Investigating the Role of Lipids for Membrane Protein Complexes using Native Mass Spectrometry and Molecular Dynamics Simulations



Daniel Quetschlich

St Hilda's College

Department of Chemistry
University of Oxford
United Kingdom

A thesis presented for the degree of
Doctor of Philosophy in Physical and Theoretical Chemistry
February 28, 2022

Investigating the Role of Lipids for Membrane Protein Complexes using Native Mass Spectrometry and Molecular Dynamics Simulations

Daniel Quetschlich
St Hilda's College

A thesis presented for the degree of
Doctor of Philosophy in Physical and Theoretical Chemistry

February 28, 2022

Abstract

The work described in this thesis is a combination native mass spectrometry experiments, molecular dynamics simulations and scientific software development. It consists of three results chapters which each chapter describing a distinct project with a focus on one of the aspects above.

In chapter 3, I employed molecular dynamics simulations in a collaborative project on using surface-induced dissociation to probe the effect of lipid binding on the stability of oligomerisation interfaces. My simulations offer a rationale of lipid binding to different locations to explain these stability differences. In addition, I contributed to experiments to probe the molecular features of these interfaces through site-directed mutagenesis to weaken these interfaces.

Chapter 4 describes my investigation on lipid binding to a bacterial potassium channel, MthK. Native mass spectrometry and molecular dynamics simulations are used to characterise lipid binding after initial experiments indicated endogenous lipids being bound to MthK. Cardiolipin has been revealed to bind specifically to MthK. These results are compared to the eukaryotic homologues through further simulations. A part of this chapter outlines different approaches to obtain the *in vivo* stoichiometry. Among these approaches I employed SoLVe-MS, which leads to very complex mass spectra. The analysis of these spectra sparked the development of the program described in chapter 5.

In the final results chapter, a web-based program for the analysis of native mass spectra, NaViA, is described. This program allows one to analyse native mass spectra through a visual user interface directly in a web browser. In addition, it offers auxiliary tools to facilitate the analysis as well as the ability to save and load sessions for sharing spectra including analysis or depositing them for publication. The ease of use and the instructive auxiliary tools make it a valuable addition to the suite of software available for the analysis of native mass spectrometry.

Acknowledgements

I would like to express my gratitude to all the people who made my time in Oxford an amazing experience.

First and foremost, I would like to thank my supervisors Carol Robinson, Phill Stansfeld and Mark Sansom for giving me the opportunity for pursuing my research in their laboratories. I am grateful for all their continuous support and encouragement during these years.

Next, I would like my industrial supervisors Zara Sands, Silvia Lovera, Rachel Davis and Idlir Liko as well as their companies UCB BioPharma and OMass Therapeutics for the funding and continued interest in my work. Presenting my research progress to them was always a great pleasure. Their continuous enthusiasm was a great encouragement.

I would also like to thank my assigned supervisors in the Robinson group. Kallol Gupta taught me a lot about how to select a good research project and strategy in science. Mark Agasid was a great teacher to learn all the techniques necessary to conduct research in the lab. Both also helped me tremendously to grow personally during my DPhil.

I would like to thank all members from the Robinson and Stansfeld/Sansom labs for their continuous advice, support and help over the years. In the Robinson lab special thanks goes to Mark Agasid, Jani Bolla, Francesco Fiorentino, Dror Chorev, Haiping Tang, Tim Esser, Wanling Song, Denis Shutin, Siyun Chen and Tarick El-Baba. Francesco and Denis were also proof readers of this thesis. I am incredibly grateful for that. In addition, I would like to thank Victoria Cullen, Todd Mize and Andy Dolan for keeping the lab running and always helping me (especially being generous with lab time in the Covid autumn last year when I urgently needed to finish my experiments). Also, I would like to thank Tom Durrant and Leo Urner for being great desk buddies in the German corner. The latter deserves extra credit for hosting numerous BBQs and being a true friend. In the MD group, I would like to express extra gratitude to Owen Vickery and Robin Corey. Robin took the time and helped me when I worried about some simulations towards the end of my DPhil.

My good friends and collaborators Joe Gault and Denis Shutin also deserve to be mentioned. It was an amazing experience to collaborate with them in the lab and in the pub. I would like to thank Joe for being a critical mind and driving the project forward, Special thanks goes to Denis for being a great friend, DotA mate, collaborator in the lab and pub buddy.

Next, I would like thank my friends outside the lab. I am grateful to Katie Truslove, Simon Nadal, Kristina Aertker, Daniel Meusburger and Yuliya Dubianok for many dinners and pub visits. All the conversations and fun experiences are invaluable. More thanks goes to Sthitadhi Roy and Francesco Fiorentio (again) for being awesome flatmates. The experiences of all "royale" dishes and pasta lunches will not be forgotten. My gratitude goes to my friends from the lab and from St Hilda's as well for all the amazing experiences during my time in Oxford.

I am also extremely grateful to my family. My brother Nils always had an open ear for my worries and I am grateful beyond words for the continuous support of my parents, Birgit and Hartmut during my life and especially during the hot phase of the thesis writing period. They have always helped me to create opportunities for me to succeed in life and encouraged me to make bold moves. My gratefulness to my family extends to the wider family for their understanding and support during the thesis writing period. One couldn't wish for more!

Last, but certainly not least, I would like to express my gratitude to Lu, my partner for her love and during our long term long distance relationship. I look forward to the future without distance.

Finally, I would like to thank everyone who I shamefully forgot to mention here but who deserved extra credit. I am truly sorry.

Declaration

The work described in this thesis was conducted between October 2017 and April 2021. The results are linked to a overall theme and each chapter focusses on a separate project. All of the data presented in this work was acquired and analysed by myself unless stated otherwise. Similarly, all figures including captions are created by me unless indicated otherwise.

The work presented in this thesis has not been submitted for a degree, diploma or any other academic certification at any other institution, in whole or in part.

Daniel Quetschlich
June 2021

Contents

List of Figures **xii**

List of Tables **xvi**

1 Introduction **1**

1.1 Cell Membranes 1

1.1.1 Membrane Lipids 1

1.1.2 Membrane Proteins 4

1.1.2.1 Transporters 5

1.1.2.2 Ion Channels 6

1.2 Protein-Lipid Interactions 10

1.3 Methods to Study Protein-Lipid Interaction 16

1.3.1 Native mass spectrometry 18

1.3.1.1 Software & algorithms for assigning native Mass Spectra 23

1.3.2 Molecular dynamics simulations 25

1.4 Thesis Outline 28

2 Theory and Methods **31**

2.1 Molecular Dynamics Simulations 32
2.1.1 Coarse-grained Molecular Dynamics Simulations 35
2.1.2 Algorithms for Molecular Dynamics Simulations 38
2.1.2.1 Thermostats and Barostats 39
2.1.2.2 Periodic Boundary Conditions and Longed-ranged Interactions 40
2.1.2.3 Integration algorithm 42
2.1.3 Performing a Timestep in Molecular Dynamics Simulations . . 42
2.2 Components of a Mass Spectrometer for Native Mass Spectrometry . 43
2.2.1 Electrospray Ionisation 44
2.2.2 Mass Analysers 48
2.2.2.1 Quadrupole Mass Analyser 48
2.2.2.2 Time-of-Flight Mass Analyser 49
2.2.2.3 Orbitrap Mass Analyser 51
2.3 Ion Mobility Mass Spectrometry 53
2.4 Dissociation Techniques 55
2.4.1 Collision-induced Unfolding & Dissociation 55
2.4.2 Surface-induced Dissociation 57
2.5 Instrumentation 59
2.5.1 Themo Fisher Scientific Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer 59
2.5.2 Themo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer 60
2.5.3 Waters Synapt G2 60

3 Chain Length Dependence of Lipid Interactions & Key Residues for the Stability of a Bacterial Sugar Transporter 63

3.1 Introduction 64

3.1.1 Protein Oligomerisation 64

3.1.2 SemiSWEET: a Bacterial Sugar Transporter 65

3.1.3 Lipid-mediated Oligomerisation of Membrane Proteins 66

3.1.4 Chapter Outline 69

3.2 Dimer Interface Stability of SemiSWEET for Different Lipids Bound . 70

3.2.1 Experimental Evidence from SID 70

3.2.2 Simulations of different CDL Species Binding to SemiSWEET 75

3.2.2.1 Initial Simulations of CDL Binding to the Occluded State of SemiSWEET 75

3.2.2.2 Binding Preferences for Different Lipids for the Occluded State 76

3.2.2.3 Creation of a Homology Model for Further Simulations 78

3.2.2.4 Definition of the Dimer Interface 80

3.2.2.5 Binding Preferences of Different Lipids for the Homology Model 82

3.2.2.6 Binding Locations for Different CDL for Homology Model 84

3.2.2.7 Atomistic Simulations to Determine the Effect of CDL binding on Dimer Stability 88

3.3 Site-directed Mutagenesis of SemiSWEET to Weaken the Dimer Interface 89

3.3.1 Mass Spectrometry of the Mutants 92
3.4 Conclusions and Future Directions 95
3.5 Materials & Methods 99
3.5.1 Expression and Purification of SemiSWEET 99
3.5.2 Native Mass Spectrometry Experiments 100
3.5.2.1 Sample preparation 100
3.5.2.2 CIU on a Synapt G2 101
3.5.3 Homology Modelling and Simulations 102
3.5.3.1 Parameters for CGMD Simulations 102
3.5.3.2 Parameters for Atomistic Simulations 103
3.5.3.3 Homology Modelling 103
3.5.3.4 ANOVA Tests 103
3.5.4 Contributions 104

4 Lipid Interactions of a Bacterial Potassium Channel 107

4.1 Introduction 108
4.1.1 BK Channel 108
4.1.2 Regulation of the BK channel 110
4.1.3 The bacterial homologue of BK:MthK 110
4.1.4 Hypothesis on Expression and Stoichiometry 112
4.2 Obtaining the *in vivo* stoichiometry 113
4.2.1 Different Cells and Vectors 114
4.2.2 Sonication of Lipid Vesicles Mass Spectrometry 117
4.2.3 Expressing TMD and RCK domain separately 119
4.2.4 Mixing purified proteins 122

4.2.5 Co-purifying MthK and RCK domain . . . 122
4.2.6 Duet Vector for Expressing MthK and RCK . . . 124
4.3 Lipid-binding of the MthK channel . . . 127
4.3.1 Endogenous Lipids and Lipidomics . . . 128
4.3.2 Lipid Titration Experiments . . . 128
4.4 Computational Studies of Lipid-binding to MthK . . . 132
4.4.1 Comparing CDL Binding to MthK to Eukaryotic Homologues 137
4.5 Conclusions and Future Directions . . . 139
4.6 Materials & Methods . . . 142
4.6.1 Expression and Purification of MthK . . . 142
4.6.1.1 Western Blots . . . 143
4.6.2 Native Mass Spectrometry Experiments . . . 143
4.6.2.1 Sample preparation . . . 143
4.6.3 Mass Spectra Collection . . . 144
4.6.4 Data Analysis . . . 144
4.6.5 Homology Modelling and Simulations . . . 144
4.6.6 Contributions . . . 145

5 A Web-based Framework for the Analysis of Native Mass Spectra147

5.1 Introduction . . . 148
5.2 Implementation . . . 150
5.3 NaViA features . . . 152
5.3.1 Charge State Selection . . . 152
5.3.2 MacSED Algorithm for Mass Assignment . . . 152
5.3.3 Comment Window . . . 156

5.3.4 Data Processing 157
5.3.5 Background Subtraction 158
5.3.6 Input Files 159
5.3.7 NaViA file format 159
5.3.8 Figure 161
5.3.9 Complex Stoichiometry 162
5.3.10 Predict Adjacent Peaks 162
5.3.11 Mass Finder 164
5.3.12 Help 165
5.4 Conclusions & Future directions 165

6 Concluding Remarks **169**

References **175**

Appendix **211**

List of Figures

1.1 Common lipid types in cell membranes. 3

1.2 Types of transporters. 6

1.3 The conservation of the selectivity filter in potassium channels. 9

1.4 Types of lipid interactions. 11

1.5 Methods to study lipid interactions with ion channels. 19

1.6 Sketch of the work flow for SoLVe-MS. 23

2.1 Sketch of contributions to a force field. In the top row the intramolecular forces are depicted in light and dark blue.In the lower row Vander-Waals forces and electrostatics are shown in green and red respectively. 36

2.2 Illustration of the Martini-Model. 38

2.3 A 2 dimensional representation of periodic boundary conditions. A square represents a unit cell. The light green moving particle exits on one side and enters the unit cell again on the other side. 41

2.4 The electrospray ionisation process in positive mode. 44

2.5 Mechanisms of ESI. 47

2.6 Quadrupole and TOF mass analysers. 50

2.7 Sketch of an Orbitrap mass analyser. 52

2.8 Different activation techniques applied in nMS 56

2.9 The instruments used in this thesis 61

3.1 Inward- and outward-open structures of SemiSWEET 66

3.2 Effect of CDL binding to to SemiSWEET. 67

3.3 Effect of CDL binding to SemiSWEET. 68

3.4 CIU of WT SemiSWEET for different lipids bound. 71

3.5 ERMS plot of SID data for different CDL binding to SemiSWEET. . 73

3.6 ANOVA of SID CDL binding results. 74

3.7 Schematic hypothesis for different CDL binding to SemiSWEET. . . . 75

3.8 Martini Models used in simulations. 76

3.9 DEI for CDL, POPG and POPE in the occluded state structure simulations. 77

3.10 Binding location of different types of CDL in initial simulations. . . . 79

3.11 Different structures of conformations of SemiSWEET and calculated CCS as well as CCS measured through IM. 79

3.12 Homology model of SemiSWEET and template (PDB 4X5N). 80

3.13 Interface of SemiSWEET dimer. 81

3.14 DEI for CDL, POPG and POPE in the homology model simulations. 83

3.15 Heatmap of the relative contact strength of residues for CDL, POPG and POPE mapped onto the sequence of the homology model. 84

3.16 Percentage of interface binding for for CDL, POPG and POPE mapped for the the homology model. 86

3.17 Percentage of contacts binding for for CDL binding to site IV. 87

3.18 Starting structure of the atomistic simulations with CDL bound & RMSD plots for CDL 14:0 and CDL 18:1 bound state of SemiSWEET for 100 ns. 89

3.19 Interface of SemiSWEET dimer. 90

3.20 Sequence alignment of SemiSWEET from different species and ontacts of conserved residues in VsSemiSWEET. 91

3.21 Mass spectra of SemiSWEET mutants upon CDL addition collected on a Waters Synapt G2. 93

3.22 CIU data of the Y62W mutant. 94

3.23 SID data of the Y62W mutant. 95

4.1 Architecture of the the BK channel and MthK. 109

4.2 Hypothesis on MthK expression and stoichiometry on the first crystal structure of MthK according to the MacKinnon paper. 113

4.3 Overview of possible stoichiometries for MthK. 114

4.4 Detergent screen, photo of precipitated MthK, western blot of expression in XL1 cells. 116

4.5 Purification and mass spectrometry of MthK purified from BL21 cells. 118

4.6 SoLVe-MS spectra of BL21 sonicated membranes with overexpressed MthK. 119

4.7 Purification and mass spectrometry of TMD of MthK. 120

4.8 Purification and mass spectrometry of TMD of MthK. 121

4.9 Mixing of purified MthK and RCK for two ratios. 123

4.10 Purification and mass spectrometry of MthK copurified with RCK. . 125

4.11 Western blot, purification and mass spectra of MthK and RCK expressed in Duet vectors. 126
4.12 Apo spectrum of partially delipidated MthK and lipidomics 129
4.13 Titration of CDL binding to MthK 130
4.14 Spectra of CDL and PG binding to MthK 131
4.15 Structure and lipid contacts of MthK 133
4.16 Normalised contacts of CDL to residues of MthK 135
4.17 Structure of $\Delta 1 - 18$ MthK used in simulations 136
4.18 Structure of BK used in simulations 138
4.19 Structure of $\Delta 1 - 115$ BK used in simulations 140

5.1 Screenshot of the interface of NaViA 151
5.2 Selection of Charge States in NaViA 153
5.3 Algorithm for Selection Overlap 154
5.4 Example of the MacSED algorithm 156
5.5 Illustration of the options to deduct the background of a spectrum. . 159
5.6 Complex Stoichiometry in NaViA 163
5.7 Adjacent Peak Prediction in NaViA 164
5.8 Mass Finder in NaViA 165

List of Tables

3.1 Parameters used in CIU experiments. 101

List of Abbreviations

5HTR serotonin receptors.

AA Ammonium acetate.

ADP adenosine diphosphate.

ANOVA Analysis of Variances Analysis.

ATP adenosine triphosphate.

BRET bioluminescence resonance energy transfer.

C8E4 Tetraethylene glycol monooctyl ether.

CCK1R cholecystokinin receptors.

CCS collisional cross section.

CDL Cardiolipin.

CEM Chain Ejection Model.

CGD Coarse-grained.

CGMD Coarse-grained molecular dynamics.

CID collisional induced dissociation.

CIU collisional induced unfolding.

CMC Critical micelle concentration.

CRM Charged Residue Model.

cryo-EM Cryogenic electron microscopy.

CSV Comma Separated Value.

DDM Dodecyl β-D-maltopyranoside.

DEI Depletion-enrichment index.

DM Decyl β-D-maltopyranoside.

DNA Deoxyribonucleic Acid.

dSTORM direct stochastic optical reconstruction microscopy.

Ec E.Coli.

ENaC Epithelial sodium channels.

ERMS energy resolved mass spectrometry.

ESI Electrospray Ionisation.

FEP Free energy perturbations.

FRET Förster resonance energy transfer.

GFP Green Fluorescent Protein.

GPCR G-Protein Coupled Receptors.

GPF green fluorescent protein.

HDX hydrogen-deuterium exchange.

IEM Ion Evaporation Model.

iFAMS Interactive Fourier-Transform Analysis for Mass Spectra.

IM-MS Ion mobility mass spectrometry.

IMAC immobilised metal affinity chromatography.

IPTG Isopropyl β-D-1-thiogalactopyranoside.

K channels Potassium ion channels.

$\mathbf{K}_{Ca}$ calcium-gated potassium channels.

$\mathbf{K}_{ir}$ inward-rectifying K channels.

$\mathbf{K}_{V}$ voltage-gated K channels.

K2P tandem pore domain potassium channels.

LB Lysogeny broth.

Lb L. biflexa.

LILBID-MS Laser Induced Liquid Bead Ion Desorption-MS.

LPS Lipopolysaccharide.

MacSED Mass and Charge State Evaluation and Determination.

MaxEnt Maximum entropy.

mGluR metabotropic glutamate receptors.

MSP Membrane scaffold protein.

MWCO Molecular weight cut-off.

NaViA Native Visual Analyser.

nESI Nanoelectrospray Ionisation.

NMR Nucleas magnetic resonance.

nMS Native Mass Spectrometry.

PA Phosphatidic acid.

PC Phosphatidylcholine.

PDB Protein Data Bank.

PE Phosphatidylethanolamine.

PG Phosphoglycerol.

PI Phosphatidylinositol.

PIP_2 Phosphatidylinositol 4,5-bisphosphate.

PME Particle Mesh Ewald.

PMF Potential of mean force.

PS Phosphatidylserine.

RCK Regulation of Conductance of potassium ions domain.

RMSD Root Mean Square Deviation.

SEC Size-exclusion column.

SF Selectivity Filter.

SID surface induced dissociation.

SMALP styrene maleic acid-lipid particle.

SoLVe-MS Sonication of Lipid Vesicles Mass Spectrometry.

SWEET Sugars Will Eventually be Exported Transporters.

TEV Tobacco Etch Virus nuclear-inclusion-a endopeptidase.

TOF time-of-flight.

Ty T. yellowstonii.

UniDec Universal Deconvolution.

Vs Vibrio sp..

VSD voltage sensor domain.

Publications

The work of chapter 5 is published in *Bioinformatics* (appended):

Quetschlich, D., Esser, T. K., Newport, T. D., Fiorentino, F., Shutin, D., Chen, S., Davis, R., Lovera, S., Liko, I., Stansfeld, P. J. & Robinson, C. V. (2021). NaViA: a program for the visual analysis of complex mass spectra. *Bioinformatics*, 37(24), 4876-4878.

A manuscript for chapter 3 is in preparation.

Chapter 1

Introduction

1.1 Cell Membranes

Cell membranes separate the inside of cells from the outside world and act as the boundary between cytoplasm and cell environment[1,2]. They carefully regulate what molecules can enter and exit the cells and to what amount[3]. Membranes are typically three to four nanometres thick and share a similar fundamental composition[4]. They consist of two main components: lipids and membrane proteins, each contributing about half of the mass of the membrane[5].

1.1.1 Membrane Lipids

Lipids are amphiphilic molecules consisting of a hydrophobic tail and a hydrophilic head group. (figure 1.1). These molecules organise themselves into a bilayer with

the hydrophobic head groups pointing out of the bilayer towards the intra and extracellular regions. The hydrophobic tails form a shielded core which provides the stability for the arrangement[1].

The chemical composition of both the headgroup and tails is diverse and there are around one thousand molecular species present in a single cell with 5% of the genome encoding for lipid synthesis[6,7]. It varies between different membranes, cell types and species[6,8]. There are four main types of lipids: glycerophospholipids, sphingolipids, glycolipids and sterols. Glycerophospholipids (also known as phospholipids) are the most common type of lipids and possess a large diversity within the class. These lipids share an architecture consisting of two linked glycerol groups connected to tails and a phosphate group. This phosphate group is connected to the group which defines the class of phospholipid. The main classes are Phosphatidylglycerol (PG), Phosphatidylethanolamine (PE), Phosphatidic acid (PA), Phosphatidylcholine (PC) Phosphatidylserine (PS) and Phosphatidylinositol (PI). In addition, there is another important phospholipid called cardiolipin (CDL). This is a special case, as it consists of two covalently linked PG lipids. The diversity of phospholipids is shown in figure 1.1. The details of protein-lipid interactions are discussed in section 1.2 after the introduction of membrane proteins in the next section.

The lipid composition of plasma membranes varies between different organisms and different domains of the tree of life[8]. For bacterial membranes, the inner membranes of *E.coli* are often used as a model organism[9]. These membranes possess a simpler composition with only three different types of phospholipids (see fig 1.1 top left), PE, PG and CDL[9,10]. These classes can be subdivided by the length and other

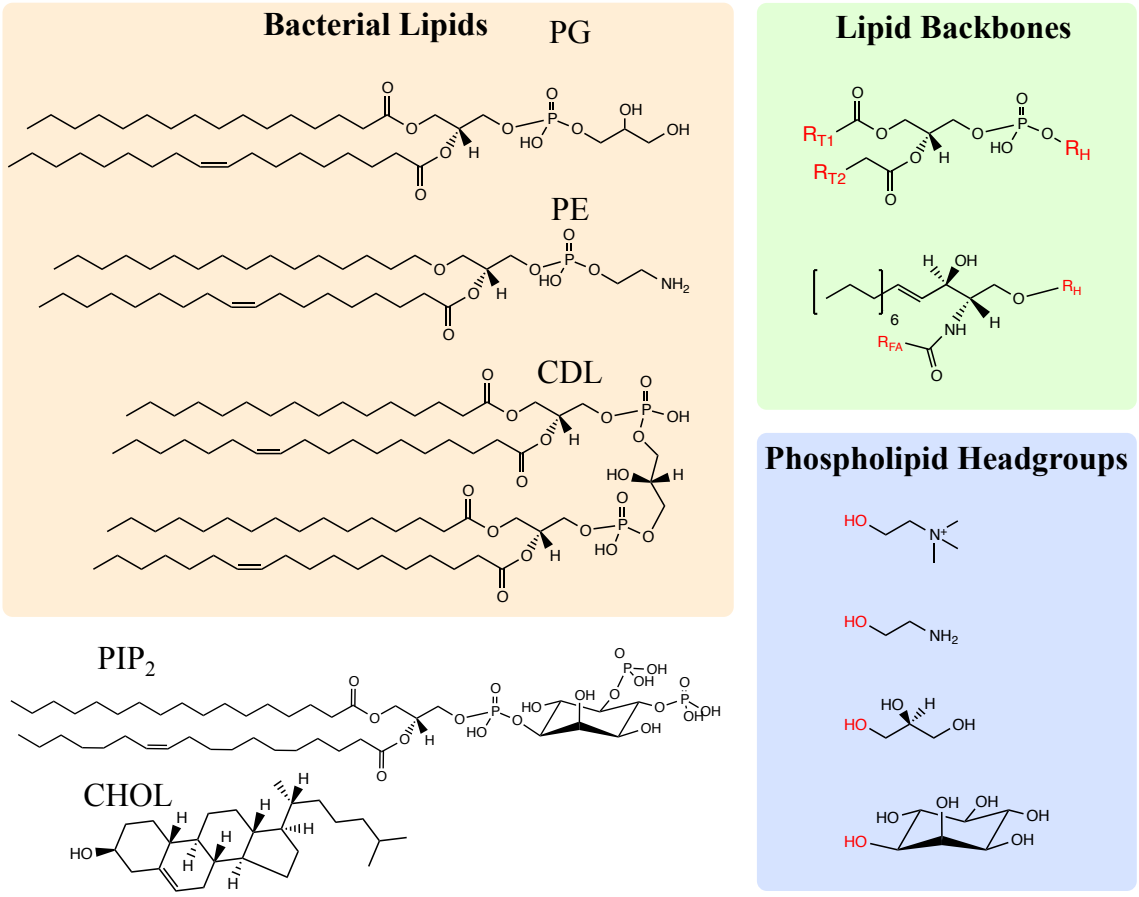

Figure 1.1. Common lipid types in cell membranes.

Different lipid types. Common lipids in bacterial membranes (yellow), lipid backbones (green) with headgroups (R_H), and tail groups R_T as well as common lipid headgroups for phospholipids (blue). In addition, a sterol, cholesterol (CHOL) and PIP_2 are shown.

properties of tails, such as degree of unsaturation. It should be noted however, that this composition only represent one case of bacterial membranes and other other bacteria possess different and often more complex membrane compositions[9].

