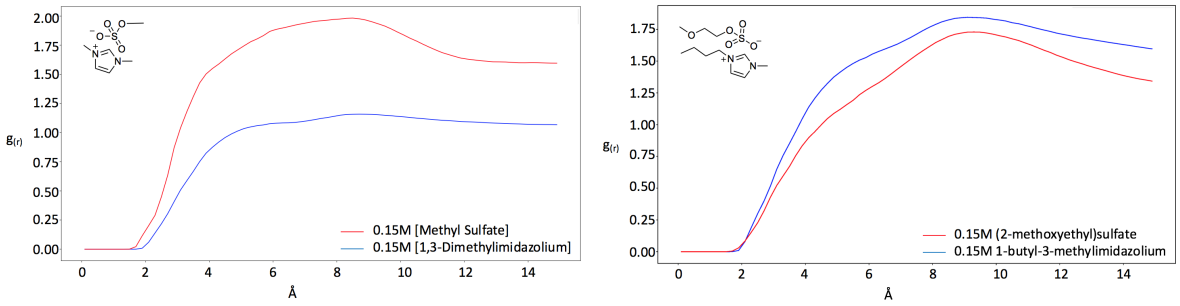


Figure 5.21: RDFs of anions and cations of 0.15 M Imidazolium ionic liquids E[EMIM][OSO_3Me] and G[BMIM][$MeOEtSO_4$] in 4 M KCl.

Figure 5.21 shows the RDFs of imidazolium-based ILs. Comparison of RDFs of **ILE** and **ILG** shows the impact of the slightly longer alkyl chain and ether functionalization of **ILG** ions. While the anion [$MeOSO_3$] of **ILE** is able to coordinate closer to the surface and shows a more frequent coordination (*g(r)* of 2), the anion [$MeOEtSO_4$] cannot since it is obstructed by the slightly longer alkyl tail, which is reflected in a *g(r)* of 1.6. Since anions seem to show in general a higher preference to coordinate closer to the potassium ions surrounding the acidic residues, [$MeOSO_3$] of **ILE** shows the more typical behaviour of excluding its paired cation to the outer hydration layers, inferred by a much lower *g(r)* of 1 for EMIM, when compared to BMIM (*g(r)* of 1.5). BMIM on the other hand appears to be entangled with the matching alkyl tail of the [$MeOEtSO_4$] anion, since both ions of **ILG** show a highly similar coordination with a maximum *g(r)* of 1.75 at around 8 Å,

however, the cation has a slightly more frequent coordination, than the anion.
The anion [$MeOSO_3$] in **ILE**, despite its above pointed out close and frequent coordination, shows a lower *g(r)* of 2, when compared to the same ion in **ILA**, where it is paired with [Me_3S] with a *g(r)* of 2.5. Interestingly, the corresponding cation in **ILA** shows an equally low coordination and confirms the exclusion of the cation by the anion.

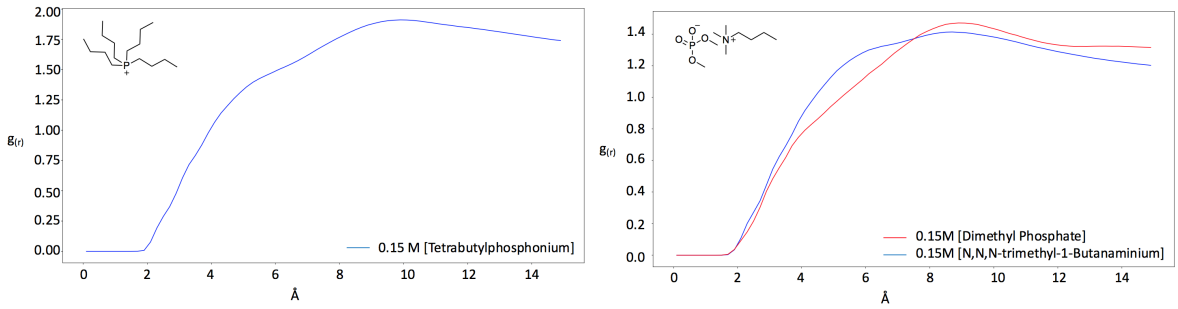


Figure 5.22: RDFs of anions and cations of 0.15 M Phospho ionic liquids I[$P_{4,4,4,4}$][Cl] and J[$(MeO)_2OPO$] in 4 M KCl.

Figure 5.22 shows the RDFs of phosphorus-based ILs. Alkyl chains of the phosphonium cation of **ILI** [$P_{4,4,4,4}$] appear to be no hindrance for the ion to increasingly accumulate in the first, second and following hydration layers with a maximum association of 1.9 at 10 Å, displaying the frequently observed gradient towards increase in association with decreasing distance. The phosphoric acid anion in **ILJ**, interestingly shows a similar behaviour to [$MeSO_3$] with two indicated peaks and a maximum *g(r)* of 1.5, and is with this less similar to [$MeOSO_3$], despite higher similarity in molecular structural arrangement. Behaviour of the cationic [$N_{1,1,1,4}$] of **ILJ** shows in comparison to its hydroxy-functionalised counterpart [Choline] of **ILC** or **ILD**, that indeed the OH-group is responsible for the direct coordination of the [Choline] cation, since the peak observed for **ILC** and **ILD** at 1.8 Å is absent for **ILJ**.

5.4.3 Structural changes of *Hv*ADH2 in ionic liquid mixtures

In the first instance, root mean square deviations (RMSDs) were plotted, and are shown in figure 5.23 for **ILA** [Me_3S][OSO_3Me], **ILB** [Tetrakis][Chloride], **ILK** [$P_{6,6,6,14}$][NTf_2] and **ILM** [Me_3S][I]. RMSDs for all other ions were moved to the appendix. A distinct

elevation of the RMSD can be observed for the lowest concentration (0.025 M) of all ILs, while the moderate concentration of 0.3 M shows a near native RMSD.

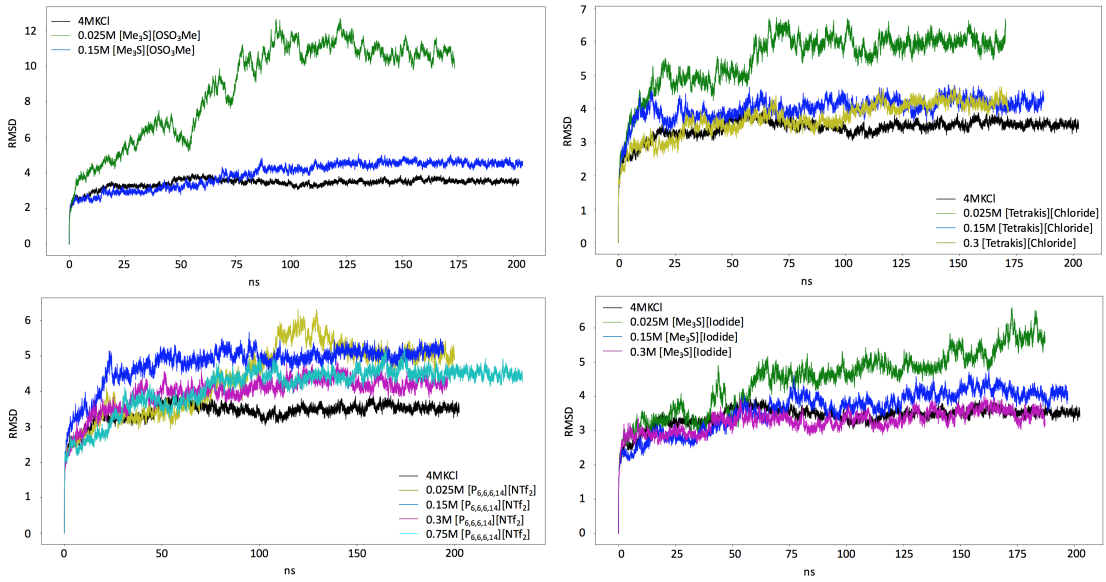


Figure 5.23: Root mean square deviations of ionic liquids A, B, K and M. For all ionic liquid systems shown an increased RMSD is observed for the lower concentrations.

Calculation of RMSD and radius of gyration (of the protein) allows plotting of free energy landscapes of the trajectories, and by using time-stamps, protein conformations can be obtained from specific wells of Gibbs free energy states. The free energy landscape of ADH2 under native conditions explores a relatively small configurational space and shows a straight descent towards the lowest energy minimum (figure 5.24), unlike most other investigated systems in this study (see *Appendix Section Free Energy Landscapes at 0.15 M Ionic liquid*).

Since unligated enzymes often exist in multiple conformations, and the binding of substrates and/or coenzymes elicit changes in conformation, it is unclear which structural reorganisations of the enzyme-substrate complexes occur by looking only at one structure of the trajectory. To understand the chronological order of the multiple conformational transitions along the reaction coordinate and their relationship to enzyme catalysis it would be necessary to investigate the structure from every well. In order to approach this in a meaningful way for this thesis, the native system (4M KCl) was investigated based on one structure from a high energy well at the start of the trajectory and the lowest en-

ergy structure towards the end of the trajectory. It was found that the enzyme converges from a closed to an open configuration, but it remains unclear which transition stages are explored in the conformational space, and whether the trajectory was long enough to allow for the enzyme to gain back a closed conformation in between.

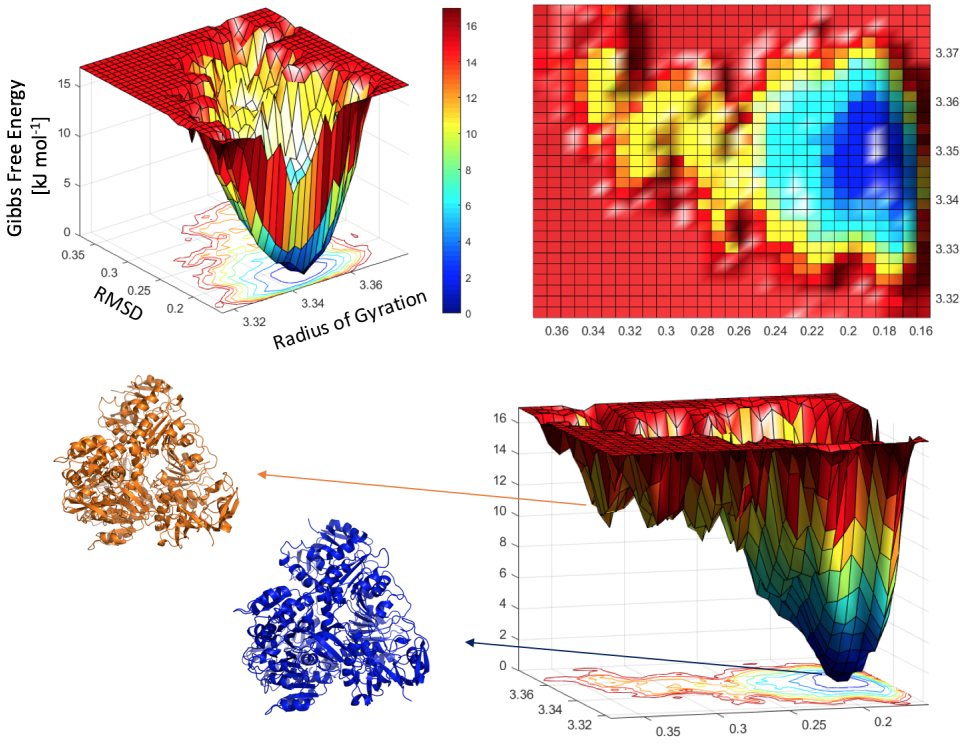


Figure 5.24: Different views on the FEL of ADH2 in 4 M KCl, showing a relatively smooth and symmetrical conformational landscape. Structures correspond to the starting conformation of higher Gibbs Free Energy (orange) and conformation of lowest energy (blue). Structures of higher energy do not diverge greatly from structures of lower energy (within 1.5 Å) and show only slight changes in loop regions towards a more compact configuration. There are no high energy barriers between the starting conformation and the conformation of lowest energy.

The conformation found at the higher energy levels at the start of the simulation is the closed configuration (shown in yellow in figure 5.25) while the conformation found at the bottom of the energy well of the ADH2 trajectory is in open configuration (shown in cyan in figure 5.25). Hence, the open structure is the more relaxed structure. The overlay of the structures in figure 5.25 demonstrates the flexibility of the N-terminus, visible as the first beta-sheet, that gets 'moved along' in the closed conformation. A flexible N-

terminus is often observed in halophilic proteins and in this case majorly participates in switching between open and closed conformation. There is an indicated interaction between residues HIS-49 and HIS-273 and it is likely, that they serve as gatekeepers for conveying the closed conformation. The most flexible regions besides the 8 N-terminal residues, are the sequences 111-HGDE, where the structure winds back on itself towards the N-terminus after capturing the structural zinc and YRGH-49E, a region between catalytic residues CYS-38 and HIS-59.

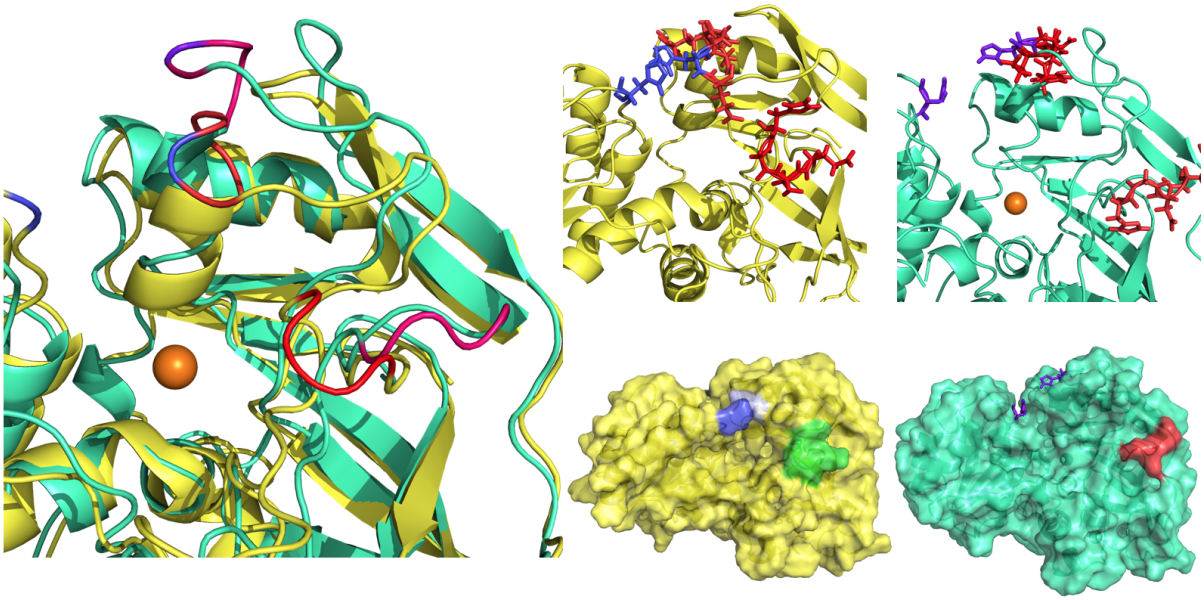


Figure 5.25: Comparison of ADH2 conformations in open(cyan) and closed(yellow) states from the trajectory in 4M KCl. Red or green coloured residues and surface regions show secondary structural elements, that are most flexible during conformational change (sequences 111-HGDE and YRGH-49E). HIS-49 and HIS-273 are coloured blue.

Based on their outstanding effect on ADH2 activity in comparison to the other IL systems, the systems 0.025M **ILA** $[Me_3S][OSO_3Me]$, 0.3 M **ILK** $[P_{6,6,6,14}][NTf_2]$, both having an activity increasing effect, and 0.15 M **ILD** [Choline][Bitartrate], having a severely detrimental effect on activity, were selected for closer inspection of their 3D fold from conformations of lowest energy. 3D-folds are shown in figure 5.26. Only the one from **ILA** can be categorised as conclusively open. The structures of the systems **ILK** and **ILD** were partially in accordance with the open and partially in accordance with the closed confirmation. However, free energy values were with around 2 kJ/mol, almost as low as the

native system with approximately 0 kJ/mol. Distinctions between open and closed conformation were made based on the position of the N-terminus, where **ILA** showed a clear concordance with the open conformation, while **ILD** and **ILK** showed concordance with the closed conformation. The position of the mobile elements 111-HGDE and YRGH-49E were considerably altered for all systems shown, however they were more in congruence with either the open or closed conformation. The structure from the **ILA** system showed the least divergence from the native system, followed by **ILK** and significant alteration is observed for the **ILD** system, where the cofactor binding domain resembles more the open conformation, than the closed one.

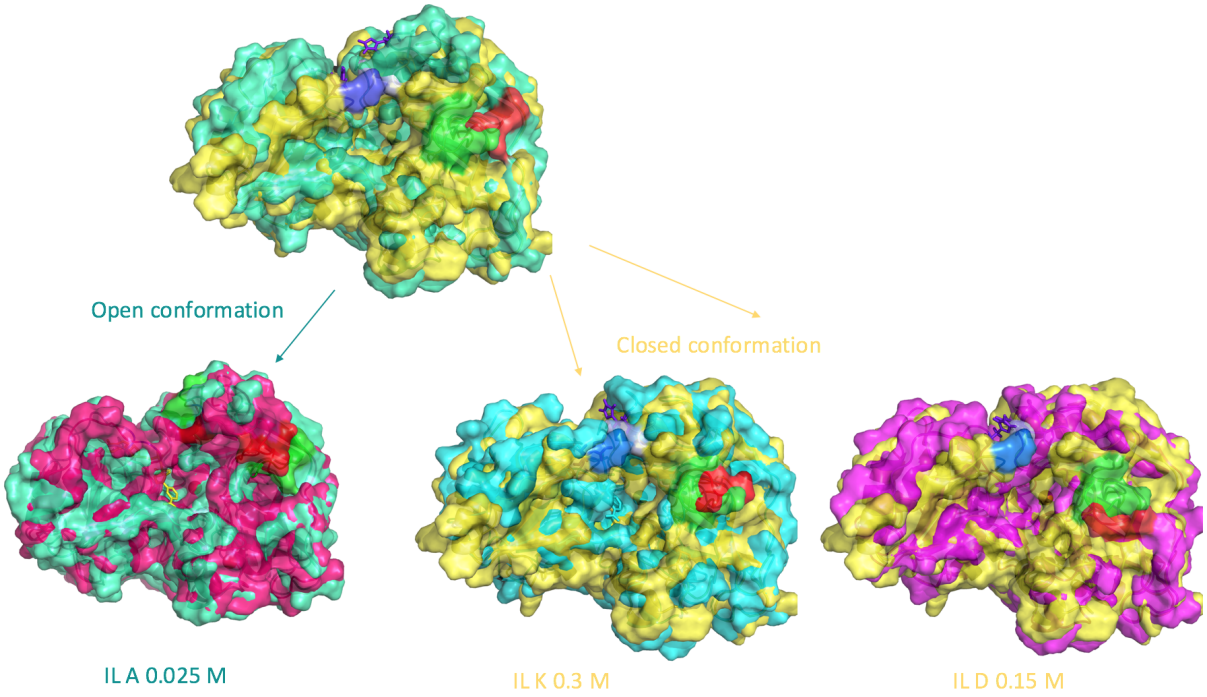


Figure 5.26: Top: Overlay of ADH2 conformations in open(cyan) and closed(yellow) states, Bottom (left to right): 0.025 M ILA $[Me_3S][OSO_3Me]$ (pink) overlayed with the open conformation of ADH2 in 4 M KCl, 0.3 M ILK $[P_{6,6,6,14}][NTf_2]$ (turquoise) and 0.15 M ILD [Choline][Bitartrate] (purple) overlayed with the closed conformation of ADH2 in 4 M KCl.

Conformational transitioning between energy wells was investigated for systems 0.025 M **ILA** $[Me_3S][OSO_3Me]$ and 0.3 M **ILK** $[P_{6,6,6,14}][NTf_2]$. Structures pulled from trajectories demonstrate a clear impact of ionic liquids on the tertiary structure of *Hv*ADH2 over the course of the simulation. Structures were most impaired at the lowest energy level.

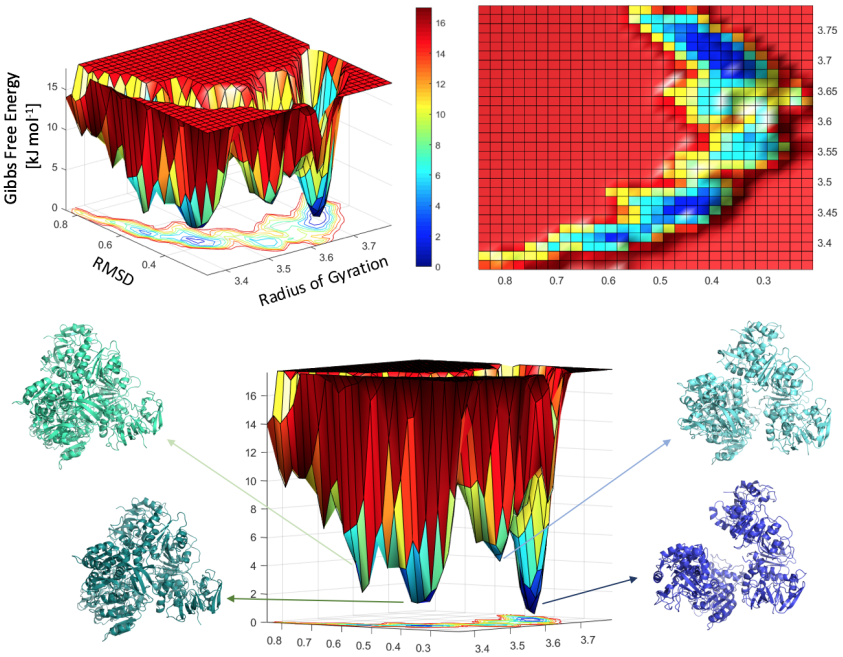


Figure 5.27: Comparison of structures from ADH2 in 0.025 M ILA $[Me_3S][OSO_3Me]$. A bimodal distribution of two distinct energy wells is observed, with two more troughs leading to them.

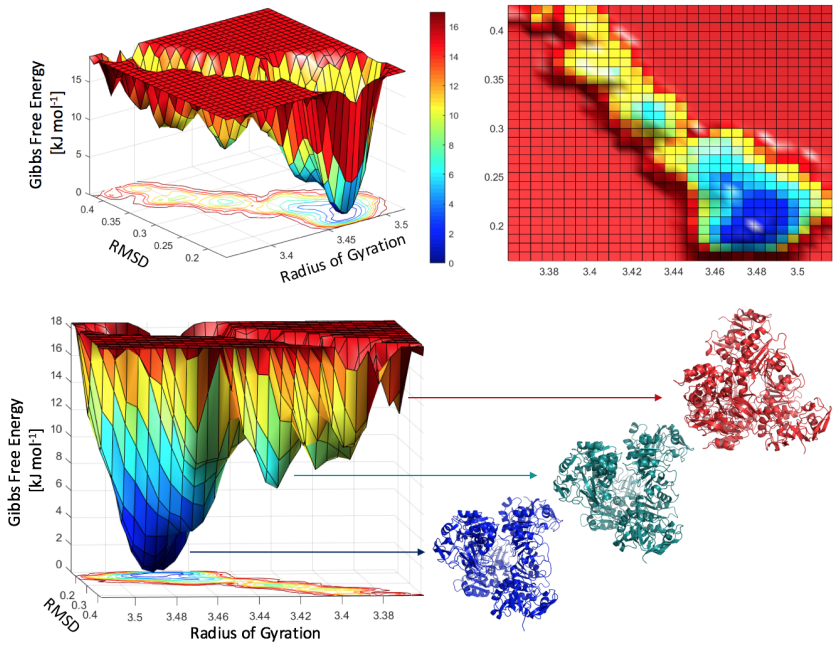


Figure 5.28: FEL of 0.3 M ILK $[P_{6,6,6,14}][NTf_2]$: the conformation progresses into multiple troughs until it reaches the energy minimum during which the tertiary structure changes.

A uniquely bistable distribution in the FEL of the 0.025 M **ILA** system was observed. This means the system has two stable equilibrium states and visualisation of structures showed the enzyme in open conformation at both bottoms of these wells (figure 5.27). The transitioning of system 0.3 M **ILK** was investigated based on three structures and showed that at a middle energy level of around 8 kJ/mol an open conformation was assumed, while at the lowest energy level a confirmation resembling a closed conformation was observed (figure 5.28). This means the structure transitions from a closed to an open to a closed conformation.

5.4.4 Docking into Lowest Energy Conformations

Lowest energy conformations of systems 0.025 M **ILA** $[Me_3S][OSO_3Me]$, 0.3 M **ILK** $[P_{6,6,6,14}][NTf_2]$ and 0.15 M **ILD** [Choline][Bitartrate] were further investigated by docking the substrate (S)-1-PheOH in the catalytic centre.

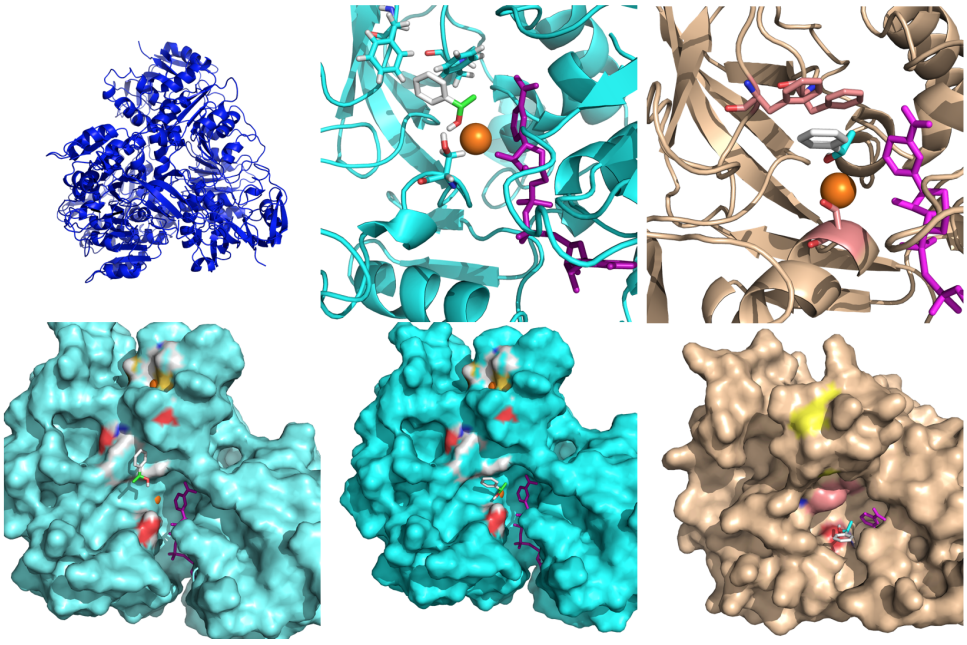


Figure 5.29: Top left: lowest energy structure of 4 M KCl trajectory. Top middle: active centre of lowest energy structure (open conformation), showing docking of the substrate (S)-1-PheOH, SER40, PHE-85 and PHE-108. Top right: active centre of closed conformation. Bottom left and middle: catalyic centre and structural zinc domain of open conformation. Bottom right: catalytic center and structural zinc domain of closed conformation.

In figure 5.29 residue SER-40 is directing the substrate (S)-1-PheOH via the alcohol group close to the active center, while PHE-85 and PHE-108 can interact with the phenyl-ring via π-π-stacking with the substrate. Sampling docking space for the open and closed conformations, showed the substrate is being coordinated via the phenyl-rings into the pocket, where it is further coordinated via SER-40. The closing of the cleft is visible in the closed conformation and shows the 'cupping' of the substrate.

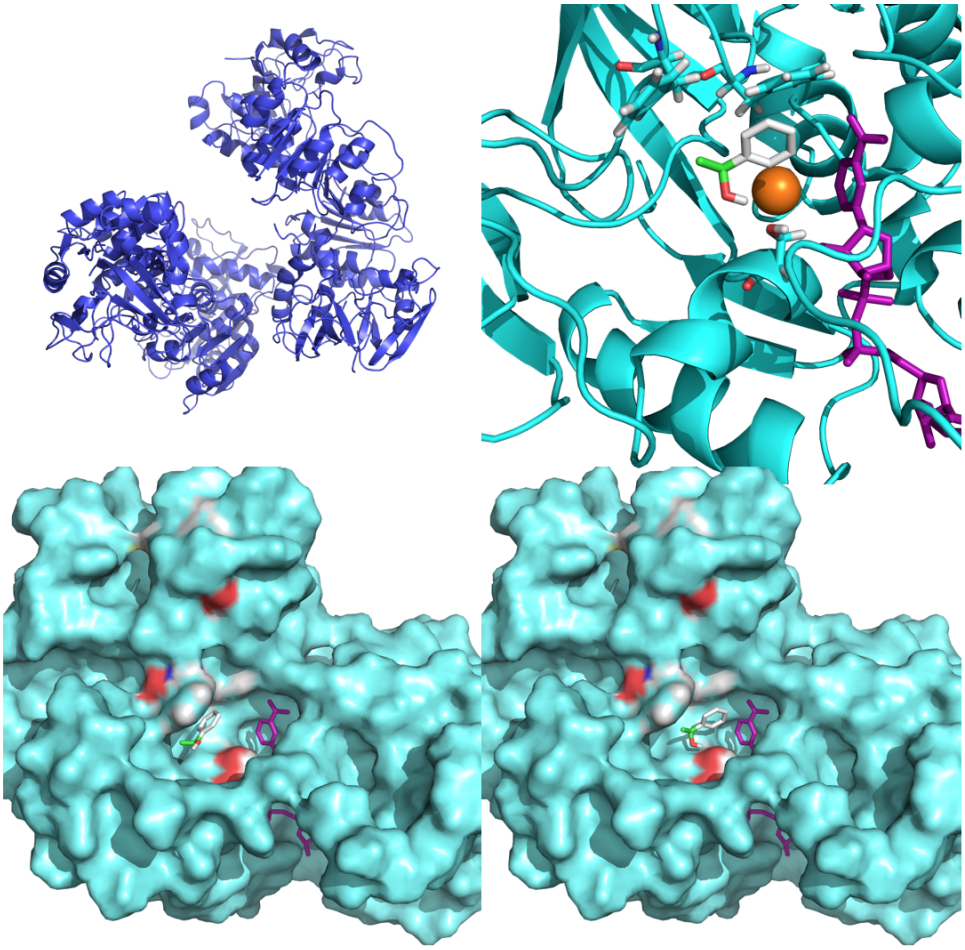


Figure 5.30: Top left: lowest energy structure pulled from 0.025 M ILA $[Me_3S][OSO_3Me]$ trajectory. Top right: the catalytic centre showing SER-40, PHE-85 and PHE-108. Bottom: Catalytic centre and structural zinc domain.

Figure 5.30 shows the coordination of the substrate in the distorted conformation of the system 0.025 M **ILA**. Both Phenylalanine residues, which serve as coordinators for phenyl-substrates, are found in altered positions compared to the 4 M KCl conformation. The structural zinc and its four coordinating Cys residues gets 'swallowed' and is not visible

in this configuration. The catalytic cleft is more open than closed.

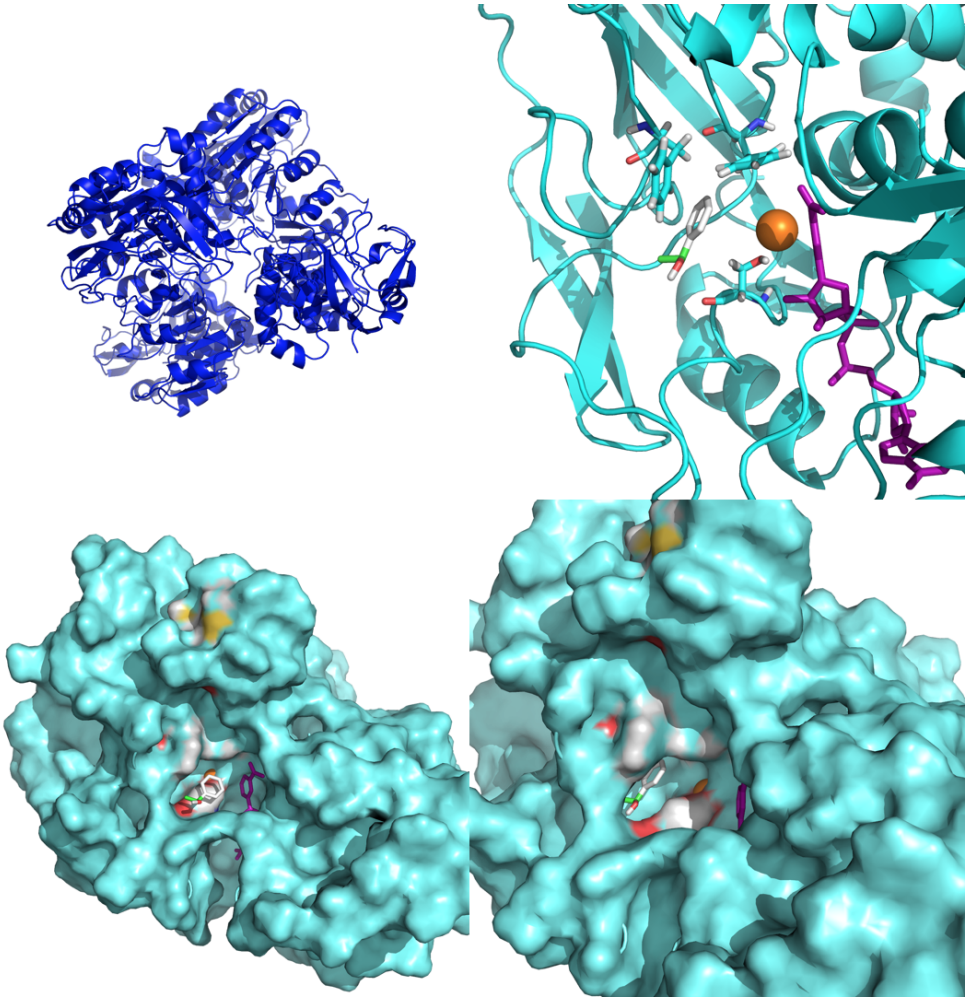


Figure 5.31: Top left: lowest energy structure pulled from 0.15 M ILD [Choline][Bitartrate] trajectory. Top right: the catalytic centre showing SER-40, PHE-85 and PHE-108. Bottom: Catalytic centre and structural zinc domain.

In figure 5.31 a novel pocket appears beneath the coordinating phenylalanine residues in the conformation of the system 0.15 M **ILD** and the substrate becomes essentially trapped in opposing direction to the cofactor. The rest of the tertiary structure diverges as well strongly compared to the 4 M KCl conformation. A channel from the structural zinc leading to the active centre appears.

Figure 5.32 shows system 0.3 M **ILK**. Despite the tertiary structures similarity to the closed conformation, the position of the coordinating residues is more in accordance to

the open conformation of the native enzyme. The catalytic zinc appears to move more towards the inside of the structure and becomes slightly obstructed. A novel channel sliding from the structural zinc to the catalytic zinc opens, but is different from the one observed in **ILD** (figure 5.31).

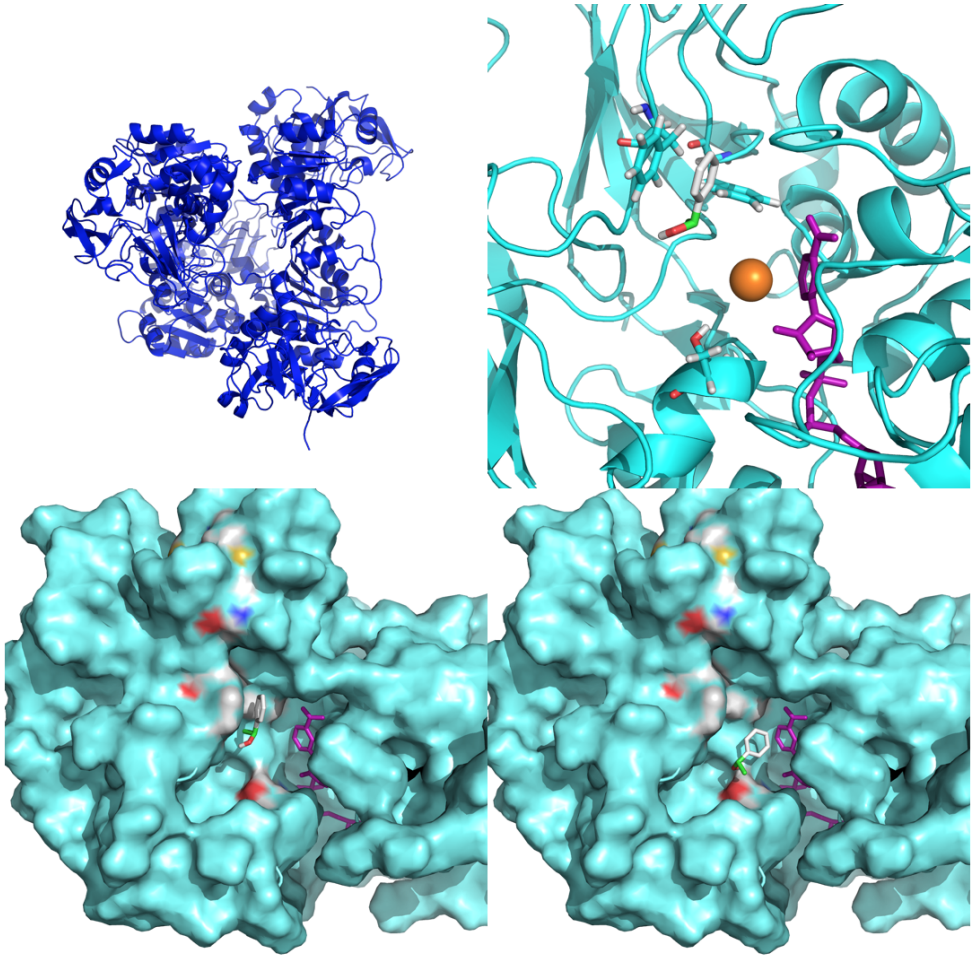


Figure 5.32: Top left: lowest energy structure pulled from 0.3 M ILK $[P_{6,6,6,14}][NTf_2]$ trajectory. Top right: the catalytic centre showing SER-40, PHE-85 and PHE-108. Bottom: Catalytic centre and structural zinc domain.

5.5 Discussion

Threading was shown to be superior to the other two methods for building the homology

model and lead to the preservation of the catalytic centre, which in the following can be used for docking. Structure compactness was shown to match halophilic characteristics, inferred from visualisation of the B-factor of the model. In a previous study a reliable homology model for the docking of flurbiprofenol was built using SwissModel [Cassidy et al., 2017], however, multiple less accurate regions located on the protein surface were reported for this model. Since in this study, the overall tertiary structure was a main focus, the model based on I-Tasser was more appropriate.

From the modelling and subsequent analysis of trajectories obtained RDFs for potassium and chloride ions in combination with the RDFs for ionic liquid cations and anions, a general picture can be drawn (figure 5.33). Compared to the buffer system all ILs led to an increased and closer coordination of potassium at the protein surface. This indicates, that the higher proportion of coordinating potassium ions essentially become trapped between the protein surface and the closely coordinating ionic liquid anions, that intrude into the first hydration shell. This intrusion is indicated by the association distance of the anions, which starts at about 2 Å and is with this similar to the association distance of potassium. The hydroxy-functionalised ILs show exceptional behaviour in this regard. Not only do their association distances exhibit a distinct first peak, that starts below 2 Å, showing a direct interaction with the acidic residues, this lower distance also indicates a much closer coordination of hydroxy-functionalised ion species to acidic residues, when compared to potassium.

ILs with ions with medium alkyl chain tails, including imidazolium-based ILs and phospho ILs tend towards a similar close coordination for both ion species (cations and anions alike) to the protein surface. A pattern of alternating patches of IL anions and cations can be envisioned. This behaviour is very much unlike the one exhibited by hydrophobic ILs, where a clear preference of closer coordination of anions over cations can be inferred from RDFs. In general, the second and third hydration layers appear to be occupied by IL anions and an increasing gradient of cations in the outer hydration layers, reflected in increasing coordination up to a distance of 8 -12 Å.

For all systems the chloride ions appear to mingle in with the cations at the outer hydration shells, and seem to be displaced from their original site of association around

hydrated potassium ions to the far side of the ionic liquid belt, that forms around the protein.

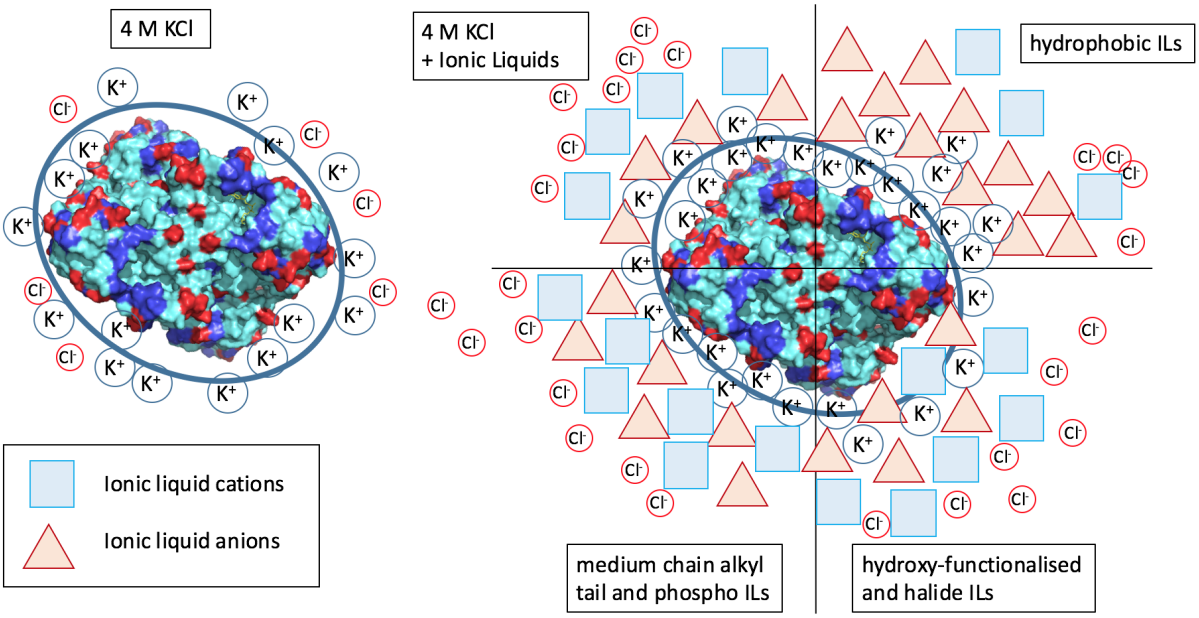


Figure 5.33: Left: Hydration shell occupation by K^+ and Cl^- ions Right: A picture showing general trends for the IL ions classified which are inferred from the RDFs observed.

Closer coordination of IL anions over cations for enzymes with engineered surface charge towards halotolerance has been observed in literature [Burney et al., 2015] and is with this in congruence with the results of this thesis. Coordination frequency, reflected in the radius of gyration, of anions over cations or cations over anions varies between systems in this study, albeit in general a tendency of anions over cations is observed. Reports on imidazolium-IL systems for mesophilic proteins, conversely, suggested a high binding affinity of the imidazolium towards the *Bacillus subtilis* Lipase-A surface via hydrophobic or π-π stacking interactions [Pramanik et al., 2019] and *C.albicans* Lipase-B [Klähn et al., 2011].

Further agreement with literature can be concluded for the shift of RDFs of potassium ions towards increased coordination to halophilic protein surfaces and shifts of RDFs of anions towards decreased coordination, when compared with mesophilic proteins, as suggested by a study on engineering a mesophilic anhydrase towards increased negative surface charge [Warden et al., 2015]. Their study only deployed common salts, whereby sodium was the sole cation paired with organic salts. Interestingly, they found multiple

coordination peaks indicated by distinct coordination distances for their common salt anions, amongst them chloride. Those distinct peaks in RDF profiles are lacking from most anions used in this study, with the exception of iodide and the hydroxy-functionalised ions, but including chloride ions of the system containing only 4 M KCl. However, their protein was engineered towards halotolerance, not halophilicity. It might be that these extra acidic residues on ADH2 lead to a more equally distributed expulsion of chloride ions, since potassium ions screen the surface. On the other hand, the studies counter cation was sodium. Sodium with a smaller circumference might allow more chloride ions to penetrate the first hydration shell. This might also explain the preference of potassium over sodium by halophilic proteins from H.volcanii [Ortega et al., 2011, Timpson et al., 2013]. Further to this it was found that water molecules in the hydration shell of potassium are orientationally more disordered than those hydrating sodium ions and are inclined to orient their dipole moments tangentially to the hydration sphere, while chloride ions form hydrogen-bonded bridges with water molecules and are readily accommodated into the H-bond network of water [Mancinelli et al., 2007]. This would further explain their preferential seclusion to the bulk solvent.

To better explain and discuss the following correlations an overview on results regarding RMSD, radius of gyration, IL-RDFs, activity and trajectory convergence is presented in figure 5.34.

Conversely, a higher RMSD at lower concentrations of ionic liquid was observed for the systems **A, B, K and M**. In agreement with this are reports from literature, where a bell-shaped dependency of enzyme structural deviations following water-content, i.e. increasing ionic liquid concentration, was reported for several enzyme-IL systems [Haberler et al., 2011, Laszlo and Compton, 2001, Latif et al., 2014, Micae?lo and Soares, 2008]. Findings in these studies suggested a stripping off of non-localised water molecules by high charge density ions, mostly anions, which are preferentially dispersed in the bulk at lower concentrations. This is especially true for hydrophilic ionic liquids, since they usually consist of small anions, which can be easily implemented in the three-dimensional water network of hydrogen bonds [Bernardes et al., 2011, Schröder et al., 2007]. While these aspects might be most influential for mesophilic proteins, RDFs in this study sug-

gest a different picture for halophilic proteins. All ionic liquid anions, with the exception of chloride, show preferential coordination to the protein surface which is even higher at lower concentrations. An explanation for this different behaviour of coordination are the potassium ions which coordinate to the surface in the first instance and thereby attract high charge density anions. Following from this, the higher RMSD for lower concentrations for the halophilic ADH2 cannot be explained by striping off essential water molecules by anions dispersed in the bulk, but rather by anions coordinating close to the protein surface and displacing the water molecules.

	0.025M	0.15M	0.3M	0.6/0.75M	IL
RMSD	Up to 12 Å	Up to 4 Å	N/A	N/A	A
Rgyr	3.45 & 3.7	3.39	N/A	N/A	
RDF	0.8 & 2.5 at 6 Å	0.6 & 1 at 6 Å	N/A	N/A	
Convergence	Bifurcated	straight	N/A	N/A	
Activity	>100%	~50%	~50%	~30%	
RMSD	Up to 7 Å	Up to 4 Å	Up to 4 Å	N/A	B
Rgyr	3.5	3.44	3.49	N/A	
RDF	2.8 at 6 Å	2 at 6 Å	2 at 6 Å	N/A	
Convergence	obstructed	obstructed	straight	N/A	
activity	~0%	~0%	N/A	N/A	
RMSD	Up to 6 Å	Up to 5 Å	Up to 4 Å	Up to 4.5 Å	K
Rgyr	1 & 1.8 at 6 Å	0.2 & 0.2 at 6 Å	0.2 & 0.2 at 6 Å	0.2 & 0.2 at 6 Å	
RDF	3.4	3.48	3.44		
Convergence	N/A	obstructed	straight	obstructed	
activity	~160%	~140%	~240%	~0%	
RMSD	Up to 6 Å	Up to 4 Å	Up to 3.5 Å	N/A	M
Rgyr	3.43	3.46	3.38	N/A	
RDF	0.6 & 2 at 6 Å	0.7 & 1.5 at 6 Å	0.6 & 1.5 at 6 Å	N/A	
Convergence	obstructed	straight	straight	N/A	
activity	>100%	~60%	~40%	~5%	
RMSD	Up to 3.5 Å	4 M KCl			
Rgyr	3.35				
RDF	N/A				
Convergence	straight				
activity	100%				

Figure 5.34: overview of RMSD, radius of gyration, IL-RDFs, activity and trajectory convergence.

A study investigating anion exchange between cations proposed a simple three step process of a system containing [OMIM][Cl]x[NTf_2]1-x. Whereby the anion needs to (I.) break its current H-bond, (II.) diffuse to the new hydrogen bonding location and (III.) displace

the incumbent anion [Avula et al., 2018]. This hypothesis was stipulated after simulating the experimental observation, that the net ion charge (NIC) of the cation varied quasi-linearly with the ratio of anion composition, while the NIC of the anion [NTf_2] remained unchanged [Villar-Garcia et al., 2014]. The higher the number of chloride ions interacting with the cation, the greater is the extent of charge transfer and hence the lower is the NIC of the cation. This means the charge of the cation is directly related to the composition of its first coordination shell [Villar-Garcia et al., 2014]. A similar replacement was found for [BMIM][Cl]x[OTf]x, whereby the stronger interacting chloride ions remove the more charge-delocalised OTf ions from the strongest H-bond donor sites [Gehrke et al., 2017]. This helps explain why chloride ions in this study preferentially confine to a distance from the protein surface which is similar to IL-cations and also helps explain the competition of IL-anions for potassium at the protein surface and the consequential displacement of essential water molecules. The IL anions find a stronger interaction partner in high charge density potassiums and locate to the surface, where they are subjected to accelerated HB-dynamics in this confined space as the H-bonds are repeatedly formed between the same ion pairs by switching back and forth between different H-bond donor/acceptor sites of the same counter-ion, which will depend on the anions number of acceptor site [Gehrke et al., 2017, Strate et al., 2018]. For both mesophilic and halophilic proteins the native ability of surface residues to mitigate between the fast bulk water dynamics and overall slower protein surface dynamics, which is based on tightly bound active water molecules, [Sterpone et al., 2012] is lost, due to the removal of dynamics-mitigating structural H_2O molecules from the protein surface into the bulk.

Based on the observation that the radius of gyration of ADH2 is highest at the lowest concentrations for **ILA, B, K and M**, indicating a slowing down or loss in compactness of the tertiary structure (or both), and IL anion coordination is simultaneously highest at lower concentrations (0.025 M), when compared to higher concentrations (0.15 - 0.3 M), it can be argued, that upon an increase of IL concentration a competition between water and anions for the HBs with cations starts to mitigate the HB dynamics on the protein surface and leads to a deceleration of the IL anion HB interactions. This slowdown of HB dynamics in the system could be more akin to protein surface native HB-dynamics and thereby enable the protein residues to behave more similar to native conditions. This

would be reflected in a similar RMSD between higher IL concentrations and the native system, which is indeed the case.

Visualisation of structures pulled from trajectories demonstrated a clear impact of ionic liquids on the tertiary structure of *Hv*ADH2. Visualisation of the structures from the native system showed a clear distinction between open and closed conformations, whilst this was not the case for the investigated systems 0.025 M **ILA**, 0.15 M **ILD** and 0.3 M **ILK**. For these systems, impairment of secondary structural elements which were shown to be important for conferring activity via tunnelling-ready states in a thermophilic *Bacillus stearothermophilus* alcohol dehydrogenase [Klinman and Kohen, 2013], was observed. These protein conformational sub-states were linked to tunnelling efficiency and peptide sequences that are mobile during switching between sub-states were identified. Results of this study show that specifically these secondary elements, namely, the 111-HGDE clamp and the YRGH-49E motif (with respect to residue numbers in the case of HvADH2) show altered conformations between the four systems (native, 0.025 M **ILA**, 0.15 M **ILD** and 0.3 M **ILK**) and thereby imply a divergent effect on activity.

Concluding the findings from the FELs, it can be theorised that the cation [Me_3S present in combination with [OSO_3Me] (**ILA**) or [I] (**ILM**) guides the transitioning towards lowest energy state in a 'straight' manner with regards to RMSD and radius of gyration progressing equally, unless paired with a hydrophobic ion, such as [NTF_2] (**ILF**). This suggests, that a slow and steady descend towards lowest energy minimum is a requirement for preserving activity, since these two systems were found to maintain activity best, with the exception of hydrophobic **ILK**. This 'straight' descent holds also true for **ILK**, where the concentration of 0.3 M showed the highest activity and shows the straightest descent compared to the other concentrations. For **ILB** a concentration of 0.3 M shows a strikingly 'straight' FEL descent, but was unfortunately never assayed, since lower concentrations showed no activity during experimental characterisation (see *Appendix, Section Free Energy Landscapes - Comparison of Concentrations*).

Interestingly, the molarity of 0.3 M correlates roughly with the molar fraction of water of 0.9 for **ILK** and **ILB**, which, as mentioned in *Chapter 1.5.1*, has been identified as a 'magic point' [Haberler et al., 2011], where a change in the organisation of the ions

and water takes place due to a transition from dipolar screening to charge screening of electrostatic forces. Based on the observation of the straightest descent and its correlation with the 'magic point', it can be theorised, that around this mol fraction a transitioning of the conformation is guided by forces, which impact the radius of gyration and RMSD in a balanced manner. This means, protein stabilising forces on the interatomic level leading to lower RMSDs and protein structure disrupting forces acting on the secondary and tertiary structural elements, which compact the structure, leading to increased radius of gyration, have to be balanced and are most likely to be balanced at this mole fraction.

All simulations started from unequilibrated Packmol/xLEaP input, including IL containing systems. An alternative approach, which was not explored due to the long and time consuming runs, would have been to perform an equilibration and production run with the system containing only the protein and 4 M KCl and returning this equilibrated system as input to Packmol/xLEaP to build systems containing the ionic liquid ions. So that runs start from a fully equilibrated ADH2 structure at its minimum.

5.6 Conclusion and Outlook

MD simulations revealed structural changes of *Hv*ADH2 residues 111-HGDE and YRGH-49E which are involved in determining the open and closed conformation. Structural rearrangements depend on the ionic liquid system and comparison of functional to non-functional *Hv*ADH2 conformations might allow for the deduction of suitable residue substitutions to obtain certain active conformations and allow further engineering of ADH2 stability for process conditions.

It was revealed that anion species of ionic liquids preferentially coordinate to halophilic protein surfaces in the presence of high concentrations of potassium chloride. This effect is even greater at lower concentrations of ionic liquid, and with this, proves to be different from mesophilic proteins where preferential exclusion of anions to the bulk is observed at lower concentrations.

An interesting correlation at a molar concentration of around 0.3 M was found between

protein activity and straight descent towards energy minimum in the free energy landscapes. This correlation was already observed in the literature [Haberler et al., 2011]. However, results are too few to back a hypothesis. Further work should be done and could involve the investigation of ADH2 activity in 0.3 M **ILB** [Tetrakis][Chloride] where a straight descent was observed from the simulations, but the system was never assayed experimentally.

In this present work only the coordination between acidic acid surface residues and ions has been analysed. Further coordination investigations should be carried out between all surface residues, except acidic ones, and ions and basic residues and ions. Future work should also evaluate with which atoms ions preferentially coordinate to the surface, e.g. with hydrophobic tails or charged head groups. Special focus of attention should be brought to the coordination of water molecules. Solvent Accessible Surface Area (SASA) calculations should be carried out. Especially coordination of water around protein surface residues which coordinate closely with IL ions could be targeted to infer displacement of water molecules. This might give first insights into HB dynamics. However, a more sophisticated water model and timesteps of finer resolution would be necessary to further investigate the nature of the hydrated potassiums influence on the broader water network. Special regions of interest for investigating water molecule coordination should also focus on the structural elements that were found to be involved in tunneling. Further regions of interest are the channel that changed and opened up more for conformations in IL system 0.3 M **ILK** and 0.15 M **ILD** between the structural zinc and binding pocket.

Chapter 6

Microcapsule Formation and Enzyme Encapsulation

6.1 Background

The aim of the final chapter is a proof of concept, which concerns the applicability of the ionic liquid-enzyme system, $[P_{6,6,6,14}][NTf_2]$-*Hv*ADH2 for encapsulation.

The system [BMIM][NTf_2] had been demonstrated previously to allow for its encapsulation [Guarnido et al., 2019]. However, the difference in the cation and its long alkyl chains are considered a challange. Microcapsule formation in the presence of surfactants had been found to lead to a large variation in the amount of encapsulated substrate dependent upon the type of surfactant and its concentration [Duquemin and Nixon, 1986]. Since the alkyl tails of $[P_{6,6,6,14}]$ could act as a quasi surfactant, resulting Microcapsules might not fully encapsulate the enzyme.

The method used for the proof of concept and according to which microcapsule formation proceeded is based on a solvent extraction and evaporation fabrication mechanism. The original method for evaporation fabrication was established by Loxley and Vincent [Loxley and Vincent, 1998], however the Marr group adapted the protocol and replaced the inner organic phase with an ionic liquid [Guarnido et al., 2019]. This method was chosen

based on the simplicity of materials and commodities required. However, the published example applied much larger quantities of materials. A challenge hereby was to downsize the approach for applicability to small-scale enzyme quantities. The other challenge presented the maintenance of the activity of the encapsulated enzyme in PVA and aqueous conditions that were used to prepare the microcapsules. The method is presented in figure 6.1. The polymer shell material, which is insoluble in the core solvent, in this case an ionic liquid, will form emulsified droplets, when presented with a small amount of surfactant in its solubilising agent (dichloromethane). These droplets form templates for the later established microcapsules, which assemble through evaporation of the volatile organic solvent.

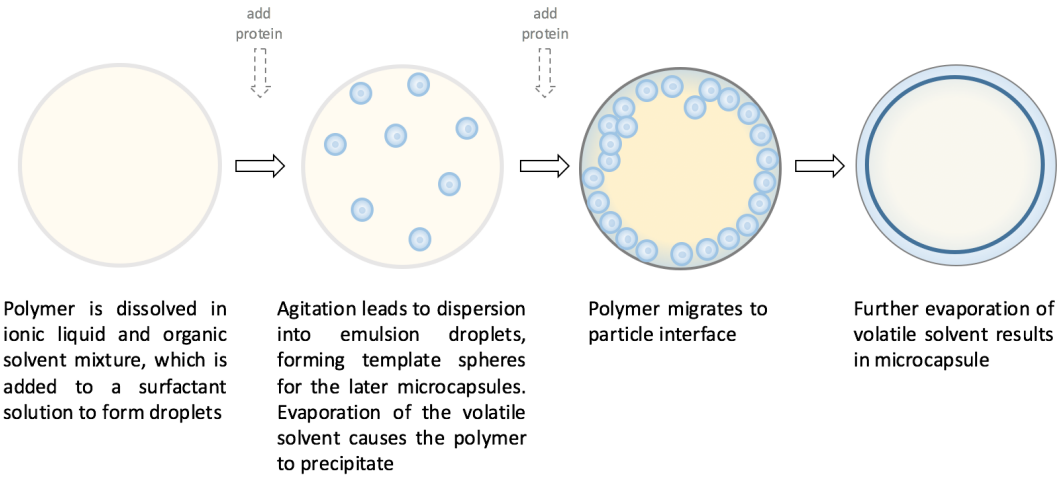


Figure 6.1: Solvent extraction and evaporation fabrication mechanism: The two components (Polystyrene and ionic liquid) are mixed with a volatile solvent (DCM) to form a single organic phase. This mixture is emulsified in water containing a small amount of surfactant (poly vinyl alcohol). The resulting droplets are templates for the microcapsules. As the volatile solvent evaporates the polymer precipitates and forms a shell surrounding the ionic liquid core. Figure was reproduced from the work by the Marr group [Guarnido et al., 2019].

Microreactors were characterised with regards to their size and shape by electron microscopy and optical microscopy as well as regarding enzymatic activity.