Eukaryotic cells usually have substantially more complex membranes than this bacterial model membrane[8]. They are composed of more classes of lipids, with glycerophospholipids (49%), sphingolipids (17%) and sterols like cholesterol (34%)[11]. This diverse dynamic composition is influenced by cell activity and varies greatly by cell type[8]. In addition, these membranes also have known asymmetries between inner and outer layers of the membrane. A common signalling lipid, Phosphatidyli-

nositol 4,5-bisphosphate (PIP_2), for example is only found in the inner leaflet[12]. The functional relevance of different of lipids is an area of active research. One of the aims of this thesis is to contribute to the current understanding the effect of CDL binding to proteins.

1.1.2 Membrane Proteins

The second component of the cell membrane are proteins. Proteins are three dimensional macromolecules made up of a sequence of amino acids encoded by deoxyribonucleic acids (DNA)[3]. These biomolecules and their assembled complexes are sometimes referred to as "nanomachines of the cell" and perform various tasks ranging from signal transduction to catalysing biochemical reactions in the cell[3,13].

Membrane proteins are a class of proteins associated in cell membranes that form around a quarter of all proteins expressed by humans[14]. They are of enormous importance to modern medicine as more than five in ten of all registered drugs target these proteins[15,16]. They fall into distinct categories: peripheral membrane proteins, which are transiently attached to the inner or outer layer of the membrane and integral membrane proteins, which are permanently located in the membrane and often span both layers of the membrane[17]. The latter are the proteins investigated in this thesis. Therefore, membrane proteins are defined as integral membrane proteins unless stated otherwise. Among integral membrane proteins there are distinctions between three different classes of proteins: receptors, ion channels and transporters which are discussed in more details below[18].

1.1.2.1 Transporters

Transporters are - as the name suggests- responsible for the selective transport of water-soluble molecules across the membrane. There is a distinction between active and passive transport. Passive transporters include ion channels and porins where no energy is required to allow molecules to diffuse through the membrane. Active transporters however are able to transport a molecule against a potential or gradient[19]. An example of an active transporters is the ATP-binding cassette transporter superfamily which is one of the largest and oldest families of proteins[20]. Energy is required for active transport, which is acquired by converting adenosine triphosphate (ATP) into adenosine diphosphate (ADP)[21]. This type of transporter is called primary transporters. Secondary transporters are fuelled by other sources of energy, such as a voltage or chemical gradient[19].

Transporters undergo a conformational change when binding their substrate in order to allow it to pass the membrane. Among transporters, there is also a distinction into uniporters (fig 1.2A) which transport a single substrate, symporters (fig 1.2B), which transport two substrates in the same direction and antiporters which transport two substrates in opposite directions (fig 1.2C)[19]. For symporters and antiporters, usually one transported molecule is the substrate while the other molecule moves along a concentration or voltage gradient to fuel the secondary active transport of the substrate[19].

Transporters are vital for many biological processes of medical relevance. Examples

include multi-drug resistance in bacteria which are key player in antibiotic resisistance and conditions such as schizophrenia caused by mutations of ATP-binding cassette subfamily A member 13[22–24].In chapter 3 lipid binding of a sugar transporter will be investigated.

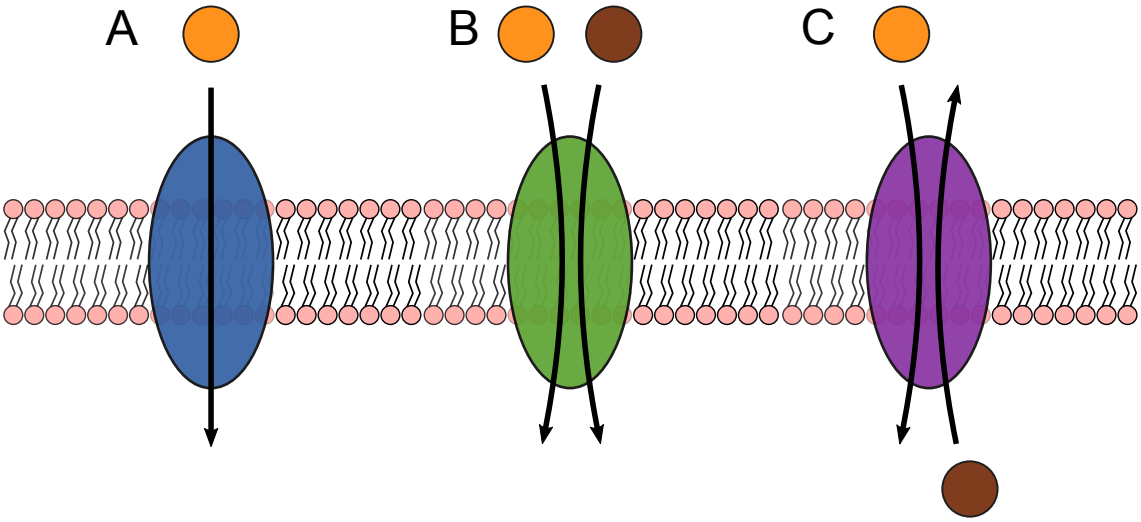

Figure 1.2. Types of transporters.
Illustration of a uniporter (A), a symporter (B) and a antiporter (C).

1.1.2.2 Ion Channels

Ion channels are proteins that allow ions to diffuse through the membrane. They are crucial for life and every known organism possesses systems to regulate the flux of ions into cells[25]. Usually, there are two states of an ion channel, open and closed[25]. In the open state ions are passively transported through these channels along a chemical or voltage gradients. In the closed state, ion diffusion is impeded. The opening and closing is highly regulated which is essential for ion channels to fulfil their biological function.

Ion channels are involved in a plethora of biological processes ranging from muscle

contraction to nerve transmission. Malfunction of ion channels is therefore linked to diseases (called 'channelopathies' because of their association with ion channels) such as nonsyndromic deafness (K_V), hyper- and hypotension (Epithelial Na^+ channel (ENaC)) or CNS disorders like epilepsy, chronic pain or mental retardation (BK)[25,26]. This key role in physiology makes them a common target for toxins and medicines[27]. After GPCRs, ion channels are the most common drug target[28,29]. Most ion channels are specific to the type of ions they allow to flow through the membrane[30]. There are ion channels for most physiologically relevant ions such as potassium, sodium, calcium or chloride[30]. In addition to a classification by ion type, they are also differentiated by their mechanism of gating[30]. There are many different gating mechanisms such as ligand-, voltage-, temperature- or light-gating. Besides, mechanosensitive channels are able to sense mechanical stimuli. Furthermore, ion channels, such as the BK channel investigated in chapter 4, can have more than one gating mechanism. In that chapter, the lipid interactions of a potassium channel will be described. Therefore, potassium channels shall be described in greater detail here.

Potassium channels (K channels) are the largest family of ion channels[31]. They can be found in all domains of life[31]. K channels are diverse in function and gating mechanism but most of them share a similar pore domain and a tetrameric architecture[32]. This pore domain consists of two TM helices linked by a pore helix and a selectivity filter (SF)[33]. This SF allows K^+ ions to selectively cross the membrane. These filter possess a highly conserved xGxGx sequence, where the glycine enables the vertical coordination groups to form a pore. This structural feature was first characterised in the KcsA potassium channel (fig 1.3A) by the MacKinnon group in 1998[32] and

has served as a model ever since[33]. This architecture has been observed in the main classes of potassium channels: it is present in inward-rectifying K channels (K_{ir}), voltage-gated potassium channels (K_V), tandem pore domain potassium channels (K2P) and calcium-gated potassium channels (K_{Ca}). This is shown in panels B-E of figure 1.3. All these potassium channels have different domains attached to the selectivity filter which enable the opening and closure of the K channels[33].

The additional domains and architecture vary by family of potassium channel. K_V possess four additional transmembrane domains with the the voltage sensor domain (VSD) at the N-terminus[35,36] (figure 1.3B). This VSD contains positively charged arginine residues in S4 transmembrane helix[37,38]. This helix moves upwards when the membrane potential changes, i.e. the membrane depolarises from a negative to a positive potential. This domain is shared by K_{Ca} (figure 1.3C). Despite the fact, that K_{Ca} is primarily gated by calcium, it is also gated by voltage[39]. In addition, these channels possess an intracellular calcium binding domain which leads to an opening of the channel when calcium is bound[40]. They will be discussed in more detail in the introduction of chapter 4.

A special class of ion channels is K2P. Their name originates from the fact that each monomer of the channel consists of two repeats of the two pore-forming transmembrane helices (figure 1.3D). Therefore, these channels operate as dimers. These channels react to a wide range of stimuli such as pH, pressure due to membrane asymmetry or signalling molecules. In addition, there are K2P members which are open at rest[41–43]. Therefore, these channels contribute to the leak current and membrane depolarisation. The final major class of K channels are inwardly-rectifying

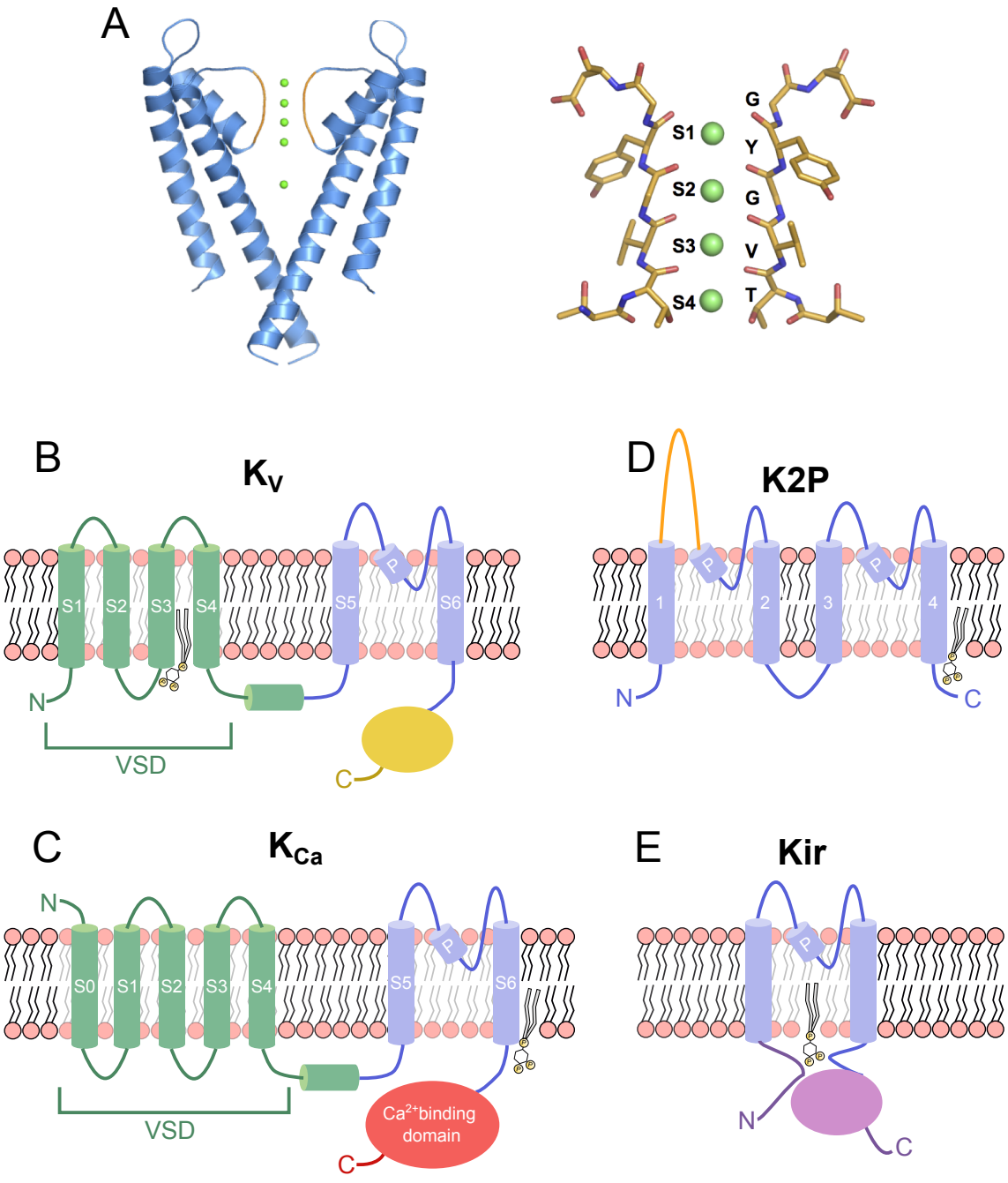

Figure 1.3. The conservation of the selectivity filter in potassium channels.
Selectivity filter in KcsA (top left) and the residues involved in the selectivity filter (top right). Conserved SF in different ion channels in the middle and bottom rows (SF in orange). Source (top row)[34].

potassium channels. K_{ir} contribute to the leak current too and are rectifying[44]. This means that they have a stronger conductance at negative membrane polarisation and less conductance after depolarisation. K_{ir} channels possess a smaller N-terminal intracellular domain and a larger C-terminal intracellular domain which

is in contact with the N-terminal domain[45] (figure 1.3). K_{ir} are regulated through a plethora of different interaction partners[45]. Most notably, they are lipid-gated ion channels. In fact, they were the first ion channels for which lipid-gating was discovered[46].

1.2 Protein-Lipid Interactions

The location of membrane proteins in lipid bilayers naturally leads to interactions of lipids and membrane proteins. These interactions involve all types of membrane proteins and many types of lipid[47]. Protein-lipid interactions can be direct or indirect and fall into three different categories: bulk lipids, annular lipids and non-annular lipids. These are shown in figure 1.4. Bulk lipids are not directly in contact with membrane proteins but they contribute to membrane properties such as viscosity, curvature or thickness and therefore indirectly affect membrane proteins[47]. Annular lipids show a preference to be close to a membrane protein (i.e. they form a shell similar to a hydration shell), but do not bind to a specific site[48]. These lipids are directly sensed by membrane proteins and can therefore influence the dynamics and function of the protein. In addition, there are non-annular lipid interactions with membrane proteins where lipids bind at specific sites. All types of lipid interactions have been shown to modulate the dynamics and function of membrane proteins. The effects of lipids onto membrane proteins are highly diverse and investigating these protein-lipid interactions is a rapidly growing field of research[47]. The interaction are specific to the protein in question[49] and all types of lipid interactions are known to modulate the behaviour of membrane proteins. Also, they are essential for modu-

lating different biological functions including medical conditions. A discussion of all known interactions would exceed the scope of this introduction. Therefore, examples of protein-lipid interactions are presented with a focus on ion channels since they are the focus of chapter 4. Comprehensive reviews by Hedger & Sansom[50], Duncan *et al.*[51] and Corradi *et al.*[47] provide a simulation focussed entry to the field while Bolla *et al.*[52] provide an overview of the evidence gathered from native mass spectrometry.

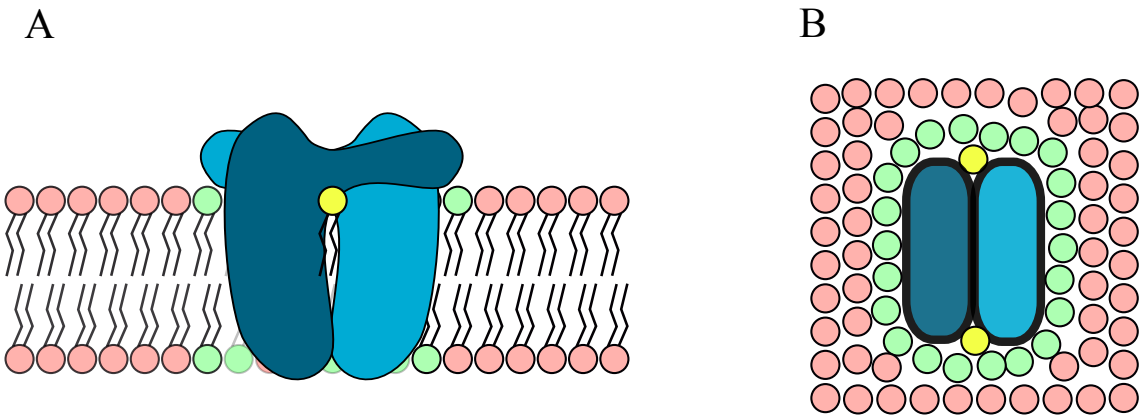

Figure 1.4. Types of lipid interactions.
Interactions of bulk lipids (red), annular (green) and non-annular (yellow) lipids with a membrane protein (blue) in side view (A) und top view (B).

Bulk lipid interactions have been shown to be present,for example, in the regulation of the epidermal growth factor receptor which requires a disordered lipid environment to be functional and was non-functional in a more viscous, ordered environment[53]. Curvature of membranes has been shown to affect the serotonin receptor (5-HT_{1A}) which exhibits an increased speed of function in membranes with a strong curvature[54]. In addition, membrane thickness influences protein structure through hydrophobic mismatch. This means, that the hydrophobic portion of membrane proteins are either exposed to water or shielded from it. This effect has been shown to influence, e.g., the activity of ATPases[55–58]. There is even a class of proteins of mechanosensitive ion channels found in all domains of life with the function of

sensing stress and deformation of membranes[59–61].

Annular lipids are in direct contact with membrane proteins and interact with them through intermolecular forces such as van der Waals interactions, electrostatics or hydrogen bonds. These interactions are sufficiently weak to allow for frequent exchange of the lipid molecules with bulk lipids. While these interactions are very dynamic and short lived. Still, there interactions can still be important for membrane protein structure of function. Examples of annular lipid interactions include TmrAB, an ABC transporter, where a composition of annular lipids tightly bound to the transporter has been observed[62], and aquaporin-0 where it has been suggested that for annular lipids the crucial interactions are acyl chains filling the gaps of the protein surface[63].

Unlike annular lipids, non-annular lipids bind to specific binding sites of membrane proteins. Therefore, the interactions are longer lived and specific to binding pockets buried deeper in the protein surface[64]. These interactions can have various effects ranging from activation of protein function to induction of dimerisation[65]. Activation through lipid-binding is a widespread phenomenon. Early work identified K channels as lipid-gated. The ATP sensitive K channel K_{ATP} is activated by PIP_2[46]. This phenomenon was then also observed for K_{ir} channels such as $K_{ir}1$, $K_{ir}2$, $K_{ir}3$ and $K_{ir}6$[66–68]. Interestingly, the opposite effect has been observed for the bacterial homologue $K_{irBAC}1.1$ when being exposed to PIP_2[69]. However, PIP_2 is not a bacterial lipid. Therefore, the significance of the observation is controversial.

Activation by PIP_2 binding is not limited to K_{ir} channels. Similarly, it was reported for other classes of potassium channels. For K2P channels it has been reported that PIP_2 increases the current influx as well as the dependency on membrane potential after the lipid bilayer experienced tension[70,71]. For TREK1 an interesting behaviour has been observed: at low concentrations, PIP_2 facilitates pore opening while at higher concentrations the opposite is the case[72]. Voltage-gated potassium channels commonly interact with PIP_2, too. An opening of the channels by PIP_2 has been observed for $K_V7.1$[73,74], $K_V11.1$ (hERG)[75], $K_V1.2$ (Shaker)[76] and $K_V2.1$[73]. In these channels, three binding sites of PIP_2 have been proposed (indicated in figure 1.3). These are located at the pore, at the linker of the VSD and at the VSD directly[77–79]. Binding events at the latter lead to a movement of the voltage-sensing S4 helix[77]. The nature of this modulation is dependent on the K_V channel and can go in opposite directions: in hERG1 the down-state is destabilised, while the opposite is true for Shaker[80,81]. This illustrates the diversity of lipid interactions even within one protein family. Since K_V and K_{Ca} share a VSD architecture, it is unsurprising that PIP_2 interactions have also been noted for members of the K_{Ca} family like the BK channel. For this channel, PIP_2 has been observed to be an additional modulator to calcium and voltage[82]. Here, PIP_2 facilitates channel opening in the presence of calcium. However, PIP_2 alone is insufficient for inducing channel opening[82]. The known binding sites of PIP_2 are also different from K_V channel. Here, a KDRDD loop between pore and calcium-binding domain has been proposed[83,84]. The interaction of PIP_2 is not limited to potassium channels. There is evidence for PIP_2 activation in sodium channels, too[85]. Epithelial sodium channels (ENaC) were shown to be activated thorough direct PIP_2 interactions with β and γ subunits[86]. In addition,

similar lipids such as phosphatidylinositol 3,4-bisphosphate and phosphatidylinositol 3,4,5-trisphosphate were indicated to have a similar effect[87]. Evidence for other sodium channels is scarce, although PIP_2 activation has been shown in a bacterial voltage-gated sodium channel while being absent in other bacterial sodium channels[88].

Other lipids also modify potassium channel function. One of those lipids is cholesterol which has been indicated to modulate K_V, K_{Ca} and K_{ir} channels[89–92]. For K_{ir} channels, the binding sites have been located at between the pore and the intracellular domains. In contrast to PIP_2, cholesterol has an inhibitory effect on most potassium channels[92]. This inhibition can manifest itself in different ways. For most K_{ir} cholesterol stabilises the closed state. This mechanism is shared by the BK channel. In K_V channels however, cholesterol was shown not to affect the opening probability but rather lead to lower unitary conductance[93]. The details of the protein-lipid interactions are sometimes not fully resolved. While early studies for K_V and K_{Ca} channels thought of bulk properties (such as fluidity, rafts and lateral stress) being altered by cholesterol and thus indirectly affecting the channel[94–96] more recent studies indicated the specific binding[97]. In fact, multiple binding sites have been identified on both types of potassium channels[92]. Also, a study using sterol with a different effect on lateral stress showed a similar inhibition[98] while chiral analogues of cholesterol had little effect[92]. For K_V channels this hypothesis has not been experimentally tested[92]. This illustrates the importance of direct specific interactions.

Protein-cholesterol interactions are not limited to potassium channels either. GPCRs for example are often activated by cholesterol. This has been shown for different receptors such as the cholecystokinin receptors (CCK1R), metabotropic glutamate receptors (mGluR) or serotonin receptors (5HTR)[99–101] where a depletion of cholesterol lead to a decrease in receptor activity. There are also indications that GPCR-cholesterol interactions can be both, direct and indirect. For CCK1R membrane fluidity has been suggested as a possible pathway for GPCR modulation as well as direct interactions[99]. For CCK1R and 5HTR, there are crystal structures available with cholesterol bound suggesting that the lipid stabilises the active conformation[102,103]. Furthermore, there are structures available where cholesterol is bound to the dimerisation interface of 5HTR and mGluR and modulates the ensemble of dimers which might point towards functional significance[92,102,104,105]. In transporters, modulation by cholesterol is known, too. For ATP-binding cassette sub-family B member 1, a multidrug resistance protein which exports harmful substances, cholesterol has been shown to change the conformation of the substrate binding site and facilitate transport[92,106,107]. Thus, cholesterol can be seen as an example of a lipid class which affects many different membrane protein classes. There are many more lipid interactions for example fatty acids, sphingolipids or other phospholipids than PIP_2 modulating membrane protein function[47].

The examples given above illustrate the diversity in protein-lipid interactions. Disturbing these interactions can also lead to medical conditions[25]. In K_{ir} channels mutations near the PIP_2 binding site can lead to an increase ($K_{ir}6.1/K_{ATP}$ in pancreatic β-cells) or a decrease ($_{ir}6.2$) in insulin production[25,108]. Other examples

include K_V channels where mutations near the PIP_2 binding site have been associated with longer QT waves[78,109]. Therefore, investigating these interactions can inform drug design and multiple techniques have been developed for that purpose.

1.3 Methods to Study Protein-Lipid Interaction

Methods to study membrane proteins fall into several different categories. There are structural techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), hydrogen-deuterium exchange (HDX), and, the most recent addition, cryogenic electron microscopy (cryo-EM). All of these techniques are capable of providing three dimensional structures of membrane proteins with and without lipids bound and provide mechanistic insight into protein-lipid interactions. However, all of them have their limitations. In X-ray crystallography, it is often challenging to determine the exact lipids from electron density. For example V-type ATPase has previously been assumed to bind to 20 PG lipids which have turned out to actually be half the number of cardiolipins[110,111]. In addition, intrinsically binding lipids are often lost during purification into a membrane mimetic[47]. Yet, in many cases tightly bound lipids could be identified[112–114]. Similarly, protein-lipid interactions can be studied by NMR[115–117]. Solution NMR has traditionally not been able to capture large proteins (above 40 kDa) and used to be limited to detergents with small micelle size[116]. However, these detergents such as foscholine-12 have been shown to have a denaturing effect on membrane proteins[118]. These obstacles have been overcome recently and allow for the study of much larger complexes[119,120]. An alternative is using with solid-state NMR, which allows to detergent free environments such as

lipid bilayers[116,121]. The latest addition to the structural toolbox is cryo-EM which has arisen in the previous decade through the resolution revolution culminating in the 2017 Nobel Prize in Chemistry[122,123]. This technique allowed to reveal structures of membrane proteins which have previously been intractable[47,124]. Examples include transient receptor potential ion channels such as TRPV1 as well other channels and classes of proteins[125–127]. HDX is a mass spectrometry technique where the effect ligands on the protein structure can be studied[128,129]. Differences in deuterium uptake of the proteins backbone allows to map which residues of proteins are more exposed to a deuterium bath in different conditions. HDX is often complementary to the other structural techniques and benefits from having an increasing number of protein structures available[130]. This technique has been successfully employed to investigate how lipids shape the conformation of secondary transporters and a secondary multidrug transporter among others[131–133].