6.2 Material and Methods

Chemicals and Material

Polystyrene MW 192.000 and 280.000, (*Sigma*, #182427 and #43002)
Mowiol Poly vinyl alcohol (PVA) (96% hydrolysed, MW 85.000-124.000 g/mol), (*Sigma*,#11773-250G)
Dichloromethane (DCM) (*ReAgent*)
$[P_{6,6,6,14}][NTF_2]$, (*Fluka*)
*Hv*ADH2
Coomassie Blue (1mg dissolved in 1ml Methanol (50% [v/v], Glacial acetic acid (10% [v/v], H2O (40%))

Experimental

A 20% (w/v) PVA solution was prepared by dissolution of 0.24 g PVA in 12 ml water at 100 ℃ and shaking, before cooling to RT. The organic phase consisting of polystyrene, IL and organic solvent was prepared in an Eppendorf tube. To this end, 0.025 g polystyrene was dissolved in 0.55 g DCM and 0.035 g IL before this mixture was added dropwise into 0.76 ml PVA solution in a 5 ml Eppendorf tube and the resulting mixture was agitated at 6000 rpm for different time lengths (1h - 16h) on a shaker. Following this, 30 μl protein solution (0.5 mg/ml) was added to 1.2 ml PVA solution, which was then added to the agitated mixture. The emulsion was left in the fumehood to evaporate DCM for 3 hrs. Microcapsules were washed multiple times with SDW before measurements commenced. Capsule size can be varied by varying the emulsification technique, whereby smaller capsules can be achieved with more vigorous stirring or ultrasound.
In case of some experiments, protein was stained with Coomassie previous to encapsulation, to be able to visualise its presence in the microcapsules.

6.2.1 Scanning Electron Microscopy

A FEI Quanta 650 Environmental Scanning Electron Microscope (ESEM) was used in Low Vacuum mode and Environmental mode. The instrument was equipped with a Schottky Field (thermal FEG) and can operate with accelerating voltages from 10 eV to 15 keV with probe currents of 4.5 micro-amps.
Low vacuum mode was used for imaging of non-conductive samples (microcapsules), that can not be coated with metal. Pressures ranging from 10 to 133 Pa can be achieved using water vapor and the anode used in this case was a low vacuum Secondary Electron (LFD) Detector. In Environmental mode pressures ranging from 133 to 2600 Pa can be achieved, also using water vapour and the anode used was a gaseous secondary electron detector (GSED). Further used for ESEM was a Peltier cooling stage, that allows imaging of hydrated specimens (up to 100 % humidity) at a temperature of 2 ℃ to prevent damage from the electron beam.

6.2.2 Optical Microscopy

A Zeiss Stemi 2000 stereo microscope was used to observe microcapsules at magnifications 4x, 10x and 20x. Plan Achromat corrections used were 0.1 (4x), 0.25 (10x) and 0.4 (20x). A mounted camera was used for imaging.

6.2.3 Spectrophotometry

An Epoch 2 Microplate Spectrophotometer (BioTek) was used with 96-well UV transparent plates to monitor the increase in absorption (slope) at 340 nm of NADPH. For calculation see *Chapter 4, Methods*. Measurements were carried out on microcapsule mixtures (20 time points) and from supernatants of microcapsule mixtures (6 times points). For assayed microcapsule mixtures, the microcapsules were resuspended carefully by inverting the tube, then 200 μl were pipetted with a slightly cut P1000 pipette tip into a 96 well plate. Reactions were started upon addition of 1 M Ethanol (20 μl) and $NADP^+$ was added to a final concentration of 2 mM and reactions were measured as for experimental assays in *Chapter 4*. In case of assays of the supernatant, Ethanol and $NADP^+$ was added directly to the whole of the microcapsule mixture, which was incubated at RT

on a shaking stand operating at 5 rpm. Samples for measurement were taken from settled microcapsule supernatants and measured in three or five minute time steps.

6.3 Results

6.3.1 Investigation of the Compostion of Microcapsules

Microcapsules derived by evaporation were firstly investigated using optical microscopy. Particle formation was observed and showed a highly inhomogenous size distribution, see figure 6.2.

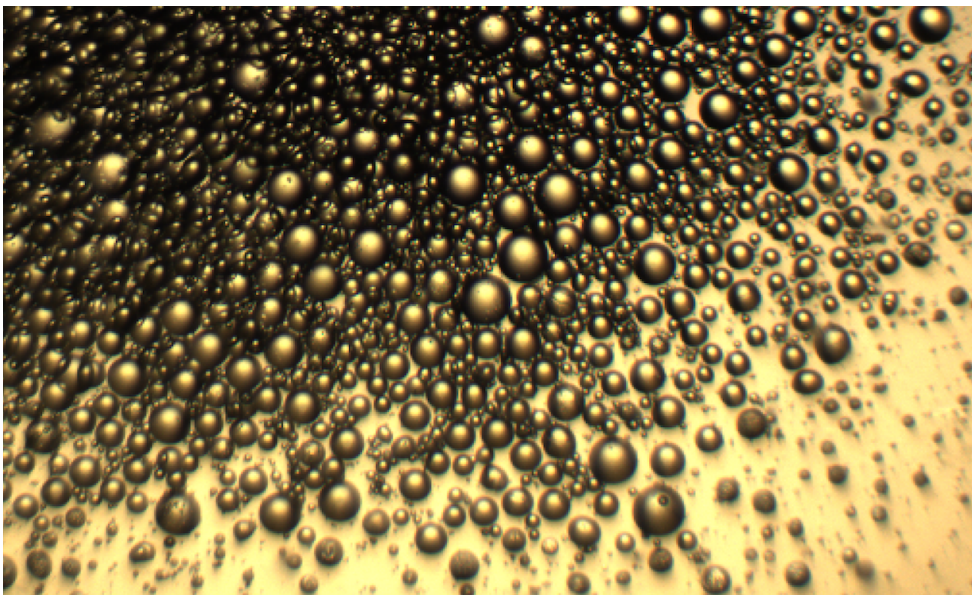


Figure 6.2: Microcapsules derived by evaporation at 4x magnification in optical microscopy. A range of inhomogenous sizes was observed.

the microscopy images indicated a size distribution between 10 and 100 μm and a wide variety of compositions of the microcapsule interiors as seen in figure 6.4 was observed. Smaller droplets were observed on many of the particles. It is not clear if these smaller droplets are inside or on the surface of the particles or whether they are ionic liquid, water or polystyrene. For some, a blue-purple colouring, possibly from Coomassie stained protein, can be observed. The colouring could equally be attributed to imaging artefacts,

such as differences in contrast. Differences in contrast could also be the reason for the observation of apparently translucent microcapsules on one hand and microcapsules with very dark appearance on the other hand, as indicated in figure 6.3 by the green, blue and red boxes. However, these differences might as well stem from different amounts of smaller droplets of ionic liquid and protein encapsulated in the interior.

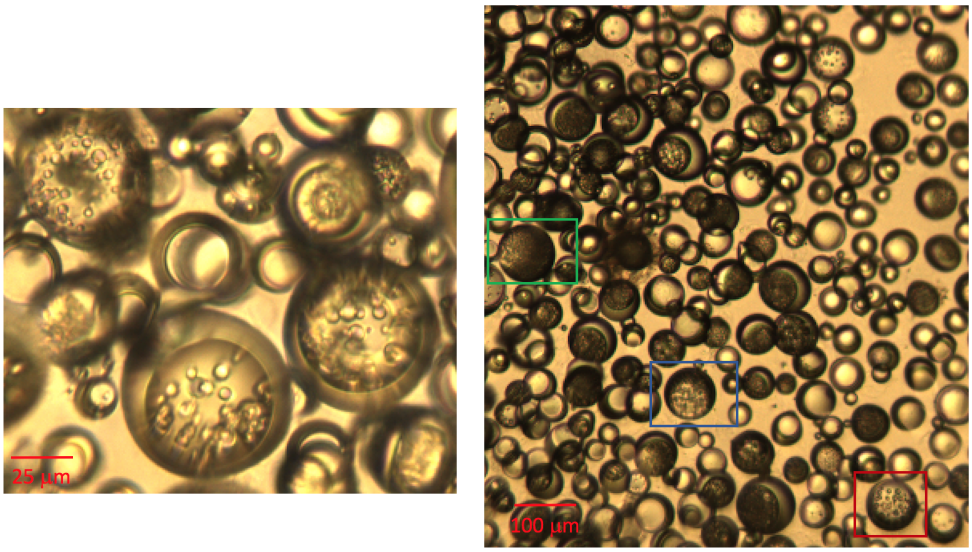


Figure 6.3: Left: Microcapsules derived by evaporation at 20x magnification (optical microscopy). Right: Microcapsules derived by evaporation at 10x magnification (optical microscopy). Boxes highlight microcapsules with apparent presence of different interiors or surfaces.

Observations of four separate experiments showed the presence of a residue outside of microcapsules as shown in figure 6.4.

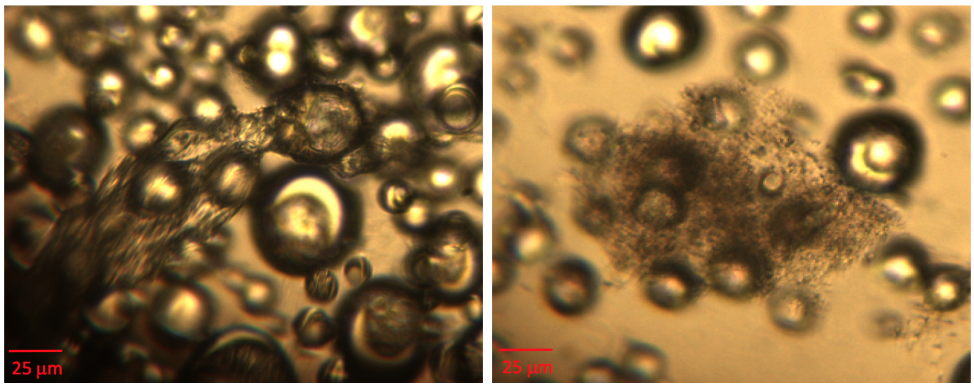


Figure 6.4: Microcapsules derived by evaporation at 20x magnification (optical microscopy). Partially, big patches of residue were observed. Patches appear to be unstained by coomassie.

6.3.2 Structural Integrity of Microcapsules

Samples prepared by evaporation were further investigated using SEM and ESEM to measure exact sizes and observe uniformity of spherecity. Microcapsules were first investigated by low vapor SEM, but the residue, already observed under the optical microscope, increasingly coated the surface of the microcapsules upon drying during measurements as can be seen in figure 6.5.

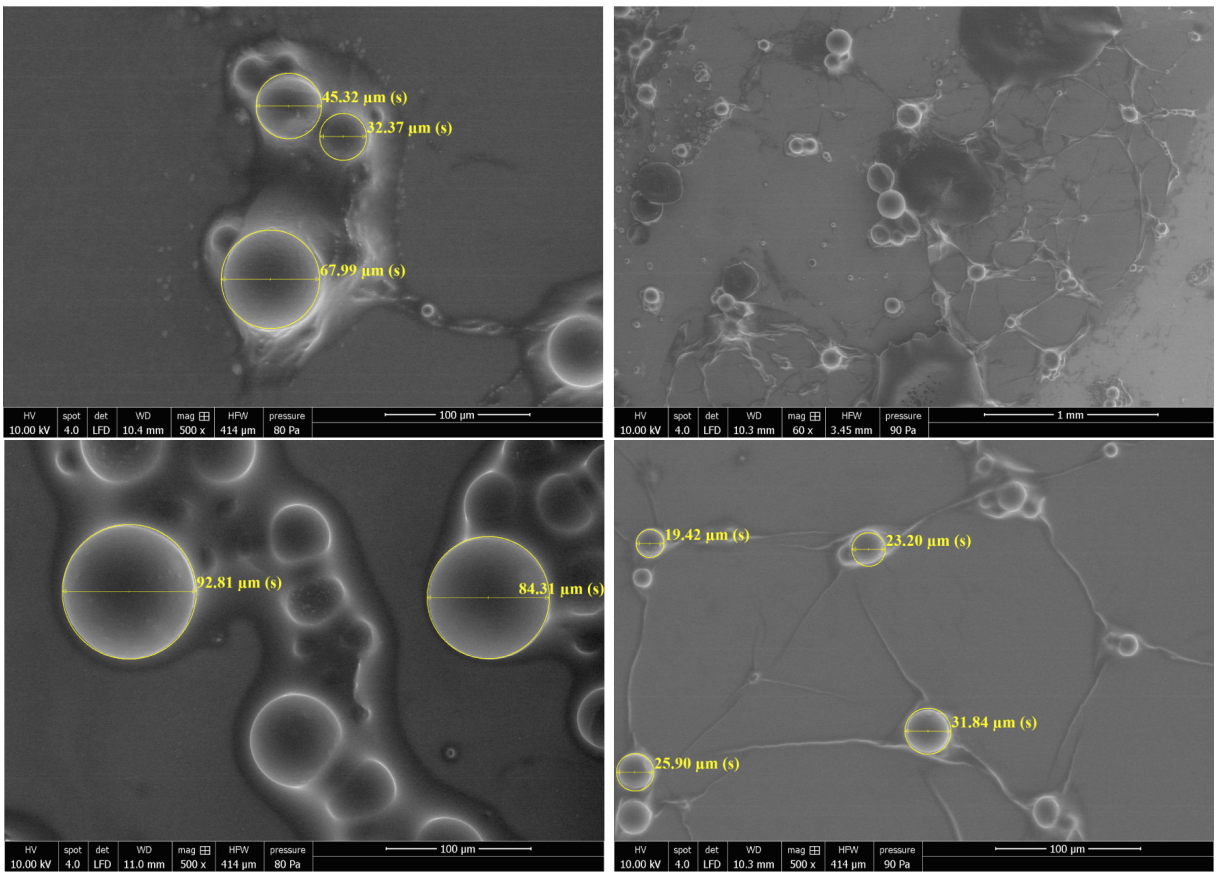


Figure 6.5: Scanning electron microscopy image taken from dried samples. Sizes of protein filled microcapsules vaired between 10 and 100 μm. Dried residue is visible in all four images, spanning between capsules either in dense packings (left) or trailed out (right).

For direct observation of microcapsule surfaces, samples were investigated using ESEM, which maintains the hydrated form of the microcapsules. Therefore, the sample was rehydrated through 100% humidity in the chamber. Solvation of the residue was observed alongside a more detailed surface of microcapsules, see figures 6.6 and 6.7.

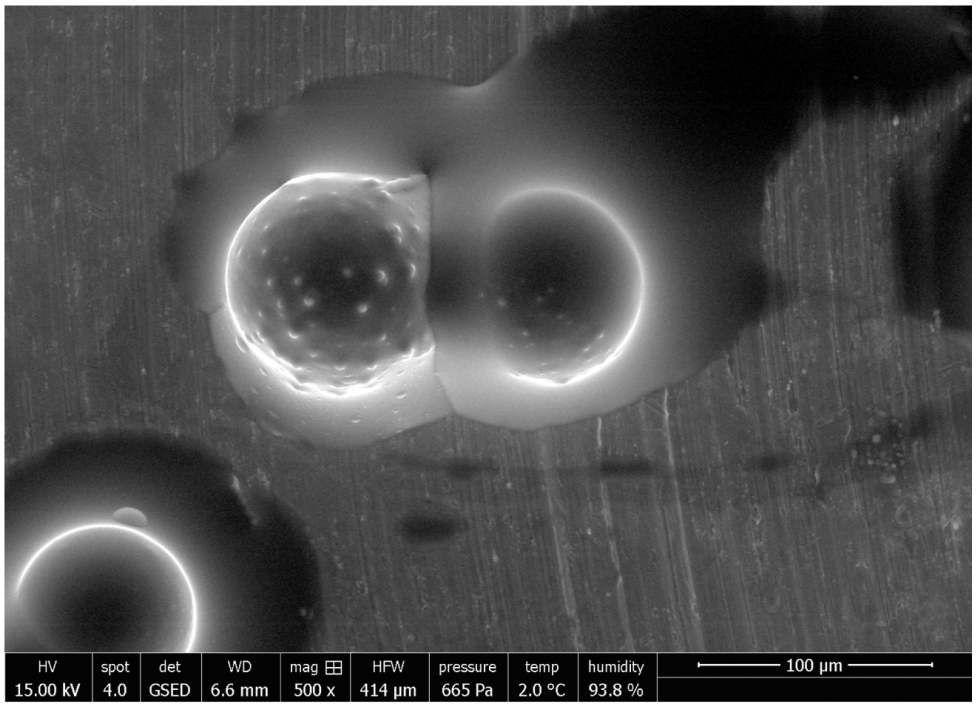


Figure 6.6: Environmental-SEM image taken from the same sample as in figure 6.6 whilst rehydration takes place. Undissolved residue immobilised to the surface of the microcapsule is visible.

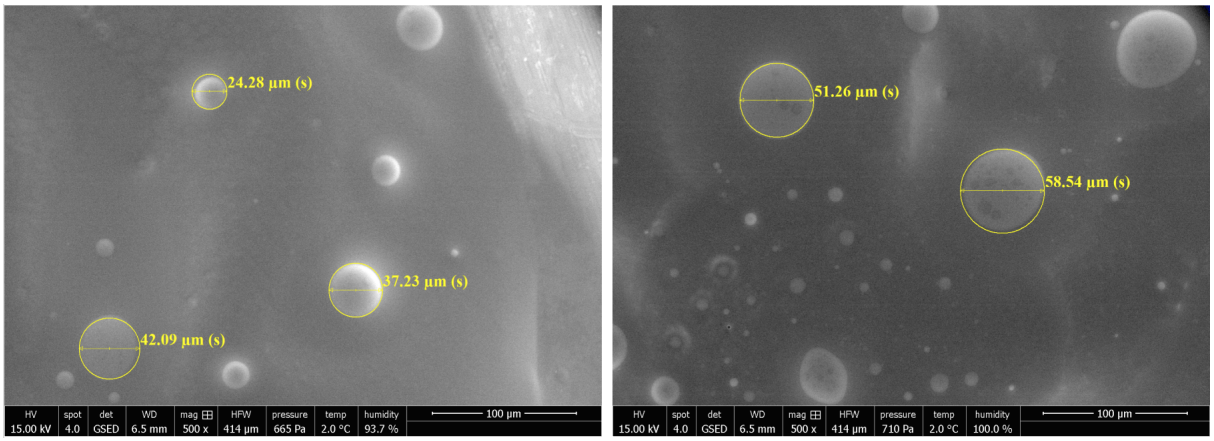


Figure 6.7: Environmental SEM images showing rehydration of the microcapsules at 93.7 % humidity (left) and 100 % humidity (right). The right image shows visible droplets inside the microcapsules close to the surface. This is possible, because secondary electron emission occurs frequently for uncoated samples and depth depends on the intensity of the beam. Operating at 15 kV will lead to about 5 keV with which electrons hit the anode. The electron beam can penetrate below the surface of up to 100 nm at this intensity and hence compositions below the surface can be detected [Thiel and Toth, 2005].

6.3.3 Activity Assays

Spectrophotometric assays were carried out on samples prepared by evaporation to observe activity of ADH2.

Activity assays of ADH2 in microcaspules prepared by evaporation

Reproducibility of measurements in case of the evaporation method was poor between replicates and even more so between replicated experiments, see figures 6.8 and 6.9. It was assumed, that the inhomogeneous size of microcapsules, especially the larger ones, obstruct the pathway of the lightbeam, which made it hard to infer if there was activity or not. The inconsistent results are likely due to large variations in the way these particles scatter and are possibly exacerbated by inhomogeneous particle distribution in the same sample volume, e.g. sedimentation could occur which may be time dependent. Further, the build up of a surface layer at the liquid-air interface was observed.

A different approach towards measuring activity was attempted. For this, prepared mi-

crocapsule samples were incubated with substrate and cofactor and only the supernatant without microcapsules was measured. The amount of time points was limited by the volume of the sample.

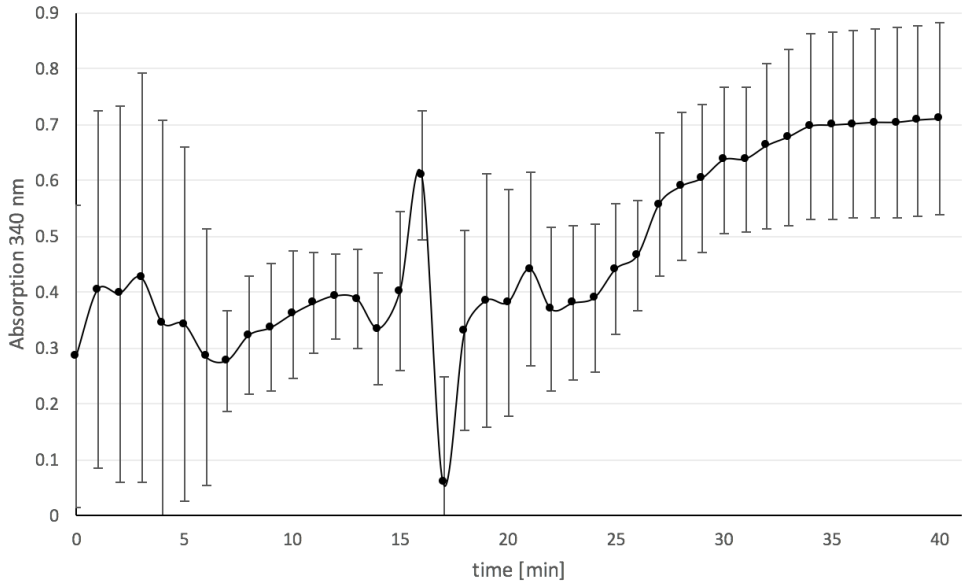


Figure 6.8: Activity assays of microcapsule mixture, showing high variability between replicates. A general trend towards an increase in activity can be inferred.

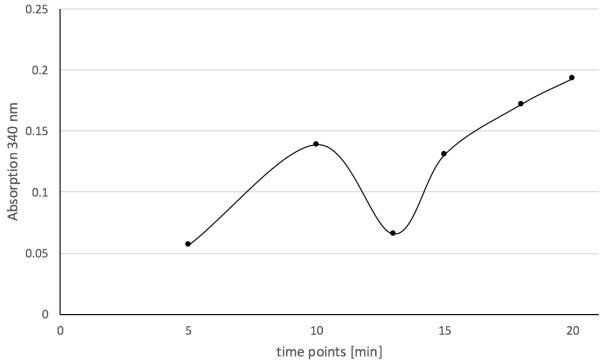


Figure 6.9: Activity assay of the supernatant of a microcapsule sample incubated with substrate and cofactor on a shaker. Volumes of 200 μl were then transfered for single measurements.

6.4 Discussion

An inhomogeneous distribution of microcapsule size was found, ranging between 10-100 μm. Further, presence of a residue was indicated in all experiments. This residue appeared in the form of a dried surface layer in case of the activity measurements and was seen in the imaging techniques, where it was found throughout the sample in smaller to larger patches. Although from the imaging techniques it is hardly possible to infer what the nature of the residue is, the 'drying-out' and build-up of a surface layer in activity assays excludes it being the ionic liquid, since it dissolves readily in both water and DCM. It rather suggests that PVA did not dissolved properly and acts as a surfactant at the liquid-air surface [Trieu and Qutubuddin, 1994] in the 96 well plates, as well as in the SEM experiments. Although it is not a chemical reaction, it might be that stoichiometrics need to be adapted during downsizing to overcome the issue with PVA residue.

Microcapsule sizes observed in the published work were centred around 2 μm [Guarnido et al., 2019]. This size difference can be explained by the difference in emulsification technique used for the small scale application in this thesis. While the published method used 76 ml at this step and hence allowed for the application of a shear mixer, the volume applied in this thesis was at this step 0.76 ml and hence the mixture containing Eppendorf tube was applied horizontally to an orbital shaker, leading to a different form of agitation. Size is further influenced by the amount of co-solvent [So and Lodge, 2016].

Despite the inconclusiveness of OM images about the encapsulation of ionic liquid and enzymes, since apparently present droplets in their interior could be attributed to contrasting effects, ESEM experiments assert the presence of droplets within vesicles. These are observable as spherical dark smudges beneath the surface of the microcapsules [Thiel and Toth, 2005].

Enzyme activity appeared to be present, however measurements showed very low reproducibility and a scarcely linear trend. Further, enzymatic activity could be attributed to the presence of the PVA residue. It is possible, that the enzyme is not encapsulated, but resides inside the PVA. Coomassie staining of the protein showed the blue colour confined to microcapsules when observed with the naked eye. Optical microscopy images indicated the presence of a blue to purple colour inside microcapsules, however, this might

be equally attributable to contrasting artefacts of the technique.

6.5 Conclusion and Outlook

Overall, formation of intact vesicles was observed. Ionic liquid is most likely to be encapsulated, however, it forms droplets within the microcapsules. This indicates, that the interior of the microcapsules is a mixture of water and ionic liquid. It is possible, that the long alkyl chains of the $[P_{6,6,6,14}]$ cation interfere with the formation of the vesicles and act themselves as a surfactant. If indeed a microenvironment of ionic liquid droplets and water mixture in the interior of microcapsules is confirmed, this might be an even greater advantage for solubilising hydrophobic substrates and maintaining enzyme activity.
The presence of the PVA residue allows no conclusive statement whether observed protein activity stems from encapsulated enzyme or PVA retained enzyme. The use of the surfactant PVA posed a major issue and obscures the meaning of data and results. This challenge needs to be addressed in future work. Lowering the amount of PVA could be tested for different ranges, since the ionic liquid itself might act as an additional surfactant to stabilise microcapsules and formation might be achieved despite lower amounts of PVA.

Chapter 7

Conclusion and Future Perspectives

Targets achieved in this sudy were the engineering of a protein expression strain of *H.volcanii* for the overexpression of the halophilic enzymes alcohol dehydrogenase 2 (ADH2) and ω-transaminase 5 (ω-TA5). Success in deriving gene deletions was verfied by Southern blots and resulted in strains eliminating co-purification of homologous proteins ADH2, ADH1 and ω-TA5, which could potentially form multimers with overexpressed His- or StrepII-tagged proteins.

ADH2 had been previously shown to be most active in its tetrameric form [Timpson et al., 2013]. For this thesis, the induction protocol was optimised and the native PAGE of purified protein suggests that the ADH2 tetramer does not form within three hours of overexpression. A prolonged induction period promotes tetramerisation leading to higher enzymatic activity. Future work could fine-tune the protocol even further or assess quaternary organisation over the course of different induction periods, to elucidate mechanisms of tetramer formation [Powers and Powers, 2003].

ADH2 tetramerisation further requires the presence of 4M KCl and the enzyme shows highest activity at around 80 ℃ [Timpson et al., 2013]. Since this temperature was likely to evaporate the maximum assay volume of 200 μl over the 20 to 25 minute measurements, a temperature of 50 ℃ was chosen in accordance with published literature [Alsafadi and Paradisi, 2013, Timpson et al., 2013]. Buffer conditions (4M KCl, 50 mM Glycine-KOH, pH10) were chosen to be the basis for all ionic liquid mixtures used in this study to enable comparison of measured relative specific activities. Thirteen different ionic liquids have been assayed: $[Me_3S][OSO_3Me]$, [Tetrakis][Cl], [Choline][Cl], [Choline][Bitartrate],

[EMIM][$MeSO_4$], [Me_3S][NTf_2], [BMIM][$MeOEtSO_4$], [$N_{1,1,8,8}$][$MeSO_3$], [$P_{4,4,4,4}$][Cl], [$N_{4,1,1,1}$] [$(MeO)_2(HO)PO$], [$P_{6,6,6,14}$][NTf_2], [Me_3SO][I] and [Me_3S][I]. Aqueous ionic liquid mixtures of those ions were used.

*Hv*ADH2 exhibited at least moderate tolerance towards almost all ionic liquids at low to moderate concentrations. The three systems [Me_3S][OSO_3Me], [$P_{6,6,6,14}$][NTf_2] and [Me_3S][I] were able to improve enzyme activity and showed promising results for future applications. Hydroxy-functionalised ionic liquids were found to diminish enzymatic activity even at the lowest concentrations and were the worst candidates.

MD simulations revealed structural changes of *Hv*ADH2 residues involved in determining the open and closed conformation of the enzyme. Regions of flexibility are in congruence with reported literature on thermophilic *B.stearothermophilus* ht-ADH, which were shown to be come rigidified at temperatures below 30 ℃ [Klinman and Kohen, 2013]. In this study, structural rearrangements observed depended on the ionic liquid system and the comparison of functional to non-functional *Hv*ADH2 conformations should be addressed in future work to identify residues for engineering IL tolerance by substitution. More specifically, residues that should be addressed in future work of engineering ADH2 stability are 111-HGDE and the YRGH-48E.

Simulations further showed that anion species of ionic liquids preferentially coordinate to halophilic protein surfaces in the presence of high concentrations of potassium chloride. This effect is even greater at lower concentrations of ionic liquid, and with this, our system proved different from mesophilic proteins [Pramanik et al., 2019].

Taking into account that hydroxy-functionalised ionic liquids were demonstrated to be good candidates for mesophilic ADH [de Gonzalo et al., 2007], the experimental strategy of maintaining the 4 M KCl buffer for these systems needs to be revised. In light of the findings from simulations, potassium might lead to an entrapment of the enzyme by the IL anions at low concentrations and this seems to be specifically detrimental when hydroxy-functionalised ionic liquids are present. On the other hand, at higher concentrations a shielding from ionic liquid ions through increased binding of potassium ions might be preferential and maintance of 4 M KCl might be beneficial. This hypothesis is based on the observation of a uniform descent towards lowest Gibbs free energy in the FEL of

the ionic liquid [Tetrakis][Chloride] at 0.3 M, which is similar to the FEL of ADH2 under native conditions. Activity assays for this system should be conducted in future work. The analysis of molecule dynamics trajectories obtained in this study has not been fully exploited yet. Future work will evaluate which atoms in ionic liquid molecules preferentially interact with surface residues, and which these are. A focus on finding water surface channels that are distrupted by ions should be guided by the literature on thermophilic *B.stearothermophilus* ht-ADH [Klinman and Kohen, 2013] and should included assessing changes in the Solvent Accessible Surface Area.

Strategies to encapsulate ADH2 and $[P_{6,6,6,14}][NTf_2]$ yielded formation of micelles which were observed by ESEM and LM. Observations were not conclusive in determining whether the protein was encapsulated or not. Future work should include marking the enzyme with a fluorescent tag [Lim et al., 2006] for observe its locality.

7.1 Concluding Remarks

This study has highlighted the importance of the quaternary organisation into tetramers or multimers in halophilic proteins for enzyme activity. Implications of conformational sub-states enabling tunneling efficiency in extremophilic alcohol dehydrogenases were found to be true for *Hv*ADH2. Similar peptide sequences that are mobile during switching between sub-states were identified and match the ones reported in literature [Klinman and Kohen, 2013]. This highlights the importance of preserving finely tuned intermolecular forces that maintain enzyme flexibility upon alteration of environmental 'hot-spot' conditions. Ionic liquids alter the electronic environment and H-bond dynamics around protein surfaces. This study has provided a glimpse at the distribution and coordination of ions to the halophilic protein surface. A hypothesis linking properties such as functionalised groups (hydroxy) or central ions (sulfoniums) or bulk properties such as hydrophobicity, to the distribution of ions around acidic acid residues was established. In a next step, further analysis of the distribution of coordinating atoms within ions and distribution of water molecules could enable a hypothesis to link distribution to activity.

Bibliography

Abraham, M. H. and Platts, J. A. Hydrogen bond structural group constants. The Journal of organic chemistry, 66(10):3484–3491, 2001.

Allers, T. Overexpression and purification of halophilic proteins in haloferax volcanii. Bioengineered bugs, 1(4):290–292, 2010.

Allers, T. and Mevarech, M. Archaeal genetics?the third way. Nature Reviews Genetics, 6(1):58, 2005.

Allers, T., Ngo, H.-P., Mevarech, M., and Lloyd, R. G. Development of additional selectable markers for the halophilic archaeon haloferax volcanii based on the leub and trpa genes. Appl. Environ. Microbiol., 70(2):943–953, 2004.

Alpizar, F., Backhaus, T., Decker, N., Eilks, I., Escobar-Pemberthy, N., Fantke, P., Geiser, K., Ivanova, M., Jolliet, O., Kim, H.-S., et al. Un environment global chemicals outlook ii-from legacies to innovative solutions: Implementing the 2030 agenda for sustainable development. 2019.

Alsafadi, D. and Paradisi, F. Effect of organic solvents on the activity and stability of halophilic alcohol dehydrogenase (adh2) from haloferax volcanii. Extremophiles, 17(1): 115–122, 2013.

Alsafadi, D., Alsalman, S., and Paradisi, F. Extreme halophilic alcohol dehydrogenase mediated highly efficient syntheses of enantiopure aromatic alcohols. Organic & biomolecular chemistry, 15(43):9169–9175, 2017.

Altis, A., Nguyen, P. H., Hegger, R., and Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. The Journal of chemical physics, 126(24): 244111, 2007.

Altschul, S. F. and Koonin, E. V. Iterated profile searches with psi-blast?a tool for discovery in protein databases. Trends in biochemical sciences, 23(11):444–447, 1998.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. Gene ontology: tool for the unification of biology. Nature genetics, 25(1):25, 2000.

Atomi, H., Imanaka, T., and Fukui, T. Overview of the genetic tools in the archaea. Frontiers in microbiology, 3:337, 2012.

Avula, N. V., Mondal, A., and Balasubramanian, S. Charge environment and hydrogen bond dynamics in binary ionic liquid mixtures: A computational study. The journal of physical chemistry letters, 9(12):3511–3516, 2018.

Ball, P. and Stillinger, F. H. H2o: A biography of water. Nature, 401(6756):850–850, 1999.

Barrett, A. Nomenclature committee of the international union of biochemistry and molecular biology (nc-iubmb). enzyme nomenclature. recommendations 1992. supplement 4: corrections and additions (1997). European journal of biochemistry, 250(1):1, 1997.

Batchelor, J. D., Olteanu, A., Tripathy, A., and Pielak, G. J. Impact of protein denaturants and stabilizers on water structure. Journal of the American Chemical Society, 126(7):1958–1961, 2004.

Baum, L. E. and Petrie, T. Statistical inference for probabilistic functions of finite state markov chains. The annals of mathematical statistics, 37(6):1554–1563, 1966.

Baxter, R. An interpretation of the effects of salts on the lactic dehydrogenase of halobacterium salinarium. Canadian journal of microbiology, 5(1):47–57, 1959.

Beker, W., Kamp, M. W., van der, Mulholland, A. J., and Sokalski, W. A. Rapid estimation of catalytic efficiency by cumulative atomic multipole moments: Application to ketosteroid isomerase mutants. Journal of chemical theory and computation, 13(2): 945–955, 2017.

Ben-Amotz, A. and Avron, M. The biotechnology of mass culturing dunaliella for products of commercial interest. Algal and cyanobacterial Biotechnology, pages 91–114, 1989.

Berendsen, H., Grigera, J., and Straatsma, T. The missing term in effective pair potentials. Journal of Physical Chemistry, 91(24):6269–6271, 1987.

Berendsen, H. J., Postma, J. P., Gunsteren, W. F., van, and Hermans, J. Interaction models for water in relation to protein hydration. In Intermolecular forces, pages 331–342. Springer, 1981.

Berendsen, H. J., Postma, J. v., Gunsteren, W. F., van, DiNola, A., and Haak, J. Molecular dynamics with coupling to an external bath. The Journal of chemical physics, 81 (8):3684–3690, 1984.

Bernardes, C. E., Piedade, M. E., Minas da, and Canongia Lopes, J. N. The structure of aqueous solutions of a hydrophilic ionic liquid: the full concentration range of 1-ethyl-3-methylimidazolium ethylsulfate and water. The Journal of Physical Chemistry B, 115 (9):2067–2074, 2011.

Bester-Rogac, M., Stoppa, A., and Buchner, R. Ion association of imidazolium ionic liquids in acetonitrile. The Journal of Physical Chemistry B, 118(5):1426–1435, 2014.

Biovia, D. S. Discovery studio visualizer. San Diego, CA, USA, 2017.

Bitan-Banin, G., Ortenberg, R., and Mevarech, M. Development of a gene knockout system for the halophilic archaeon haloferax volcanii by use of the pyre gene. Journal of Bacteriology, 185(3):772–778, 2003.

Bordoli, L. and Schwede, T. Automated protein structure modeling with swiss-model workspace and the protein model portal. In Homology Modeling, pages 107–136. Springer, 2011.

Bornhorst, J. A. and Falke, J. J. [16] purification of proteins using polyhistidine affinity tags. In Methods in enzymology, volume 326, pages 245–254. Elsevier, 2000.

Brandts, J. F. The thermodynamics of protein denaturation. ii. a model of reversible denaturation and interpretations regarding the stability of chymotrypsinogen. Journal of the American Chemical Society, 86(20):4302–4314, 1964.

Bräsen, C., Urbanke, C., and Schönheit, P. A novel octameric amp-forming acetyl-coa synthetase from the hyperthermophilic crenarchaeon pyrobaculum aerophilum. FEBS letters, 579(2):477–482, 2005.

Brena, B., González-Pombo, P., and Batista-Viera, F. Immobilization of enzymes: a literature survey. In Immobilization of enzymes and cells, pages 15–31. Springer, 2013.

Breuert, S., Allers, T., Spohn, G., and Soppa, J. Regulated polyploidy in halophilic archaea. PloS one, 1(1):e92, 2006.

Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M.-J., Ferrer, J., Pire, C., Esclapez, J., and Rice, D. W. Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile haloferax mediterranei. Proceedings of the National Academy of Sciences, 103(13):4846–4851, 2006.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. Mesophilic crenarchaeota: proposal for a third archaeal phylum, the thaumarchaeota. Nature Reviews Microbiology, 6(3):245, 2008.

Bruce, D. W., Cabry, C. P., Lopes, J. N. C., Costen, M. L., D?Andrea, L., Grillo, I., Marshall, B. C., McKendrick, K. G., Minton, T. K., Purcell, S. M., et al. Nanosegregation and structuring in the bulk and at the surface of ionic-liquid mixtures. The Journal of Physical Chemistry B, 121(24):6002–6020, 2017.

Buenger, J. and Driller, H. Ectoin: an effective natural substance to prevent uva-induced premature photoaging. Skin pharmacology and physiology, 17(5):232–237, 2004.

Burney, P. R., Nordwald, E. M., Hickman, K., Kaar, J. L., and Pfaendtner, J. Molecular dynamics investigation of the ionic liquid/enzyme interface: Application to engineering enzyme surface charge. Proteins: Structure, Function, and Bioinformatics, 83(4):670–680, 2015.

Cahn, R. S., Ingold, C., and Prelog, V. Specification of molecular chirality. Angewandte Chemie International Edition in English, 5(4):385–415, 1966.

Canutescu, A. A. and Dunbrack Jr, R. L. Cyclic coordinate descent: A robotics algorithm for protein loop closure. Protein science, 12(5):963–972, 2003.

Cao, S., Xu, P., Ma, Y., Yao, X., Yao, Y., Zong, M., Li, X., and Lou, W. Recent advances in immobilized enzymes on nanocarriers. Chinese Journal of Catalysis, 37 (11):1814–1823, 2016.

Cardona, C., Quintero, J., and Paz, I. Production of bioethanol from sugarcane bagasse: status and perspectives. Bioresource technology, 101(13):4754–4766, 2010.

Case, D. A., Darden, T., Cheatham, T., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., et al. Amber 10. Technical report, University of California, 2008.

Cassidy, J., Bruen, L., Rosini, E., Molla, G., Pollegioni, L., and Paradisi, F. Engineering substrate promiscuity in halophilic alcohol dehydrogenase (hvadh2) by in silico design. PloS one, 12(11):e0187482, 2017.

Catchpole, R. and Forterre, P. Positively twisted: The complex evolutionary history of reverse gyrase suggests a non-hyperthermophilic last universal common ancestor. bioRxiv, page 524215, 2019.

Cavagnero, S., Debe, D. A., Zhou, Z. H., Adams, M. W., and Chan, S. I. Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. Biochemistry, 37(10):3369–3376, 1998.

Chapman, J., Ismail, A., and Dinu, C. Industrial applications of enzymes: Recent advances, techniques, and outlooks. Catalysts, 8(6):238, 2018.

Charlebois, R. L., Schalkwyk, L. C., Hofman, J. D., and Doolittle, W. F. Detailed physical map and set of overlapping clones covering the genome of the archaebacterium haloferax volcanii ds2. Journal of molecular biology, 222(3):509–524, 1991.

Che, H. and Hest, J. C., van. Stimuli-responsive polymersomes and nanoreactors. Journal of Materials Chemistry B, 4(27):4632–4647, 2016.

Chen, C.-C., Walia, R., Mukherjee, K. J., Mahalik, S., and Summers, D. K. Indole generates quiescent and metabolically active e. coli cultures. 2015.

Chothia, C. and Lesk, A. M. The relation between the divergence of sequence and structure in proteins. The EMBO journal, 5(4):823–826, 1986.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., et al. Caver 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS computational biology, 8(10):e1002708, 2012.

Clark, J. H. Green chemistry: challenges and opportunities. Green Chemistry, 1(1):1–8, 1999.

Clark, J. H. Chemistry goes green. Nature chemistry, 1(1):12, 2009.

Cline, S. W., Schalkwyk, L. C., and Doolittle, W. F. Transformation of the archaebacterium halobacterium volcanii with genomic dna. Journal of bacteriology, 171(9): 4987–4991, 1989.

Cole-Hamilton, D. J. Homogeneous catalysis–new approaches to catalyst separation, recovery, and recycling. Science, 299(5613):1702–1706, 2003.

Coles, S. W., Smith, A. M., Fedorov, M. V., Hausen, F., and Perkin, S. Interfacial structure and structural forces in mixtures of ionic liquid with a polar solvent. Faraday discussions, 206:427–442, 2017.

Colovos, C. and Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. Protein science, 2(9):1511–1519, 1993.

Cook, J. L., Hunter, C. A., Low, C. M., Perez-Velasco, A., and Vinter, J. G. Solvent effects on hydrogen bonding. Angewandte Chemie International Edition, 46(20):3706–3709, 2007.

Cooley, J. W. and Tukey, J. W. An algorithm for the machine calculation of complex fourier series. Mathematics of computation, 19(90):297–301, 1965.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society, 117(19):5179–5197, 1995.

Cornish-Bowden, A. The effect of natural selection on enzymic catalysis. Journal of molecular biology, 101(1):1–9, 1976.

Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R., and Embley, T. M. The archaebacterial origin of eukaryotes. Proceedings of the National Academy of Sciences, 105(51):20356–20361, 2008.

Curran, M. A. Strengths and limitations of life cycle assessment. In Background and future prospects in Life Cycle Assessment, pages 189–206. Springer, 2014.

Dabirmanesh, B., Khajeh, K., Akbari, J., Falahati, H., Daneshjoo, S., and Heydari, A. Mesophilic alcohol dehydrogenase behavior in imidazolium based ionic liquids. Journal of Molecular Liquids, 161(3):139–143, 2011.

Dabirmanesh, B., Khajeh, K., Ranjbar, B., Ghazi, F., and Heydari, A. Inhibition mediated stabilization effect of imidazolium based ionic liquids on alcohol dehydrogenase. Journal of Molecular Liquids, 170:66–71, 2012.

Dahl, K. H. and McKinley-McKee, J. S. The imidazole-promoted inactivation of horse-liver alcohol dehydrogenase. European journal of biochemistry, 120(3):451–459, 1981.

Dalziel, K. Kinetic studies of liver alcohol dehydrogenase and ph effects with coenzyme preparations of high purity. Journal of Biological Chemistry, 238(8):2850–2858, 1963.

Darden, T., York, D., and Pedersen, L. Particle mesh ewald: An n? log (n) method for ewald sums in large systems. The Journal of chemical physics, 98(12):10089–10092, 1993.

Das, S., Mukherjee, B., and Biswas, R. Orientational dynamics in a room temperature ionic liquid: Are angular jumps predominant? The Journal of chemical physics, 148 (19):193839, 2018.

Gonzalo, G., de, Lavandera, I., Durchschein, K., Wurm, D., Faber, K., and Kroutil, W. Asymmetric biocatalytic reduction of ketones using hydroxy-functionalised water-miscible ionic liquids as solvents. Tetrahedron: Asymmetry, 18(21):2541–2546, 2007.

Dekker, F., Bliek, A., Kapteijn, F., and Moulijn, J. Analysis of mass and heat transfer in transient experiments over heterogeneous catalysts. Chemical Engineering Science, 50 (22):3573–3580, 1995.

DeLano, W. L. Pymol, 2002.

Dewan, S. Global markets for enzymes in industrial applications. bcc research, bio030g. 2012, 2012.

DiCosimo, R., McAuliffe, J., Poulose, A. J., and Bohlmann, G. Industrial use of immobilized enzymes. Chemical Society Reviews, 42(15):6437–6474, 2013.

Discher, B. M., Won, Y.-Y., Ege, D. S., Lee, J. C., Bates, F. S., Discher, D. E., and Hammer, D. A. Polymersomes: tough vesicles made from diblock copolymers. Science, 284(5417):1143–1146, 1999.

Doerffel, K. and Eckschlager, K. Optimal strategies in analysis. Deutscher Verlag fr Grundstoffindustrie, 1981.

Domańska, U. and Marciniak, A. Phase behaviour of 1-hexyloxymethyl-3-methylimidazolium and 1, 3-dihexyloxymethyl-imidazolium based ionic liquids with alcohols, water, ketones and hydrocarbons: The effect of cation and anion on solubility. Fluid Phase Equilibria, 260(1):9–18, 2007.

Dong, K. and Zhang, S. Hydrogen bonds: a structural insight into ionic liquids. Chemistry–A European Journal, 18(10):2748–2761, 2012.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of computational chemistry, 24(16):1999–2012, 2003.

Duplais, C., Krasovskiy, A., Wattenberg, A., and Lipshutz, B. H. Cross-couplings between benzylic and aryl halides ?on water?: synthesis of diarylmethanes. Chemical Communications, 46(4):562–564, 2010.

Duquemin, S.-J. and Nixon, J. The effect of surfactants on the microencapsulation and release of phenobarbitone from gelatin-acacia complex coacervate microcapsules. Journal of microencapsulation, 3(2):89–93, 1986.

Dwars, T., Paetzold, E., and Oehme, G. Reactions in micellar systems. Angewandte Chemie International Edition, 44(44):7174–7199, 2005.

Dyall-Smith, M. The halohandbook: protocols for halobacterial genetics. Mark Dyall-Smith, Martinsried, Germany, pages 1–144, 2008.

Dym, O., Mevarech, M., and Sussman, J. L. Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. Science, 267(5202):1344–1346, 1995.

Earle, M. J. and Seddon, K. R. Ionic liquids. green solvents for the future. Pure and applied chemistry, 72(7):1391–1398, 2000.

Eckstein, M., Villela Filho, M., Liese, A., and Kragl, U. Use of an ionic liquid in a two-phase system to improve an alcohol dehydrogenase catalysed reduction. Chemical communications, (9):1084–1085, 2004.

Eisenberg, D., Lüthy, R., and Bowie, J. U. [20] verify3d: assessment of protein models with three-dimensional profiles. In Methods in enzymology, volume 277, pages 396–404. Elsevier, 1997.

Eisenberg, H., Mevarech, M., and Zaccai, G. Biochemical, structural, and molecular genetic aspects of halophilism. In Advances in protein chemistry, volume 43, pages 1–62. Elsevier, 1992.

Elias, M., Wieczorek, G., Rosenne, S., and Tawfik, D. S. The universality of enzymatic rate–temperature dependency. Trends in Biochemical Sciences, 39(1):1–7, 2014.

Eme, L., Spang, A., Lombard, J., Stairs, C. W., and Ettema, T. J. Archaea and the origin of eukaryotes. Nature Reviews Microbiology, 15(12):711, 2017.

Ewald, P. P. Die berechnung optischer und elektrostatischer gitterpotentiale. Annalen der physik, 369(3):253–287, 1921.

Eyring, H. and Hirschfelder, J. The theory of the liquid state. Journal of physical chemistry, 41(2):249–257, 1937.

Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J., and Oesterhelt, D. Living with two extremes: conclusions from the genome sequence of natronomonas pharaonis. Genome research, 15(10):1336–1343, 2005.

Fell, D. and Cornish-Bowden, A. Understanding the control of metabolism, volume 2. Portland press London, 1997.

Feller, G. Protein stability and enzyme activity at extreme biological temperatures. Journal of Physics: Condensed Matter, 22(32):323101, 2010.

Fischer, D. 3d-shotgun: a novel, cooperative, fold-recognition meta-predictor. Proteins: Structure, Function, and Bioinformatics, 51(3):434–441, 2003.

Freire, M. G., Santos, L. M., Marrucho, I. M., and Coutinho, J. A. Evaluation of cosmo-rs for the prediction of lle and vle of alcohols+ ionic liquids. Fluid Phase Equilibria, 255 (2):167–178, 2007.

Frolow, F., Harell, M., Sussman, J. L., Mevarech, M., and Shoham, M. Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2fe-2s ferredoxin. Nature structural biology, 3(5):452, 1996.

Fukuchi, S. and Nishikawa, K. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. Journal of molecular biology, 309(4): 835–843, 2001.

Fumino, K., Wulf, A., and Ludwig, R. The potential role of hydrogen bonding in aprotic and protic ionic liquids. Physical Chemistry Chemical Physics, 11(39):8790–8794, 2009.

Gaci, N., Borrel, G., Tottey, W., O?Toole, P. W., and Brugère, J.-F. Archaea and the human gut: new beginning of an old story. World journal of gastroenterology: WJG, 20(43):16062, 2014.

Gagniuc, P. A. Markov chains: from theory to implementation and experimentation. John Wiley & Sons, 2017.

Gaitzsch, J., Huang, X., and Voit, B. Engineering functional polymer capsules toward smart nanoreactors. Chemical reviews, 116(3):1053–1093, 2015.

Gasteiger, J. and Engel, T. Chemoinformatics: a textbook. John Wiley & Sons, 2006.

Gehrke, S., Domaros, M., von, Clark, R., Hollóczki, O., Brehm, M., Welton, T., Luzar, A., and Kirchner, B. Structure and lifetimes in ionic liquids and their mixtures. Faraday discussions, 206:219–245, 2017.

Giver, L., Gershenson, A., Freskgard, P.-O., and Arnold, F. H. Directed evolution of a thermostable esterase. Proceedings of the National Academy of Sciences, 95(22): 12809–12813, 1998.

Glimm, E. Fisher, ni: Statistical analysis of circular data. cambridge university press, cambridge, uk 1995. 277 pp.,£ 16.95. Biometrical Journal, 38(3):314–314, 1996.

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J., Melo, M. N., Seyler, S. L., Domański, J., Dotson, D. L., Buchoux, S., Kenney, I. M., et al. Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations. In Proceedings of the 15th Python in Science Conference, volume 98. SciPy Austin, TX, 2016.

Grant, W. D. Introductory chapter: half a lifetime in soda lakes. In Halophilic microorganisms, pages 17–31. Springer, 2004.

Greaves, T. L. and Drummond, C. J. Protic ionic liquids: properties and applications. Chemical reviews, 108(1):206–237, 2008.

Greco, C., Fourmond, V., Baffert, C., Wang, P.-h., Dementin, S., Bertrand, P., Bruschi, M., Blumberger, J., Gioia, L., de, and Léger, C. Combining experimental and theoretical methods to learn about the reactivity of gas-processing metalloenzymes. Energy & Environmental Science, 7(11):3543–3573, 2014.

Guarnido, I. L., Routh, A. F., Mantle, M. D., Serrano, M. F., and Marr, P. C. Ionic liquid microcapsules: Formation and application of polystyrene microcapsules with ionic liquid cores. ACS Sustainable Chemistry & Engineering, 7(2):1870–1874, 2019.

Gutteridge, A. and Thornton, J. Conformational changes observed in enzyme crystal structures upon substrate binding. Journal of molecular biology, 346(1):21–28, 2005.

Guy, L. and Ettema, T. J. The archaeal ?tack?superphylum and the origin of eukaryotes. Trends in microbiology, 19(12):580–587, 2011.

Haberler, M. and Steinhauser, O. On the influence of hydrated ionic liquids on the dynamical structure of model proteins: a computational study. Physical Chemistry Chemical Physics, 13(40):17994–18004, 2011.

Haberler, M., Schröder, C., and Steinhauser, O. Solvation studies of a zinc finger protein in hydrated ionic liquids. Physical Chemistry Chemical Physics, 13(15):6955–6969, 2011.

Haberler, M., Schrder, C., and Steinhauser, O. Hydrated ionic liquids with and without solute: the influence of water content and protein solutes. Journal of chemical theory and computation, 8(10):3911–3928, 2012.

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of cheminformatics, 4(1):17, 2012.

Harms, M. J. and Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nature Reviews Genetics, 14(8):559, 2013.

Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T. M., et al. The complete genome sequence of haloferax volcanii ds2, a model archaeon. PloS one, 5(3):e9605, 2010.

Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. Accelerated growth in the absence of dna replication origins. Nature, 503(7477):544, 2013.

Hayes, R., Warr, G. G., and Atkin, R. Structure and nanostructure in ionic liquids. Chemical reviews, 115(13):6357–6426, 2015.

Hayes, S. T., Assaf, G., Checksfield, G., Cheung, C., Critcher, D., Harris, L., Howard, R., Mathew, S., Regius, C., Scotney, G., et al. Commercial synthesis of (s, s)-reboxetine succinate: a journey to find the cheapest commercial chemistry for manufacture. Organic Process Research & Development, 15(6):1305–1314, 2011.

Heiss, C., Laivenieks, M., Zeikus, J. G., and Phillips, R. S. The stereospecificity of secondary alcohol dehydrogenase from thermoanaerobacter e thanolicus is partially deter-

mined by active site water. Journal of the American Chemical Society, 123(2):345–346, 2001.

Hertwich, E. G., Gibon, T., Bouman, E. A., Arvesen, A., Suh, S., Heath, G. A., Bergesen, J. D., Ramirez, A., Vega, M. I., and Shi, L. Integrated life-cycle assessment of electricity-supply scenarios confirms global environmental benefit of low-carbon technologies. Proceedings of the National Academy of Sciences, 112(20):6277–6282, 2015.

Hildebrand, J. A history of solution theory. Annual review of physical chemistry, 32(1): 1–24, 1981.

Hinkson, I. V. and Elias, J. E. The dynamic state of protein turnover: It's about time. Trends in cell biology, 21(5):293–303, 2011.

Hiraga, Y., Duereh, A., and Smith Jr, R. L. Aspects of solvent polarity and solvent properties in developing efficient systems for processing biomass with ionic liquid mixtures and supercritical co2. The Journal of Supercritical Fluids, 134:12–20, 2018.

Hockney, R., Goel, S., and Eastwood, J. Quiet high-resolution computer models of a plasma. Journal of Computational Physics, 14(2):148–158, 1974.

Holmes, M. L., Nuttall, S. D., and Dyall-Smith, M. L. Construction and use of halobacterial shuttle vectors and further studies on haloferax dna gyrase. Journal of Bacteriology, 173(12):3807–3813, 1991.

Holmes, M. L., Scopes, R. K., Moritz, R. L., Simpson, R. J., Englert, C., Pfeifer, F., and Dyall-Smith, M. L. Purification and analysis of an extremely halophilic β-galactosidase from haloferax alicantei. Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, 1337(2):276–286, 1997.

Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., and Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: Tip4p-ew. The Journal of chemical physics, 120(20):9665–9678, 2004.

Huang, X. and Miller, W. A time-efficient, linear-space local similarity algorithm. Advances in Applied Mathematics, 12(3):337–357, 1991.

Hunt, P. A. Quantum chemical modeling of hydrogen bonding in ionic liquids. In Ionic Liquids II, pages 225–246. Springer, 2017.

Hunt, P. A., Ashworth, C. R., and Matthews, R. P. Hydrogen bonding in ionic liquids. Chemical Society Reviews, 44(5):1257–1288, 2015.

Hussain, W., Pollard, D. J., Truppo, M., and Lye, G. J. Enzymatic ketone reductions with co-factor recycling: Improved reactions with ionic liquid co-solvents. Journal of Molecular Catalysis B: Enzymatic, 55(1-2):19–29, 2008.

Hutchinson, E. G. and Thornton, J. M. A revised set of potentials for β-turn formation in proteins. Protein Science, 3(12):2207–2216, 1994.

Huyghues-Despointes, B. M., Pace, C. N., Englander, S. W., and Scholtz, J. M. Measuring the conformational stability of a protein by hydrogen exchange. In Protein Structure, Stability, and Folding, pages 69–92. Springer, 2001.

Ichiye, T. and Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins: Structure, Function, and Bioinformatics, 11(3):205–217, 1991.

Inczedy, J., Lengyel, T., Ure, A. M., Gelencsér, A., Hulanicki, A., et al. Compendium of analytical nomenclature. The Orange Book, 3rd Edn., 1998.

Israelachvili, J. and Wennerström, H. Role of hydration and water structure in biological and colloidal interactions. Nature, 379(6562):219, 1996.

Ivanov, I., Lira, R. B., Tang, T.-Y. D., Franzmann, T., Klosin, A., Silva, L. C., da, Hyman, A., Landfester, K., Lipowsky, R., Sundmacher, K., et al. Directed growth of biomimetic microcompartments. Advanced Biosystems, page 1800314, 2019.

Jaenicke, R. Stability and folding of ultrastable proteins: eye lens crystallins and enzymes from thermophiles. The FASEB journal, 10(1):84–92, 1996.

Jaenicke, R. and Lilie, H. Folding and association of oligomeric and multimeric proteins. Advances in protein chemistry, 53:329–401, 2000.

Jain, S. and Bates, F. S. On the origins of morphological complexity in block copolymer surfactants. Science, 300(5618):460–464, 2003.

Jegannathan, K. R. and Nielsen, P. H. Environmental assessment of enzyme use in industrial production–a literature review. Journal of cleaner production, 42:228–240, 2013.

Jékely, G. Origin and evolution of the self-organizing cytoskeleton in the network of eukaryotic organelles. Cold Spring Harbor perspectives in biology, 6(9):a016030, 2014.

Jorgensen, W. L. and Madura, J. D. Temperature and size dependence for monte carlo simulations of tip4p water. Molecular Physics, 56(6):1381–1392, 1985.

Jorgensen, W. L. and Tirado-Rives, J. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. Journal of the American Chemical Society, 110(6):1657–1666, 1988.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics, 79(2):926–935, 1983.

Kajitani, M. and Ishihama, A. Determination of the promoter strength in the mixed transcription system: promoters of lactose, tryplophan and ribosomal protein l10 operons from excherichia coli. Nucleic acids research, 11(3):671–686, 1983.

Kandler, O. Thermophiles: The keys to molecular evolution and the origin of life? Athens: Taylor and Francis, pages 19–31, 1998.

Kar, M., Plechkova, N. V., Seddon, K. R., Pringle, J. M., and MacFarlane, D. R. Ionic liquids–further progress on the fundamental issues. Australian Journal of Chemistry, 72(2):3–10, 2019.

Kastritis, P. L., Papandreou, N. C., and Hamodrakas, S. J. Haloadaptation: insights from comparative modeling studies of halophilic archaeal dhfrs. International Journal of Biological Macromolecules, 41(4):447–453, 2007.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. The phyre2 web portal for protein modeling, prediction and analysis. Nature protocols, 10(6):845, 2015.

Khullar, P., Singh, V., Mahal, A., Kumar, H., Kaur, G., and Bakshi, M. S. Block copolymer micelles as nanoreactors for self-assembled morphologies of gold nanoparticles. The Journal of Physical Chemistry B, 117(10):3028–3039, 2013.

Klähn, M., Lim, G. S., Seduraman, A., and Wu, P. On the different roles of anions and cations in the solvation of enzymes in ionic liquids. Physical Chemistry Chemical Physics, 13(4):1649–1662, 2011.

Klinman, J. P. and Kohen, A. Hydrogen tunneling links protein dynamics to enzyme catalysis. Annual review of biochemistry, 82:471–496, 2013.

Kohlmann, C., Robertz, N., Leuchs, S., Dogan, Z., Lütz, S., Bitzer, K., Na?amnieh, S., and Greiner, L. Ionic liquid facilitates biocatalytic conversion of hardly water soluble ketones. Journal of Molecular Catalysis B: Enzymatic, 68(2):147–153, 2011.

Konar, S., Sharma, A., Banerjee, S., Ghorai, P. K., and Biswas, R. Water-mediated weakening of inter-ionic interactions in aqueous mixtures of ionic liquid: An investigation combining quantum chemical calculations and molecular dynamics simulations. Chemical Physics, 524:31–39, 2019.

Koshland Jr, D. E. Conformational changes: how small is big enough? Nature medicine, 4(10):1112, 1998.