An array of orthogonal techniques is available to study protein-lipid interactions *in vitro* and *in vivo* (figure 1.5). Some of these techniques are specific to a particular protein class. One example of this is single-channel patch-clamp electrophysiology[134,135]. This is shown for PIP_2 interactions in figure 1.5 (bottom left upper panel). In this technique, the conductance of single ion channels is recorded in response to adding a ligand such as PIP_2 (or a truncated version like diC8-PIP2 with only eight carbon atoms in the acetyl tails)[136]. This allows quantifying the activation by lipids. Alternatively, a competing agent such as neomycin can be added to compete with the activating ligand (figure 1.5 bottom left lower panel)[137]. Alternatives to this technique are flux assays based on a fluorescent dye[138]. In such assays

the exit of the dye out of a compartment is measured as a response to adding a channel opening ligand through a decaying florescence signal (1.5 - top left). To counter the changing charge, a protonophore is added to allow the influx of protons. There are many variants of these assays with some working the opposite way by activating a dye upon channel opening[139]. Fluorescent dyes or labels are also used in other approaches such as patch clamp fluorometry (figure 1.5 - bottom right lower panel). This technique combines electrophysiological measurements with measuring conformational changes in an ion channel or an ion transporter by fluorescent labels through Förster resonance electron transfer (FRET)[140]. Similarly, fluorescent binding assays allow to measure binding of a fluorescent-labelled lipid such as PIP_2 to a protein labelled by a green fluorescent protein (GPF) again through FRET or, alternatively, bioluminescence resonance energy transfer (BRET) (figure 1.5 - top right upper panel)[141]. More recently, super resolution microscopy has also entered the field of protein-lipid interactions, too (figure 1.5 - bottom right upper panel). For techniques such as direct stochastic optical reconstruction microscopy (dSTORM) it is possible to measure the clustering of fluorescently labelled lipids or a co-localisation with equally labelled proteins[142].

1.3.1 Native mass spectrometry

Mass spectrometry is an analytical chemistry technique to measure the mass of molecules. For this purpose, molecules are ionised and transferred into the gas phase where the mass to charge ratio of a mixture of analytes is determined. From this, the mass of the molecules can be deduced.

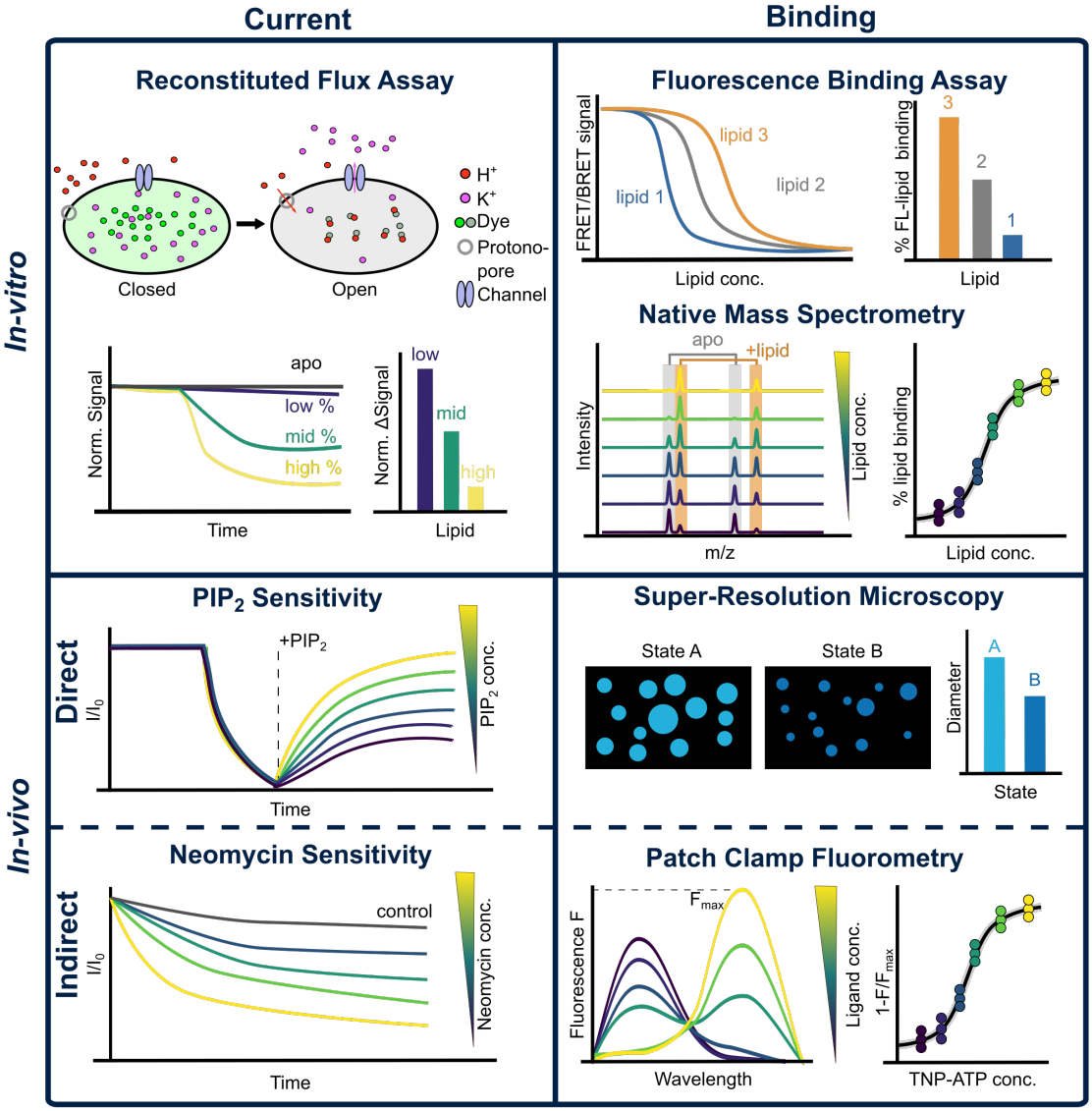

Figure 1.5. Methods to study lipid interactions with ion channels.

Different methods used to study lipid interactions in ion channels. The top row are in vitro methods while the bottom row are more in vivo methods.

After early ideas in the late nineteenth century the first modern mass spectrometers were constructed in 1918 by Arthur Jeffrey Dempster and 1919 by Francis William Aston[143,144]. The latter's works allowed the investigation of chemical isotopes leading to a Nobel Prize in Chemistry in 1922[145]. Later, mass spectrometry played a crucial role in separating isotopes in the Manhattan project[146]. Since then, mass spectrometry developed rapidly so it became known as "a complete chemical laboratory"[147].

Mass spectrometry entered the realm of biology with the development of soft ionisation techniques in the 1980s. A breakthrough was the development of electrospray ionisation (ESI) by Fenn *et al.* in 1984 paving the way to investigate the large biomolecules through mass spectrometry[148,149]. Native mass spectrometry (nMS) started in the 1990s when non-denatured protein complexes were transferred into the gas phase (although the name only appeared in 2004 and refers to the fact that the observed state matches the *in vitro* state before ionisation)[150–154]. Despite early scepticism[155], it became evident, that ESI it became evident that ESI is able to preserve the non-covalent interactions of proteins and ligands, i.e. that stoichiometry of protein complexes can be studied in the gas phase[154,156]. These development lead to the emergence of the field of native mass spectrometry which investigates the structure, stoichiometry and dynamics of intact protein complexes in the gas phase[52,154,157]. In combination with other mass spectrometry techniques such as HDX, dissociation techniques (section 2.4), ion mobility mass spectrometry (section 2.3), solution-phase techniques like titration assays, PAGE gels and mutagenesis as well as computational techniques[52,154] it is possible to study various aspects of proteins and their interaction partners: ligand identification, kinetics and affinities of ligands, binding sites, protein stability, subunit connectivity and protein conformation as well as their change upon ligand binding[52].

In fact, native mass spectrometry has been established as an invaluable tool to investigate protein-lipid interactions[52,128,158,159] (1.5 - top right lower panel). Alongside molecular dynamics simulations, this technique forms the main part of this thesis.

Therefore, technical details and the associated physical processes are explained in chapter 2. nMS is often used as a complementary technique to e.g. X-ray crystallography[52]. Native mass spectrometry is a valuable addition to existing tools of structural biology because it neither requires sample labelling nor sample homogeneity. In addition, it offers great sensitivity with a few microlitres of sample at micromolar concentration being sufficient for an experiment[160]. This is particularly useful in the field of membrane proteins where low yields are a significant restraint to research[161].

The study of membrane proteins in nMS started in 2008 with the first spectrum of an intact membrane protein complex in detergent being collected by Barrera *et al.*[162]. After this initial breakthrough on ATPases, other membrane proteins have been studied as well and lipids have been identified to be bound to membrane proteins[111,163–165]. With the development of suitable nMS platforms it became possible to characterise more and more protein-lipid interactions[165–167].

In key studies, lipids have been identified to influence the stoichiometry of membrane proteins[168]. There it was shown, that some membrane proteins, including a bacterial sugar transporter semiSWEET, require lipids as a 'molecular glue' to form functional dimers (chapter 3 follows this study up). Another example of is the work of Yen *et al.* where PIP_2 was shown to stabilise the active state of a three class GPCR and to improve the selectivity of G-protein binding[169]. As GPCRs are key drug targets, nMS allows to inform drug discovery for these proteins[170]. Implications for drug discovery could also be derived from a work on the Lipopolysaccharide (LPS) system where binding mechanism of an antimicrobial peptide has been

studied through a combination of nMS, MD and HDX[22]. Other work illustrates the close relationship of membrane proteins with their environment: for example, a mechanosensitive channel MScL exhibited different stoichiometries based on the used detergent and lipids[171,172]. This is one of many examples of the influence of detergent that illustrate the need to investigate other membrane mimetics for nMS (although new detergents can be used to fine tune these lipidation and charge of the protein complexes[173,174]).

Popular detergent-free methods for nMS include using lipid-nanodiscs or styrene maleic acid-lipid particle (SMALP)) as alternative membrane mimetics, and Sonication of Lipid Vesicles Mass Spectrometry (SoLVe-MS)[175,176]. Nanodiscs are artificial lipid bilayers encircled by two amphipathic membrane scaffold protein (MSP)[177]. These proteins in 'nanomembranes' were introduced in nMS to provide a more natural environment for membrane proteins because no detergent is involved (except from membrane solublisation)[178,179]. This led to advances in instrumentation and software (see UniDEC sections 1.3.1.1)[130]. A similar alternative are SMALPs. They go a step further and eliminate the detergent in the membrane solublisation step[180,181]. In SMALPs the MSP is replaced by an SMA polymer which helps to extracts membrane proteins from the membrane with native lipids[182]. This membrane mimetic has been used in combination with Laser Induced Liquid Bead Ion Desorption-MS (LILBID-MS) to ionise SMALPs to investigate the native oligomeric state of embedded membrane proteins[175].

The final method (which does not require any chemical) for membrane mimetics

is SoLVe-MS (figure 1.6). This technique was developed by Chorev *et al.*[176,183]. In this technique purified membrane samples are sonicated for several minutes to create lipid vesicles containing membrane protein complexes. These vesicles are dissociated in the mass spectrometer to enable a investigation of "protein assemblies ejected directly from native membranes"[176]. While this technological breakthrough lead to intense debates in the community[184,185] it definitely offers an exciting route for investigating protein complexes where purification through overexpression of protein is not possible such as proteins of blood platelets[186].

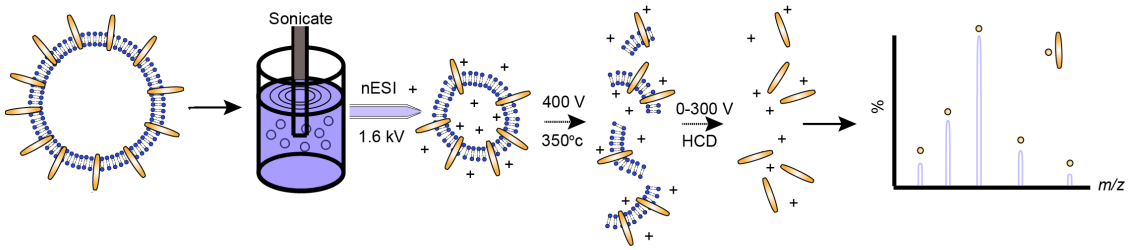

Figure 1.6. Sketch of the work flow for SoLVe-MS.
A sample of purified membranes containing proteins is sonicated to create smaller vesicles. These enter the mass spectrometer through nESI where they will be dissociated is the source region and cleaned of most lipids in the collision cell. As a result protein complexes including native ligands are being resolved. Image from Chorev et al.[176].

1.3.1.1 Software & algorithms for assigning native Mass Spectra

Native mass spectra of macromolecules such as proteins contain ions of multiple molecular species and charge states[187]. This is particularly true when the analytes have several oligomeric states or a large mass range, or include adducts such as salt, lipids or detergent[187]. Therefore, analysis algorithms for these spectra need to deconvolute the charge distribution and the charge state of an ion. Over time, multiple approaches have been developed. Early attempts based on the principle of maximum entropy (MaxEnt)[188,189] were employed but most instrument manufacturers[187].

These algorithms calculate theoretical charge-neutral spectrum through optimising an objective function based on Shannon entropy and quality of fit. However, the specifics of the used scores and criteria are not publicly known[187,190]. In addition, manufacturers include software like MassLynx (Waters) an Xcalibur (Thermo Fisher Scientific) provide tools are useful to select spectra from a histogram of the data acquisition but offer little for the extended analysis of the spectra.

To address this shortcoming, multiple other approaches have been presented over the years. Universal Deconvolution (UniDec) was published in 2015 and employs a Baysian framework which fits a theoretical spectrum to the raw data in an iterative manner[191]. This framework was designed with polydisperse spectra from membrane proteins in nanodiscs as well as ion mobility data in mind. It is widely used in the nMS community with more than 300 citations since publication (according to Google Scholar as of April 2021). Since its publication, this framework has been extended by modules for high-throughput MetaUniDec[192], improved algorithms[193] as well as a scoring function[190] UniScore. This score provides a "intuitive" assessment of the quality of fitting, including criteria such as consistency of peak shape for different charge states, smoothness of the charge distribution and the symmetry of the peak distribution[190]. A commercial alternative to UniDEC is Protein Metric Intact which uses a "parsimonious" algorithm[187,194]. This algorithm tries to fit a spectrum with as few masses as possible and a smooth charge state distribution[187]. Similarly to MaxEnt, details of the exact criteria for this algorithm remain undisclosed[190].

While the examples mentioned above form the main software for nMS, there are

further publications focussing on specific aspects. Interactive Fourier-Transform Analysis for Mass Spectra (iFAMS) focusses on spectra of many repeated subunits or salt adducts to proteins and has been incorporated into UniDec[195–197]. Additional work aims at improving peak detection[198] or providing useful algorithms for background subtraction[199]. There are further software packages such as ORIGAMI which was developed for Waters instruments[200] and packages for ion mobility (PULSAR) or collision-induced unfolding data (CIUSuite)[201,202].

1.3.2 Molecular dynamics simulations

The second technique used in this thesis are molecular dynamics simulations. MD simulations make it possible to observe the "wiggling and jiggling" of atoms and molecules in atomistic resolution which the Nobel Prize winning physicist Richard Feynman was refererring to when he introduced the atomic hypothesis as the most powerful thought of science in his famous lectures on Physics[203]:

> *"All things are made of atoms — little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another. In that one sentence ... there is an enormous amount of information about the world."* – Richard Feynman

This technique was proposed in 1955[204] and further developed to study the behaviour of hard spheres[205,206]. The next major breakthrough was the study of liquid argon in 1964[207]. As the function and dynamics of biochemistry are also covered by

the laws of physics, this technique has been applied to numerous examples ranging from early simulations of bovine pancreatic trypsin inhibitor in 1977 to the study of the crowded environment of whole cells four decades later[208–210]. This broad range of applications culminated in the Nobel Prize in Chemistry 2013 being awarded to the leading scientists in the field[211]. The field is still rapidly developing with the use of graphical processing units and specialised hardware being developed for MD[212,213].

The simulation of lipid bilayers began in the 1980s and a decade later the first proteins in bilayers were investigated[214,215]. The scarcity of membrane protein structures was initially a limitation to MD simulations of this class of proteins. But with the dawn of cryo-EM the number of membrane protein structures doubled over the last decade leading to a significant rise in popularity of MD simulations of membrane proteins[216,217]. MD simulations made many contributions to the understanding of membrane proteins. Lipid binding sites can be determined, conformation changes can be investigated (to a degree if they happen fast enough), binding affinities and kinetics can be estimated as well as functional annotation[50,218–220]. In the paragraphs below a few examples illustrate these contributions.

Atomistic simulations of membrane protein are computationally expensive. Therefore, coarse-graining approaches have been developed. The approach used in this work is the Martini force field, which is described in section 2.1.1. This force field originates from the simulation of lipid bilayers and then transitioned to including proteins. It has been widely applied to membrane proteins including studying lipid interactions of ion channels, G-Protein Coupled Receptors (GPCR) and trans-

porters[22,169,221–224]. This makes it a suitable choice for studying lipid interactions of ion channels and sugar transporters in this thesis. In addition, it has been employed to describe a fingerprint of lipid interactions for different protein classes[49]. Coarse-grained CGD MD simulations can be integrated with even more coarse approaches such as dynamically triangulated surfaces to study membrane deformations and their effect on membrane proteins[225–227]. Martini is also commonly used in multiscale simulation studies with atomistic simulations[50,228,229]. CGMD can be employed to determine which lipids bind at which location on the protein and consecutive transformation into atomistic coordinates is able to provide a more detailed picture of the forces involved in the binding process[229]. An example for functional annotation through CGMD are K_{ir} channels, where Pipatpolkai *et al.* characterised the PIP_2 binding affinities for residues and mutations of clinical relevance[222].MD simulations can also be applied to complement other structural techniques. One example is the refinement of the open state of the glycine receptor containing problematic regions which are important for molecules of medical interest[230].

Energetics of ligands (including lipids) binding can be addressed though different routes in atomistic as well as coarse-grained molecular dynamics[219,231]. To date, there are two methods to determine protein-lipid interactions: potential of mean force calculations (PMF) and free energy perturbations (FEP). Both have been applied to numerous examples[219,232]. For PMF calculations the ligand is pulled along a collective variable (usually a linear coordinate perpendicular to the binding site) and short simulations are performed in small windows inside an umbrella potential. The results of these simulations are combined via weighted histograms to explore

the one dimensional free energy profile along the collective variable. This technique has first been applied to lipid binding with CGMD for CDL binding to the cytochrome c oxidase and complex IV[233,234] but it has been employed for other classes since[219,235,236]. An alternative to PMFs are FEP calculations where a ligand or a subgroup is "alchemically" morphed into a different molecule[219,237]. Therefore, the energy difference of binding of these states can be calculated though thermodynamic cycles[232]. This technique offers similar values to PMF but have the advantage of lower computational costs.[219]. In addition, this approach can be applied to measure the effect of site-directed mutagenesis as well by morphing key residues in lipid binding[222].

Finally, MD simulations prove to be a useful tool for complementing other structural techniques. This has been demonstrated in numerous examples. For new protein structures acquired by cryo-EM or X-ray crystallography, MD simulations often yield additional insight into the system in question[238–240]. The same is true for nMS or HDX studies[22,168,241]. Thus, MD and other techniques can synergise to achieve great insight into biomolecular systems.

1.4 Thesis Outline

This thesis presents research from three projects. To this end, theory and methods used in both native mass spectrometry and molecular dynamics simulations are introduced in chapter 2. There, components of a mass spectrometer are explained as well as methods activating membrane protein complexes and instrumentation.

Furthermore, foundations of molecular dynamics including the used coarse-graining methods are presented.

The first results chapter is chapter 3 which explains how molecular dynamics are used as a complementary tool for studying lipid binding influencing dimer stability probed through surface-induced dissociation data for a bacterial sugar transporter, semiSWEET, provided by a collaborator, Denis Shutin. In addition, the influence of mutations to interrupt the interface has been collected. Chapter 4 represents the bulk of the work done for this thesis. It investigates lipid interactions for a bacterial potassium ion channel MthK in mass spectrometry experiments as well as simulations. In addition, the data is compared to lipid binding to its eukaryotic homologue, the BK channel. This chapter includes work to achieve the in-vivo stoichiometry of the channel. The final chapter describes a novel program for the analysis of mass spectra. The creation of this program has been sparked by complex mass spectra acquired for investigating MthK stoichiometry by SoLVe-MS spectra.

Chapter 2

Theory and Methods

The work in this thesis is based on the integration of two complementary techniques: molecular dynamics simulations and native mass spectrometry. Native mass spectrometry is employed to determine protein stoichiometry and ligands. However, nMS does not provide structural information on binding sites. Therefore, information obtained by native mass spectrometry can be used as input for the setup MD simulations. These simulations can serve as a "computational microscope" to investigate binding locations. Examples of the combination of MD and nMS include PIP_2 binding to G-Protein Coupled Receptors[169], lipid-induced oligomerisation[65] and lipid interactions of OmpF[242]. Both techniques will be explained in the following sections to a level of depth necessary for the understanding of the following chapters.

2.1 Molecular Dynamics Simulations

Molecular dynamics simulations allow the dynamics of biomolecules to be investigated by computing trajectories of their movement in time. These dynamics are covered by the time-dependent Schrödinger equation[243]

$$\frac{i\hbar\partial\Psi(\vec{r},\vec{R})}{\partial t} = H\Psi(\vec{r},\vec{R}). \tag{2.1}$$

Here, $\hbar$ is Planck's constant divided by 2π, H is the Hamiltonian and Ψ the wave function. The wave function is determined by the electron coordination $\vec{r}$ and the nucleus coordinates of the atoms of the molecules $\vec{R}$. Solving this equation directly is computationally not feasible. Hence, three approximations allow us to map this equation to the classical equation of motion.

Born-Oppenheimer Approximation The Born-Oppenheimer Approximation is based on the observation that the mass of a nucleus is four orders of magnitude heavier than electrons in the atomic shell[244]. Therefore, the dynamics of electrons is much faster than the dynamics of the nucleus. This allows one to separate the dynamics of the nucleus and electrons from one another. In molecular dynamics simulations the positions of electrons are assumed to be equilibrated in their ground state for each timestep which allows for larger timesteps and therefore, reduces computational cost[245].

Classical Approximation The second approximation of molecular dynamics simulation is that Newton's equation of motion is a sufficient description for the behaviour of biomolecules. Newton's equation of motion can be expressed by

$$\ddot{\vec{r}} = \frac{1}{m}\nabla U(\vec{r}), \tag{2.2}$$

where $\vec{r}$ is a set of atomic coordinates, m the mass and $U(\vec{r})$ is a potential. This potential arises due to interactions such as electrostatics, dipole interactions and Van-der-Waals interactions. Solving this equation enables the computation of trajectories for the system. Assuming that the time average can equals the ensemble average (i.e. the system is ergodic), the simulation of long trajectories enables enables the probing of thermodynamic and kinetic behaviour of the system in question.

Force Fields The final approximation are force fields[246]. They approximate $U(\vec{r})$ by a combination of analytical functions of nuclei coordinates. This approximation of the true potentials is commonly called a "force field", despite the fact that they describe a potential and not forces. Force fields generally describe the potential in terms of intramolecular and intermolecular interactions,

$$U(\vec{r}) = U_{Intra.}(\vec{r}) + U_{Inter.}(\vec{r}). \tag{2.3}$$

The first term of this equation describes the interactions that keep a series of covalently bound atoms of a molecule in their respective position and consists of the following terms

$$U_{Intra.}(\vec{r}) = U_{Bonds}(\vec{r}) + U_{Angles}(\vec{r}) + U_{Dihedrals}(\vec{r}). \tag{2.4}$$

Here $U_{Bonds}(\vec{r})$ and $U_{Angles}(\vec{r})$ are modelled as harmonic potentials around equilibrium bond length r^{eq} for specific bonds and equilibrium angles θ^{eq}, respectively. These spring potentials also have a spring constant $k^{b/\theta}$ that depends on the atoms involved. For the dihedral angles, the torsion energy is described by a cosine function of the rotation angle χ around a bond with n and χ^0 determining energy minima and a force constant k^{χ} describing the strength of the interaction. The shape of these potentials as well as their mathematical form are given in the upper row of figure 2.1.

The intermolecular forces are captured by the second term of equation 2.3, $U_{Inter.}(\vec{r})$. They consist of the electrostatic potential, i.e. attractive or repulsive Coulomb interactions $U_{Coulomb}(\vec{r})$ and a Lennard-Jones potential $U_{LJ}(\vec{r})$ to model dipole interactions. Therefore,

$$U_{Inter.}(\vec{r}) = U_{LJ}(\vec{r}) + U_{Coulomb}(\vec{r}). \tag{2.5}$$

The form of these potentials is sketched in the lower row of figure 2.1. These parameters of these potentials, such as force and equilibrium constants form the foundation of a classical force field. Their values are based on quantum mechanical calculations, experiments or a combination of both, depending on the force field. Commonly used force field are AMBER and CHARMM, but there is a range of force fields available today and the development of new and improved force fields remains an area of active research[247–249]. Advanced force field might include effects such as polarisation[250].

2.1.1 Coarse-grained Molecular Dynamics Simulations

Molecular dynamics simulations are computationally expensive and therefore only shorter timescales (up to a microseconds) and smaller systems are feasible to simulate with limited computing power[251]. This is especially true for all-atom simulations where each atom is treated explicitly. To overcome this limitation, united atom force fields have been developed to calculate the movements of hydrogen atoms implicitly[252]. Coarse-grained molecular dynamics (CGMD) goes a step further as they merge heavy atoms (i.e. non-hydrogens) in order to improve simulation speed[253].