Kumar, A. and Venkatesu, P. Does the stability of proteins in ionic liquids obey the hofmeister series? International journal of biological macromolecules, 63:244–253, 2014.

Laage, D., Stirnemann, G., Sterpone, F., Rey, R., and Hynes, J. T. Reorientation and allied dynamics in water and aqueous solutions. Annual review of physical chemistry, 62:395–416, 2011.

Lake, J. A., Henderson, E., Oakes, M., and Clark, M. W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proceedings of the National Academy of Sciences, 81(12):3786–3790, 1984.

Lam, S. Y., Yeung, R. C., Yu, T.-H., Sze, K.-H., and Wong, K.-B. A rigidifying salt-bridge favors the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity. PLoS biology, 9(3):e1001027, 2011.

Lam, W. L. and Doolittle, W. F. Shuttle vectors for the archaebacterium halobacterium volcanii. Proceedings of the National Academy of Sciences, 86(14):5478–5482, 1989.

Lane, N. Bioenergetic constraints on the evolution of complex life. Cold Spring Harbor perspectives in biology, 6(5):a015982, 2014.

Lanyi, J. K. Salt-dependent properties of proteins from extremely halophilic bacteria. Bacteriological reviews, 38(3):272, 1974.

Lanyi, J. K. and Stevenson, J. Studies of the electron transport chain of extremely halophilic bacteria iv. role of hydrophobic forces in the structure of menadione reductase. Journal of Biological Chemistry, 245(16):4074–4080, 1970.

Large, A., Stamme, C., Lange, C., Duan, Z., Allers, T., Soppa, J., and Lund, P. A. Characterization of a tightly controlled promoter of the halophilic archaeon haloferax volcanii and its use in the analysis of the essential cct1 gene. Molecular microbiology, 66(5):1092–1106, 2007.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. Procheck: a program to check the stereochemical quality of protein structures. Journal of applied crystallography, 26(2):283–291, 1993.

Laszlo, J. A. and Compton, D. L. α-chymotrypsin catalysis in imidazolium-based ionic liquids. Biotechnology and bioengineering, 75(2):181–186, 2001.

Latif, M. A. M., Micaêlo, N. M., and Rahman, M. B. A. Influence of anion–water interactions on the behaviour of lipases in room temperature ionic liquids. RSC Advances, 4(89):48202–48211, 2014.

Leach, A. Molecular modelling: principles and applications. 2001, harlow, england. New York: Prentice Hall, 24(744):16.

Lee, C.-F., Makhatadze, G. I., and Wong, K.-B. Effects of charge-to-alanine substitutions on the stability of ribosomal protein l30e from thermococcus celer. Biochemistry, 44 (51):16817–16825, 2005.

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., et al. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. Journal of chemical theory and computation, 12(1):405–413, 2015.

Lee, T.-S., Cerutti, D. S., Mermelstein, D., Lin, C., LeGrand, S., Giese, T. J., Roitberg, A., Case, D. A., Walker, R. C., and York, D. M. Gpu-accelerated molecular dynamics and free energy methods in amber18: performance enhancements and new features. Journal of chemical information and modeling, 58(10):2043–2050, 2018.

Lennard-Jones, J. The liquid state. Proceedings of the Physical Society, 52(6):729, 1940.

Lenton, S., Walsh, D. L., Rhys, N. H., Soper, A. K., and Dougan, L. Structural evidence for solvent-stabilisation by aspartic acid as a mechanism for halophilic protein stability in high salt concentrations. Physical Chemistry Chemical Physics, 18(27):18054–18062, 2016.

Lerner, N., Peterson, E., and Chang, S. The strecker synthesis as a source of amino acids in carbonaceous chondrites: Deuterium retention during synthesis. Geochimica et Cosmochimica Acta, 57(19):4713–4723, 1993.

Leuko, S., Rettberg, P., Pontifex, A., and Burns, B. On the response of halophilic archaea to space conditions. Life, 4(1):66–76, 2014.

Levitt, M. Accurate modeling of protein conformation by automatic segment matching. Journal of molecular biology, 226(2):507–533, 1992.

Li, G. and Young, K. D. Indole production by the tryptophanase tnaa in escherichia coli is determined by the amount of exogenous tryptophan. Microbiology, 159(2):402–410, 2013.

Li, T., Liang, J., Ambrogelly, A., Brennan, T., Gloor, G., Huisman, G., Lalonde, J., Lekhal, A., Mijts, B., Muley, S., et al. Efficient, chemoenzymatic process for manu-

facture of the boceprevir bicyclic [3.1. 0] proline intermediate based on amine oxidase-catalyzed desymmetrization. Journal of the American Chemical Society, 134(14):6467–6472, 2012.

Li, Y. and Zhang, Y. Remo: A new protocol to refine full atomic protein models from c-alpha traces by optimizing hydrogen-bonding networks. Proteins: Structure, Function, and Bioinformatics, 76(3):665–676, 2009.

Liew, W. H., Hassim, M. H., and Ng, D. K. Review of evolution, technology and sustainability assessments of biofuel production. Journal of Cleaner Production, 71:11–29, 2014.

Lim, G. S. and Klhn, M. On the stability of proteins solvated in imidazolium-based ionic liquids studied with replica exchange molecular dynamics. The Journal of Physical Chemistry B, 122(39):9274–9288, 2018.

Lim, Y. T., Lee, K. Y., Lee, K., and Chung, B. H. Immobilization of histidine-tagged proteins by magnetic nanoparticles encapsulated with nitrilotriacetic acid (nta)-phospholipids micelle. Biochemical and biophysical research communications, 344(3): 926–930, 2006.

Lipshutz, B. H., Ghorai, S., Leong, W. W. Y., Taft, B. R., and Krogstad, D. V. Manipulating micellar environments for enhancing transition metal-catalyzed cross-couplings in water at room temperature. The Journal of organic chemistry, 76(12):5061–5073, 2011.

Littlechild, J. A. Archaeal enzymes and applications in industrial biocatalysts. Archaea, 2015, 2015.

Liu, Y.-F., Zhang, N., Liu, X., Wang, X., Wang, Z.-X., Chen, Y., Yao, H.-W., Ge, M., and Pan, X.-M. Molecular mechanism underlying the interaction of typical sac10b family proteins with dna. PloS one, 7(4):e34986, 2012.

Lo, J., Zheng, T., Hon, S., Olson, D. G., and Lynd, L. R. The bifunctional alcohol and aldehyde dehydrogenase gene, adhe, is necessary for ethanol production in clostridium

thermocellum and thermoanaerobacterium saccharolyticum. Journal of bacteriology, 197(8):1386–1393, 2015.

Lobanov, M. Y., Bogatyreva, N., and Galzitskaya, O. Radius of gyration as an indicator of protein structure compactness. Molecular Biology, 42(4):623–628, 2008.

Longo, L. M., Lee, J., and Blaber, M. Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. Proceedings of the National Academy of Sciences, 110(6):2135–2139, 2013.

Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. The penultimate rotamer library. Proteins: Structure, Function, and Bioinformatics, 40(3):389–408, 2000.

Loxley, A. and Vincent, B. Preparation of poly (methylmethacrylate) microcapsules with liquid cores. Journal of colloid and interface science, 208(1):49–62, 1998.

MacElroy, R. D. Some comments on the evolution of extremophiles. Biosystems, 6(1): 74–75, 1974.

MacFarlane, D. R. and Seddon, K. R. Ionic liquids?progress on the fundamental issues. Australian journal of chemistry, 60(1):3–5, 2007.

MacFarlane, D. R., Chong, A. L., Forsyth, M., Kar, M., Vijayaraghavan, R., Somers, A., and Pringle, J. M. New dimensions in salt–solvent mixtures: a 4th evolution of ionic liquids. Faraday discussions, 206:9–28, 2017.

MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. The journal of physical chemistry B, 102(18):3586–3616, 1998.

Madern, D. and Ebel, C. Influence of an anion-binding site in the stabilization of halophilic malate dehydrogenase from haloarcula marismortui. Biochimie, 89(8):981–987, 2007.

Madern, D., Ebel, C., and Zaccai, G. Halophilic adaptation of enzymes. Extremophiles, 4(2):91–98, 2000.

Mahoney, M. W. and Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. The Journal of Chemical Physics, 112(20):8910–8922, 2000.

Maiorov, V. N. and Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. 1994.

Mancinelli, R., Botti, A., Bruni, F., Ricci, M., and Soper, A. Hydration of sodium, potassium, and chloride ions in solution and the concept of structure maker/breaker. The Journal of Physical Chemistry B, 111(48):13570–13577, 2007.

Mandai, T., Yoshida, K., Ueno, K., Dokko, K., and Watanabe, M. Criteria for solvate ionic liquids. Physical Chemistry Chemical Physics, 16(19):8761–8772, 2014.

Marcus, Y. Thermodynamics of solvation of ions. part 5.?gibbs free energy of hydration at 298.15 k. Journal of the Chemical Society, Faraday Transactions, 87(18):2995–2999, 1991.

Marg, B.-L., Schweimer, K., Sticht, H., and Oesterhelt, D. A two-α-helix extra domain mediates the halophilic character of a plant-type ferredoxin from halophilic archaea. Biochemistry, 44(1):29–39, 2005.

Martínez, L., Andrade, R., Birgin, E. G., and Martínez, J. M. Packmol: a package for building initial configurations for molecular dynamics simulations. Journal of computational chemistry, 30(13):2157–2164, 2009.

McDonald, G. D. and Storrie-Lombardi, M. C. Biochemical constraints in a protobiotic earth devoid of basic amino acids: the ?baa (-) world? Astrobiology, 10(10):989–1000, 2010.

McGuffin, L. J., Bryson, K., and Jones, D. T. The psipred protein structure prediction server. Bioinformatics, 16(4):404–405, 2000.

Meeron, E. Theory of potentials of average force and radial distribution functions in ionic solutions. The Journal of Chemical Physics, 28(4):630–643, 1958.

Melo, F. and Feytmans, E. Assessing protein structures with a non-local atomic interaction energy. Journal of molecular biology, 277(5):1141–1152, 1998.

Mesecar, A. D., Stoddard, B. L., and Koshland, D. E. Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. Science, 277(5323):202–206, 1997.

Mevarech, M., Frolow, F., and Gloss, L. M. Halophilic enzymes: proteins with a grain of salt. Biophysical chemistry, 86(2-3):155–164, 2000.

Meyer, L.-E., Langermann, J., von, and Kragl, U. Recent developments in biocatalysis in multiphasic ionic liquid reaction systems. Biophysical reviews, 10(3):901–910, 2018.

Micae?lo, N. M. and Soares, C. M. Protein structure and dynamics in ionic liquids. insights from molecular dynamics simulation studies. The Journal of Physical Chemistry B, 112 (9):2566–2572, 2008.

Miran, M. S., Yasuda, T., Susan, M. A. B. H., Dokko, K., and Watanabe, M. Binary protic ionic liquid mixtures as a proton conductor: high fuel cell reaction activity and facile proton transport. The Journal of Physical Chemistry C, 118(48):27631–27639, 2014.

Missimer, J. H., Steinmetz, M. O., Baron, R., Winkler, F. K., Kammerer, R. A., Daura, X., and Van Gunsteren, W. F. Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. Protein science, 16(7):1349–1359, 2007.

Monstadt, G. M. and Holldorf, A. Arginine deiminase from halobacterium salinarium. purification and properties. Biochemical journal, 273(3):739–745, 1991.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. Journal of computational chemistry, 30(16):2785–2791, 2009.

Mu, Y., Nguyen, P. H., and Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. Proteins: Structure, Function, and Bioinformatics, 58(1):45–52, 2005.

Mullakhanbhai, M. F. and Larsen, H. Halobacterium volcanii spec. nov., a dead sea halobacterium with a moderate salt requirement. Archives of Microbiology, 104(1): 207–214, 1975.

Müller-Santos, M., Souza, E. M., de, Pedrosa, F. d. O., Mitchell, D. A., Longhi, S., Carrière, F., Canaan, S., and Krieger, N. First evidence for the salt-dependent folding and activity of an esterase from the halophilic archaea haloarcula marismortui. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids, 1791(8): 719–729, 2009.

Musa, M. M., Ziegelmann-Fjeld, K. I., Vieille, C., and Phillips, R. S. Activity and selectivity of w110a secondary alcohol dehydrogenase from thermoanaerobacter ethanolicus in organic solvents and ionic liquids: mono-and biphasic media. Organic & biomolecular chemistry, 6(5):887–892, 2008.

Nallani, M., De Hoog, H.-P. M., Cornelissen, J. J., Palmans, A. R., Hest, J. C., van, and Nolte, R. J. Polymersome nanoreactors for enzymatic ring-opening polymerization. Biomacromolecules, 8(12):3723–3728, 2007.

Nestl, B. M., Nebel, B. A., and Hauer, B. Recent progress in industrial biocatalysis. Current opinion in chemical biology, 15(2):187–193, 2011.

Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N. S., Thorsson, V., Sbrogna, J., et al. Genome sequence of halobacterium species nrc-1. Proceedings of the National Academy of Sciences, 97(22):12176–12181, 2000.

Noé, F., Horenko, I., Schütte, C., and Smith, J. C. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. The Journal of chemical physics, 126(15):04B617, 2007.

Nordwald, E. M. Engineering ionic liquid tolerant enzymes. 2015.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.-J., Hattori, M., Kanai, A., Atomi, H., et al. Insights into the evolution of archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic acids research, 39(8):3204–3223, 2010.

Omta, A. W., Kropman, M. F., Woutersen, S., and Bakker, H. J. Negligible effect of ions on the hydrogen-bond structure in liquid water. Science, 301(5631):347–349, 2003.

Onuchic, J. Zl schulten and pg wolynes: Ann. Rev. Phys. Chem, 48:545–600, 1997.

Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. Journal of computational chemistry, 25(13):1656–1676, 2004.

Oren, A. Industrial and environmental applications of halophilic microorganisms. Environmental technology, 31(8-9):825–834, 2010.

Ortega, G., Laín, A., Tadeo, X., López-Méndez, B., Castano, D., and Millet, O. Halophilic enzyme activation induced by salts. Scientific reports, 1:6, 2011.

Ortega, G., Diercks, T., and Millet, O. Halophilic protein adaptation results from synergistic residue-ion interactions in the folded and unfolded states. Chemistry & biology, 22(12):1597–1607, 2015.

Ortenberg, R., Rozenblatt-Rosen, O., and Mevarech, M. The extremely halophilic archaeon haloferax volcanii has two very different dihydrofolate reductases. Molecular microbiology, 35(6):1493–1505, 2000.

Ovchinnikov, S., Kamisetty, H., and Baker, D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. Elife, 3: e02030, 2014.

Pace, C. N. and Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. Biophysical journal, 75(1):422–427, 1998.

Pace, C. N., Scholtz, J. M., and Grimsley, G. R. Forces stabilizing proteins. FEBS letters, 588(14):2177–2184, 2014.

Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., and De Gioia, L. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. Journal of molecular graphics and modelling, 27(8):889–899, 2009.

Parfrey, L. W., Lahr, D. J., Knoll, A. H., and Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proceedings of the National Academy of Sciences, 108(33):13624–13629, 2011.

Park, J.-T., Song, H.-N., Jung, T.-Y., Lee, M.-H., Park, S.-G., Woo, E.-J., and Park, K.-H. A novel domain arrangement in a monomeric cyclodextrin-hydrolyzing enzyme from the hyperthermophile pyrococcus furiosus. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1834(1):380–386, 2013.

Parra-Cruz, R., Jger, C. M., Lau, P. L., Gomes, R. L., and Pordea, A. Rational design of thermostable carbonic anhydrase mutants using molecular dynamics simulations. The Journal of Physical Chemistry B, 122(36):8526–8536, 2018.

Perro, A., Nicolet, C., Angly, J., Lecommandoux, S., Le Meins, J.-F., and Colin, A. Mastering a double emulsion in a simple co-flow microfluidic to generate complex polymersomes. Langmuir, 27(14):9034–9042, 2010.

Peters, M. B., Yang, Y., Wang, B., Fsti-Molnr, L., Weaver, M. N., and Merz Jr, K. M. Structural survey of zinc-containing proteins and development of the zinc amber force field (zaff). Journal of chemical theory and computation, 6(9):2935–2947, 2010.

Peters, R. J., Marguet, M., Marais, S., Fraaije, M. W., Van Hest, J. C., and Lecommandoux, S. Cascade reactions in multicompartmentalized polymersomes. Angewandte Chemie International Edition, 53(1):146–150, 2014.

Pikal-Cleland, K. A., Cleland, J. L., Anchordoquy, T. J., and Carpenter, J. F. Effect of glycine on ph changes and protein stability during freeze–thawing in phosphate buffer systems. Journal of pharmaceutical sciences, 91(9):1969–1979, 2002.

Pike, S. J., Hutchinson, J. J., and Hunter, C. A. H-bond acceptor parameters for anions. Journal of the American Chemical Society, 139(19):6700–6706, 2017.

Pike, S. J., Lavagnini, E., Varley, L. M., Cook, J. L., and Hunter, C. A. H-bond donor parameters for cations. Chemical Science, 2019.

Pire, C., Esclapez, J., Diaz, S., Perez-Pomares, F., Ferrer, J., and Bonete, M. Alteration of coenzyme specificity in halophilic nad (p)+ glucose dehydrogenase by site-directed mutagenesis. Journal of Molecular Catalysis B: Enzymatic, 59(4):261–265, 2009.

Plapp, B. V. Conformational changes and catalysis by alcohol dehydrogenase. Archives of biochemistry and biophysics, 493(1):3–12, 2010.

Platts, L. and Falconer, R. J. Controlling protein stability: Mechanisms revealed using formulations of arginine, glycine and guanidinium hcl with three globular proteins. International journal of pharmaceutics, 486(1-2):131–135, 2015.

Powers, E. T. and Powers, D. L. A perspective on mechanisms of protein tetramer formation. Biophysical journal, 85(6):3587–3599, 2003.

Prabhu, N. V. and Sharp, K. A. Heat capacity in proteins. Annu. Rev. Phys. Chem., 56: 521–548, 2005.

Pramanik, S., Dhoke, G. V., Jaeger, K.-E., Schwaneberg, U., and D. Davari, M. How to engineer ionic liquids resistant enzymes? insights from combined molecular dynamics and directed evolution study. ACS Sustainable Chemistry & Engineering, 2019.

Privalov, P. L. and Makhatadze, G. I. Contribution of hydration to protein folding thermodynamics: Ii. the entropy and gibbs energy of hydration. Journal of molecular biology, 232(2):660–679, 1993.

Qvist, J., Ortega, G., Tadeo, X., Millet, O., and Halle, B. Hydration dynamics of a halophilic protein in folded and unfolded states. The Journal of Physical Chemistry B, 116(10):3436–3444, 2012.

Rabideau, B. D. and Ismail, A. E. Mechanisms of hydrogen bond formation between ionic liquids and cellulose and the influence of water content. Physical Chemistry Chemical Physics, 17(8):5767–5775, 2015.

Ramachandran, G. T. and Sasisekharan, V. Conformation of polypeptides and proteins. In Advances in protein chemistry, volume 23, pages 283–437. Elsevier, 1968.

Rao, F. Local transition gradients indicating the global attributes of protein energy landscapes. The Journal of Physical Chemistry Letters, 1(10):1580–1583, 2010.

Rao, S. T. and Rossmann, M. G. Comparison of super-secondary structures in proteins. Journal of molecular biology, 76(2):241–256, 1973.

Reed, C. J., Lewis, H., Trejo, E., Winston, V., and Evilia, C. Protein adaptations in archaeal extremophiles. Archaea, 2013, 2013.

Reid, M. F. and Fewson, C. A. Molecular characterization of microbial alcohol dehydrogenases. Critical reviews in microbiology, 20(1):13–56, 1994.

Remmert, M., Biegert, A., Hauser, A., and Söding, J. Hhblits: lightning-fast iterative protein sequence searching by hmm-hmm alignment. Nature methods, 9(2):173, 2012.

Ribot, J. C., Guerrero-Sanchez, C., Greaves, T. L., Kennedy, D. F., Hoogenboom, R., and Schubert, U. S. Amphiphilic oligoether-based ionic liquids as functional materials for thermoresponsive ion gels with tunable properties via aqueous gelation. Soft Matter, 8 (4):1025–1032, 2012.

Richard, S. B., Madern, D., Garcin, E., and Zaccai, G. Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 å resolution structures of the wild type and a mutant of malate dehydrogenase from haloarcula marismortui. Biochemistry, 39(5):992–1000, 2000.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature, 499(7459):431, 2013.

Roca, M., Liu, H., Messer, B., and Warshel, A. On the relationship between thermal stability and catalytic power of enzymes. Biochemistry, 46(51):15076–15088, 2007.

Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. Protein structure prediction using rosetta. In Methods in enzymology, volume 383, pages 66–93. Elsevier, 2004.

Rossi, M., Ciaramella, M., Cannio, R., Pisani, F. M., Moracci, M., and Bartolucci, S. Extremophiles 2002. Journal of bacteriology, 185(13):3683–3689, 2003.

Rost, B. Twilight zone of protein sequence alignments. Protein engineering, 12(2):85–94, 1999.

Roy, A., Kucukural, A., and Zhang, Y. I-tasser: a unified platform for automated protein structure and function prediction. Nature protocols, 5(4):725, 2010.

Rupley, J. A. and Careri, G. Protein hydration and function. In Advances in protein chemistry, volume 41, pages 37–172. Elsevier, 1991.

Šali, A. and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology, 234(3):779–815, 1993.

Sanger, F., Nicklen, S., and Coulson, A. R. Dna sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences, 74(12):5463–5467, 1977.

Santos, A. R., Hanson-Heine, M. W., Besley, N. A., and Licence, P. The impact of sulfur functionalisation on nitrogen-based ionic liquid cations. Chemical communications, 54 (81):11403–11406, 2018.

Sawle, L. and Ghosh, K. How do thermophilic proteins and proteomes withstand high temperature? Biophysical journal, 101(1):217–227, 2011.

Schindl, A., Hagen, M. L., Muzammal, S., Gunasekera, H. A., and Croft, A. K. Proteins in ionic liquids: Reactions, applications, and futures. Frontiers in chemistry, 7, 2019.

Schoonen, L. and Hest, J. C., van. Compartmentalization approaches in soft matter science: from nanoreactor development to organelle mimics. Advanced Materials, 28 (6):1109–1128, 2016.

Schröder, C. Proteins in ionic liquids: current status of experiments and simulations. In Ionic Liquids II, pages 127–152. Springer, 2017.

Schröder, C., Rudas, T., Neumayr, G., Benkner, S., and Steinhauser, O. On the collective network of ionic liquid/water mixtures. i. orientational structure. The Journal of chemical physics, 127(23):234503, 2007.

Schröder, U., Wadhawan, J. D., Compton, R. G., Marken, F., Suarez, P. A., Consorti, C. S., Souza, R. F., de, and Dupont, J. Water-induced accelerated ion diffusion: voltammetric studies in 1-methyl-3-[2, 6-(s)-dimethylocten-2-yl] imidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate and hexafluorophosphate ionic liquids. New Journal of Chemistry, 24(12):1009–1015, 2000.

Schuliger, J., Brown, S., Baross, J., and Kelly, R. Purification and characterization of a novel amylolytic enzyme from es 4, a marine hyperthermophilic archaeum. Molecular Marine Biology and Biotechnology, 2(2):76–87, 1993.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. Swiss-model: an automated protein homology-modeling server. Nucleic acids research, 31(13):3381–3385, 2003.

Selkoe, D. J. The cell biology of β-amyloid precursor protein and presenilin in alzheimer's disease. Trends in cell biology, 8(11):447–453, 1998.

Sheldon, R. A. Enzyme immobilization: the quest for optimum performance. Advanced Synthesis & Catalysis, 349(8-9):1289–1307, 2007.

Sheldon, R. A. Characteristic features and biotechnological applications of cross-linked enzyme aggregates (cleas). Applied microbiology and biotechnology, 92(3):467–477, 2011.

Shi, J., Blundell, T. L., and Mizuguchi, K. Fugue: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. Journal of molecular biology, 310(1):243–257, 2001.

Shi, X.-A., Zong, M.-H., and Lou, W.-Y. Effect of ionic liquids on catalytic characteristics of horse liver alcohol dehydrogenase. Chinese Journal of Chemistry, 24(11):1643–1647, 2006.

Shibue, R., Sasamoto, T., Shimada, M., Zhang, B., Yamagishi, A., and Akanuma, S. Comprehensive reduction of amino acid set in a protein suggests the importance of prebiotic amino acids for stable proteins. Scientific Reports, 8(1):1227, 2018. doi: 10.1038/s41598-018-19561-1. URL `https://doi.org/10.1038/s41598-018-19561-1`.

Shukla, M., Srivastava, N., and Saha, S. Theoretical and spectroscopic studies of 1-butyl-3-methylimidazolium iodide room temperature ionic liquid: Its differences with chloride and bromide derivatives. Journal of Molecular Structure, 975(1-3):349–356, 2010.

Siddaramappa, S., Challacombe, J. F., DeCastro, R. E., Pfeiffer, F., Sastre, D. E., Giménez, M. I., Paggi, R. A., Detter, J. C., Davenport, K. W., Goodwin, L. A.,

et al. A comparative genomics perspective on the genetic content of the alkaliphilic haloarchaeon natrialba magadii atcc 43099 t. BMC genomics, 13(1):165, 2012.

Siglioccolo, A., Paiardini, A., Piscitelli, M., and Pascarella, S. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. BMC structural biology, 11(1):50, 2011.

Sippl, M. J. Calculation of conformational ensembles from potentials of mena force: an approach to the knowledge-based prediction of local structures in globular proteins. Journal of molecular biology, 213(4):859–883, 1990.

Skerra, A. and Schmidt, T. G. [18] use of the strep-tag and streptavidin for detection and purification of recombinant proteins. In Methods in enzymology, volume 326, pages 271–304. Elsevier, 2000.

So, S. and Lodge, T. P. Size control and fractionation of ionic liquid filled polymersomes with glassy and rubbery bilayer membranes. Langmuir, 32(19):4959–4968, 2016.

Söding, J. Protein homology detection by hmm–hmm comparison. Bioinformatics, 21(7): 951–960, 2004.

Somero, G. N. Temperature adaptation of enzymes: biological optimization through structure-function compromises. Annual Review of Ecology and Systematics, 9(1):1–29, 1978.

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., Eijk, R., van, Schleper, C., Guy, L., and Ettema, T. J. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature, 521(7551):173, 2015.

Sprenger, K., Jaeger, V. W., and 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, 119(18):5882–5895, 2015.

Sterner, R. h. and Liebl, W. Thermophilic adaptation of proteins. Critical Reviews in Biochemistry and Molecular Biology, 36(1):39–106, 2001.

Sterpone, F., Stirnemann, G., and Laage, D. Magnitude and molecular origin of water slowdown next to a protein. Journal of the American Chemical Society, 134(9):4116–4119, 2012.

Stillinger, F. H. and Rahman, A. Improved simulation of liquid water by molecular dynamics. The Journal of Chemical Physics, 60(4):1545–1557, 1974.

Strate, A., Neumann, J., Overbeck, V., Bonsa, A.-M., Michalik, D., Paschek, D., and Ludwig, R. Rotational and translational dynamics and their relation to hydrogen bond lifetimes in an ionic liquid by means of nmr relaxation time experiments and molecular dynamics simulation. The Journal of chemical physics, 148(19):193843, 2018.

Street, A. G. and Mayo, S. L. Intrinsic β-sheet propensities result from van der waals interactions between side chains and the local backbone. Proceedings of the national academy of sciences, 96(16):9074–9076, 1999.

Strillinger, E., Grötzinger, S. W., Allers, T., Eppinger, J., and Weuster-Botz, D. Production of halophilic proteins using haloferax volcanii h1895 in a stirred-tank bioreactor. Applied microbiology and biotechnology, 100(3):1183–1195, 2016.

Stroud, A., Liddell, S., and Allers, T. Genetic and biochemical identification of a novel single-stranded dna-binding complex in haloferax volcanii. Frontiers in microbiology, 3:224, 2012.

Synek, V. Evaluation of the standard deviation from duplicate results. Accreditation and Quality Assurance, 13(6):335–337, 2008.

Taherzadeh, M. J. and Karimi, K. Enzyme-based hydrolysis processes for ethanol from lignocellulosic materials: a review. BioResources, 2(4):707–738, 2007.

Tariq, M., Ali, S., and Khalid, N. Activity of homogeneous and heterogeneous catalysts, spectroscopic and chromatographic characterization of biodiesel: A review. Renewable and Sustainable Energy Reviews, 16(8):6303–6316, 2012a.

Tariq, M., Freire, M. G., Saramago, B., Coutinho, J. A., Lopes, J. N. C., and Rebelo, L. P. N. Surface tension of ionic liquids and ionic liquid solutions. Chemical Society Reviews, 41(2):829–868, 2012b.

Taupin, C. M.-J., Härtlein, M., and Leberman, R. Seryl-trna synthetase from the extreme halophile haloarcula marismortui: Isolation, characterization and sequencing of the gene and its expression in escherichia coli. European journal of biochemistry, 243(1-2):141–150, 1997.

Theorell, H. and Chance, B. Studies on liver alcohol dehydrogenase 2 the kinetics of the compound of horse liver alcohol dehydrogenase and reduced diphosphopyridine nucleotide. Acta Chemica Scandinavica, 5(7-8):1127–1144, 1951.

Thiel, B. L. and Toth, M. Secondary electron contrast in low-vacuum/ environmental scanning electron microscopy of dielectrics. Journal of applied physics, 97(5):4, 2005.

Thomas, V., Namdeo, M., Murali Mohan, Y., Bajpai, S., and Bajpai, M. Review on polymer, hydrogel and microgel metal nanocomposites: A facile nanotechnological approach. Journal of Macromolecular Science, Part A: Pure and Applied Chemistry, 45 (1):107–119, 2007.

Thompson, M. P., Peñafiel, I., Cosgrove, S. C., and Turner, N. J. Biocatalysis using immobilized enzymes in continuous flow for the synthesis of fine chemicals. Organic Process Research & Development, 23(1):9–18, 2018.