One approach to CGMD is the Martini Force Field[253,254]. This model merges four heavy atoms on average to create a bead which interacts with other beads. The advantage of merging atoms is that it reduces the number of interactions. In addition, the fastest interactions such as the movement of hydrogen atoms are integrated. Therefore, the Martini Force field is up to three orders of magnitude faster than

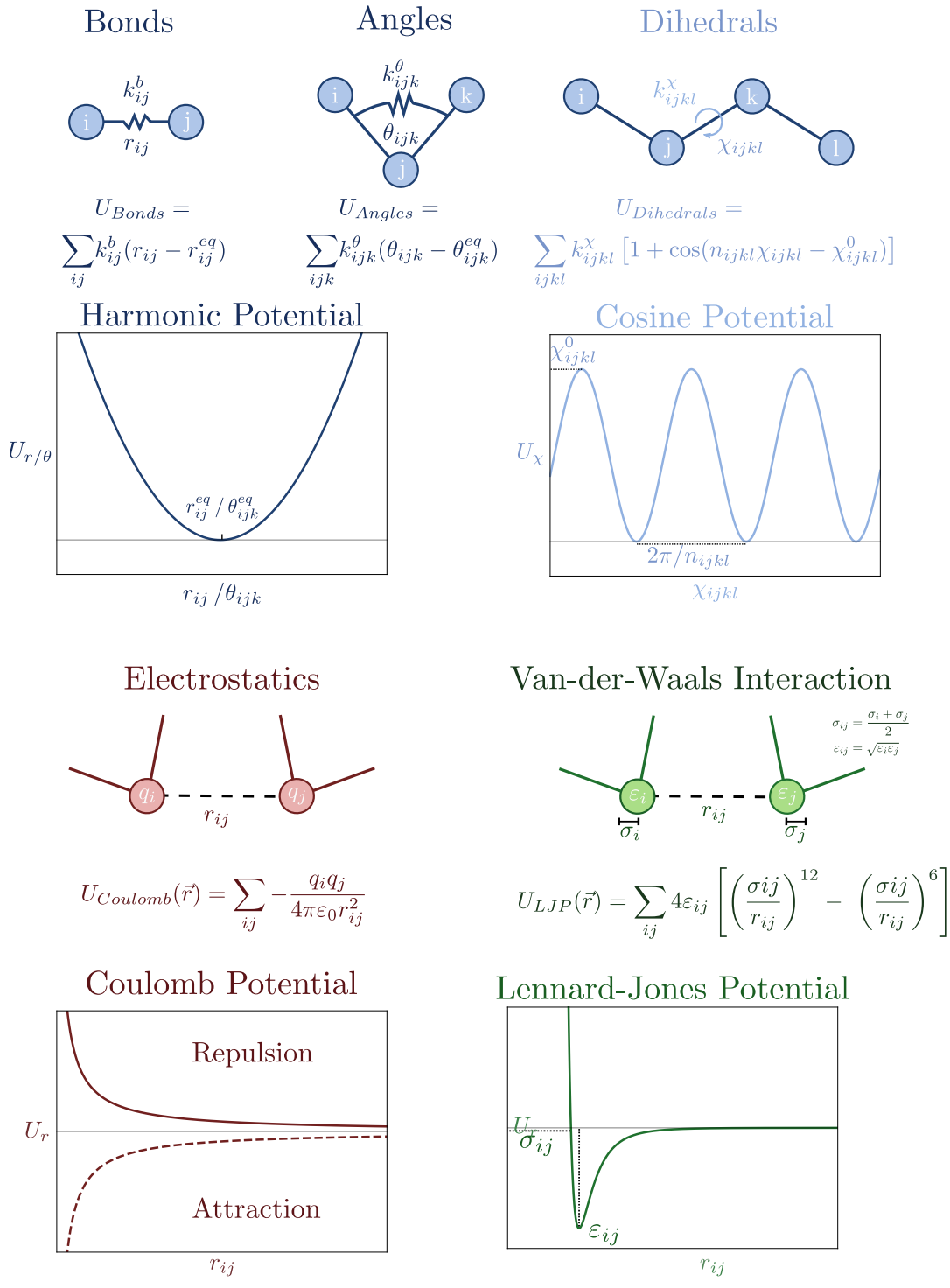

***Figure 2.1.** Sketch of contributions to a force field. In the top row the intramolecular forces are depicted in light and dark blue.In the lower row Van-der-Waals forces and electrostatics are shown in green and red respectively.*

atomistic molecular dynamics simulations. However, the price of this simplification is a loss of complexity, e.g. hydrogen bonds, in CGMD leading to an instability of the protein structure. To rectify this problem, the protein coordinates are maintained with an elastic network holding the structure together[255]. This elastic network consists of a series of harmonic potentials between backbone beads, within a given cut-off, to stabilise protein structure. A downside of this approach is that elastic network prevents proteins from adopting different large scale conformations. Thus, CGMD with Martini and an elastic network are unable to detect larger conformational change[255]. This limitation remains unresolved but recent work on on exchanging the harmonic models for Lennard-Jones potentials might overcome this obstacle[256].

The Martini Model was published in 2004 for modelling lipids in a bilayer[254]. Since then, the force field parameters where improved and parameters for proteins have been added[257,258]. In addition, further attempts for multiscale modelling as well as speeding up simulations through implicit solvent were conducted[256]. Despite its use in the study of protein-lipid interactions the Martini model is not without critique and pitfalls[259,260]. A common critique is that certain lipids are too attractive and appear to be bind non-specifically[259]. Another concern is the problem of different sized beads interacting because mixing of different sizes Lennart-Jones potential might lead to artificial energy barriers[260]. In response to these pitfalls, Martini 3.0 has been developed, with these interactions reduced, along with other improvements[261]. This version also aims to broaden the scope of applications of the Martini model to areas such as material sciences, where it has been used to investigate sys-

tems like organic electronics or nanoparticles[262].

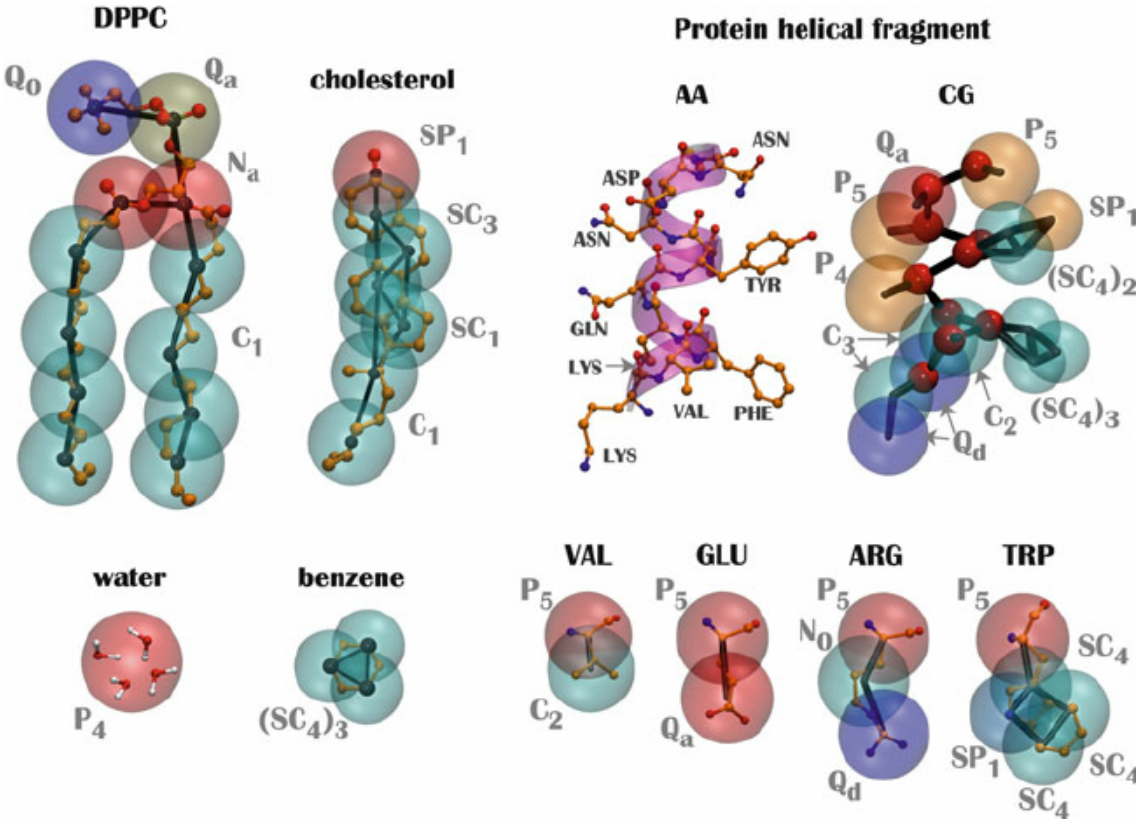

Figure 2.2. Illustration of the Martini-Model.

Coarse-graining of different molecules overlaying atomistic level. The molecules shown are the lipids (DPPC, cholesterol), water, benzene a protein fragment and several amino acids. Hydrogens are only shown in the case of water. Source[263]

2.1.2 Algorithms for Molecular Dynamics Simulations

In order to perform molecular dynamics simulations on a computer, Newtons equation of motion needs to be solved for consecutive time steps. In addition to the force field modelling the physics of the biomolecular system a few more concepts are needed to perform simulations. These are thermo- and barostats for keeping temperature and pressure constant, periodic boundary conditions for preventing the system to be dominated by the "walls" of the system and an algorithm to actually

solve Newton's equation of motion.

2.1.2.1 Thermostats and Barostats

Molecular dynamics simulation ensure that the energy and particle number of the molecules in the system are constant. However, biological processes usually take place at constant pressure and temperature. For a realistic representation of these processes, it is therefore important to preserve these quantities. In MD, this is done though thermostats and barostats. While there are many options, we focus here on the thermo- and barostats used in this thesis. For CGMD pressure coupling was done with a Berendsen barostat[264]. This method couples the system weakly to an external bath and rescales the pressure by a time constant τ_p so

$$\left(\frac{dp}{dt}\right)_{ext} = \frac{p_0 - p}{\tau_p}. \tag{2.6}$$

The box and velocities are scaled proportionally to this term. While this barostat has been tested to be sufficient for CGMD, a more advanced technique for pressure coupling is the Parinello-Rahman barostat[265,266]. This method has been employed for atomistic simulations and CGMD after equilibration. It allows the box to adapt in size during the simulation. Therefore, the box edges are organised into a 3×3 matrix $\mathbf{B}$ of the three vectors spanning the edges of the box. The coupling through the thermostat enter are given by

$$\left(\mathbf{W}^{-1}\right)_{ij} = \frac{4\pi^2 \beta_{ij}}{3\tau_p L}. \tag{2.7}$$

Where β_{ij} are the isothermal compressibilities (usually given as $\beta_{H_2O} \approx 4.510^{-5}$ bar), $L = max_{ij}\,(B_{ij})$ (largest box element) and the coupling constant τ_p. With this the box vectors evolve according to

$$\frac{d^2\mathbf{B}}{dt^2} = V\mathbf{W}^{-1}\left(\mathbf{B}^T\right)^{-1}(P - P_{ref})\,. \tag{2.8}$$

Here, P and P_{ref} are the current and reference pressures. Both types of simulations use V-rescale (also known as Bussi thermostat) for temperature coupling[267]. This algorithm rescales all velocities according to the temperature $T_{resc.}$ set by the simulation setup.

$$\vec{v}_{resc.} = \sqrt{\frac{E_{kin,resc.}(T_{resc.})}{E_{kin}(\vec{r})}} \quad \text{with} \quad E_{kin,resc.}(T_{resc.}) = \frac{N_f k_b T_{resc.}}{2}. \tag{2.9}$$

Here, $E_{kin,resc.}(T_{resc.})$ and $E_{kin}(\vec{r})$ are the kinetic energies expected from the set temperature with N_f as the degrees of freedom of the system.

2.1.2.2 Periodic Boundary Conditions and Longed-ranged Interactions

Molecular dynamics simulations usually employ periodic boundary conditions (PBC). PBC state that the sides of the simulation box are connected. This means that a particle exiting the box to the right will re-enter on the left. This principle is illustrated in figure 2.3 below. Periodic boundary conditions are used to avoid box size and box walls dominating the behaviour of the biomolecular system.

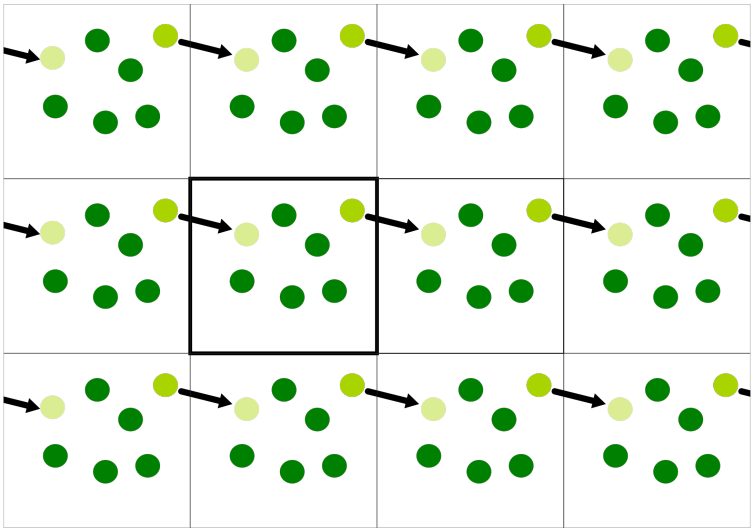

Figure 2.3. A 2 dimensional representation of periodic boundary conditions. A square represents a unit cell. The light green moving particle exits on one side and enters the unit cell again on the other side.

While intramolecular forces as well as Lennard-Jones potentials are short-ranged and are thus largely unaffected by periodic boundary conditions, Coulomb potentials are longer ranged and need to be addressed. The problem is, that an infinite number of boxes adjacent to one another leads to a non-convergent infinite sum of electrostatic potentials. This problem is addressed by the Particle Mesh Ewald (PME)[268,269]. This method combines Ewald summation and "smearing out" of particles charges over a grid to achieve computational efficiency. Ewald summation makes use of the fact that these infinite sums are easier to calculated in reciprocal space. Hence, the potential is split up into direct interaction and self-interaction, which are computed in real space and long-ranged potentials which are calculated in reciprocal space following a Fourier transformation. This makes long-ranged interactions feasible to compute.

2.1.2.3 Integration algorithm

To compute trajectories the position and velocities of particles in the system need to be updated in intervals of the timestep Δt. This timestep needs to be chosen such that it is small enough to capture the fastest movements of the system. This is usually the vibration of hydrogen atoms around covalent bounds. To improve computational speed by doubling the timestep, these bonds are commonly restricted around the equilibrium length. This allows for a timestep of 2 fs instead of 1 fs. A commonly used scheme in MD is the leapfrog-algorithm, where the positions $\vec{r}(t)$ and velocities $\vec{r}(t)$ are updated according to the following equations[270]:

$$\vec{v}(t + \Delta t/2) = \vec{v}(t - \Delta t/2) + \Delta t\vec{a}(t+) \tag{2.10}$$

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \Delta t\vec{v}(t + \Delta t/2). \tag{2.11}$$

2.1.3 Performing a Timestep in Molecular Dynamics Simulations

To prepare molecular dynamics simulations, the initial conditions of the system need to be setup. This is done by placing the protein structure in a box and afterwards adding additional molecules such as water, ions and in the case of membrane proteins into the box as well[271]. For membrane proteins, the lipids can either be placed randomly into a box following the MemProtMD pipeline or placed in onto a two dimensional grid using the *insane* (short for INSert membrANE) script[271,272]. Afterwards an energy minimisation is performed.

To perform one timestep in MD simulations all of the steps described above need to be performed in order. First, the initial conditions need to be used as input to compute forces though the force field as a second step. Afterwards, positions $\vec{r}$ and velocities $\vec{v}$ need to be updated. This is the key step which consists of multiple substeps: the thermo- and barostats variables need to be determined. Next, the velocities will be updated though the leapfrog algorithm. Afterwards, they will be corrected by the thermostat and positions will be updated. As a penultimate step they will be checked for position constrains and the barostat will be applied. Then, the output is used as the input for the next timestep if the simulation is not finished.

Once the simulation is finished, the trajectory can be analysed through various techniques ranging from Markov State models, over Root-Mean Square Deviation (RMSD) to contact analysis[273–275].

2.2 Components of a Mass Spectrometer for Native Mass Spectrometry

The simplest mass spectrometer consists of an ionisation source for the creation of the analyte ions, a mass analyser to separate these ions by m/z and an ion detector to detect them. Here, the components used for nMS shall be introduced.

2.2.1 Electrospray Ionisation

The first step of mass spectrometry is ion generation. In native mass spectrometry, this is done through ESI. This concept was first proposed for mass spectrometry in 1968 by Dole *et al.* to transport large molecules into a mass spectrometer[148]. However, this idea was experimentally realised by Fenn as late as 1984[149]. ESI is based on connecting a thin capillary holding the analyte liquid to a downstream electrode with high potential (figure 2.4). In positive mode, the potential leads to a build-up of positive charges at the Taylor cone of the sample solvent surface such that charged droplets are ejected from the capillary forming a jet[156]. In the electric

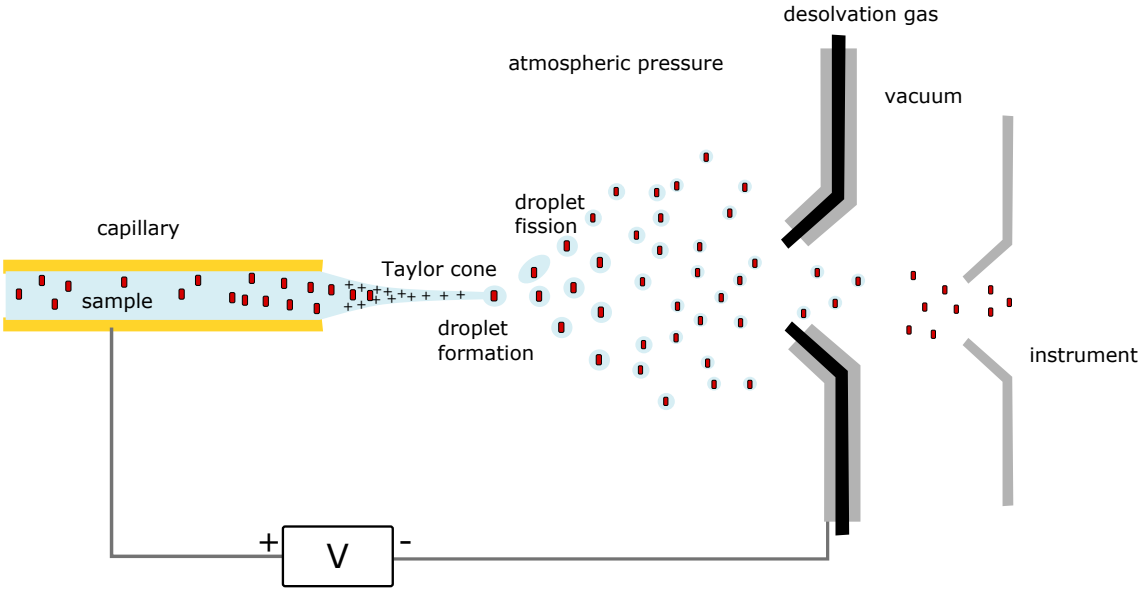

Figure 2.4. The electrospray ionisation process in positive mode.
Samples are loaded into a capillary connected to a high voltage power supply (V). Droplets containing proteins (red) form at the tip of the Taylor cone and fission due to Coulombic repulsion. Desolvated ions of proteins enter the instrument.

field solvent evaporates reducing the droplet size. This causes an increase of surface charge density. The surface charges repel one another leading to droplet fission. The equilibrium for the maximum stable charges in a droplet, Q_{Ry}, is given by the

Rayleigh equation

$$Q_{Ry} = 8\pi\sqrt{\varepsilon_0 \gamma R^3}, \tag{2.12}$$

where R is the the radius of the droplet, γ the surface tension and ε_0 the electric permittivity of the local environment. To facilitate the evaporation a gas such as N_2 can be added, with a flow direction opposite to droplet motion. This process is sketched in figure 2.4.

The precise mechanism of ion formation from droplets eluded understanding to date. However, there are two main opposing theories for explaining the process (figure 2.5. The first is the ion evaporation model (IEM) published in 1976 by Iribarne and Thomson[276]. In this model a droplet containing a charged analyte shrinks until the repulsion of charges becomes sufficiently large for the charged analyte to exit the droplet as an ion. This mechanism applies to molecules as analytes[277]. However, this process has also been observed for small protein ions when the protein was supercharged by lowering the pH value of the solution[278]. This requires high stability of protein structure in acidic conditions and is therefore not a common mechanism for the ionisation of proteins.

The competing theory of the charged residue model (CRM) on the other hand is thought to apply to large molecules such as globular proteins[148,279]. In this model, Coulombic fission and IEM occur until only a single molecule of interest is present in a droplet. These droplets undergo fission and evaporate until the analyte with

almost no solvent molecules is left and the analyte ends up with charges similar to the Rayleigh limit. Thus, the charge of a protein ion depends on the ionisation process. When contrasting these two mechanisms for ion creation, the IEM states that charged analyte leaves the droplet while the CRM proposed that the solvent leaves the analyte. CRM has been observed for both soluble and membrane proteins.

For unfolded and denatured proteins, a third mechanism is proposed, the chain ejection model (CEM)[277,280]. In this model, an unfolded chain is partially ejected from the droplet. During this process, the charges are equilibrated until the chain leaves the droplet through CRM. This mechanism typically produces a larger distribution of charges than folded proteins emitted through CRM.

The next step in the development of electrospray ionisation was the development of nanoelectrospray ionisation (nESI) in 1993[281–283]. This technique is a miniature version of the original electrospray technique and was developed because only 10% of species could be resolved in previous ESI. The key change is reducing capillary orifice diameter from about half a millimetre to the low micrometer range. This change offered multiple advantages over standard ESI. While ESI often required organic solvents to be added, nESI allow for the use of aqueous buffers which are more likely to preserve intermolecular interactions in proteins. In addition, lower flow rates due to smaller orifice reduced the amount of sample needed from millilitres to microlitres.

The smaller orifice also leads to the formation of smaller droplets. As a consequence less solvent needs to evaporate from analyte ion. This allows to lower the energies

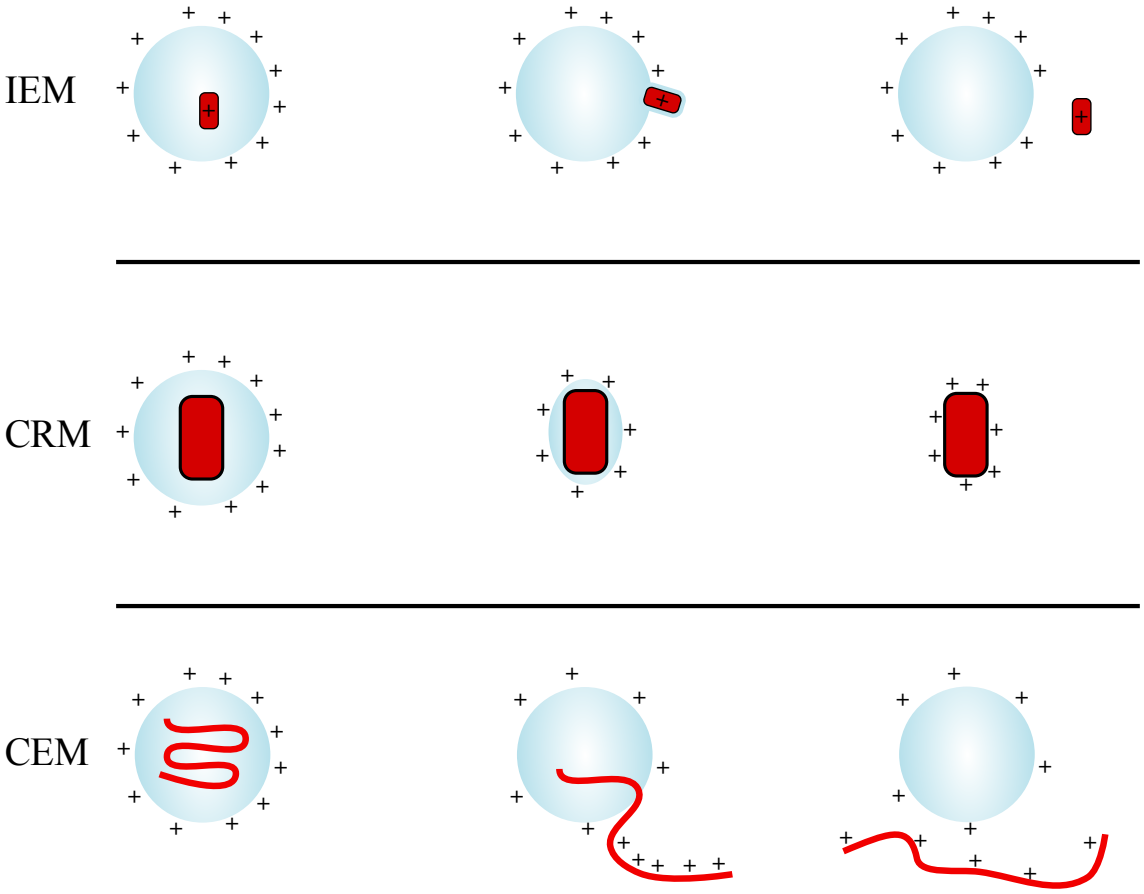

Figure 2.5. Mechanisms of ESI.

***IEM:**: A charged particle leaves the droplet and enters the mass spectrometer. **CRM:** Solvent evaporates from the droplet until the analyte (e.g. a protein) remains desolvated as an ion. **CEM:** A protein slowly unfolds and leaves the droplet. An equilibrium of charge on the chain occurs until it is ejected as an unfolded protein ion.*

applied to to achieve sufficient desolvation. These gentler conditions inn combination with aqueous buffers enable the preservation of non-covalent interactions[153]. Furthermore, the number of salt adducts and the number of oligomerisation artefacts are reduced[284,285]. The former occurs because fewer fission events lead to less increase of the salt concentration while the latter is a consequence of fewer proteins in one droplets due to a smaller size of the initial droplet. These advantages make nESI more viable for studying protein interactions.

2.2.2 Mass Analysers

A mass analysers is the core of a mass spectrometry experiments. It is the component that separates ions according to their ratio of mass and charge. There are different types of mass analysers which are typically used at different stages for nMS experiments.

2.2.2.1 Quadrupole Mass Analyser

Quadrupole mass analysers consist of four parallel rods of round or hyperbolic shape[286]. These rods are arranged in a quadratic geometry and connected by a fixed DC voltage U as well as an alternating AC voltage V(figure 2.6A) at the opposite rods. Therefore the potentials are given by

$$\Phi_{\pm} = \pm\left(U + V\cos(2\pi f t)\right) \tag{2.13}$$

Here, f is the oscillation frequency and t is the time. The combination of the potential from the four rods leads to a quadrupole electric field, with the motion of ions is described by the Mathieu equation[287]. This equation allows for stable trajectories for a certain m/z range depending on the choice of V, U and f. The fact that there is a minimal m/z for passing through the quadrupole can be understood through the notion that are too heavy ions are not affected sufficiently by the AC potential V and thus divert from a stable trajectory by the DC potential U. For ios that are too light the inverse is the case. They follow the oscillations of AC potential V until they are come into contact with a rod and discharge. This property makes quadrupole mass analysers suitable to serve as a mass filter in native mass

spectrometry[288]. Light ions with a high transmission rates, such as salts, can be filtered out as well as ions large unspecific clusters of molecules. Thus, proteins and protein assemblies in the expected mass to charge to charge ratio dcan be studied through native mass spectrometry with less noise. This property makes them also useful tools for the dissociation experiments described in section 2.4 because it is possible to use them to select a molecular species of interest[156]. When applying only radio frequency, a quadrupole can serve as an ion guide[289]. Quadrupole mass analysers are not used as mass analysers in nMS experiments because their resolution is inferior to other analysers described below at the large m/z values required for nMS experiments.

2.2.2.2 Time-of-Flight Mass Analyser

The second type of mass analysers are time-of-flight (TOF) mass analysers[290]. TOF mass analysers distinguish ions by measuring the name giving time of flight t thorugh a drift region of a known length L. The m/z can now be calculated by

$$m/z = \frac{2eVt^2}{L^2}, \tag{2.14}$$

where e is the electron charge and V the potential. This type of mass analyser is shown in figure 2.6B. In usual technical realisations the ions are accelerated with a pusher to enter a enter a drift region where they approach a reflectron. The reflectron consists of a of number lenses with a fixed electrostatic potential. This potential has the same sign as the ions leading to a deceleration when the ions approach through repulsion. Once the ions stop, the repulsion accelerates them towards the

detector. For measuring t correctly, knowledge of the starting time is essential. This makes TOF mass analyser naturally incompatible with nESI which produces a continuous stream of ions. To overcome this problem, the ion stream enters the pusher perpendicularly to the acceleration. Thus, the difference in initial velocities is minimised. However, in practise it is still present leading to a distribution in measured t and therefore m/z.

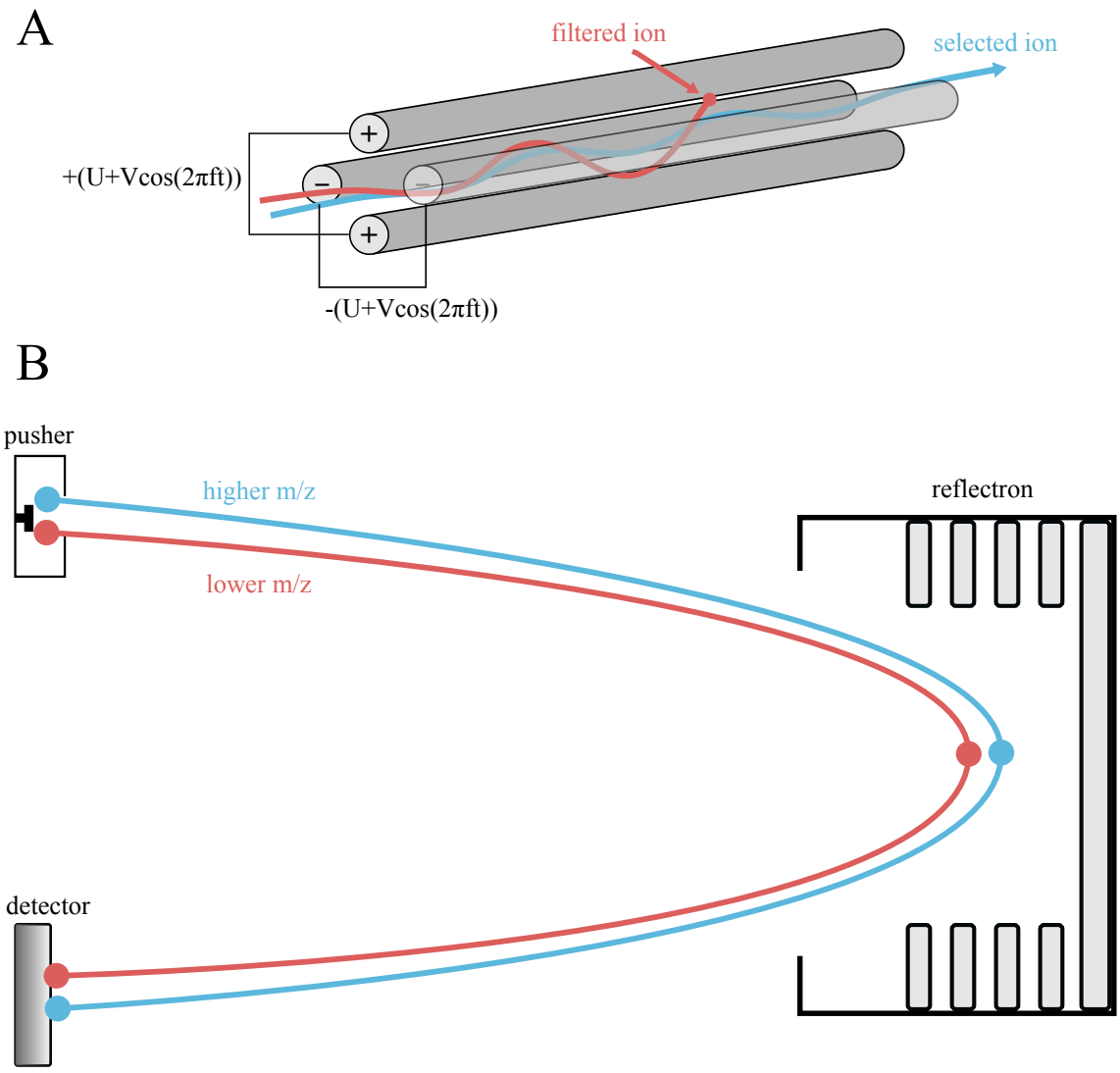

Figure 2.6. Quadrupole and TOF mass analysers.

A *Quadrupole mass analyser. Ions enter pass through four parallel rods and are selected based on their m/z* ***B*** *TOF mass analyser. Ions are accelerated in the pusher towards the reflectron that decelerates them before accelerating them towards the detector.*

2.2.2.3 Orbitrap Mass Analyser

The third type of mass analysers is the Orbitrap mass analyser. These analysers have been first realised by Makarov *et al.* in 2000 despite the idea going back to the Kingdon trap in 1923[291,292]. The first commercially available Orbitrap instrument was developed by ThermoFisher Scientific in 2005[293]. The Orbitrap consists of an inner electrode shaped as a spindle with an outer electrode covering the inner electrode. This geometry induces an electrostatic potential

$$U(r,z) = \frac{kz^2}{2} - \frac{kr^2}{4} + \frac{k}{2}\left(R_m\right)^2 \ln\left(\frac{r}{R_m}\right) + C, \tag{2.15}$$

where k is the field curvature, R_m the characteristic radius. r and z are defined in figure 2.7. This means, that movement along the z-axis and movement around the z-axis are independent of one another. Oscillations in the z-direction around the spindle are a stable trajectory in such a device. The frequency of such an oscillation is linked to the m/z of an ion[291].

$$\omega = \sqrt{\frac{k}{(m/z)}}. \tag{2.16}$$

Thus, this frequency is used to determine the m/z value of ions. In contrast to TOF and quadrupole mass analysers, an Orbitrap mass analyser also functions as well as as a mass detectors. Orbitrap instruments have recently gained popularity for native mass spectrometry due to their ability to maintain high resolution across a wide m/z range[294]. Despite being only recently introduced, Orbitraps have al-

ready had profound impact on the field[294–296]. For membrane proteins this type of mass analyser has enabled the characterisation of lipid binding up to the acyl chain length[166]. To make Orbitrap-based spectrometers compatible with native mass spectrometry, the instruments have been modified specifically for the study of large molecules[297]. A drawback of Orbitraps are the requirement for longer acquisition times for mass spectra. This property made them not readily compatible with ion mobility experiments, although there are attempts to combine these techniques in recent years[298–300].

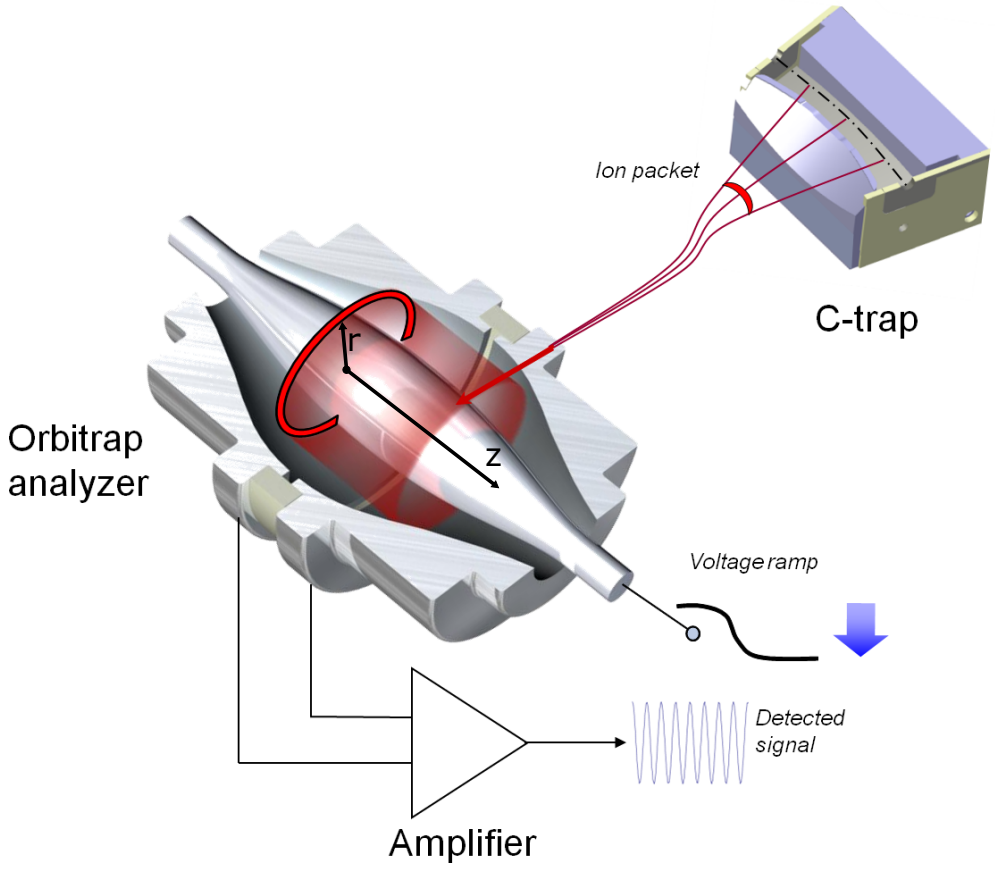

Figure 2.7. Sketch of an Orbitrap mass analyser.
Ions enter the Orbitrap perpendicular to the z-axis and perform oscillations in z-direction around the spindle. These oscillations are recorded and amplified for mass measurement. Image created by Thermo Fisher Scientific.

2.3 Ion Mobility Mass Spectrometry

Ion mobility mass spectrometry (IM-MS) is an extension of nMS which allows investigating the size and geometry of a protein assembly[301]. Proteins are separated when moving through a cell of inert gas in a weak electric field. In this cell, larger proteins will collide with more gas molecules and are hence slowed down more than smaller proteins given a similar mass to charge ratio. Therefore, protein with the same m/z can be distinguished if e.g. one protein is unfolded and one is folded. The mobility of proteins can be used to calculate a collisional cross section (CCS) which can be compared to theoretical calculations based on protein structures to estimate the conformation that the protein adopts in the gas phase[201]. This makes ion mobility mass spectrometry a useful tool in structural biology.

There are two main types of ion mobility cells used in native mass spectrometry: drift tubes and travelling wave ion mobility. The drift tube is the conceptionally simpler implementation of the idea. In a drift tube with a constant electric field, the protein collides with gas molecules. The velocity of the protein is constant and determined by its mobility and the electric field E. This can be used to calculate the CCS in a drift tube (CCS_{DT}) through the Mason-Schrampp equation[302]

$$CCS_{DT} = \frac{3ze}{16NK_0}\sqrt{\frac{2\pi}{\mu k_b T}}. \tag{2.17}$$

Here, e is the elementary charge, z the charge state of the protein, N the number density of the gas in the drift tube, k_B the Boltzmann constant, T the temperature, μ the reduced mass of the gas molecule and K_0 is the reduced mobility. The electric

mobility can be calculated from the drift velocity and the electric field. The reduced mobility is the mobility where the pressure and the temperature are each divided by 760 Torr and 273.2 K, respectively.

An alternative to the drift tube is the travelling wave ion mobility mass spectrometry[303]. This type of ion mobility is based on a stacked ring ion guide linked to a RF voltage. A potential is sequentially applied to the pair of rings leading to a moving potential "wave" which drags the proteins along. The advantages of travelling wave ion mobility over the drift tube pendant are an improved sensitivity as well as resolving power and higher transmission rate[303–305]. However, the drawback of the travelling wave is that the observed arrival times can not be directly compared to CCS but need to be checked against a calibration curve of CCS provided by drift tube ion mobility[306].

A key application of IM-MS for structural biology is the comparison of calculated CCS measured to expected CCS from protein structures[201]. This can done by using the projection approximation which averages the projection of a protein structure over all directions[307]. In addition, this methods corrects for a mass difference of e.g. cleaved complexes and is multiplied by an empirically derived factor of 1.14. Thus, the CCS is given by[308]

$$CCS_{calc} = 1.14 \times CCS_{PA} \times \sqrt[3]{\left(\frac{M_{exp}}{M_{struc}}\right)^2}. \tag{2.18}$$

Where CCS_{PA} is the calculated CCS by the projection average, M_{exp} the mass of

the experimental protein and M_{struc} the mass of the protein structure. This formula allows comparing experimental and structural CCS values.

2.4 Dissociation Techniques

One of the main advantages of native mass spectrometry is its ability to access stoichiometry and ligand binding of protein complexes. In practise, this is achieved by disrupting the non-covalent interactions in the gas phase. The most common method to accomplish this is by repeated collisions of the protein complex with inert gas molecules for collision induced unfolding (CIU) and collision induced dissociation (CID). Less common techniques include the collision of the protein complex with a surface known as surface induced dissociation (SID). These techniques are explained below.

2.4.1 Collision-induced Unfolding & Dissociation

The most common way of activating protein ions in native mass spectrometry is to collide protein ions into neutral gas molecules[309]. For this purpose, the ions enter a specific cell filled with an inert gas such as argon or nitrogen, accelerated by an applied electrical potential. At low voltages these collisions "clean" the protein ion of adducts such as membrane mimetics in the case of membrane proteins[165]. While in IM-MS the collisions are tuned to occur at low energy in order to preserve interactions, the opposite is the case for CIU and CID. Here, the voltage of of the electrodes is increased until the structure of the protein becomes unstable. There-

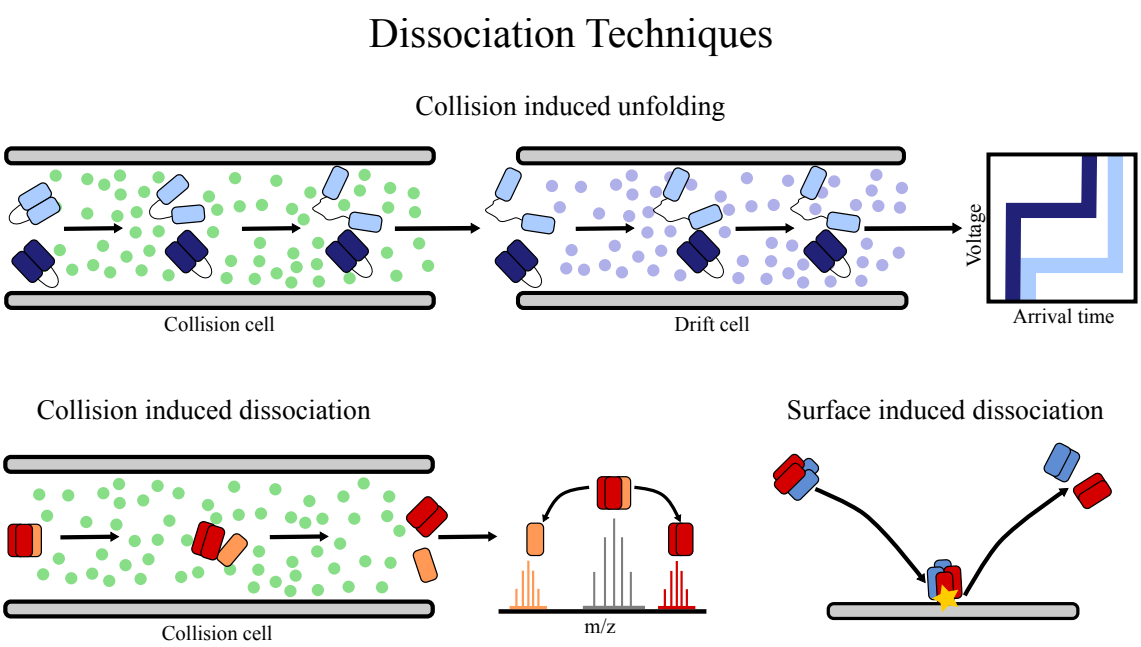

Figure 2.8. Different activation techniques applied in nMS

Collision induced unfolding (top) *Ions collide with neutral gas ions in a collision cell leading to an unfolding of the protein. The CCS is then measured through ion mobility mass spectrometry. Unfolding occurs abruptly at a certain activation voltage. This results in a plot as indicated on the right by a sharp increase in arrival time. The more stable complex (dark blue) does so at higher activation voltage then the less stable protein (light blue).* ***Collision induced dissociation (bottom, left)*** *Protein ions pass through the collision cell. The voltage is ramped up until dissociation occurs. When it does, a subunit is ejected with high charge (orange) showing in the low m/z region while the remaining complex loses charge and moves into the higher m/z region.* ***Surface induced dissociation(bottom right)*** *A dimer of dimers is collided with a surface leading to a dissociation into dimers.*

fore, the voltage can be used as a measurement of the complex stability.

In CIU, the collision cell is linked to an IM-MS cell (figure 2.8, top). This means, that after colliding the protein ions with gas molecules they are sorted by size. The activation voltage is increased in a controlled manner. At a certain voltage, the protein starts to unfold[301]. This is a self-enforcing process as it involves the transfer of protons towards the surface area made available by the unfolding process. Thus, the protein unfolding is accelerated further[310–313]. As a consequence the CCS (and arrival time) increases abruptly once unfolding occurs (see figure 2.8 top right).

Therefore, the collision voltage at which unfolding occurs can be employed as a gauge for protein stability. While theoretical models exist to link this value to a physical quantity, there is no consensus yet and research is ongoing[165]. However, it is possible to use this voltage for relative comparisons. This means it is possible to investigate the amount of stabilisation or destabilisation of ligand binding or mutagenesis for a protein relative to its apo state[165].

Collision induced dissociation uses a similar principle to CIU but goes a step further. The applied voltage is increased until a dissociation of a non-covalently bound subunit occurs. This is monitored by collecting native mass spectra (figure 2.8, bottom left). Interestingly, this process mostly leads to both dissociation fragments having the same charge because this minimises the energy due to maximum charge repulsion[310]. However, the process of CID is usually independent of the stoichiometry of the protein complex[309]. This means, that regardless of how many units an oligomer has, an unfolded monomer is ejected in most cases[301]. Therefore, the amount of structural information provided by CID is limited. Multiple sequential ejections of subunits can be observed at higher activation voltages[301].

2.4.2 Surface-induced Dissociation

An alternative to CID is SID. In this method the high number of collision with gas molecules is replaced with one high energy collision with a plane solid surface. SID was developed in 1975 and initially applied to small organic molecules[314]. In the

following decades, the technique has been refined to become a more common (yet still niche) operation[315,316]. SID was applied to native mass spectrometry only in the late 2000s[317–319]. The idea behind this dissociation technique is that a single fast high energy collision favours the pathway with the breaking of the oligomeric interface instead of subunit unfolding, unlike its CID counterpart (see figure 2.8). Therefore, it is more likely to probe the interface stability of an oligomer[320]. SID is also able to provide more structural information than CID. This was demonstrated by the work on tetrameric proteins[321]. Proteins consisting of two dimers (D2 symmetry) were dissociated into dimers by breaking the weakest subunit interface, while a tetrameric protein of four equivalent monomers (C4 symmetry) has been dissociated into monomers, dimers and trimers[321,322]. SID is therefore an extremely useful method to inform protein models where complex structures at high resolution cannot be achieved[323]. Here, it can inform the stoichiometry and connectivity of subunits[324].

Probing the effect of ligand binding on interface stability is another possible use case of SID, in contrast to CID, which probes the individual subunit stabilities with respect to unfolding. This is what SID will be used for in this thesis in chapter 3 where the effect of CDL taillength on the interface of a dimeric sugar transporter will be investigated.

2.5 Instrumentation

In this section the instrument used in this thesis shall be introduced. Both, ToF and Orbitrap instruments have been used in this work. Most of the spectra have been collected on an Orbitrap instruments, a Q Exactive UHMR Hybrid Quadrupole-OrbitrapMass Spectrometer (figure 2.9 (top)). These experiments were augmented by proteomics and lipidomics experiments have been performed on a LTQ Orbitrap XL hybrid Ion Trap-Orbitrap Mass Spectrometer (LTQ Orbitrap XL). Since these orbitrap instruments are not equipped with ion mobility cells, the dissociation and ion mobility experiments in chapter 3 have been carried out on a modified Synapt G2 with an ion mobility cell (figure 2.9 (bottom)).

2.5.1 Themo Fisher Scientific Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer

Most of the spectra used in this thesis have been collected by a Q Exactive instruments. For these instruments, the capillary containing the analyte ions is aligned with the spectrometer entrance. Once inside, the ions are focussed through a stacked-ring ion guide (S-lens). Afterward,s they enter a bent flatapole which consists of a square array of flat metal rods. These rods form a 90-degree ion guide and filter out neutral particles. Next, they enter a quadrupole where a particular m/z range can be selected. The selected analyte ions enter a Higher Energy Collisional Dissociation cell (HCD cell) where they can be activated by collision with neutral gas molecules. Finally, ions are collected in a curved linear trap (C-trap) before they enter the Orbitrap mass analyser.

2.5.2 Themo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer

The LTQ Orbitrap XL instrument architecture resembles the Q Exactive architecture. However, the bent flatapole is replaced by a linear quadrupole and an octapole to facilitate focusing. In addition, the ions trapped in the linear ion trap can be activated by a subsequent HCD or CID cell.

2.5.3 Waters Synapt G2

The modified Synapt G2 is a Q-ToF (quadrupole time-of-flight) mass spectrometer. Ions enter the instrument from a z-spray (90-degree angle) electrospray ion source and are focussed through a set of ion lenses. The z-spray allows filtering out neutral species. The pressure at the frontend is kept at $\approx 5\,\mathrm{mbar}$ for collisional focussing in soft ionisation conditions. The next stage is a quadrupole to filter ions for a set m/z range.. Next, the ions enter a stacked ring ion guide collision cell (Trap), which is filled with a neutral gas (commonly argon or nitrogen) for activation through CID and CIU. In order to enable SIDIM experiments, a truncated trap is used to make space for an SID-cell. Afterwards, the ions enter an IM cell or are directly analysed by a ToF analyser before detection.

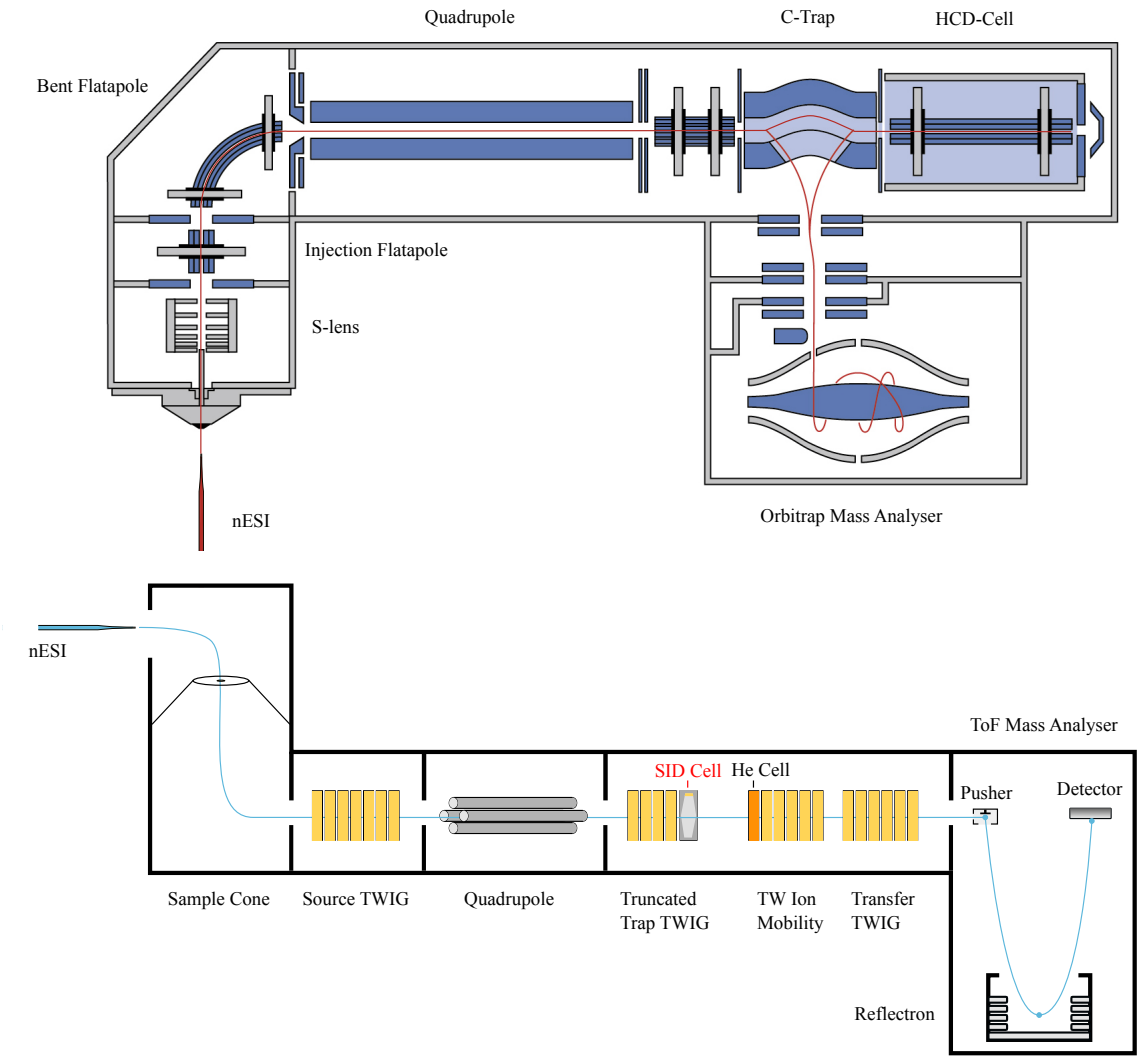

Figure 2.9. The instruments used in this thesis

Illustration of an Thermo Fisher Scientific Q Exactive UHMR Hybrid Quadrupol-Orbitrap mass spectrometer (top) and a Waters Synapt G2 (bottom) Image source (top) Thermo Fisher Scientific, modified.