Timpson, L. M., Liliensiek, A.-K., Alsafadi, D., Cassidy, J., Sharkey, M. A., Liddell, S., Allers, T., and Paradisi, F. A comparison of two novel alcohol dehydrogenase enzymes (adh1 and adh2) from the extreme halophile haloferax volcanii. Applied microbiology and biotechnology, 97(1):195–203, 2013.

Tom, L. I., Jorge, M., Gomes, J. R., and Coutinho, J. A. Molecular dynamics simulation studies of the interactions between ionic liquids and amino acids in aqueous solution. The Journal of Physical Chemistry B, 116(6):1831–1842, 2012.

Toral, A. R., Antonia, P., Hernández, F. J., Janssen, M. H., Schoevaart, R., Rantwijk, F., van, and Sheldon, R. A. Cross-linked candida antarctica lipase b is active in denaturing ionic liquids. Enzyme and Microbial Technology, 40(5):1095–1099, 2007.

Trieu, H. and Qutubuddin, S. Polyvinyl alcohol hydrogels i. microscopic structure by

freeze-etching and critical point drying techniques. Colloid and Polymer Science, 272 (3):301–309, 1994.

Trott, O. and Olson, A. J. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry, 31(2):455–461, 2010.

Valencia, A. and Pazos, F. Computational methods for the prediction of protein interactions. Current opinion in structural biology, 12(3):368–373, 2002.

Vossenberg, J. L., van de, Driessen, A. J., and Konings, W. N. The essence of being extremophilic: the role of the unique archaeal membrane lipids. Extremophiles, 2(3): 163–170, 1998.

Gunsteren, W., van, Billeter, S., Eising, A., Hünenberger, P., Krüger, P., Mark, A., Scott, W., and Tironi, I. Biomolecular simulations: the gromos96 manual and user guide biomos bv. Züurich, Groningen, 1996.

Gunsteren, W. F., van, Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glättli, A., Hünenberger, P. H., et al. Biomolecular modeling: goals, problems, perspectives. Angewandte Chemie International Edition, 45(25):4064–4092, 2006.

Rantwijk, F., van and Sheldon, R. A. Biocatalysis in ionic liquids. Chemical reviews, 107 (6):2757–2785, 2007.

Vega, C., Abascal, J. L., Conde, M., and Aragones, J. What ice can teach us about water interactions: a critical comparison of the performance of different water models. Faraday discussions, 141:251–276, 2009.

Verlet, L. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. Physical review, 159(1):98, 1967.

Vieille, C. and Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol. Mol. Biol. Rev., 65(1):1–43, 2001.

Vijayaraghavan, R., Izgorodin, A., Ganesh, V., Surianarayanan, M., and MacFarlane, D. R. Long-term structural and chemical stability of dna in hydrated ionic liquids. Angewandte Chemie International Edition, 49(9):1631–1633, 2010.

Villar-Garcia, I. J., Lovelock, K. R., Men, S., and Licence, P. Tuning the electronic environment of cations and anions using ionic liquid mixtures. Chemical Science, 5(6): 2573–2579, 2014.

Volkamer, R., Jimenez, J. L., San Martini, F., Dzepina, K., Zhang, Q., Salcedo, D., Molina, L. T., Worsnop, D. R., and Molina, M. J. Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected. Geophysical Research Letters, 33(17), 2006.

Vriezema, D., Aragones, M., Elemans, J., and Cornelissen, J. Lm; rowan, ae; nolte, rjm. Chem. Rev, 105:1445, 2005.

Wales, D. Energy landscapes (cambridge molecular science), 2003.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. Development and testing of a general amber force field. Journal of computational chemistry, 25(9): 1157–1174, 2004.

Wang, J., Wang, W., Kollman, P. A., and Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. Journal of molecular graphics and modelling, 25(2):247–260, 2006.

Warden, A. C., Williams, M., Peat, T. S., Seabrook, S. A., Newman, J., Dojchinov, G., and Haritos, V. S. Rational engineering of a mesohalophilic carbonic anhydrase to an extreme halotolerant biocatalyst. Nature communications, 6:10278, 2015.

Welton, T. Ionic liquids: a brief history. Biophysical reviews, 10(3):691–706, 2018.

Wendoloski, D., Ferrer, C., and Dyall-Smith, M. A new simvastatin (mevinolin)-resistance marker from haloarcula hispanica and a new haloferax volcanii strain cured of plasmid phv2. Microbiology, 147(4):959–964, 2001.

Wiedenhofer, D., Fishman, T., Lauk, C., Haas, W., and Krausmann, F. Integrating material stock dynamics into economy-wide material flow accounting: Concepts, modelling, and global application for 1900–2050. Ecological Economics, 156:121–133, 2019.

Woese, C. R. and Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences, 74(11):5088–5090, 1977.

Woese, C. R., Kandler, O., and Wheelis, M. L. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. Proceedings of the National Academy of Sciences, 87(12):4576–4579, 1990.

Wojnarowska, Z. and Paluch, M. Recent progress on dielectric properties of protic ionic liquids. Journal of Physics: Condensed Matter, 27(7):073202, 2015.

Woodley, J. M. Accelerating the implementation of biocatalysis in industry. Applied microbiology and biotechnology, 103(12):4733–4739, 2019a.

Woodley, J. M. Reaction engineering for the industrial implementation of biocatalysis. Topics in Catalysis, pages 1–6, 2019b.

Wu, S. and Zhang, Y. Lomets: a local meta-threading-server for protein structure prediction. Nucleic acids research, 35(10):3375–3382, 2007.

Wu, S. and Zhang, Y. Muster: improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins: Structure, Function, and Bioinformatics, 72(2):547–556, 2008.

Wu, S., Skolnick, J., and Zhang, Y. Ab initio modeling of small proteins by iterative tasser simulations. BMC biology, 5(1):17, 2007.

Xie, W. and Sahinidis, N. V. Residue-rotamer-reduction algorithm for the protein side-chain conformation problem. Bioinformatics, 22(2):188–194, 2005.

Xu, Y. and Xu, D. Protein threading using prospect: design and evaluation. Proteins: Structure, Function, and Bioinformatics, 40(3):343–354, 2000.

Yang, J. and Zhang, Y. Protein structure and function prediction using i-tasser. Current protocols in bioinformatics, 52(1):5–8, 2015.

Yang, Z. The power of phylogenetic comparison in revealing protein function. Proceedings of the National Academy of Sciences, 102(9):3179–3180, 2005.

Yang, Z. Hofmeister effects: an explanation for the impact of ionic liquids on biocatalysis. Journal of biotechnology, 144(1):12–22, 2009.

Zangi, R. Can salting-in/salting-out ions be classified as chaotropes/kosmotropes? The Journal of Physical Chemistry B, 114(1):643–650, 2009.

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature, 541 (7637):353, 2017.

Zhang, G. and Ge, H. Protein hypersaline adaptation: insight from amino acids with machine learning algorithms. The protein journal, 32(4):239–245, 2013.

Zhang, S., Zhou, Q., Lu, X., Song, Y., and Wang, X. Physicochemical properties of ionic liquid mixtures. Springer, 2017.

Zhang, T., Datta, S., Eichler, J., Ivanova, N., Axen, S. D., Kerfeld, C. A., Chen, F., Kyrpides, N., Hugenholtz, P., Cheng, J.-F., et al. Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance. Green Chemistry, 13 (8):2083–2090, 2011.

Zhang, Y. and Skolnick, J. Spicker: a clustering approach to identify near-native protein folds. Journal of computational chemistry, 25(6):865–871, 2004.

Zhang, Y., Kihara, D., and Skolnick, J. Local energy landscape flattening: parallel hyperbolic monte carlo sampling of protein folding. Proteins: Structure, Function, and Bioinformatics, 48(2):192–201, 2002.

Zhang, Y., Huang, X., and Li, Y. Negative effect of [bmim][pf6] on the catalytic activity of alcohol dehydrogenase: mechanism and prevention. Journal of Chemical Technology &

Biotechnology: International Research in Process, Environmental & Clean Technology, 83(9):1230–1235, 2008.

Zhou, H. and Zhou, Y. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. Proteins: Structure, Function, and Bioinformatics, 55(4):1005–1013, 2004.

Zhou, H. and Zhou, Y. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. Proteins: Structure, Function, and Bioinformatics, 58(2):321–328, 2005.

Zhu, D. and Hua, L. How carbonyl reductases control stereoselectivity: approaching the goal of rational design. Pure and Applied Chemistry, 82(1):117–128, 2010.

Chapter 8

Appendix

8.1 Free Energy Landscapes at 0.15 M Ionic liquid

In the following free energy landscapes of all ionic liquid systems of concentration 0.15M are shown in an overview for comparison of RMSD and radius of gyration distribution. The transitioning paths through the conformational space are unlike the native system not characterised by a straight descend towards energy minimum, but show distorted energy wells and an unbalanced exploration torn between RMSD and radius of gyration. The change in radius of gyration from starting structures towards low energy structures is significantly increased in all ionic liquids compared to the buffer system and thereby indicates a loss in compactness of the structure during the exploration of configurational space. All structures converge towards a lower RMSD, which is to be expected for lower energy configurations and end up at around 0.2 - 0.18 Å. The FEL of the system ILA $[Me_3S][OSO_3Me]$ is very similar to the FEL of the native system. The trough shows an even smoother descend, but is about a third increased in radius for ILA and the initial conformation needs to cross a high energy barrier for the descend.

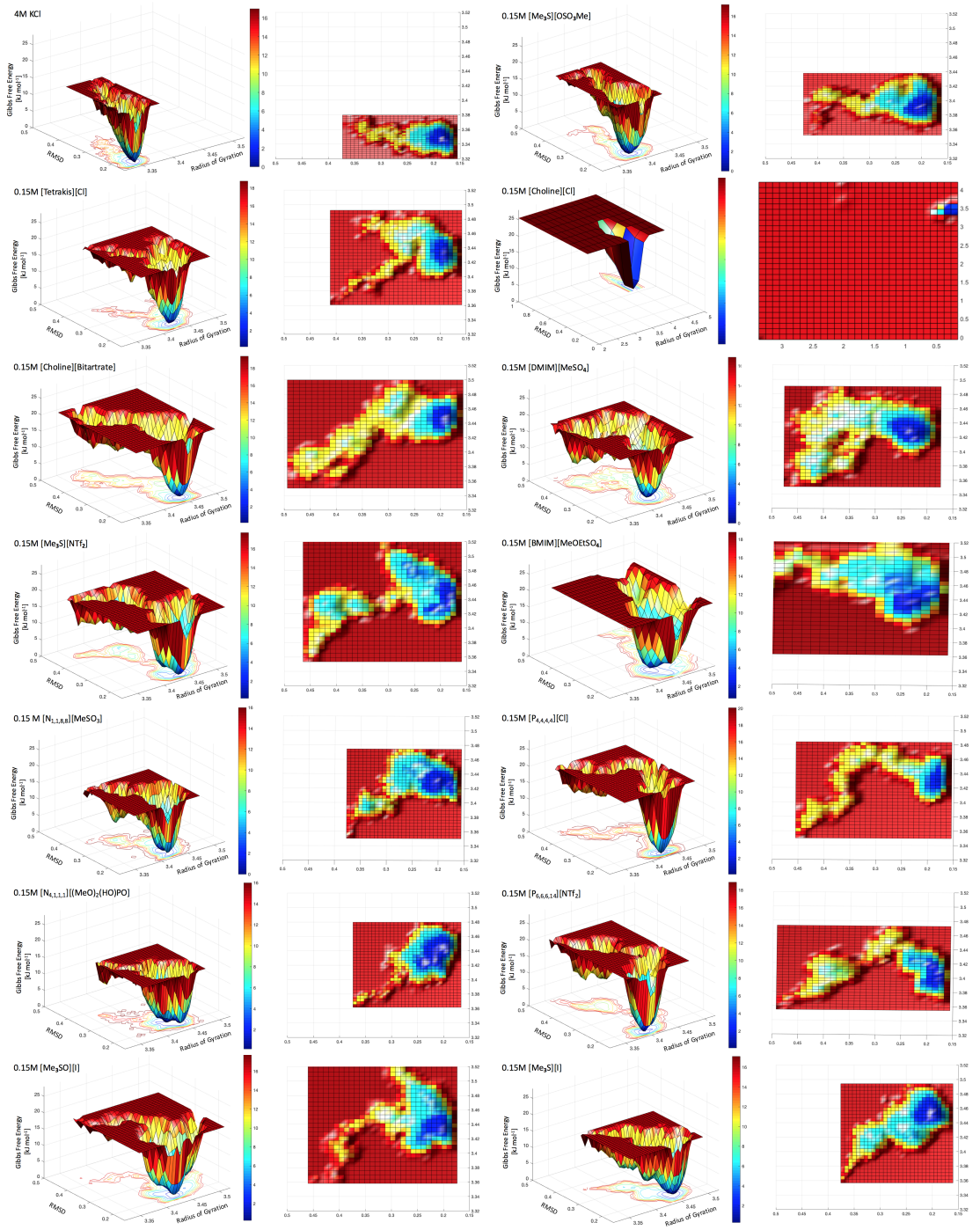


Figure 8.1: Side view and top view of FEL of Ionic liquids (A to M, in this order), from left to right and top to bottom: 4M KCl, $[Me_3S][OSO_3Me]$, [Tetrakis][Cl], [Choline][Cl], [Choline][Bitartrate], $[EMIM][MeSO_4]$, $[Me_3S][NTf_2]$, $[BMIM][MeOEtSO_4]$, $[N_{1,1,8,8}][MeSO_3]$, $[P_{4,4,4,4}][Cl]$, $[N_{4,1,1,1}][(MeO)_2(HO)PO]$, $[P_{6,6,6,14}][NTf_2]$, $[Me_3SO][I]$, $[Me_3S][I]$

8.1.1 Unscaled Free Energy Landscapes

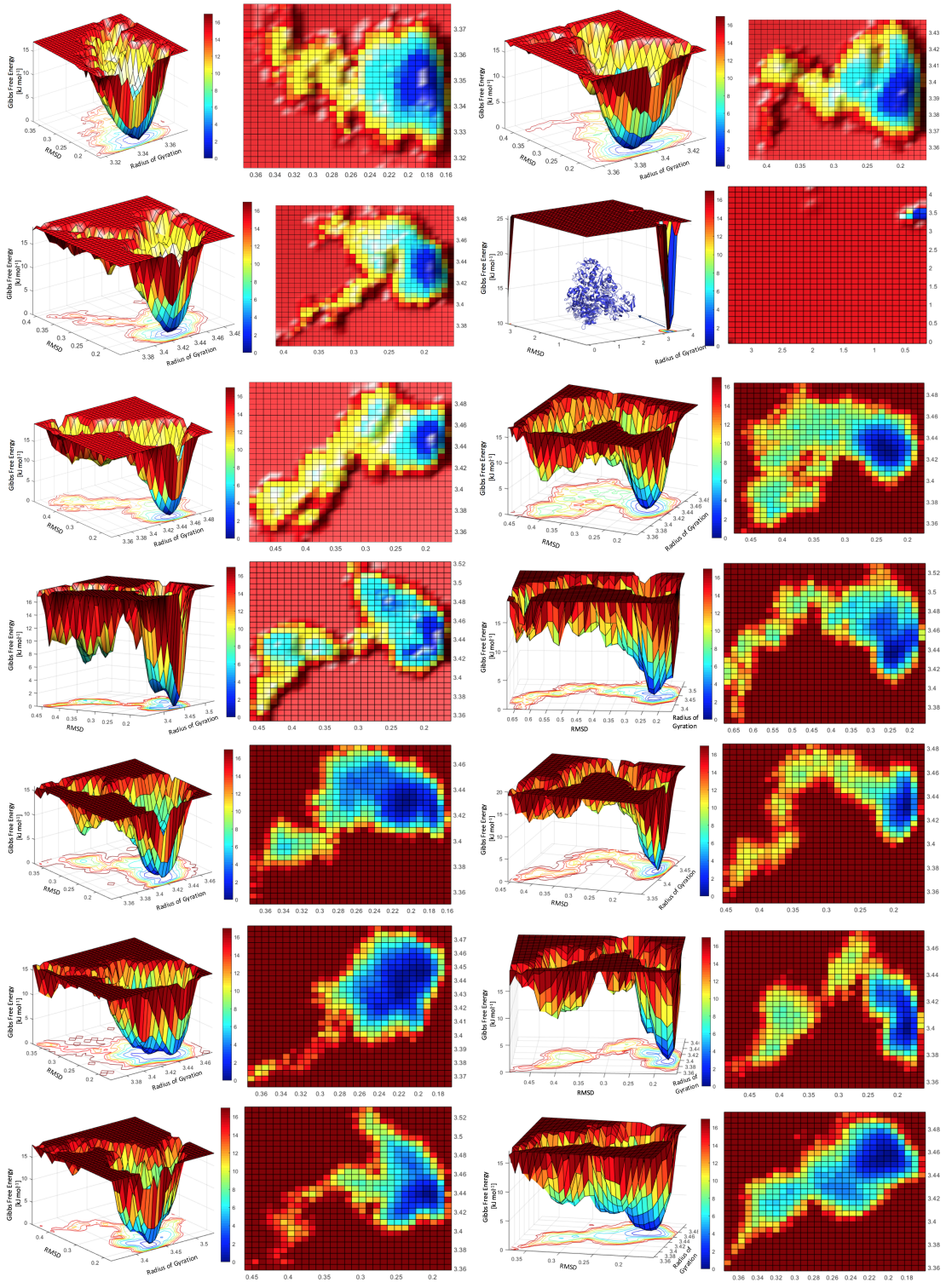


Figure 8.2: Side view and top view of FEL of Ionic liquids (A to M, in this order), from left to right and top to bottom: 4M KCl, $[Me_3S][OSO_3Me]$, [Tetrakis][Cl], [Choline][Cl], [Choline][Bitartrate], $[DiMIM][MeSO_4]$, $[Me_3S][NTf_2]$, $[BMIM][MeOEtSO_4]$, $[N_{1,1,8,8}][MeSO_3]$, $[P_{4,4,4,4}][Cl]$, $[N_{4,1,1,1}][(MeO)_2(HO)PO]$, $[P_{6,6,6,14}][NTf_2]$, $[Me_3SO][I]$, $[Me_3S][I]$

8.1.2 Free Energy Landscapes - Comparison of Concentrations

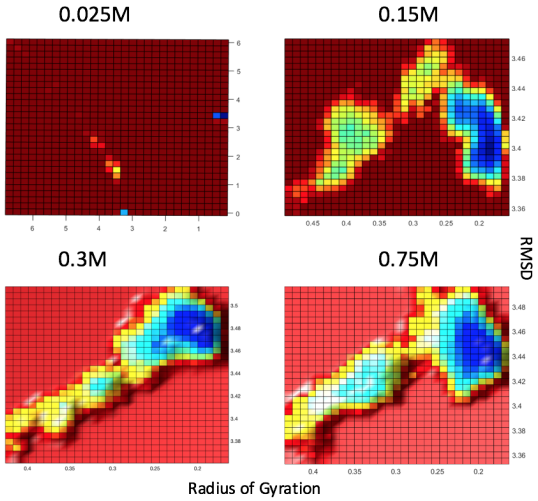


Figure 8.3: Comparison of ILK $[P_{6,6,6,14}][NTf_2]$ FELs at different concentrations. At 0.025M the enzyme showed the same striking behaviour as seen in 0.15M ILC. The FEL at 0.75M ILK shows clearly a very high energy barrier between the starting conformation and the lowest energy minimum, whilst this is absent at 0.3M, despite the otherwise similar path of exploration.

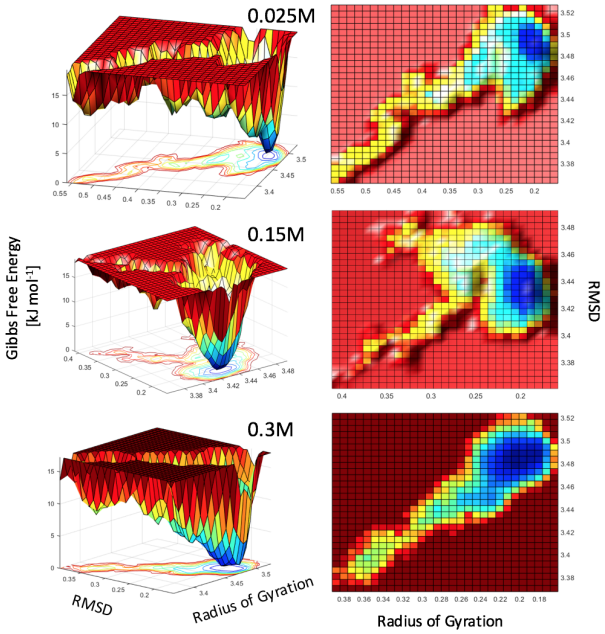


Figure 8.4: Concentrations 0.025 and 0.15M ILB show long and winding paths towards free energy minimum, whereas at 0.3 M an exceptionally straight behaviour is observed

8.2 Relative Specific Activity Assays ILA - ILM

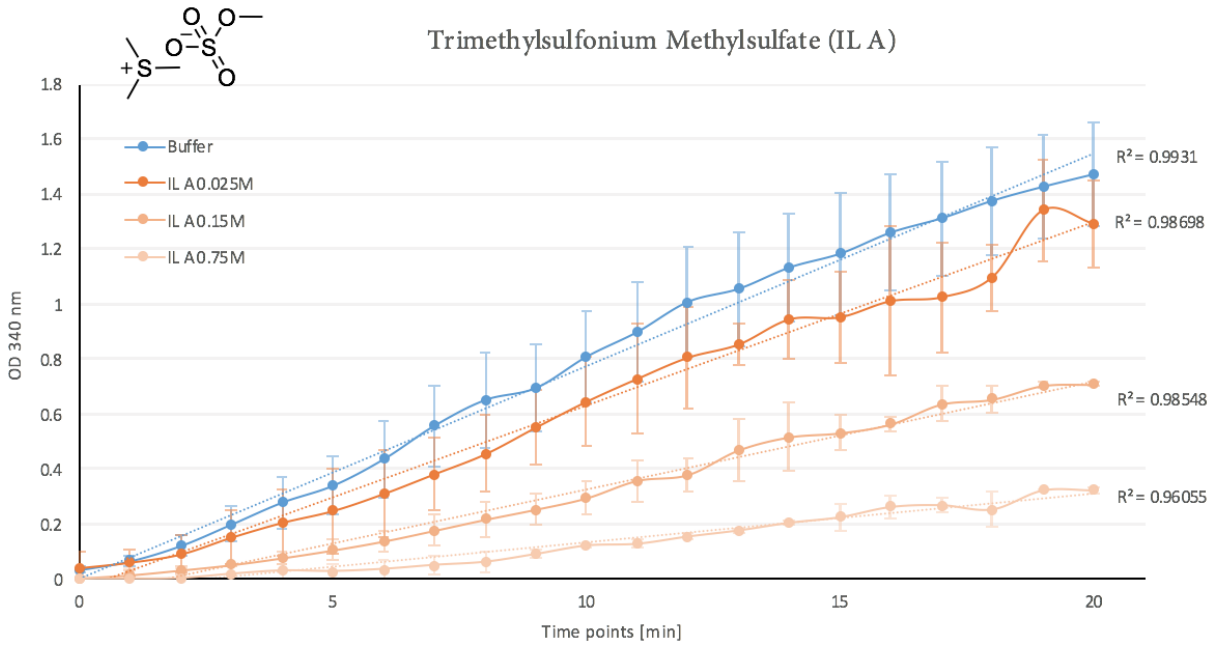


Figure 8.5: IL A [Me_3S][OSO_3Me]: All IL-mixture concentrations were lower in activity, when compared to the reference buffer and activity decreased with increasing IL concentration. Trends were mainly linear, with one outlier for ILA 0.025 M.

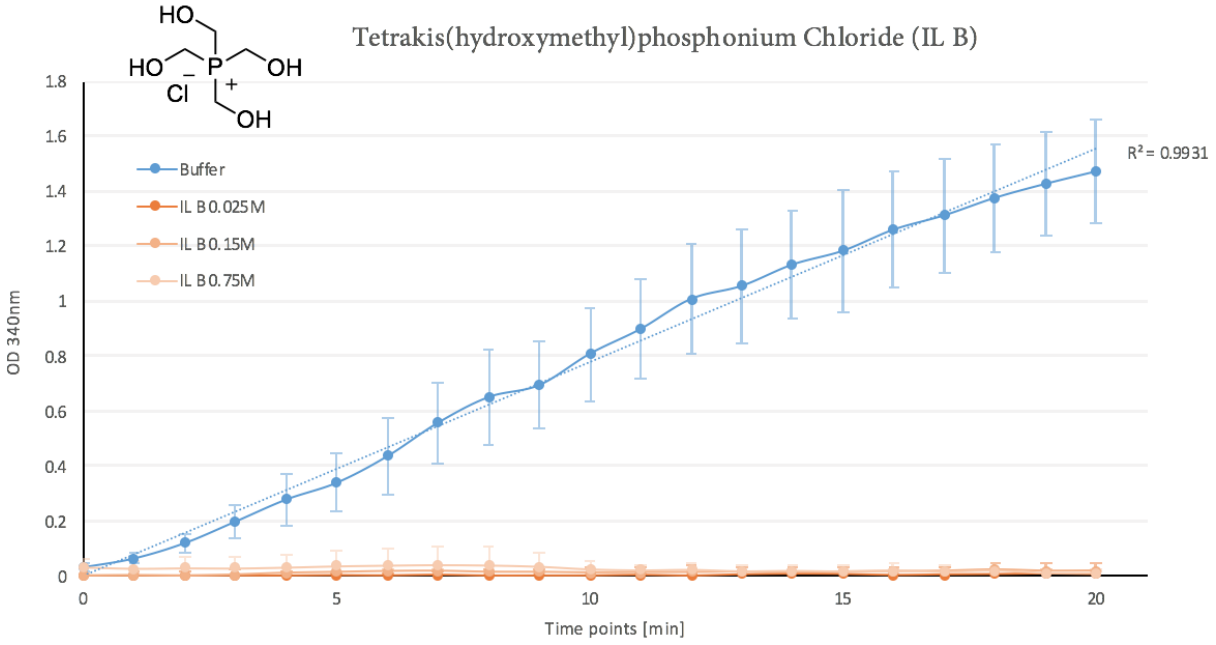


Figure 8.6: ILB [Tetrakis][Cl]: The enzyme showed no activity in any of the IL-mixtures.

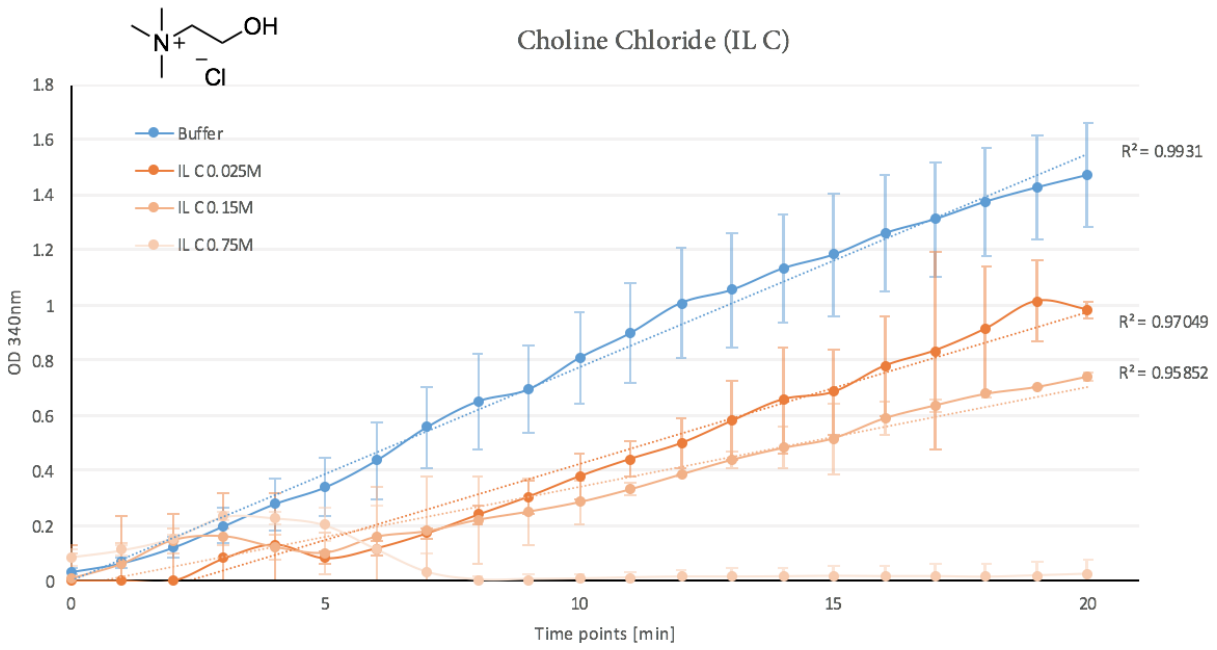


Figure 8.7: ILC [Choline][Cl]: All IL-mixture concentrations were lower in activity, when compared to the reference buffer, whereby ILC 0.75 M resulted in no activity. Trends for ILC 0.025 and 0.15 M were mostly linear. For ILC 0.75 M an increase in absorption is detected over the first 7 minutes, which is again reduced to zero thereafter.