Chapter 3

Chain Length Dependence of Lipid Interactions and Key Residues for the Stability of a Bacterial Sugar Transporter

Abstract

Membrane proteins are key drug targets but difficult to characterise by biophysical techniques[16]. This also applies to protein-lipid interactions that modulate membrane protein complexes in a plethora of ways (see section 1.2). Native mass spectrometry combined with molecular dynamics simulations established itself as a powerful approach to study these interactions and their effect on the stability of the complex[52,325]. While traditional dissociation techniques such as CIU probe the stability against unfolding, SID is employed here to study the influence of lipid binding on the interface stability of a dimeric sugar transporter. This was a highly collaborative project where Denis Shutin performed nMS experiments of lipid binding. I contributed CGMD simulations to rationalise the experimental observations for the

first part of this chapter. In the second part site-directed mutagenesis experiments conducted by me to selectively weaken the interface are presented.

3.1 Introduction

3.1.1 Protein Oligomerisation

Oligomerisation is a widespread phenomenon among proteins. Around a third of proteins exhibit oligomerisation[326] and the average oligomerisation state for a protein has been estimated to be a tetramer[327]. This is unsurprising given that the cell is a crowded environment. While most of these interactions are random and short-lived, many others are specific and long-lived interactions and vital for proteins to fulfil their biological functions[328].

Membrane proteins are often found as oligomers. Examples for this include ion channels and transporters such as SemiSWEET (studied in this chapter)[30,329]. For these proteins, an assembly of subunits is necessary to form a functional complex. Other examples of oligomerisation are less clear, such as GPCRs where oligomerisation and its functional relevance remains debated (although they are also seen as a possible therapeutic route)[330,331]. Studying the oligomerisation of membrane proteins is challenging for two biochemical reasons. They are expressed with lower yields compared to their soluble counterparts and are difficult to solubilise due to their amphiphilic nature, requiring a membrane mimetic[332]. Also, it has been shown that different lipids modulate the oligomerisation process, further increasing complexity

(section 3.1.3). This makes the study of these phenomena challenging but important to inform the drug discovery process given the crucial role membrane proteins play in it[16].

3.1.2 SemiSWEET: a Bacterial Sugar Transporter

Sugars play an essential role in biology. They are not only a source and storage of energy for many organisms but also a key structural component of nucleic acids or glycolipids[333]. Therefore, the transport of sugars across the biological membranes is crucial for the survival of cells. One family of sugar transporters is the Sugars Will Eventually be Exported Transporters (SWEET) family discovered in 2010[334]. These proteins are common throughout most domains of life and transport mono and disaccharides across the membrane[335–337]. SWEET transporters possess an architecture of 3+1+3 transmembrane helices consisting of two repeats of a 3-TM helices motif connected by a linker helix. The prokaryotic homologues of SWEET transporters share the 3-TM helices motive. However, they form a dimer of two monomers consisting of three transmembrane helices each. As they form half a SWEET transporter, they are called SemiSWEET[338,339]. These were identified in 2013 and thought to form the evolutionary ancestor to 7-TM SWEET transporters[329]. Crystal structures of SemiSWEET appeared in 2014 and 2015[338,339] in multiple conformations. These conformations included inward- and outward-open conformations as well as occluded states (figure 3.1.)

SemiSWEET contains a PQ-loop (proline-glutamine) which is conserved for many types of transporters[340]. This loop is essential for the conformational change upon

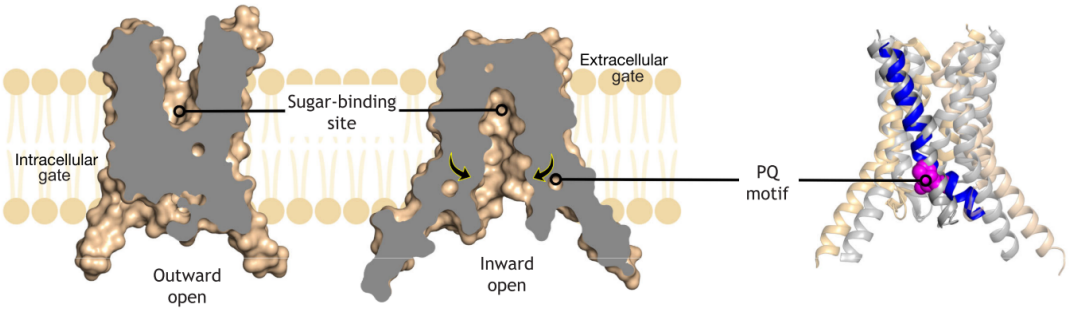

Figure 3.1. Inward- and outward-open structures of SemiSWEET
Adapted from[340].

sugar binding and hence for the transport[338,339]. Proline is known as a helix breaker and enables the flexibility for the structural rearrangement of the interrupted helices[340,341]. In the PQ-family, it allows for a bending motion of the first transmembrane helix, which results in closure of the intracellular gate while opening to the extracellular side[340].

SemiSWEET is among the shortest membrane proteins, with a monomer mass of around 11 kDa for 97 amino acids. This made it an ideal system to probe the modulation of dimerisation through lipids described in the next section that is the foundation of the work in this chapter. The reasons are twofold: as a light membrane protein the peaks are expected in the lower m/z regions leading to a better resolution and a homodimer is more suitable to be studied as there is only one symmetric dimerisation interface, which avoids further complications due to multiple potentially different interfaces[342].

3.1.3 Lipid-mediated Oligomerisation of Membrane Proteins

The interactions of membrane proteins and their surrounding lipids are diverse and often functionally relevant (section 1.2). These specific interactions have been in-

dicated to facilitate oligomerisation processes of membrane proteins by various biophysical techniques, including nMS[166–168]. In their work, Gupta *et al.* combined of nMS and the analysis of membrane protein structures to demonstrate that many membrane proteins need lipids to oligomerise. Such proteins were shown to have relatively weak subunit interfaces, as indicated by low buried surface area of the interface and the low number of salt bridges (figure 3.2A). Although some discussions about the predictive power of this plot remain and suggestions to improve the plot have been made[342]. SemiSWEET has been found to have a subunit interface of an intermediate strength, where lipids facilitating oligomerisation but not being necessary for it (figure 3.2B)[168]. This was demonstrated by adding CDL to apo SemiSWEET and observing the shift in the monomer-dimer equilibrium towards the larger oligomeric species (figure 3.2C).

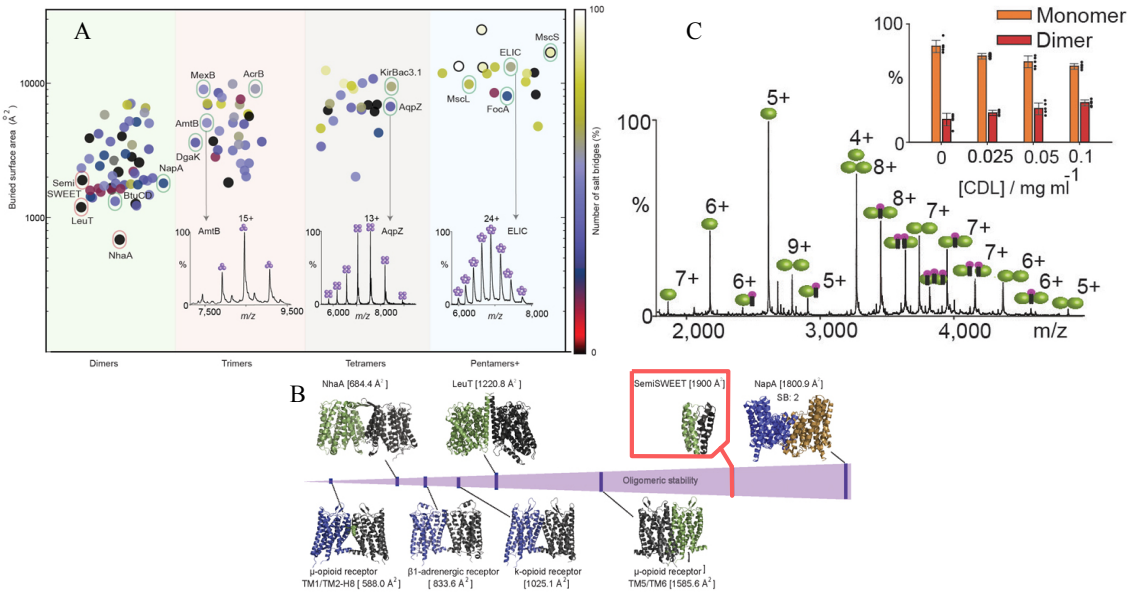

Figure 3.2. Effect of CDL binding to to SemiSWEET.

A *Burried interface area and the number of salt bridges for different membrane proteins.* ***B*** *SemiSWEET interface stability in comparison to other membrane proteins. SemiSWEET is highlighted in red.* ***C*** *CDL binding to SemiSWEET affects the monomer-dimer equilibrium. Adapted from Gupta et al. with permission from Springer Nature*[168].

After this initial finding it was shown that different CDL species present in cell

membranes bind to SemiSWEET[166]. Notably, SemiSWEET seemed to exhibit a preference for CDL molecules with longer hydrophobic chain tails compared to the distribution of lipids in the native membrane (figure 3.3A). This fact was also investigated in a recent MD study[224] where Deng *et al.* established four binding sites for CDL and its crucial role in coordinating dimer formation(figure 3.3B).

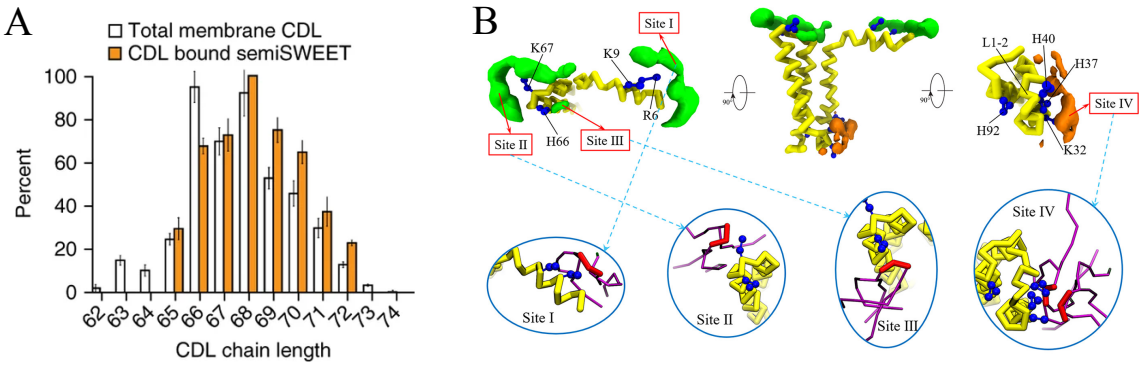

Figure 3.3. Effect of CDL binding to SemiSWEET.

A *Distribution of CDL in the membrane and bound to SemiSWEET. Adapted from Gault et al. with permission from Springer Nature*[340]. ***B*** *Four binding sites of CDL to SemiSWEET determined from CGMD. Adapted from Deng et al.*[224]

Despite the recent advances, several aspects of the SemiSWEET-CDL interactions remain unknown, including the reason for the shift between the natural abundance of shorter chainlength CDL in the membrane compared to the CDL bound to SemiSWEET. In addition, the exact nature of stabilisation CDL provides to SemiSWEET dimers by different CDL species still needs to be investigated. These questions can be addressed by using nMS due to its ability to study individual protein-lipid interactions. However, the most common activation method combined with nMS, CIU, generally probes stability against unfolding 2.4) and is therefore unable to measure the influence of lipids on interface stability. This fact limited previous studies to use indirect methods, such as observing the ratio of different oligomeric states of SemiSWEET at different lipid concentrations[168,176,343]. However, the strength of the subunit interface can be investigated by nMS more directly

by using an alternative dissociation technique, SID, which has recently been shown to be able to probe interface stability for soluble proteins and identify stabilising ligands[344].

3.1.4 Chapter Outline

The work done in this chapter consists of two interlinked projects. In the first part. SID experiments and MD simulations of SemiSWEET have been carried out to understand the influence of chain-length and saturation of different CDL species on the stability of SemiSWEET dimers. The aim of this part is to establish that SID is able to investigate these types of problems for membrane proteins and to find out what causes the different stabilisation strength of different species of CDL. For this part, I conducted MD simulations to complement work by Denis Shutin in the Robinson group who carried out SID experiments of SemiSWEET including lipid binding to probe the interface stability of the protein and its modulation through lipids. For this part, I my work focussed exclusively on simulations in close collaboration with the experiments.

In the second part, this work is extended to probe the suitability for site-directed mutagenesis of the dimer interface of the same protein. Therefore, conserved residues of SemiSWEET have been identified and mutated to selectively weaken the interface of SemiSWEET. Here, my work comprised of the design of the mutations, protein expression and purification (together with Denis Shutin) as well as CID experiments. In addition, Denis Shutin performed SID experiments on the mutants.

3.2 Dimer Interface Stability of SemiSWEET for Different Lipids Bound

3.2.1 Experimental Evidence from SID

The project started with dissociation work on SemiSWEET. To this is end, spectra of SemiSWEET have been collected. First, a particular charge state for dissociation must be determined as different charge states of a protein can exist in various initial conformations that can affect the experimental outcome[164]. This charge state should ideally include the protein in its folded state with high intensity. In addition, interpretation of experimental results can be hindered if several oligomeric species overlap in m/z space. This problem can be circumvented by avoiding peaks where the charge is divisible by the number of subunits in an oligomer. For dimeric SemiSWEET this means that even charge states should be avoided (for example, an 8+ dimer has the same m/z value as a 4+ monomer). Here, we proceed with the 7+ charge state for all CIU and SID experiments to enable comparison between different experiments.

The CIU experiments showed a difference is stability for CDL and other bacterial phospholipids with respect to unfolding (figure 3.4). However, no change in stabilisation between different CDL tails was observed. This could either indicate the absence of the preference of SemiSWEET monomers for particular CDL species or the insufficient resolution of CIU technique to detect any small differences in unfolding stability that could exist.

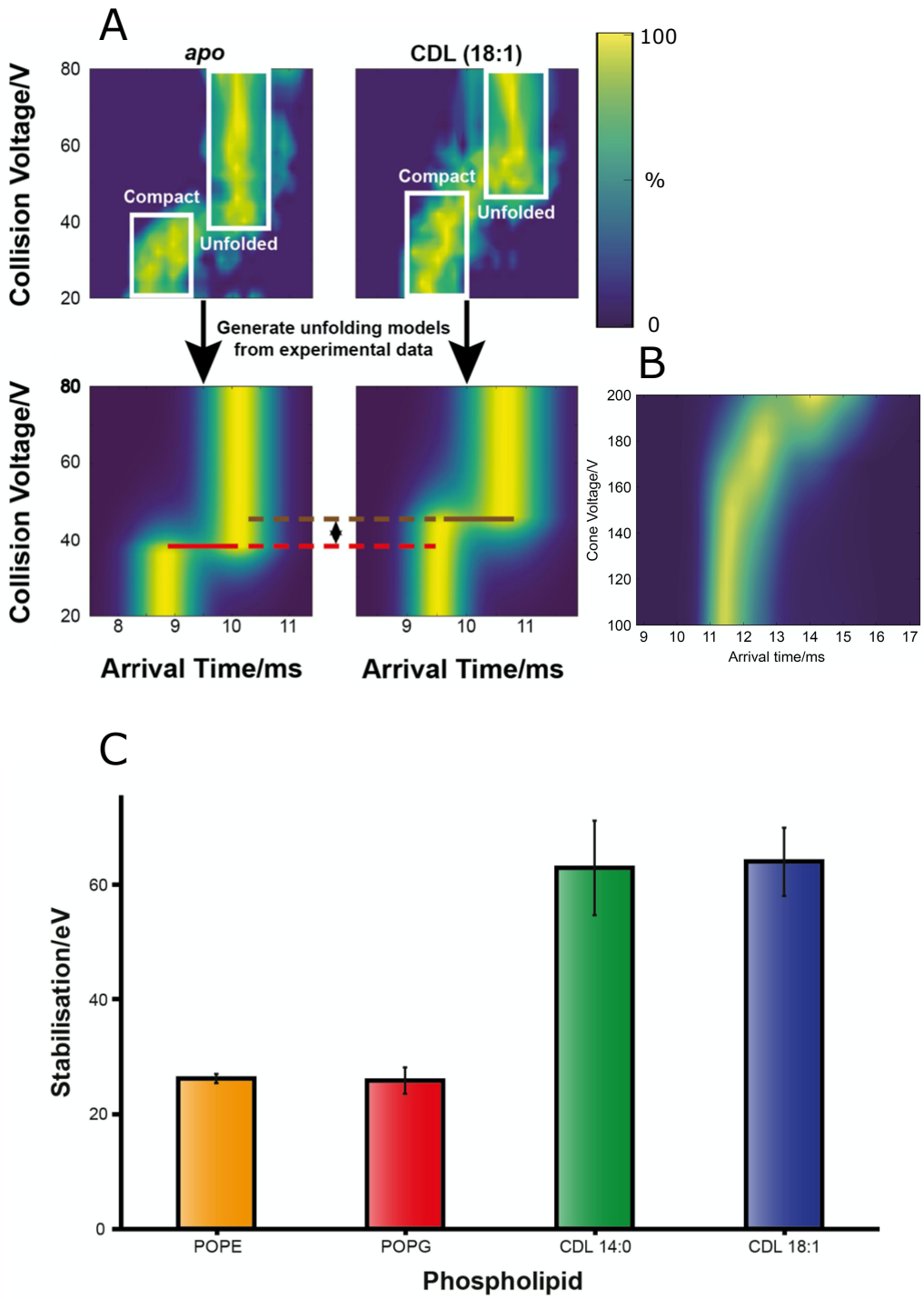

Figure 3.4. CIU of WT SemiSWEET for different lipids bound.

A *CIU trap ramp data for SemiSWEET collected from the 7+ state and a schematic model fitted. The difference between brown and red line yield the stabilisation provided by lipid-binding.* ***B*** *CIU cone ramp of the 7+ state.* ***C*** *Stabilisation by different phospholipids. This stabilisation is calculated by multiplying the difference in voltage of the red and brown line in panel A with the charge selected (7+ in this case). POPG and POPE could be distinguished from CDL binding. But no difference between them or between different CDLs were significant. Figure & data provided by Denis Shutin.*

As was mentioned above, SID can be used to obtain information about the dimer interface strength, rather than the monomer unfolding probed by CIU. Data for SID is is typically presented as an energy resolved mass spectrometry (ERMS) plot. In this plot the x-axis describes the stepwise increase in collision energy used to activate the protein and the y-axis shows the relative percentage of the undissociated precursor. For SemiSWEET the voltage has been varied in 2 V steps and three repeats have been collected (figure 3.5 top). This has been fitted with a sigmoidal curve for the apo and different CDL. The CDL with 18:1 tails is visibly more stable than the other species, especially in the 40-60 V region (figure 3.5 bottom and inlet).

To analyse this data in a more rigorous manner, an Analysis of Variances (ANOVA) test has been carried out for different voltages[345,346]. This type of test determines whether the difference between multiple data sets is statistically significant. Hence, it is a generalisation of a t-test for multiple sets of data[347].

The results are given in figure 3.6. A one-way ANOVA test shows the statistically significant results in figure 3.6 top. Here, a stabilising effect was noted for CDL in general and also between some pairs of different CDL species, but the differences were inconsistent between voltages. To get a general picture, a two-way mixed ANOVA test was performed on the range of 43-55V, for which the stabilisation differences were the most pronounced (figure 3.5 bottom) and the range from 15-63 V a larger range from 15-63 V (figure 3.6, bottom left and bottom right respectively). Here a statistically significant difference was noted for apo vs all CDL and between 14 C-atoms chain length and 18:1 tail.

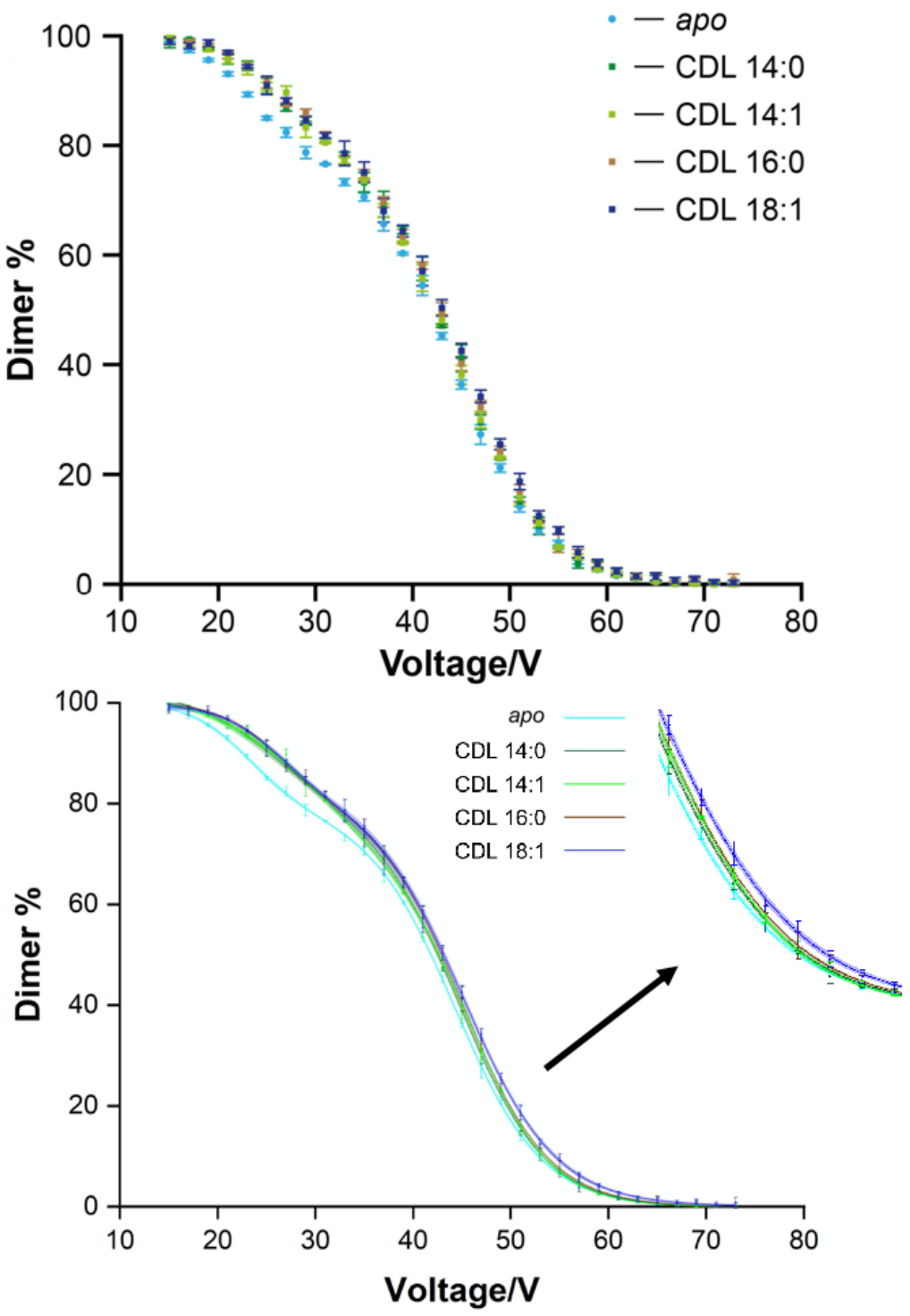

Figure 3.5. ERMS plot of SID data for different CDL binding to SemiSWEET. ***Top*** *SID data for apo and different CDL bound. Error bars indicate standard deviation for three technical repeats.* ***Bottom*** *SID data fitted to a sigmoidal. Figure, data and data analysis provided by Denis Shutin.*

Therefore, we can conclude that different species of CDL have a statistically significant difference in stabilising the interface. However, the exact nature of the observed

Voltage/V	43	45	47	49	51	53	55
	16:0>*apo*	14:0>*apo*	16:0>*apo*	14:1>*apo*	18:1>*apo*	18:1>*apo*	18:1>14:0
	18:1>*apo*	18:1>*apo*	18:1>*apo*	16:0>*apo*	18:1>14:0		18:1>14:1
		18:1>14:1	18:1>14:0	18:1>*apo*			18:1>16:0
			18:1>14:1	18:1>14:0			
				18:1>14:1			

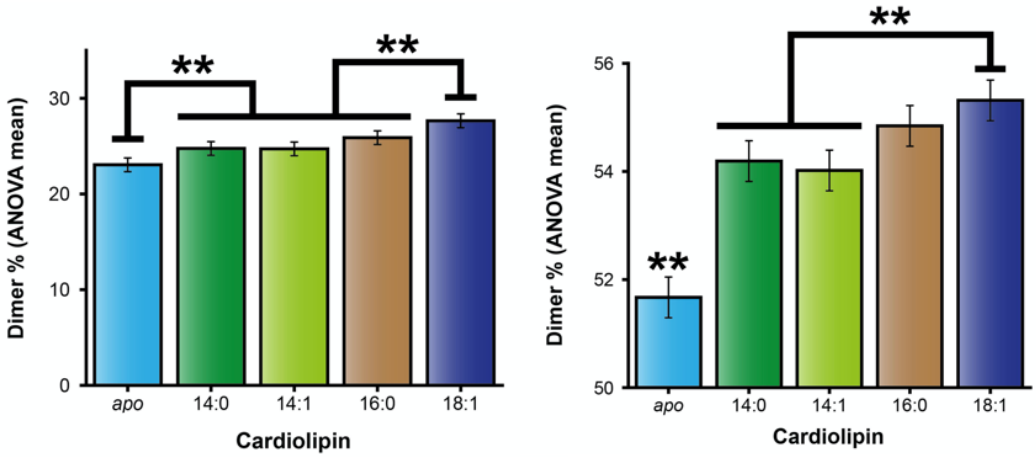

Figure 3.6. ANOVA of SID CDL binding results.

Top. *One-way ANOVA test results for different voltages.* ***Bottom*** *Two-way mixed ANOVA for the data if 43-55V (**left**) and for the range 15-63V (**right**). Double stars (**) imply statistically significant differences. The data is based on three technical repeats. Figure, data and data analysis provided by Denis Shutin.*

differences between the lipid species is unclear. Mass spectrometry alone is unable to answer this question. Therefore, I ran CGMD to determine molecular details of the interaction. The initial hypothesis was that there might be a noticeable difference in binding location for different CDL depending on chainlength and saturation with short unsaturated CDL binding more on the sides of SemiSWEET (figure 3.7 left with green CDL) while longer chainlength CDL tends to bind more on the dimer interface and therefore stabilises the interface more (figure 3.7 right with blue CDL). Initial work of SID only compared CDL species with certain combinations of tail lengths and the number of double bonds for practical reasons. Therefore, the additional aim of some of the simulations was to investigate other combinations in taillength and degree of unsaturation.

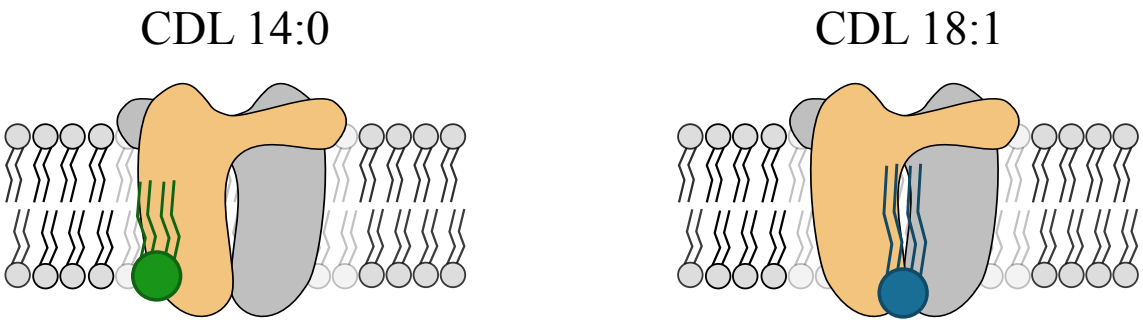

Figure 3.7. Schematic hypothesis for different CDL binding to SemiSWEET. *CDL binds more away from the dimer interface for shorter tails (left, green CDL) and more on the dimer interface for longer tails (right, blue CDL).*

3.2.2 Simulations of different CDL Species Binding to SemiSWEET

The first step for setting up CGMD simulations was preparing suitable lipid parameters for the simulations. As a starting point CDL parameters from the Martini collection has been used[348]. For each of the six CDL models created (figure 3.8) the tails and their bonds have been modelled based on lipid tails of CDL 18:1 and saturated tails respectively. For each CDL, four identical acyl chains have been used, as this matches the lipids purchased from Avanti for the SID experiments.

3.2.2.1 Initial Simulations of CDL Binding to the Occluded State of SemiSWEET

I started this project simulating 14:0 and 18:1 CDL in a mixed bilayer binding to SemiSWEET in an occluded state (PDB: 4QND). This setup has been chosen because the structure matches the one used in previous work and these were the CDL species for which initial SID data was available[168]. Also, these CDL represent the extreme cases for the effect that is being investigated as they differ in both the saturation and the length of the tail. The simulations were run for 5 μs each with four repeats. For each repeat, MemProtMD pipeline was executed[271].

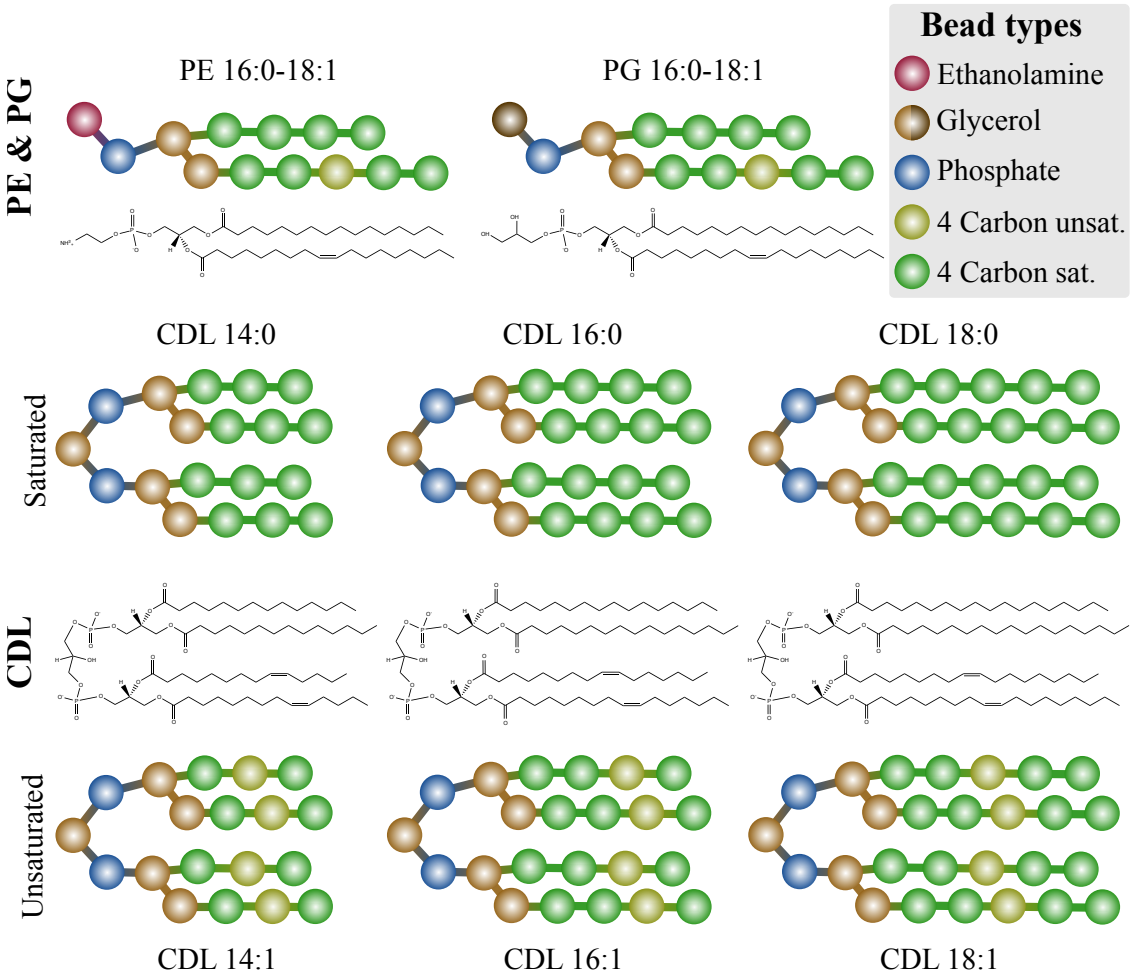

Figure 3.8. Martini Models used in simulations.

Standard Martini models have been used for PE and PG lipids. For CDL different lipids have been created. Starting from the 18:1 six different versions of CDL have been created modelling the chemical structures shown as hybrid of saturated and unsaturated tails in the middle between the saturated and unsaturated rows. The dark and light glycerol beads indicate a different parametrisation based on the fact, that one is a linker while the other is the end of a molecule in a headgroup.

3.2.2.2 Binding Preferences for Different Lipids for the Occluded State

The first step was to determine binding preferences for different lipids to SemiSWEET. This can help to compare the results to experiments. To this aim, the depletion-enrichment index (DEI) has been chosen. This index divides the concentration of lipids of each species found in a shell of 7 Å around the protein by the concentration of lipids found in bulk[49]. Thus, it normalises for lipid concentration and indicates how much a lipid binds to specific residues compared to expected random interac-

tions caused by different concentrations. A DEI below one indicates that the lipid tends not to bind to the protein while an DEI above one indicates a preference of this lipid to bind to a protein. The results are shown in figure 3.9.

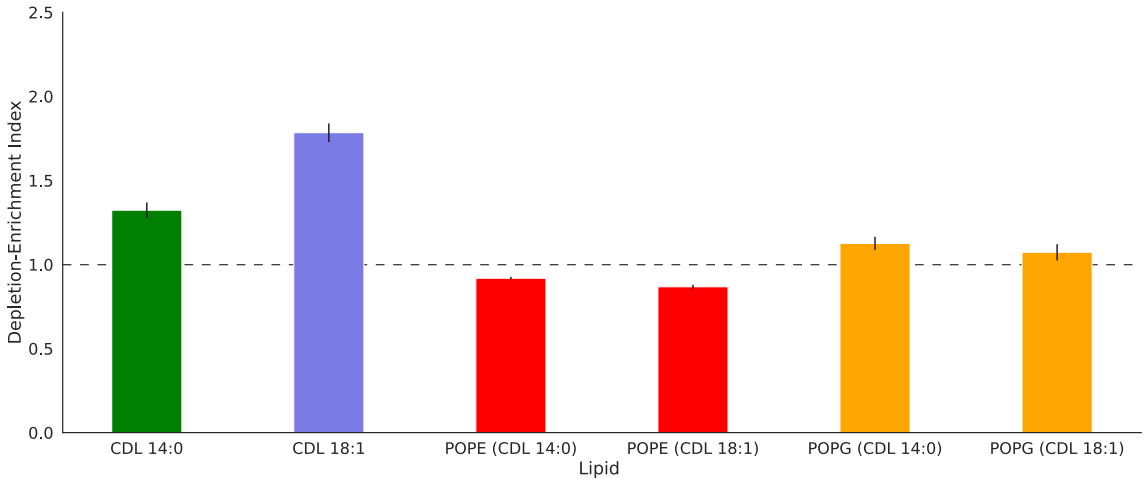

Figure 3.9. DEI for CDL, POPG and POPE in the occluded state structure simulations.

Labels of the lipids indicate the species investigated. For POPE and POPG the information in the brackets show which CDL was used in the simulations.

Here, a higher DEI for CDL is visible: with 1.32 ± 0.05 for CDL 14:0 and 1.78 ± 0.06 for CDL 18:1. Also, CDL 18:1 appears to be binding stronger to SemiSWEET than CDL 14:0. PE lipids exhibit a DEI below one with 0.87 ± 0.01 and 0.92 ± 0.01 in shorter and longer tail CDL simulations. PG shows a slight preference with 1.08 ± 0.05 and 1.13 ± 0.04. As a side note, these sets of simulation show that the DEI seems to be fairly constant for the same lipid types but it is still sensitive enough to distinguish different tails of lipids. Also, it is worth noting, that the values for PE and PG appear significantly different due to low error bars from different replica of the simulations. Therefore, the DEI should be seen as a relative number and it is hard to compare it over different sets of simulations with vastly lipid compositions. For this work it is appropriate, because I work with constant lipid concentrations which only differ in the length of the hydrophobic chains of CDL.

SemiSWEET appears to be binding more to negatively charged lipids such as CDL and PG over zwitterionic lipids like PE. As a next step the location of CDL binding is investigated to explain the difference in binding strength.

To find binding locations, contacts have been determined by a headgroup distance of less than 6 A as is standard for CGMD[274]. The results are shown in figure 3.10. While the contacts vary to a degree in strength, the same types of residues account for the contacts. Therefore, it is impossible to decide from this whether there is a significant difference. This leads to the question whether there is no difference or whether this observation is due to the conformation chosen in simulations or due to analysis. To address this question, a homology model based on CCS data from the experiment has been created (see section 3.2.2.3). In addition, the dimerisation interface has been defined manually for a more rigorous analysis (see section 3.2.2.4).

3.2.2.3 Creation of a Homology Model for Further Simulations

With the data from initial simulations being inconclusive, the question arose whether using the occluded state was the correct modelling of the structure investigated in the experiments. For SemiSWEET this is not trivial for two reason. SemiSWEET has three different conformations: inward-open, occluded (i.e. closed to both sides) and outward-open (figure 3.11). As a passive transporter, it interconverts freely between them[338,339,349]. To align simulations and experiments the CCS of SemiSWEET has been determined through ion mobility mass spectrometry. This value has been compared to CCS calculated using IMPACT[350] software of different states of SemiSWEET shown in figure 3.11.

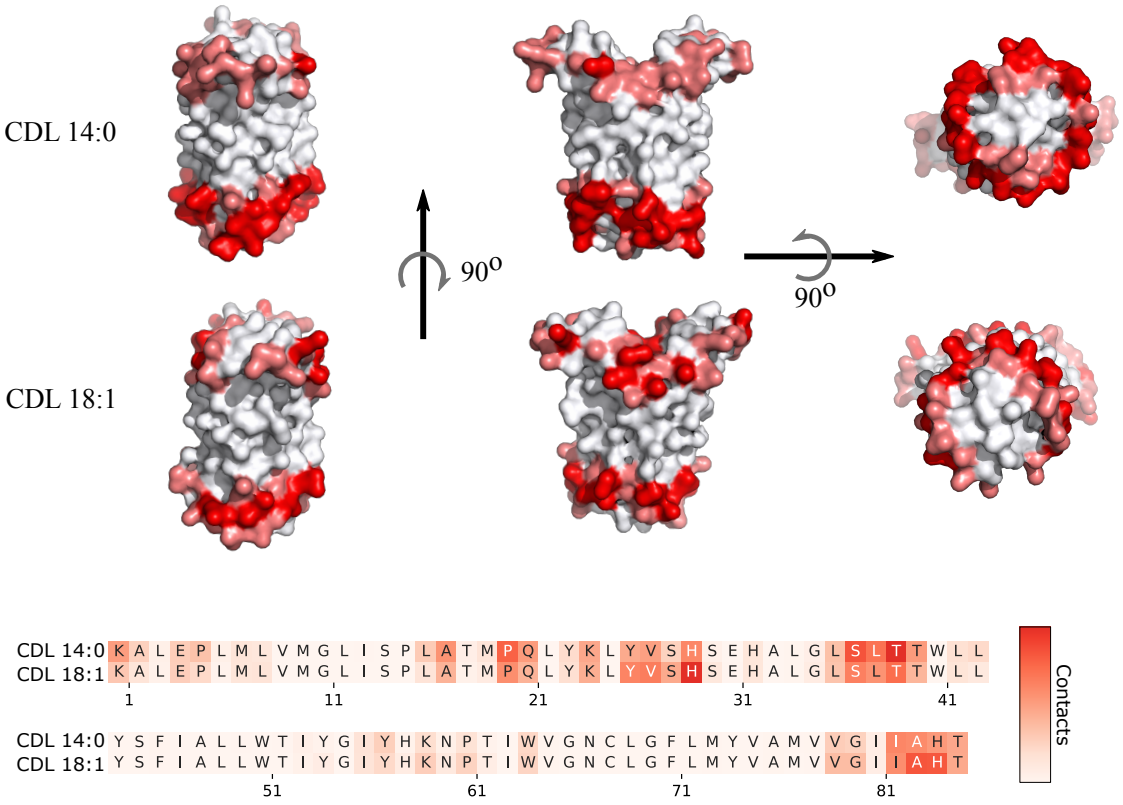

Figure 3.10. Binding location of different types of CDL in initial simulations. ***Top*** *CDL contacts mapped onto the 4QND SemiSWEET structure.* ***Bottom*** *Contacts mapped onto the sequence.*

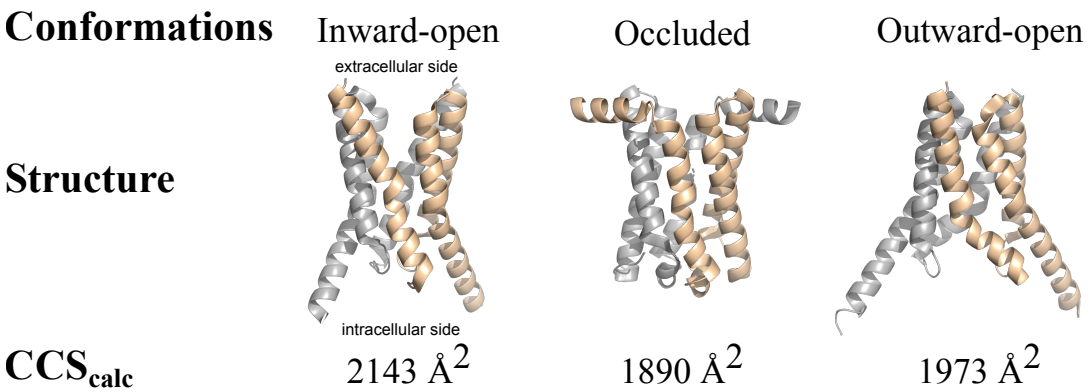

Figure 3.11. Different structures of conformations of SemiSWEET and calculated CCS as well as CCS measured through IM.

CCS was calculated by the IMPACT software package[350]*. The inward open conformation shows the best agreement.*

The inward-open conformation matched the experimental value best. This fact is in agreement with the expectation, since Lee *et al.* implied the inward-open conformation is more stable than the outward-open conformation[339]. On this basis, I decided to create a homology model of *Vibrio sp.* SemiSWEET to ensure a close link to experiments. The model was created with Swiss Model using the inward-open structure of *E.coli* as a template. The structure is shown in figure 3.12 below and was used for the remainder of the simulations.

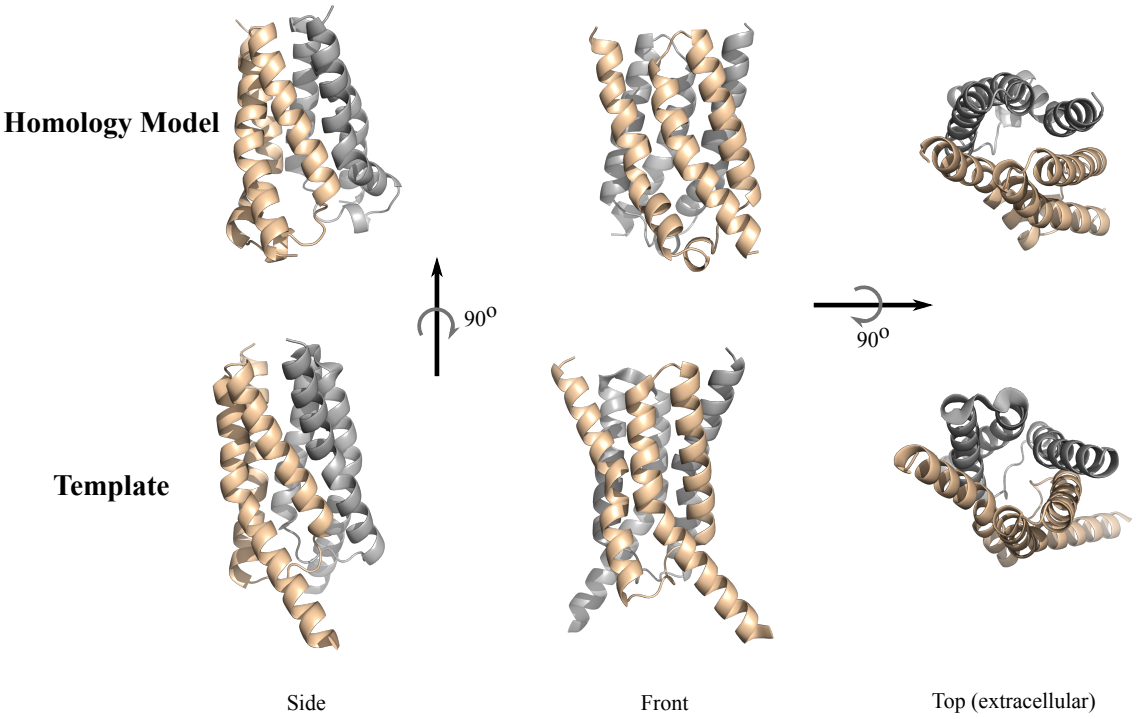

Figure 3.12. Homology model of SemiSWEET and template (PDB 4X5N). *The homology model shares the structure of the template but lacks the helices pointing far into the intracellular side as well as an ampiphilic helix found in the occluded-state structure 4QND (see figure 3.11 above).*

3.2.2.4 Definition of the Dimer Interface

The initial analysis made it hard to determine whether CDL binds to the interface or away from it. Therefore, two binding regions have been defined. The first approach is defining the interface through a minimum distance of 6 Å because this distance

is commonly used in CGMD to define contacts[274]. This is shown in figure 3.13 in the upper row. It is clearly visible that most of the residues pointing towards the pore of the transporter are classified as forming the interface. While this is true, this definition is not helpful for the analysis of CDL binding as the lipid only sees the outer regions of the transporter. Hence, the interface has been defined visually by hand (3.13 lower row). The manually defined interface is used in the subsequent sections to discuss binding at the interface and non-interfacial binding.

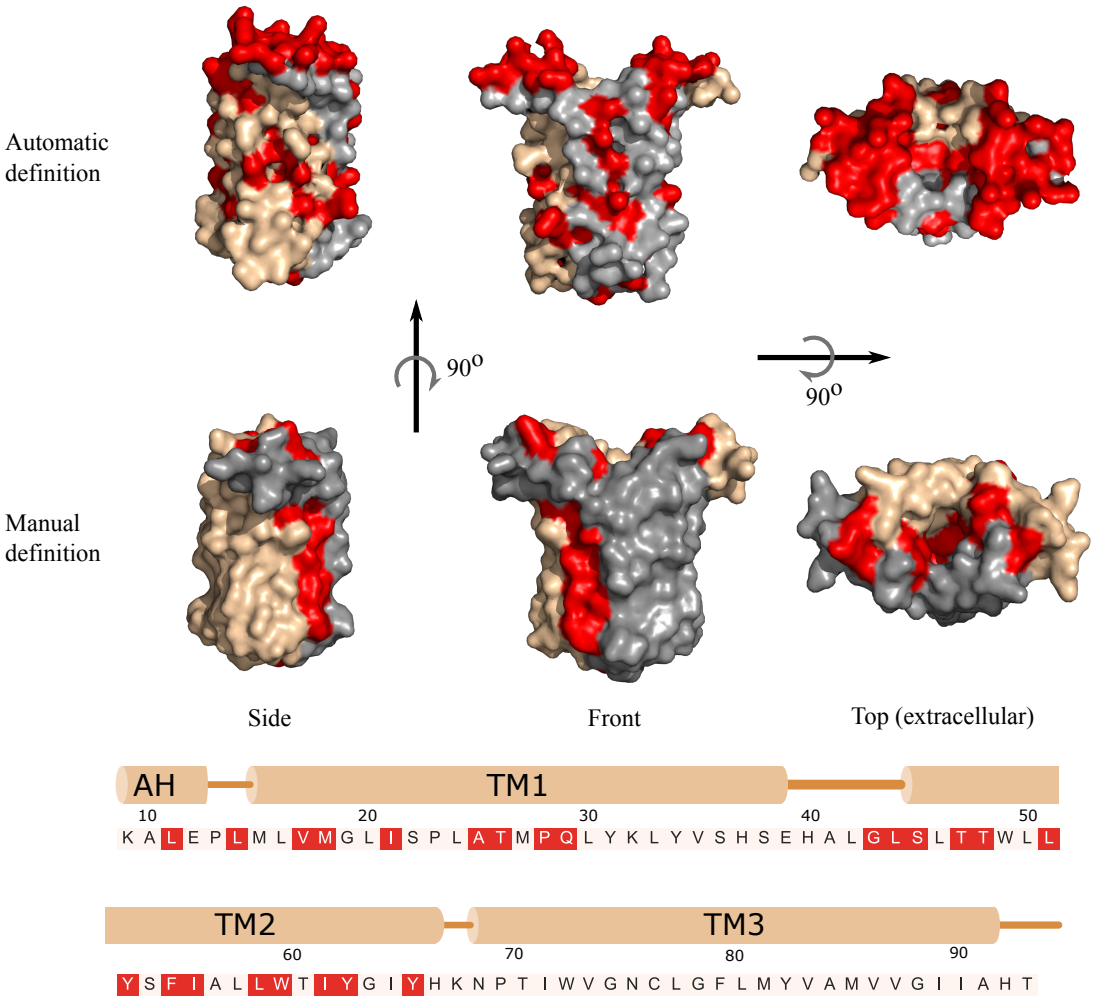

Figure 3.13. Interface of SemiSWEET dimer.