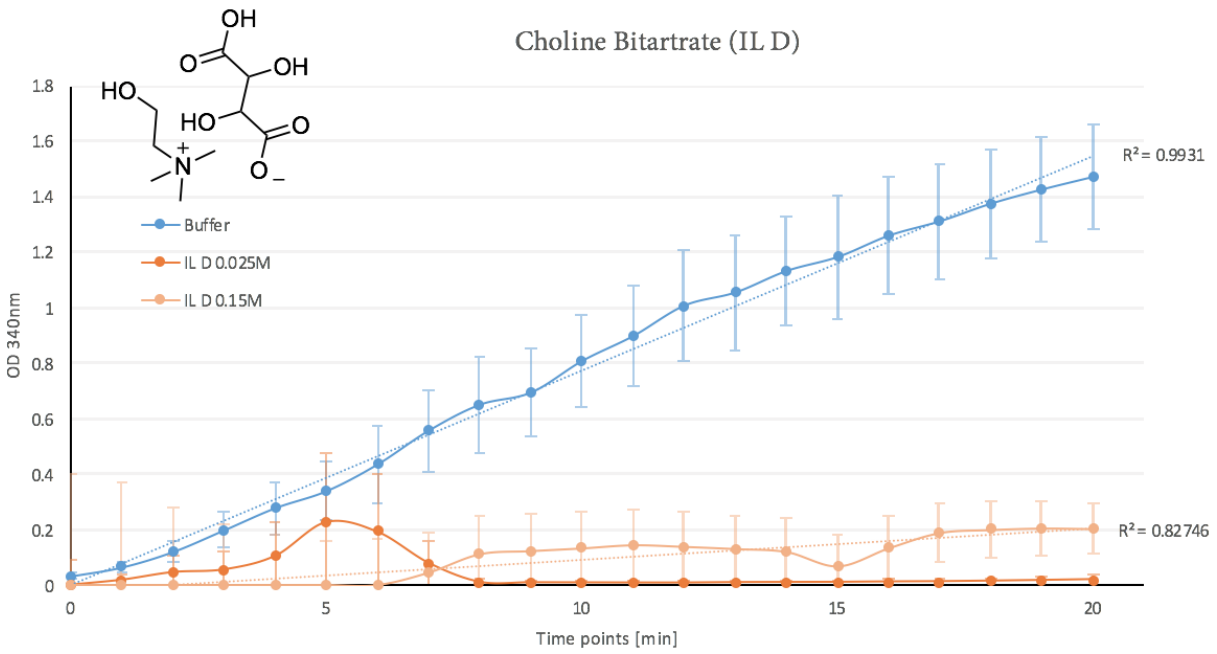


Figure 8.8: ILD [Choline] [Bitartrate]: All concentrations were considerably lower in activity, when compared to the reference buffer. For ILD 0.025 M an increase in absorption is detected over the first 7 minutes, which is again reduced to zero thereafter. The enzyme appears to start being active at around 7 minutes in ILD 0.15 M with stagnating activity thereafter.

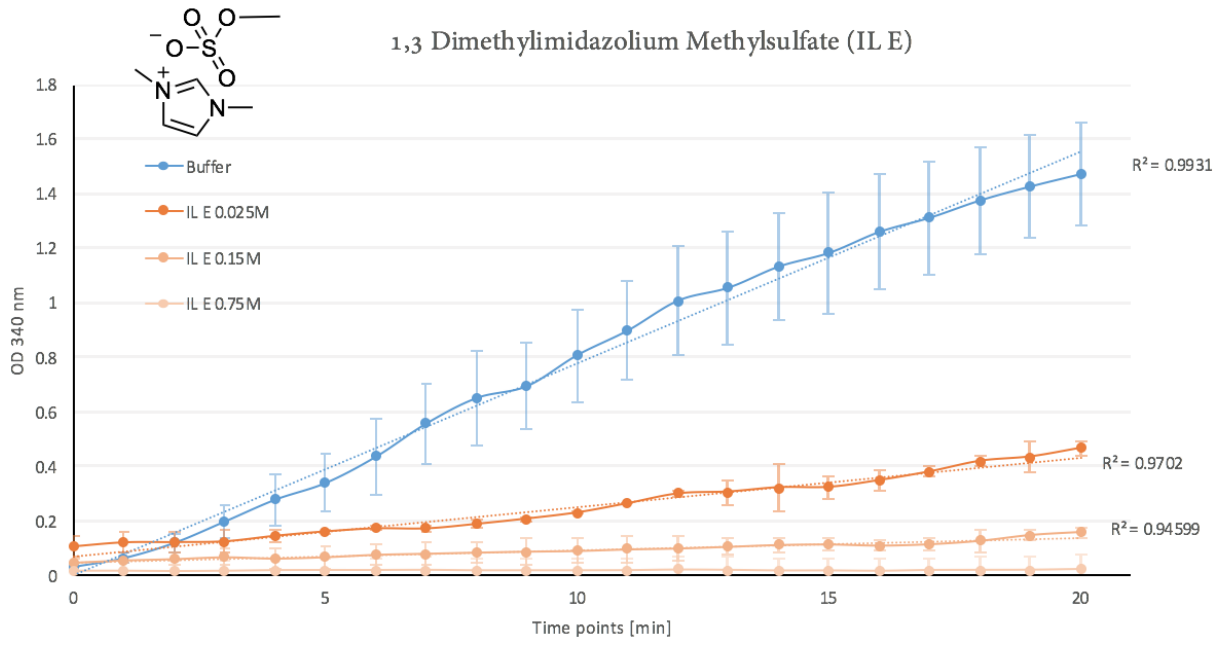


Figure 8.9: ILE [DiMIM][$MeSO_4$]: For all concentrations of IL-mixtures activity was considerably lower, when compared to the reference. At 0.75 M there was almost no activity.

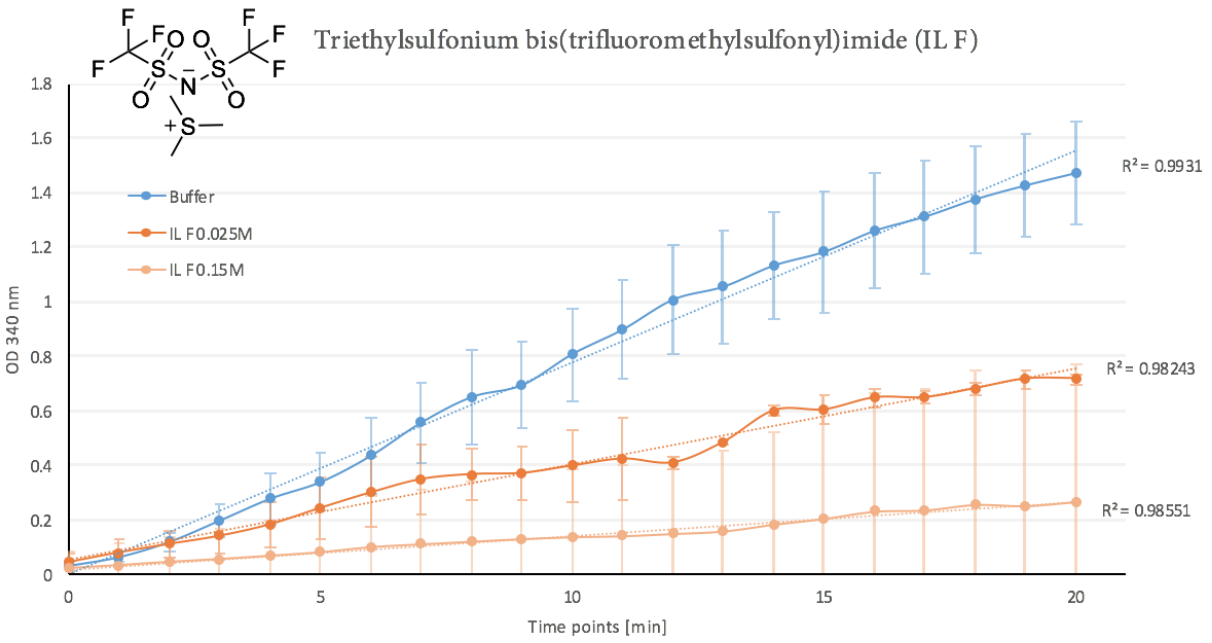


Figure 8.10: ILF $[Me_3S][NTf_2]$: In both measured IL concentrations enzyme activity was lower than for the buffer. The concentration of 0.15 M showed low precision between measurements.

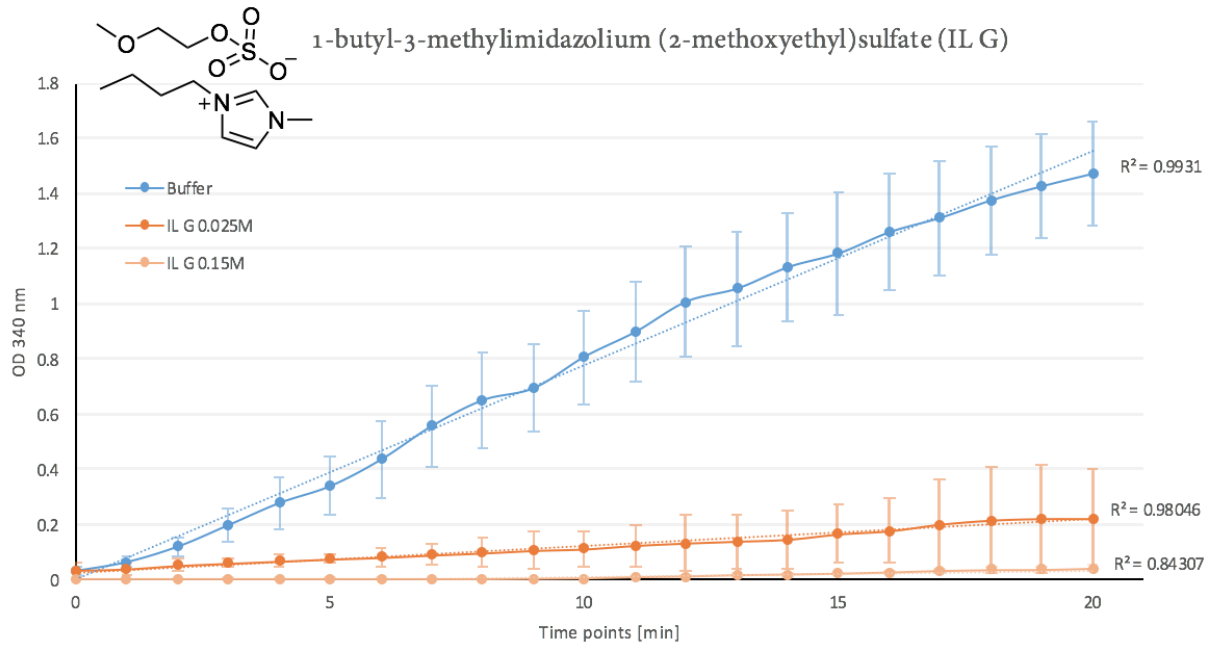


Figure 8.11: ILG $[BMIM][MeOEtSO_4]$: For all concentrations of IL-mixtures activity was considerably lower, when compared to the reference. At 0.75 M there was almost no activity.

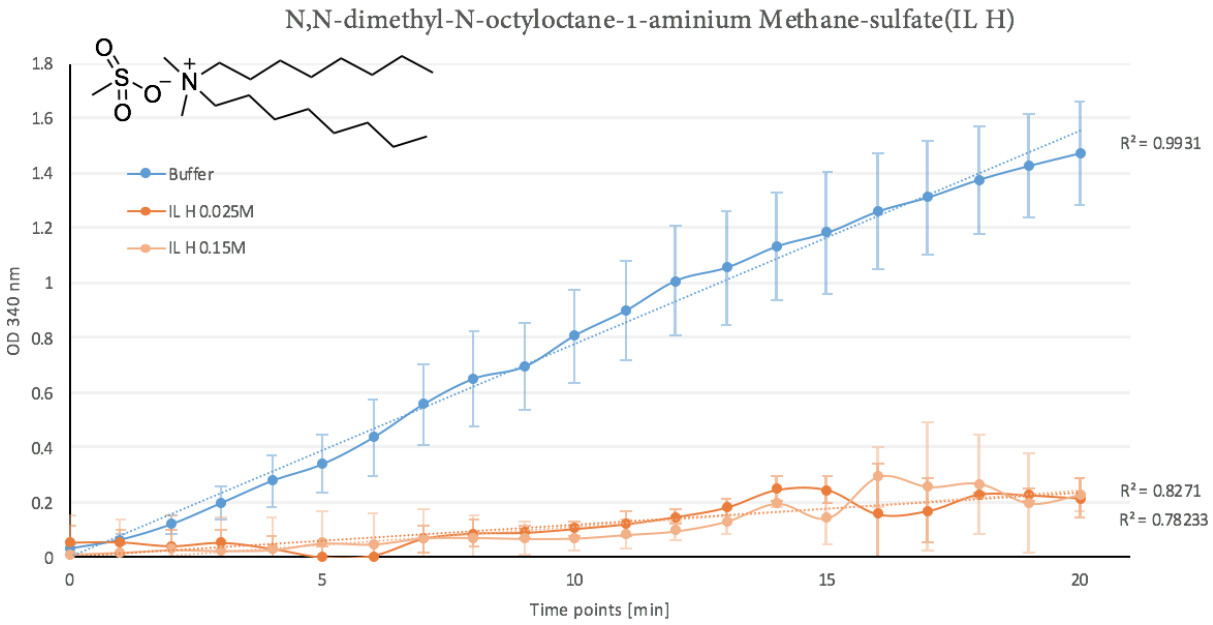


Figure 8.12: ILH $[N_{1,1,8,8}][MeSO_3]$: Both concentrations were considerably lower in activity, when compared to the reference buffer. Both trends show low linearity.

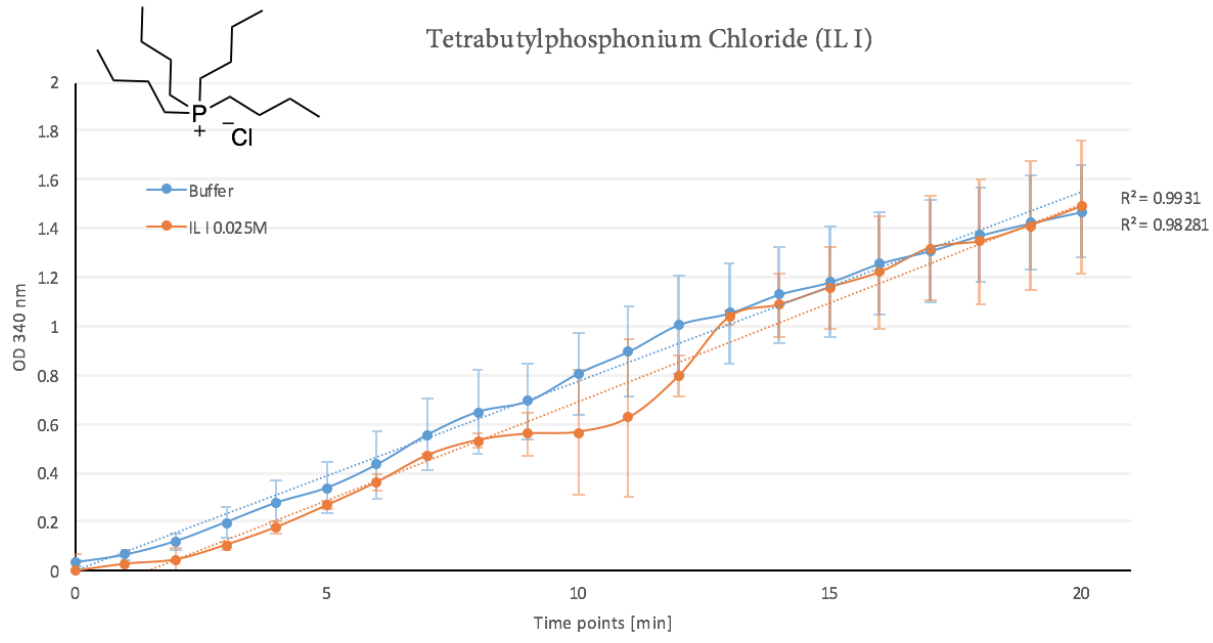


Figure 8.13: ILI $[P_{4,4,4,4}][Cl]$: Only the lowest concentration was measured, but the enzyme showed similar activity compared to the reference.

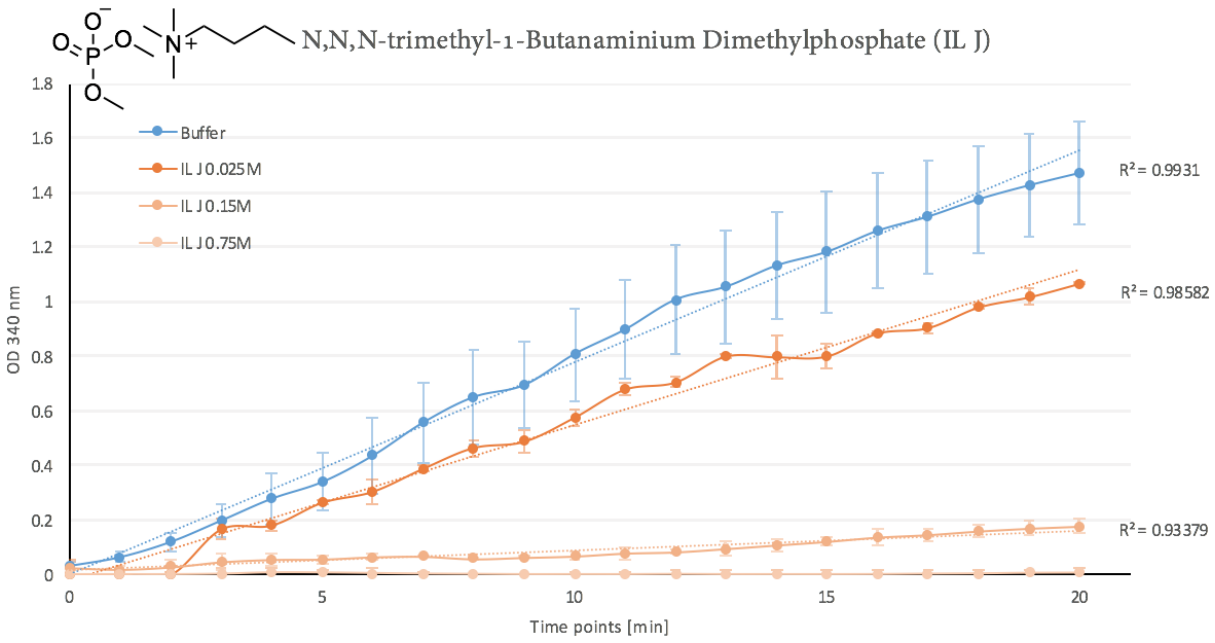


Figure 8.14: ILJ $[N_{4,1,1,1}][(MeO)_2(HO)PO]$: Enzyme activity was lower in all concentrations, when compared to the reference, and followed a decrease with increasing IL concentration.

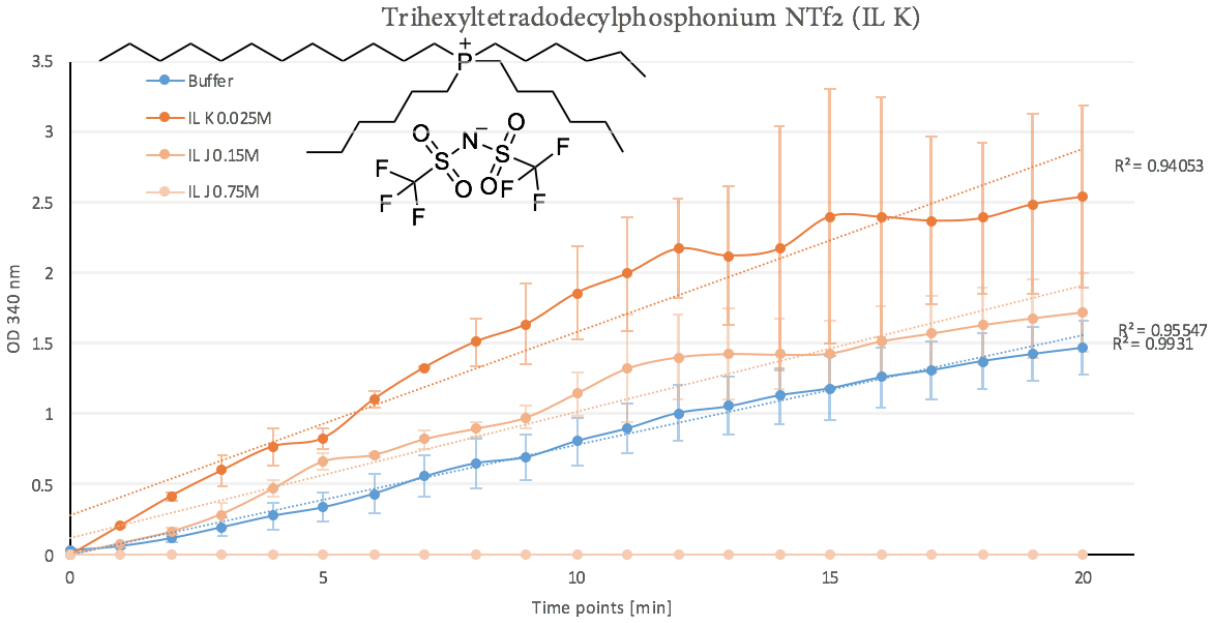


Figure 8.15: ILK $[P_{6,6,6,14}][NTf_2]$: At concentrations 0.025 and 0.15 M IL enzyme activity was higher compared to the reference system. In 0.75 M enzyme activity was diminished.

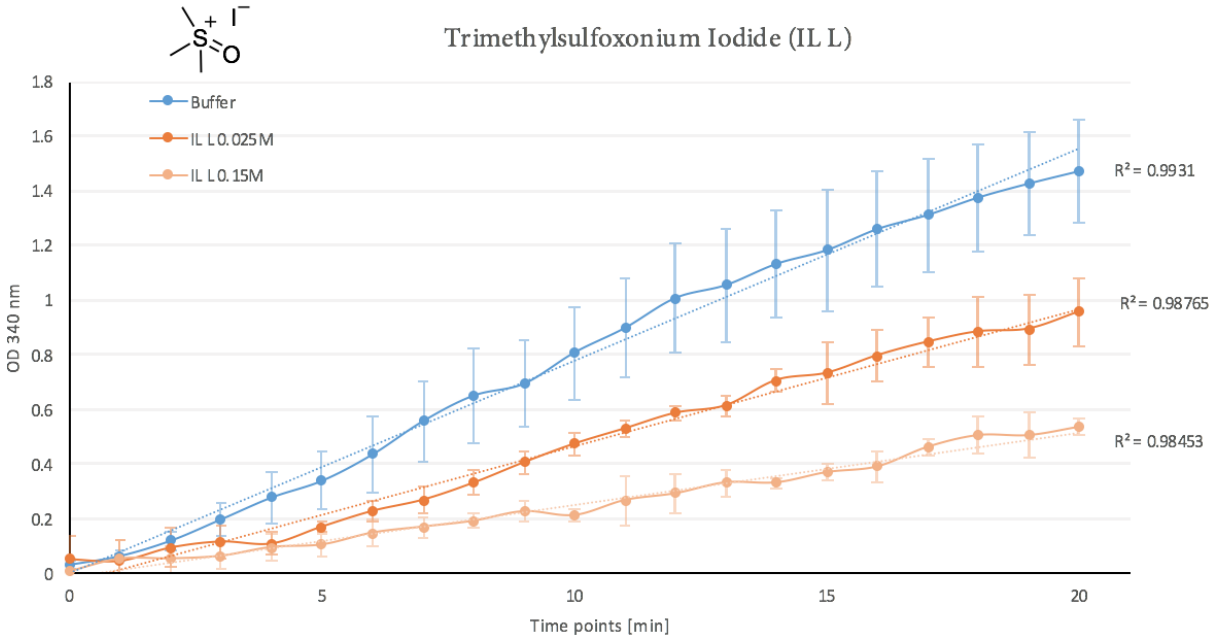


Figure 8.16: ILL [Me_3SO][I]: For both assayed concentrations enzyme activity was lower in comparison to the reference.

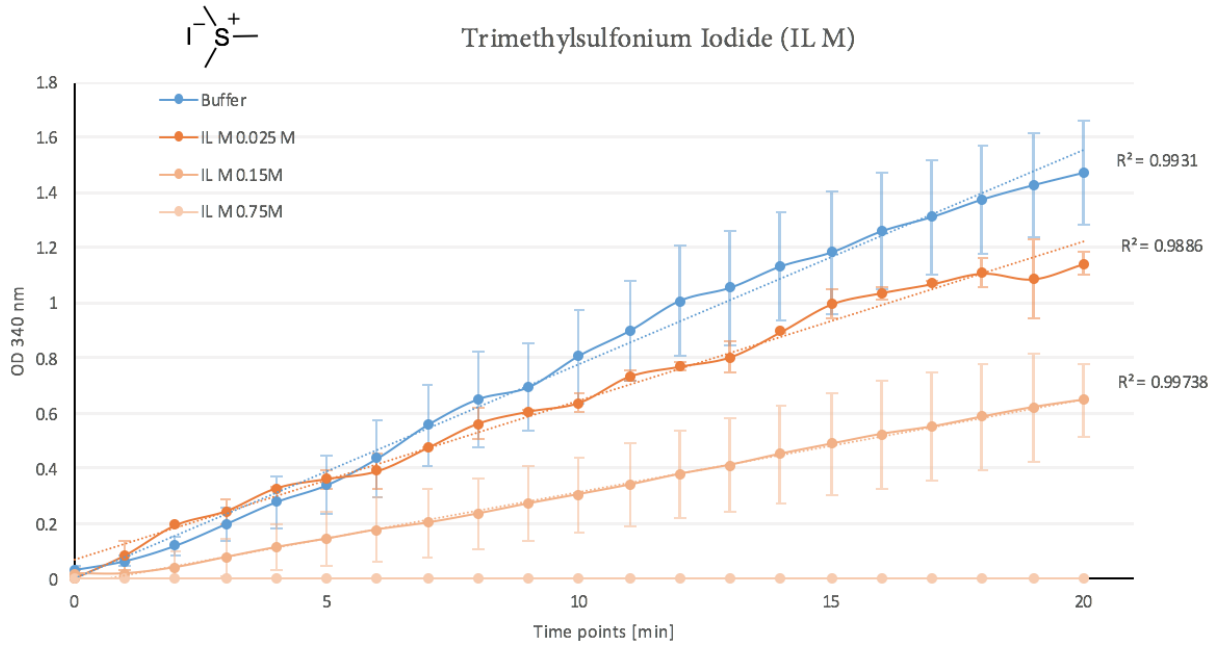


Figure 8.17: ILM [Me_3S][I]: For assayed concentrations 0.025 M and 0.15 M enzyme activity was lower in comparison to the reference. In 0.75 M enzyme activity was diminished.

8.3 Relative Specific Activity Assays ILA, ILK and ILM

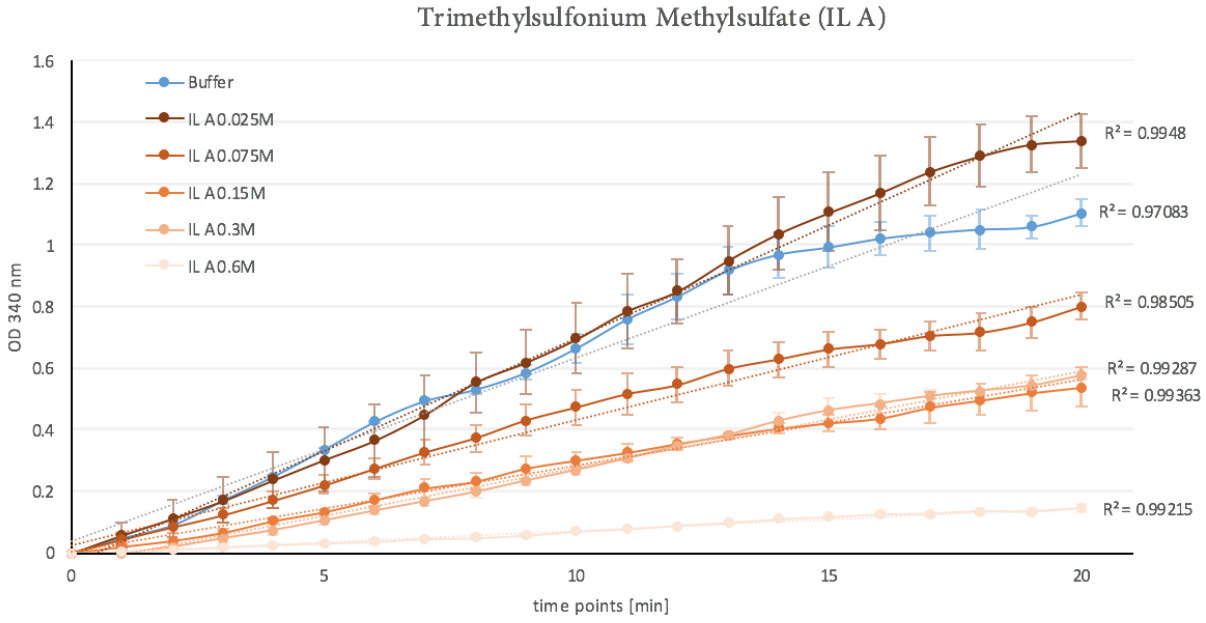


Figure 8.18: ILA $[Me_3S][OSO_3Me]$: Buffer plateaued early in comparison with the lowest concentration 0.025M. Concentration 0.3 M showed similar results to concentration 0.15 M and in 0.6 M IL A activity was still observed.

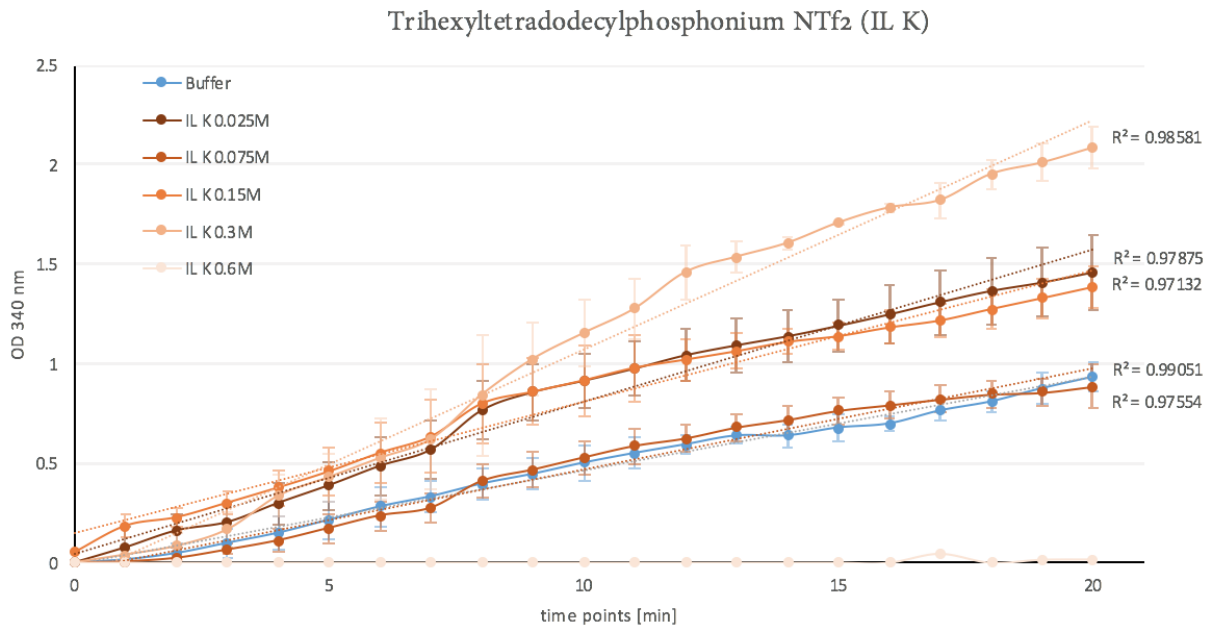


Figure 8.19: ILK $[P_{6,6,6,14}][NTf_2]$: Orbital shaking was not repeated with double-orbital, hence the discussed 'kink' of section *Investigation of absorption curve* is observable. 0.3 M showed the double the amount of activity compared to the buffer system, whilst no activity was observed at 0.6 M.

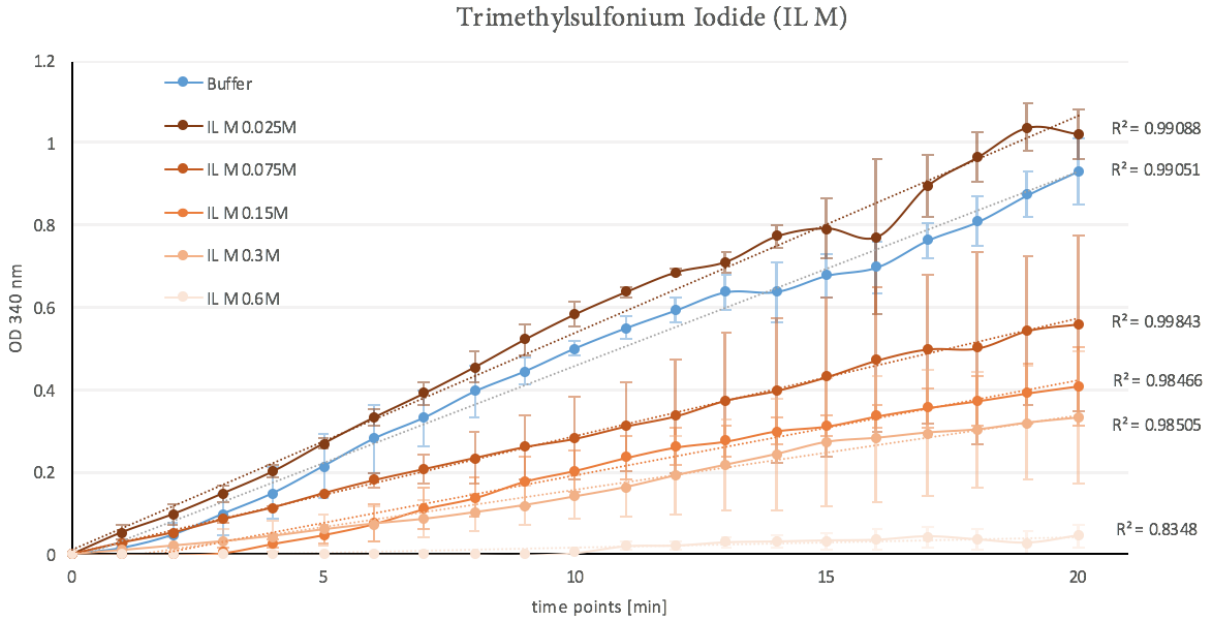


Figure 8.20: ILM $[Me_3S][I]$: a measurement artefact is observed for concentration 0.025 M. Activity in this concentration is higher compared to the buffer, with all other concentrations showing a dependent decline for enzyme activity.

8.4 RDFs of Potassium and Chloride in ILs

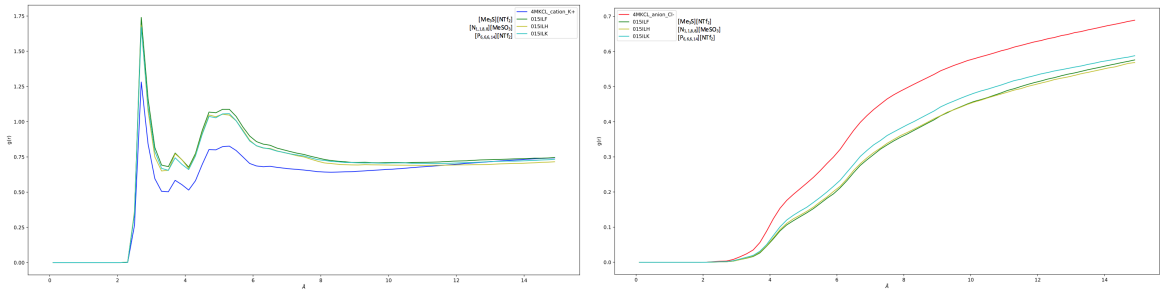


Figure 8.21: RDF of K^+ and Cl^- ions in 4 M KCl in the presence of 0.15 M Hydrophobic ionic liquids F[Me_3S][NTF_2], H[$N_{1,1,8,8}$][$MeSO_3$] and K[$P_{6,6,6,14}$][NTF_2]

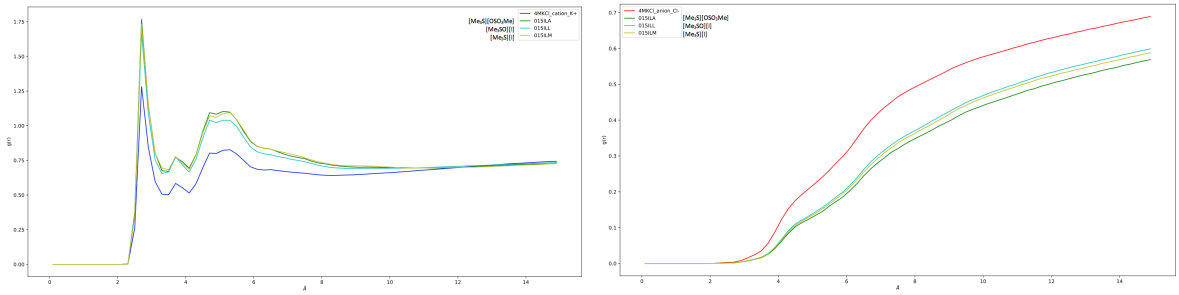


Figure 8.22: RDF of K^+ and Cl^- ions in 4 M KCl in the presence of 0.15 M Sulfo ionic liquids A[Me_3S][OSO_3Me], L[Me_3SO][I] and M[Me_3S][I]

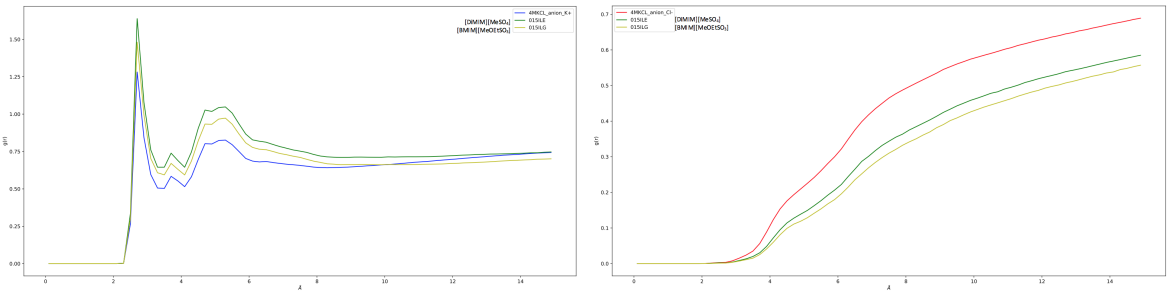


Figure 8.23: RDF of K^+ and Cl^- ions in 4 M KCl in the presence of 0.15 M Imidazolium ionic liquids E[DiMIM][OSO_3Me] and G[BMIM][$MeOEtSO_4$]

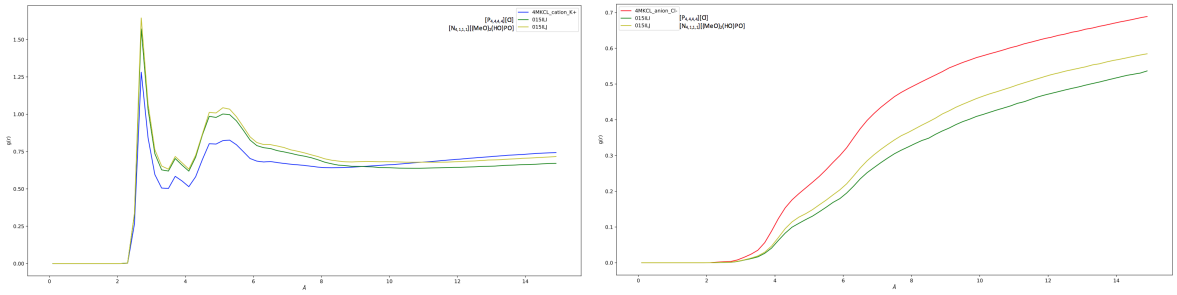


Figure 8.24: RDF of K^+ and Cl^- ions in 4 M KCl in the presence of 0.15 M Phospho ionic liquids I[$P_{4,4,4,4}$][Cl] and J[$(MeO)_2OPO$][$N_{4,1,1,1}$]

8.5 RMSDs of 0.15M ILC - ILJ, ILL

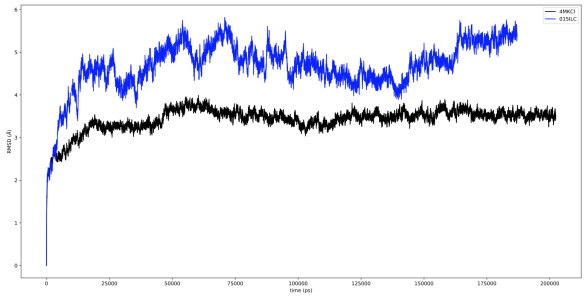


Figure 8.25: RMSD ILC 0.15M

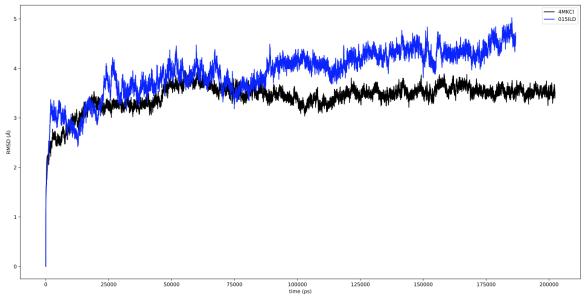


Figure 8.26: RMSD ILD 0.15M

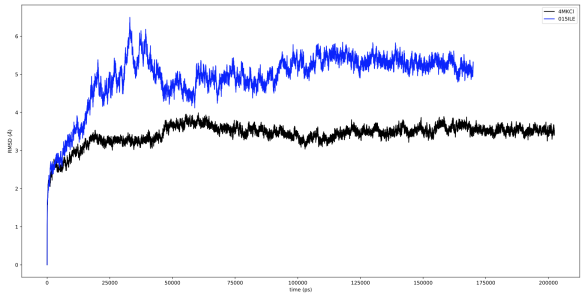


Figure 8.27: RMSD ILE 0.15M

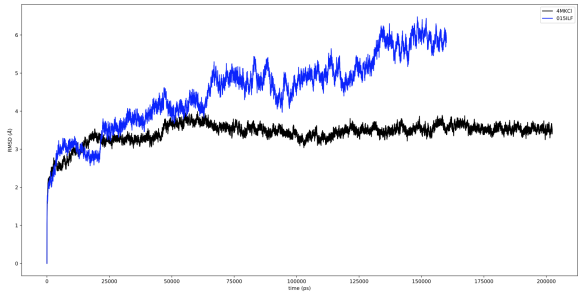


Figure 8.28: RMSD ILF 0.15M

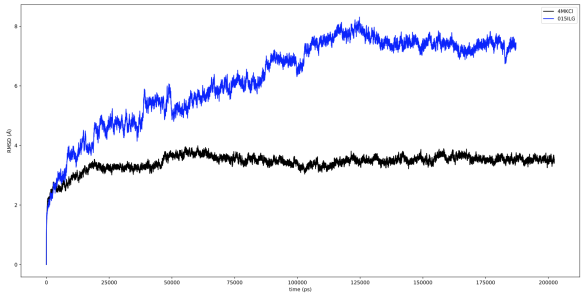


Figure 8.29: RMSD ILG 0.15M

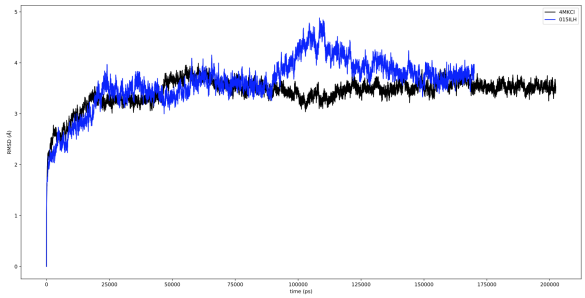


Figure 8.30: RMSD ILH 0.15M

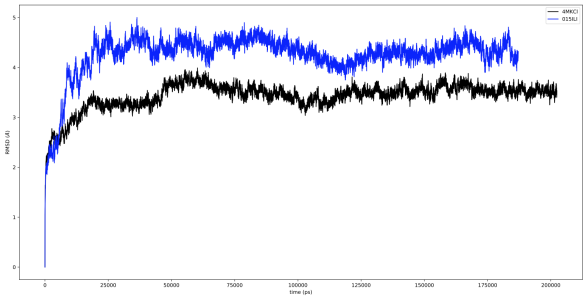


Figure 8.31: RMSD ILI 0.15M

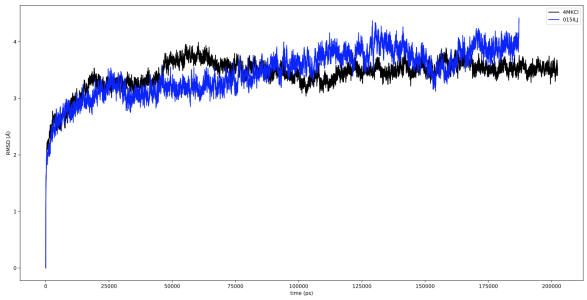


Figure 8.32: RMSD ILJ 0.15M

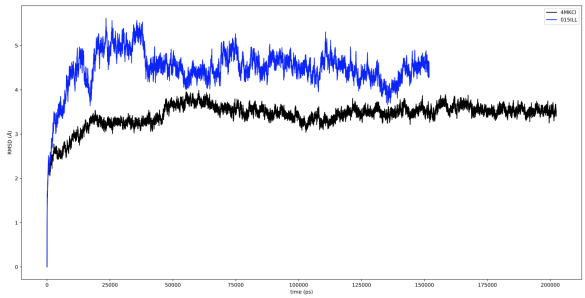


Figure 8.33: RMSD ILL 0.15M

8.6 Scripts for MD Simulations

```
tolerance 2.0
filetype pdb
output 190415_075ILM_box.pdb

structure 190415_ADH.pdb
  number 1
  fixed 0. 0. 0. 0. 0. 0.
  centerofmass
end structure

structure mes.pdb
  number 204
  inside box -50. -50. -50. 50. 50. 50.
  outside sphere 0. 0. 0. 60.
end structure

structure iod.pdb
  number 204
  inside box -50. -50. -50. 50. 50. 50.
  outside sphere 0. 0. 0. 60.
end structure
```

```
source leaprc.ff14SB
addAtomTypes {{"ZN" "Zn" "sp3"} {"s3" "s" "sp3"} {"N2" "N" "sp3"}}
source leaprc.gaff
addAtomTypes {{"OE" "HE" "HF" "HG"}}
loadamberparams frcmod.ionsjc_tip4pew
loadoff atomic.ions.lib
loadamberprep ZAFF.prep
loadamberparams ZAFF.frcmod
loadamberparams frcmod.ionsjc_tip3p
loadoff NAD.lib
loadoff mes.lib
loadoff iod.lib
loadamberparams NAD.frcmod
loadamberparams mes.frcmod
loadamberparams iod.frcmod
box = loadpdb 190415_03ILM_box.pdb
bond box.1397.ZN box.38.SG
bond box.1397.ZN box.59.NE2
bond box.1398.ZN box.103.SG
bond box.1398.ZN box.89.SG
bond box.1398.ZN box.92.SG
bond box.1398.ZN box.95.SG
bond box.1399.ZN box.387.SG
bond box.1399.ZN box.408.NE2
bond box.1400.ZN box.438.SG
bond box.1400.ZN box.441.SG
bond box.1400.ZN box.452.SG
bond box.1400.ZN box.444.SG
bond box.1401.ZN box.736.SG
bond box.1401.ZN box.757.NE2
bond box.1402.ZN box.787.SG
bond box.1402.ZN box.790.SG
bond box.1402.ZN box.793.SG
bond box.1402.ZN box.801.SG
bond box.1403.ZN box.1085.SG
bond box.1403.ZN box.1106.NE2
bond box.1404.ZN box.1142.SG
bond box.1404.ZN box.1150.SG
bond box.1404.ZN box.1136.SG
bond box.1404.ZN box.1139.SG
setbox box centers
#addions box K+ 0
#addions box Cl- 0
solvatebox box TIP3PBOX 0
addionsrand box K+ 800 Cl- 800 4
savepdb box 190415_03ILM.pdb
saveamberparm box 03ILM190415.prmtop 03ILM190415.inpcrd
check box
quit
```

```
#!/bin/sh
#SBATCH --job-name=06Mm1
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --gres=gpu
#SBATCH --mem=4gb
#SBATCH --time=160:00:00
#SBATCH --partition=bectgtxq
#SBATCH --qos=bectgtx
#SBATCH --mail-type=end
#SBATCH --mail-user=ezxas9@nottingham.ac.uk


module purge
module load amber-uon/cuda9.2/18

export CUDA_VISIBLE_DEVICES=0

cd /gpfs01/home/ezxas9/min1_sim/min12

pmemd.cuda -i min1wat.in -p 06ILM190415.prmtop -c 06ILM190415.inpcrd
-r min106ILM190415.rst -x min106ILM190415.nc -ref 06ILM190415.inpcrd
-o min106ILM190415.out -O
```

```
Initial minimisation min1
 &cntrl
  imin    = 1, ntmin = 1, nmropt = 0, drms = 0.1,
  iwrap  = 0,
  maxcyc = 50000000,
  ncyc   = 30000000,
  cut    = 10.0, nsnb = 20,
  ntpr   = 1000, ntwr = 1000,
  ntx = 1, irest = 0,
  ntf = 1, ntb = 1,
  igb = 0,
  ibelly = 0, ntr = 1,
  restraintmask="!:WAT", restraint_wt=10.0,
 &end
/
END
END
```

```
LET WATER MOVE
&cntrl
  nmropt = 0, imin = 0, ntrx = 1, ntxo = 1, ntwe = 0, ntwv =0,
  ntx    = 1,       irest  = 0,
  ntpr   = 500,
  ntwr   = 500,
  ntf    = 2,       ntb    = 2,
  cut    = 10.0,    nsnb   = 20,
  nstlim = 10000,   nscm   = 2500,   iwrap = 1,
  t      = 0.0,     dt     = 0.001,
  temp0  = 323.15,   tempi  = 0.0,    tautp=0.5,
  ntt    = 1,
  ntp    = 1 ,       taup = 1.0,
  ntc    = 2,       tol    = 0.00001,
  ibelly=0, ntr=1,
  restraintmask="!:WAT" , restraint_wt=10.0,
&end
```

```
Minimize all atoms
  System minimization:
&cntrl
   imin=1, ntmin=1, nmropt=0, drms=0.1
   maxcyc=20000000, ncyc=15000000,
   ntx=1, irest=0,
   ntpr=100, ntwr=100, iwrap=0,
   ntf=1, ntb=1, cut=10.0, nsnb=20,
   igb=0,
   ibelly=0, ntr=0,
&end
/
END
END
```

```
heat NPT 0.5ps
  Heating System
&cntrl
   imin=0, nmropt=1,
   ntx=1, irest=0,
   ntpr=500, ntwr=500, ntwx=500, iwrap=1,
   ntf=2, ntb=1, cut=10.0, nsnb=20,
   igb=0,
   ibelly=0, ntr=1,
   nstlim=250000, nscm=500, dt=0.001,
   ntt=1, temp0=0.0, tempi=0.0, tautp=0.5
   ntc=2,restraintmask=':1-1407',
   restraint_wt=10.0,
&end

&wt type='REST', istep1=0, istep2=0, value1=1.0, value2=1.0, &end
&wt type='TEMP0', istep1=0, istep2=250000, value1=0.0,
value2=323.15, &end
&wt type='END'

 &end
/
END
END
```

```
equil NTP 0.5ns
 heat
 &cntrl
  imin=0,irest=1,ntx=5,
  nstlim=2500000,dt=0.001,
  ntc=2,ntf=2,
  cut=10.0, ntb=2, ntp=1, taup=1.0,
  ntpr=100, ntwx=1000, ntwr=100,
  ntt=3, gamma_ln=2.0,
  temp0=323.15,iwrap=1,
  ntr=1, restraintmask=':1-1407',
  restraint_wt=10.0,
/
END
END
```

```
2ADH production
 &cntrl
  imin   = 0,
  ig     = -1,
  iwrap  = 1,
  irest  = 1, ntx = 5,
  ntp    = 1, pres0 = 1.0, taup = 2.0,
  cut    = 10.0,
  tempi  = 323.15, temp0 = 323.15,
  ntc    = 2, ntf = 2,
  ntt    = 3, gamma_ln = 2.0,
  nstlim = 29520000,
  ntwx   = 5000, ntwe = 5000, ntpr = 5000, ntwr = 5000,
  dt     = 0.002,
/
END
END
```

```
equil NTP 0.5ns
 heat
 &cntrl
  imin=0,irest=1,ntx=5,
  nstlim=2500000,dt=0.001,
  ntc=2,ntf=2,
  cut=10.0, ntb=2, ntp=1, taup=1.0,
  ntpr=100, ntwx=1000, ntwr=100,
  ntt=3, gamma_ln=2.0,
  temp0=323.15,iwrap=1,
  ntr=1, restraintmask=':1-1407',
  restraint_wt=10.0,
/
END
END
```

8.7 Scripts for processing gromacs output

```
#!/usr/bin/env python

"""
Given two values, looks in a 3 column-file (output of sham.pl)
which time frame matches closest.
"""

import sys

USAGE = "USAGE: get_timestamp.py -f <sham output> -1 <value 1> -2
<value 2>\n"

# Parse arguments
read_input, read_value1, read_value2 = False, False, False
input_file, value1, value2 = None, None, None
for arg in sys.argv[1:]:
    if read_input:
        read_input = False
        input_file = arg
    elif read_value1:
        read_value1 = False
        value1 = float(arg)
    elif read_value2:
        read_value2 = False
        value2 = float(arg)
    if arg[0] == "-":
        if arg == "-f":
            read_input = True
            continue
        elif arg == "-1":
            read_value1 = True
            continue
        elif arg == "-2":
            read_value2 = True
        else:
            print USAGE
            sys.stderr.write('ERROR: Option not recognized: %s\n' %arg)
            sys.exit(1)

if not input_file:
    print USAGE
    sys.stderr.write('ERROR: You forgot to provide an input file.\n')
    sys.exit(1)

# Open sham output
x_values, y_values, time_values = [], [], []
fhandle = open(input_file)
for line in fhandle:
    if line[0] != "#" and len(line.split()) == 3:
        t,x,y = line.split()
        x_values.append(float(x))
        y_values.append(float(y))
        time_values.append(float(t))
fhandle.close()

def find_common_frame(min_x, min_y):
    for xval in min_x:
        xframe = xval[0]
```

```
        for yval in min_y:
            yframe = yval[0]
            if xframe == yframe:
                return (xframe, xval[1], yval[1])
    return (None, None, None)

# If you cannot find anything, try increasing the nval variable
nval = 50
min_x = sorted(enumerate(x_values), key=lambda x: abs(x[1]-
value1))[:nval]
min_y = sorted(enumerate(y_values), key=lambda x: abs(x[1]-
value2))[:nval]

frame, x, y = find_common_frame(min_x, min_y)

if not frame:
    print "No timestamp found.."
    sys.exit(0)
print "## T = %s (%s, %s)" %(time_values[frame], x, y)
```

```
#!/usr/bin/perl -w

# This program reads in two .xvg files (specified on the command line)
and prints out a
# concatenated output file suitable for input into g_sham for
generating free energy landscapes.
# The output line contains, for example: time value1 value2
#
# Caveat 1 (should be obvious): both input files must contain the same
number of data points
# Caveat 2: the desired quantity should be the second item on the line
in the .xvg file (after time)
#
# Values of -data1 and -data2 are the number of the data set in the
input .xvg file(s), i.e. 0 would
# be the first column (usually time), 1 would be the first data set,
etc.

use strict;

unless (@ARGV) {
    die "Usage: perl sham.pl -i1 input1.xvg -i2 input2.xvg -data1 1 -
data2 1 -o graph.xvg\n";
}

my %args = @ARGV;

# define input and output files
my $input1;
my $input2;
my $output;

if (exists($args{"-i1"})) {
    $input1 = $args{"-i1"};
} else {
    die "No -i1 specified!\n";
}

if (exists($args{"-i2"})) {
    $input2 = $args{"-i2"};
} else {
    die "No -i2 specified!\n";
}

if (exists($args{"-o"})) {
    $output = $args{"-o"};
} else {
    $output = "graph.xvg";  # default output name
}

# define input data sets
my $d1;
my $d2;

if (exists($args{"-data1"})) {
    $d1 = $args{"-data1"};
} else {
    $d1 = 1;    # default to first non-time data column
```

```
}

if (exists($args{"-data2"})) {
    $d2 = $args{"-data2"};
} else {
    $d2 = 1;
}

open(IN1, "<$input1") || die "$!: Cannot open $input1";
my @in1 = <IN1>;
close(IN1);

open(IN2, "<$input2") || die "$!: Cannot open $input2";
my @in2 = <IN2>;
close(IN2);

# need to chop off the headers first, since they may cause files with
the same number of data
# points to have different lengths (one file has more header info than
another

my $size1 = scalar(@in1);

for (my $i=0; $i<$size1; $i++) {
    if ($in1[$i] =~ /^[#@]/) {
        splice(@in1, $i, 1);
        $i--;
        $size1--;
    }
}

my $size2 = scalar(@in2);

for (my $j=0; $j<$size2; $j++) {
    if ($in2[$j] =~ /^[#@]/) {
        splice(@in2, $j, 1);
        $j--;
        $size2--;
    }
}

# debug
# print "@in1\n";
# print "@in2\n";

open(OUT, ">>$output") || die "$!: Cannot open $output\n";
print OUT "# graph.xvg for g_sham\n";

for (my $i=0; $i<scalar(@in1); $i++) {
    my $line1 = $in1[$i];
    my $line2 = $in2[$i];

    if (($line1 =~ /^[#@]/) || ($line2 =~ /^[#@]/)) {
        # do nothing, it's a comment - probably not necessary after
above cleanup
    } else {
        my @data1 = split(" ", $line1);
        my @data2 = split(" ", $line2);
```

```
            my $time = $data1[0];
            my $val1 = $data1[$d1];
            my $val2 = $data2[$d2];

            printf OUT "%8.3f\t%f\t%f\n", $time, $val1, $val2;
        }
    }

    close(OUT);

    exit;
```

```
#!/usr/bin/env python

import sys

"""
Utility tool to convert xpm files generated by GROMACS to a 3-column
text file.
"""

USAGE = "USAGE: xpm2txt.py -f <input xpm file> -o <output txt file> [-
s]\n"
USAGE+= "Options:\n"
USAGE+= "\t-s\t(int)\tSorts the output by a given column"
USAGE+= "\n" # always keep this line

# Parse arguments
read_input, read_output, sort = False, False, False
xpm_file, out_file, column_sort = None, None, None
for arg in sys.argv[1:]:
    if read_input:
        read_input = False
        xpm_file = arg
    elif read_output:
        read_output = False
        out_file = arg
    elif sort:
        sort = False
        column_sort = int(arg)
    if arg[0] == "-":
        if arg == "-f":
            read_input = True
            continue
        elif arg == "-o":
            read_output = True
            continue
        elif arg == "-s":
            sort = True
        else:
            print USAGE
            sys.stderr.write('ERROR: Option not recognized: %s\n' %arg)
            sys.exit(1)

if not xpm_file:
    print USAGE
    sys.stderr.write('ERROR: You forgot to provide an input file.\n')
    sys.exit(1)
if not out_file:
    out_file = "out.txt"

# Parse XPM file
xpm_handle = open(xpm_file)
xpm_data = []
x_axis, y_axis = [], []
letter_to_value = {}
for line in xpm_handle:
    if line.startswith("/* x-axis"):
        x_axis = map(float, line.split()[2:-2]) # We trim the last
value
```

```
    if line.startswith("/* y-axis"):
        y_axis = map(float, line.split()[2:-2]) # We trim the last
value

    if line.startswith('"') and x_axis and y_axis: # Read data
        xpm_data.insert(0, line.strip().strip(',')[1:-1])

    if line.startswith('"') and len(line.split()) > 4:
        letter = line.split()[0][1:]
        value = float(line.split()[-2][1:-1])
        letter_to_value[letter] = value
xpm_handle.close()

# Match x/y/data
txt_values = []
for y_index, data_value in enumerate(xpm_data):
    y_value = y_axis[y_index]
    for x_index, x_value in enumerate(x_axis):
        txt_values.append([x_value, y_value,
letter_to_value[data_value[x_index]]])

# Apply sorting if requested
if column_sort:
    try:
        txt_values.sort(key=lambda x: x[column_sort-1])
    except IndexError:
        print USAGE
        sys.stderr.write('ERROR: Column not found (%s)\n'
%(column_sort))
        sys.exit(1)

# Print to file
out_handle = open(out_file, 'w')
for x, y, z in txt_values:
    out_handle.write("%3.5f\t%3.5f\t%3.5f\n" %(x,y,z))
out_handle.close()
```