Top *Automatic definition of the interface (red) based on a 6 Å minimum distance similar to the binding definition of lipids. Due to the small size of SemiSWEET a lot of residues pointing towards the por of the transporter have been selected.* ***Bottom*** *A visual, manual selection of the interfacial residues (red) selecting for outer residues where CDL might bind.*

3.2.2.5 Binding Preferences of Different Lipids for the Homology Model

Similar to the initial simulations, a DEI for different types of lipids has been determined for the systematic simulations of the homology model (figure 3.14). Like the occluded state, a clear binding preference of CDL is visible in all cases, matching experimental data. An unexpected observation however, is the relative strength of the DEI for different types of CDL. While there is a trend towards stronger binding of unsaturated CDL to SemiSWEET, the trend for the chain length is not as clearly visible as the experimental data from Gault *et al.* suggests[166]. While the tendency to bind longer chain CDL is true for 16:X to 18:1 CDL, the observation for 14:X (especially the unsaturated CDL) defies this trend. A possible explanation is the position of the double bond. For CDL 14:1 this is at the second position of the tail, while it is in the third position for CDL 16:1 and 18:1. Therefore, this might be a decisive factor. However, different positions of the double bonds are not covered by SID experiments. Therefore, this route was not pursued any further. But it would be an interesting direction to pursue in future studies.

While the DEI for different CDL might be an interesting avenue to pursue, the core interest of this project is to find the relative affinities to locations in CDL binding to SemiSWEET. This is because in SID experiments, we select SemiSWEET with CDL bound state and assess the stability of this complex. While the DEI above one indicates that CDL is present (in line with the literature[168]) but the the binding location is unknown.

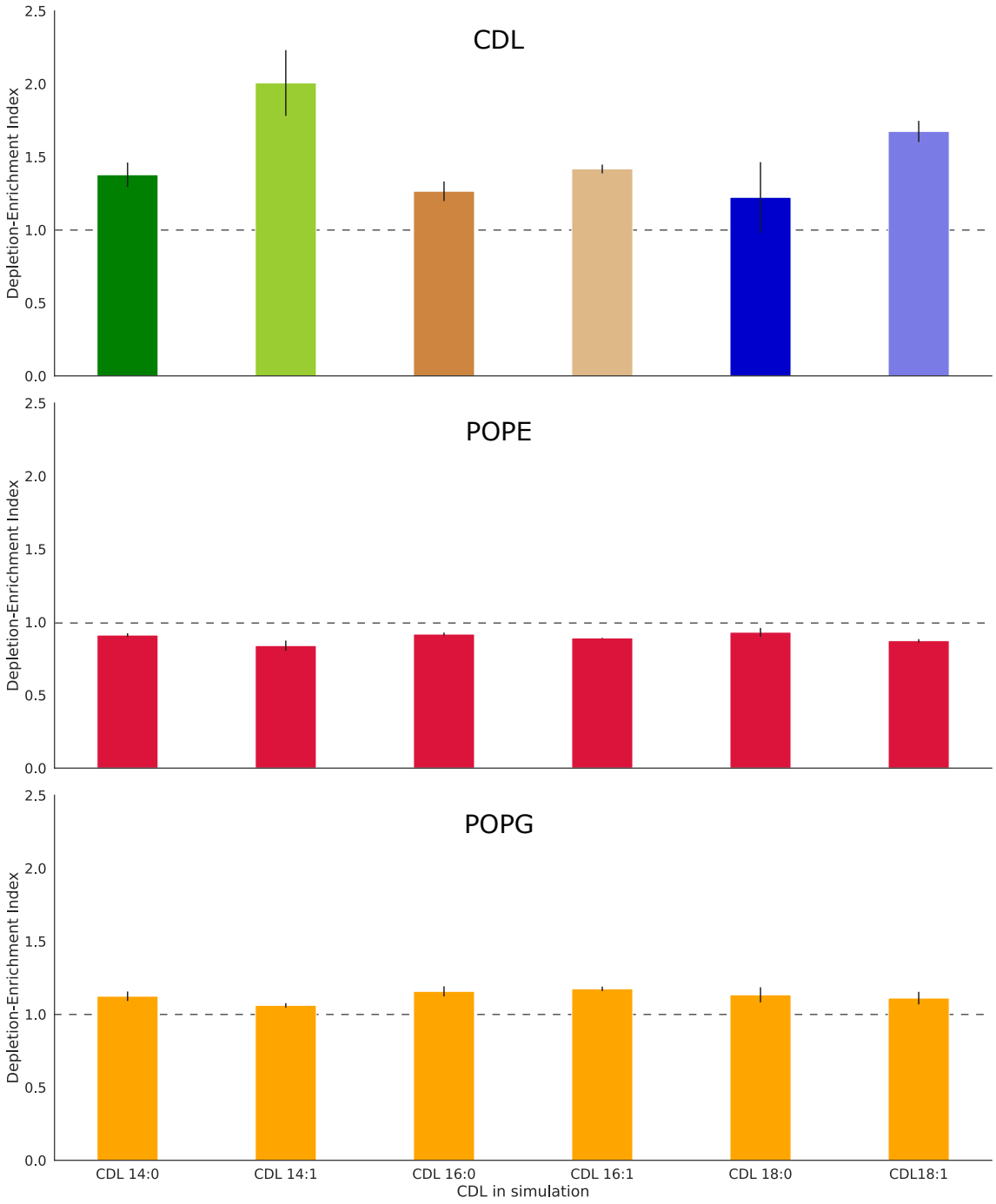

Figure 3.14. DEI for CDL, POPG and POPE in the homology model simulations.

Top *DEI for different types of CDL in different set of simulations. Clear differences between lipids are visible.* ***Middle*** *DEI for POPE in different simulations. Here, very little difference has been observed for different CDL simulations and all simulations exhibit a DEI below one indicating that it is less present near the protein than in bulk.* ***Bottom*** *DEI for POPG. A binding above one indicates a slight binding preference, although it is lower than CDL in all cases.*

3.2.2.6 Binding Locations for Different CDL for Homology Model

An analysis of binding location has been carried by studying headgroup contacts of CDL to SemiSWEET, similar to the occluded state. The results are shown in figure 3.15. Here, the contacts are normalised through division by the largest number of contacts for a lipid type in a set of simulation. Here, CDL tends to bind to similar residues in all simulations. This matches the expectation since it is the same headgroup in all cases with only the tails differing.

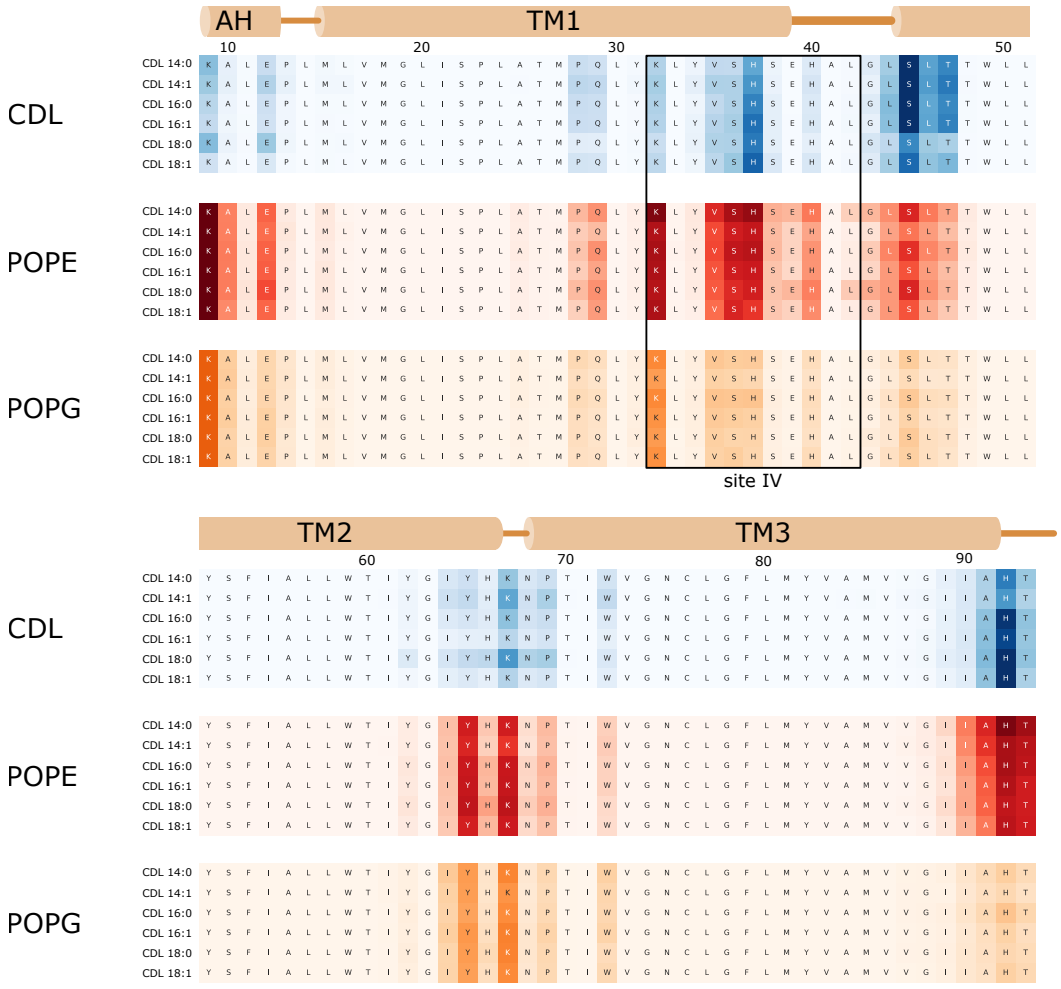

Figure 3.15. Heatmap of the relative contact strength of residues for CDL, POPG and POPE mapped onto the sequence of the homology model.

Contacts of different lipid classes are shown in different colours (CDL in blue, POPE in red and POPG in orange). A deeper colour indicates stronger binding. For the scale the number of contacts of a lipid with a residue in the simulation has been normalised by division through the sum of all contacts.

Therefore, I decided to probe the interface binding hypothesis by investigating the percentage of contacts taking place at the interface for different lipid classes (figure 3.16). Here, a similar trend as for the DEI: CDL 14:1 binds strongest to the interface while the other values are rather similar.

To probe whether there is a statistically significant difference, an ANOVA test has been performed. This test assumes that the sample originate from the same distribution an discards this hypothesis if a p-value is below 0.05. This test has been performed for the interface percentage for different lipid classes. It yielded no significant difference for POPE (p=0.37) and POPG (p=0.74). However, for CDL a significant difference was noticed(p=0.018). This means, at least one sample is significantly different from another. To see which samples are significantly different, a post-hoc Tukey test has been performed on all possible pairs: only CDL 14:1 and 18:0 can be distiguished with a significance of p=0.0078. However, against the hypothesis CDL 14:1 appears to be binding stronger to the interface. While this difference appears to be valid, the question is whether it accurately represents the underlying reality. Given that the error bars for CDL 18:0 are rather large and larger than other CDL, it is unclear, whether this is a valid comparison. Also, the uncertainty are much smaller for CDL 18:1 or CDL 16:0, the closest related types of CDL in these simulations. Therefore, despite finding a difference for these percentages I conclude that the hypothesis of interfacial binding for longer CDL to provide the stabilisation for SID cannot be answered from these CGMD simulations.

As mentioned in the introduction of this paper, recently Deng *et al.* identified four binding sites of CDL though extensive CGMD of the occluded state (figure 3.3B). While sites I-III appear less strong binding, site IV (figure 3.15 box) at the

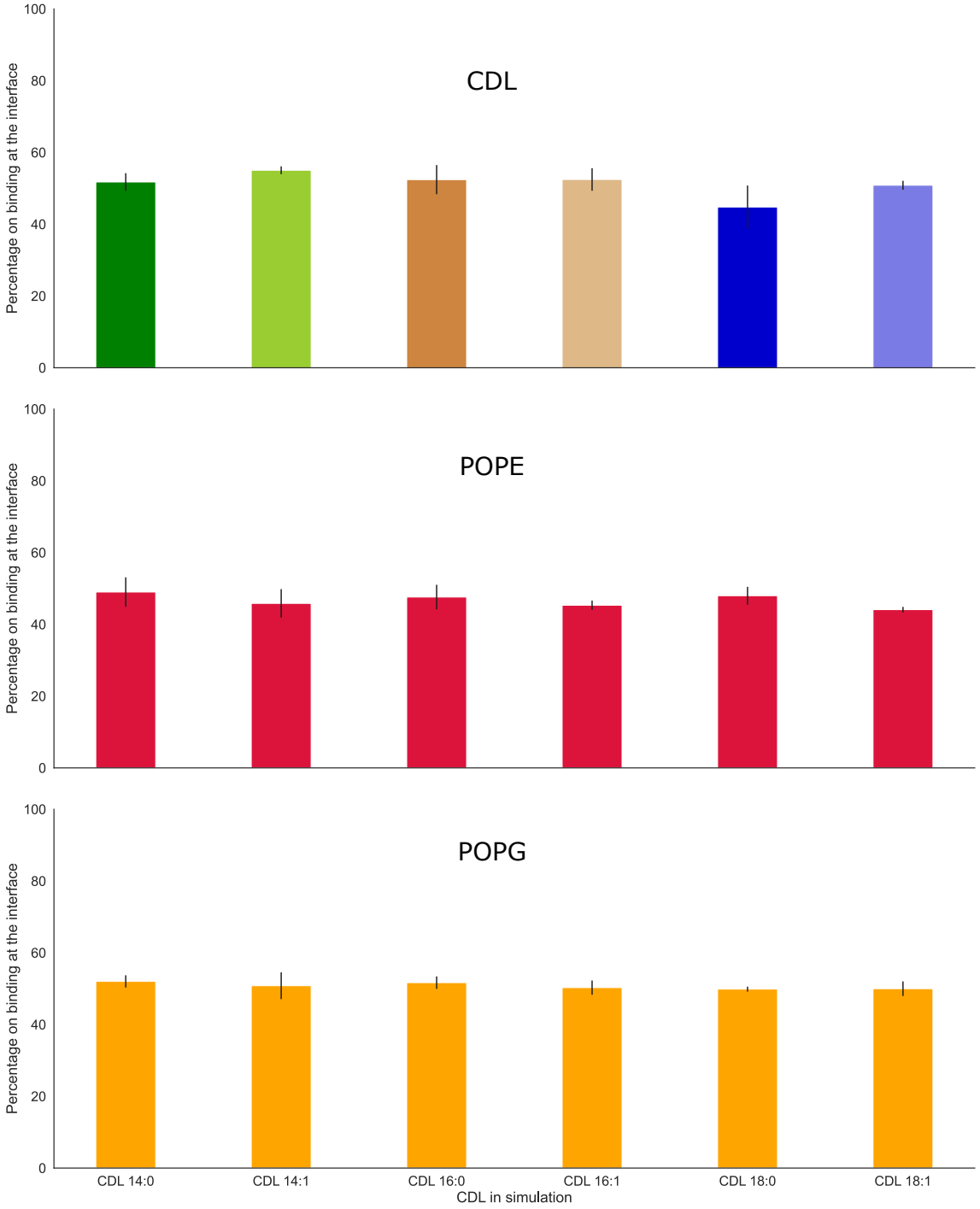

Figure 3.16. Percentage of interface binding for for CDL, POPG and POPE mapped for the the homology model.

Top *Percentage of binding at the interface for different types of CDL in different set of simulations. These percentage of interface binding does not vary significantly.* ***Middle*** *Percentage of binding at the interface for POPE in different simulations. Similar to CDL, differences are minor.* ***Bottom*** *Percentage of binding at the interface for POPG. For both POPG and POPE very little difference is visible.*

intracellular site showed strong binding (88.9% occupation and a residence time of 220 ns) and is most conserved in different SemiSWEETs[224]. In addition, they

performed FEP calculations leading to a binding energy of −15 kJ/mol. This site was also indicated to regulate dimer formation through observing the dimerisation process though extensive simulations of monomers. This makes it worthwhile to probe whether CDL binding to this site leads to the difference in the stabilisation effect. Similar to the interface residues, the percentage of contacts of CDL with the residues indicated in site IV is used to quantify the the relative binding preference (figure 3.17). A clear tendency towards binding for longer tails is visible as well as a tendency for unsaturated tails to improve binding.

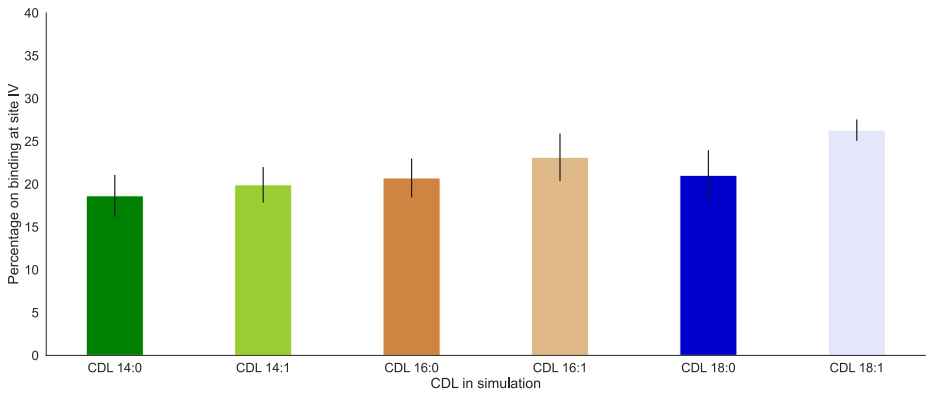

Figure 3.17. Percentage of contacts binding for for CDL binding to site IV.

To probe whether the observations are statistically significant, ANOVA tests are performed again. This yielded a significant difference (p=0.0092). Therefore, post-hoc Tukey tests have been performed on all pairs. This allowed to distinguish the short CDL 14:X from long unsaturated CDL 18:1 with significance (p=0.0071 for CDL 14:0 and p=0.021 for CDL 14:1). Therefore, binding to site IV appears to be the decisive factor for different CDL tails binding to SemiSWEET. CDL binding to site IV is shown to facilitate the correct orientation in the dimersation process. The CGMD data here suggest, that it is more likely to see CDL bound to site IV

for longer lipid tails. The stabilisation effect observed in SID could indicate that CDL binding to this site also helps to stabilise the dimer interface. This leads to the questions of what molecular details are different to stabilise the binding. It is challenging to obtain insight into these subtle effects though CGMD. Therefore, I resorted to atomistic simulations in the next section.

3.2.2.7 Atomistic Simulations to Determine the Effect of CDL binding on Dimer Stability

To probe stability of the dimer, I selected CDL-bound state snapshots of CDL 14:0 and CDL 18:1 CGMD simulations to create starting points for atomistic simulations (figure 3.18 top). These atomistic simulations were run for four repeats of 100 ns each (for more details see the methods section below). The first observation is, that in the course of this simulation span CDL does not leave the binding location. Hence, these simulations investigate only the CDL-bound state. As a measurement for stability the RMSD for the C_α atoms of the monomers relative to one another has been chosen. The reason for this choice is that SID measures interface stability and should be reflected in the choice of the measured quantity. The RMSD plots are shown in figure 3.18 bottom.

Averaging over the RMSD after an equilibration time of 20 ns (included in figure 3.18), yields an RMSD of 0.58 ± 0.09 nm for the CDL 14:0 simulations and an RMSD of 0.53 ± 0.06 nm for the CDL 18:1 simulations. This is highly similar and therefore inconclusive. The difference is probably due to a single run exhibiting more flexibility for CDL 14:0. To investigate the effect of lipid-binding PMF calculations of pulling the dimer apart could be carried out. However, these simulations would

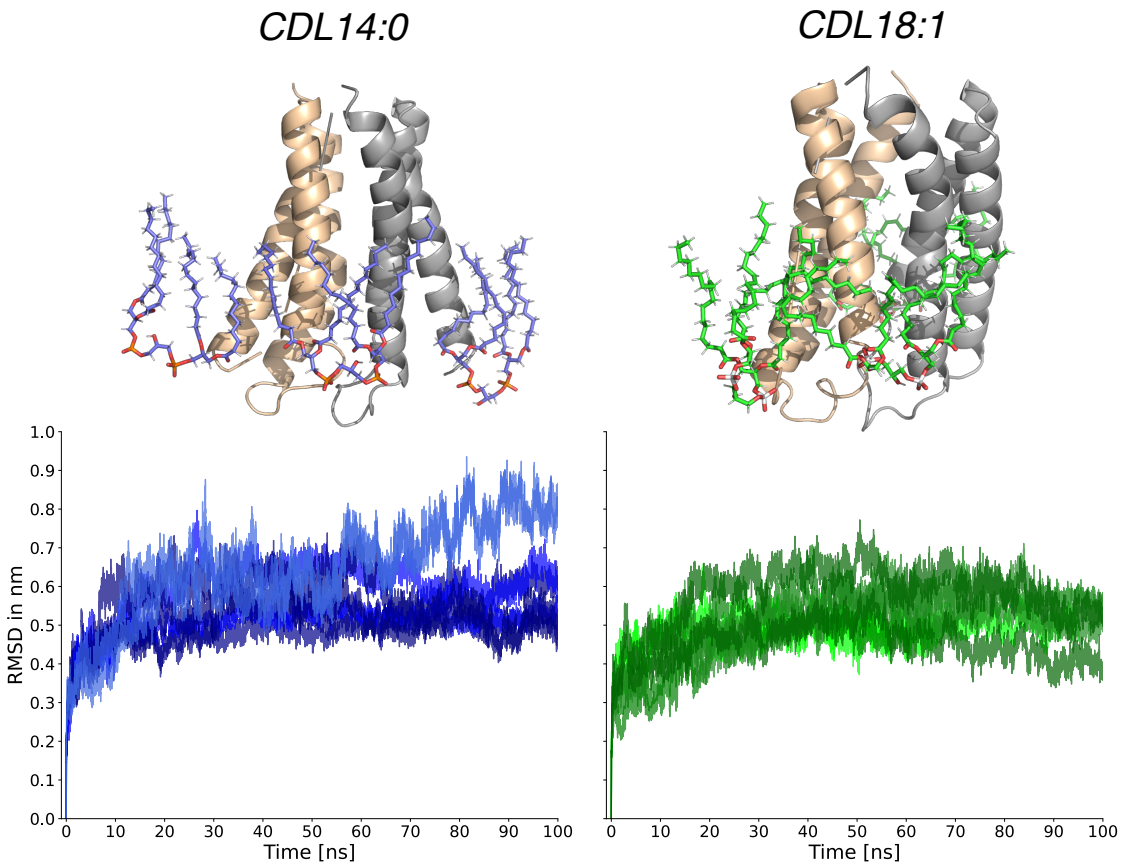

Figure 3.18. Starting structure of the atomistic simulations with CDL bound & RMSD plots for CDL 14:0 and CDL 18:1 bound state of SemiSWEET for 100 ns.

Top *Starting snapshot obtained from CG simulations and converted by CG2AT. CDL 14:0 is shown in blue and CDL 18:1 in green.* ***Bottom*** *RMSD of the dimers respective to one another of the C_α atoms of the dimer.*

likely require long sampling times due to the high number of degrees of freedom. This makes them very hard to impossible to perform.

3.3 Site-directed Mutagenesis of SemiSWEET to Weaken the Dimer Interface

The second part of this chapter investigates mutations of the dimer interface of SemiSWEET. The selection of suitable mutants was done through integrating orthogonal approaches. First, residues at the dimer interface were selected through

PSN-Ensemble, a network based approach that determines residues involved in protein-protein interactions through side-chain interactions of the occluded state of *E.coli* SemiSWEET in the occluded state (PDB: 4QND)[351]. This yielded two networks of interaction in the dimer: one at the lower half centred around the PQ motive and another one near the amphiphilic helix at the extracellular side (figure 3.19A). The residues with most contacts are shown mapped onto the sequence (figure 3.19B). In total, six residues have been defined as being in contact with two residues from the other monomer and twelve residues as being in contact with one residue from the other monomer.

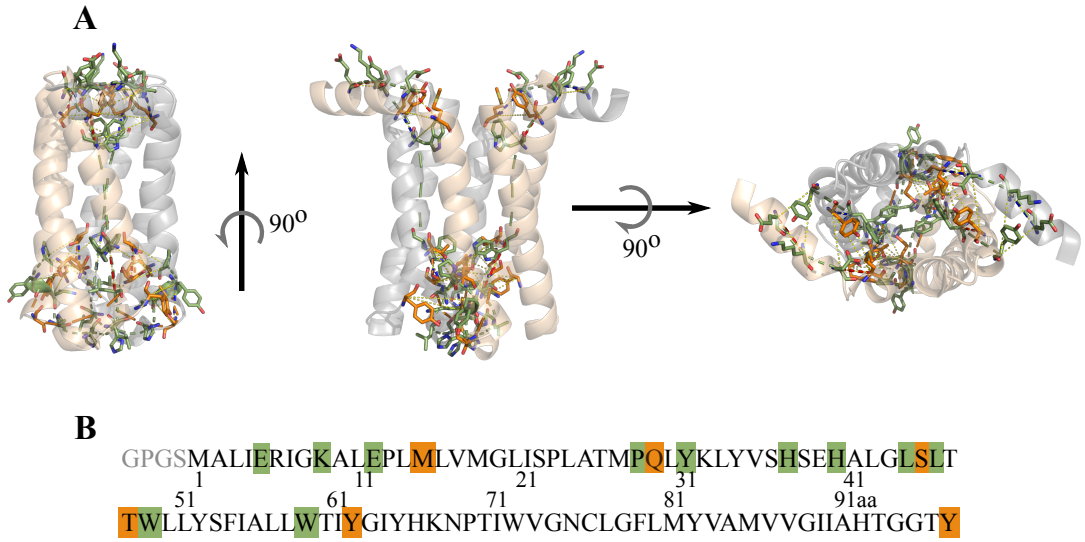

Figure 3.19. Interface of SemiSWEET dimer.

A *Network of interactions defined by PSN-Ensemble. Two networks are visible: one near the extracellular side and one centred around the PQ loop. Residues with one contact are shown in green, residues with two contacts in orange.* ***B*** *Mapping of residues onto the SemiSWEET sequence.*

Thus, to narrow down the selection, sequence alignment of the SemiSWEET from different species (*Vibrio Sp.* (Vs), *E.coli* (Ec), *T. yellowstonii* Ty and *L. biflexa* (Lb) has been carried out (figure 3.20A). While there are several conserved residues, there are two regions of particular conservation: the PQ loop and the YG motif. Here, two residues are conserved indicating biological relevance (especially for the

PQ-loop described above). Further analysis of the position of two residues in the occluded state of VsSemiSWEET revealed that they are essential for forming polar contacts of the dimerisation interface in different structures (figure 3.20B).

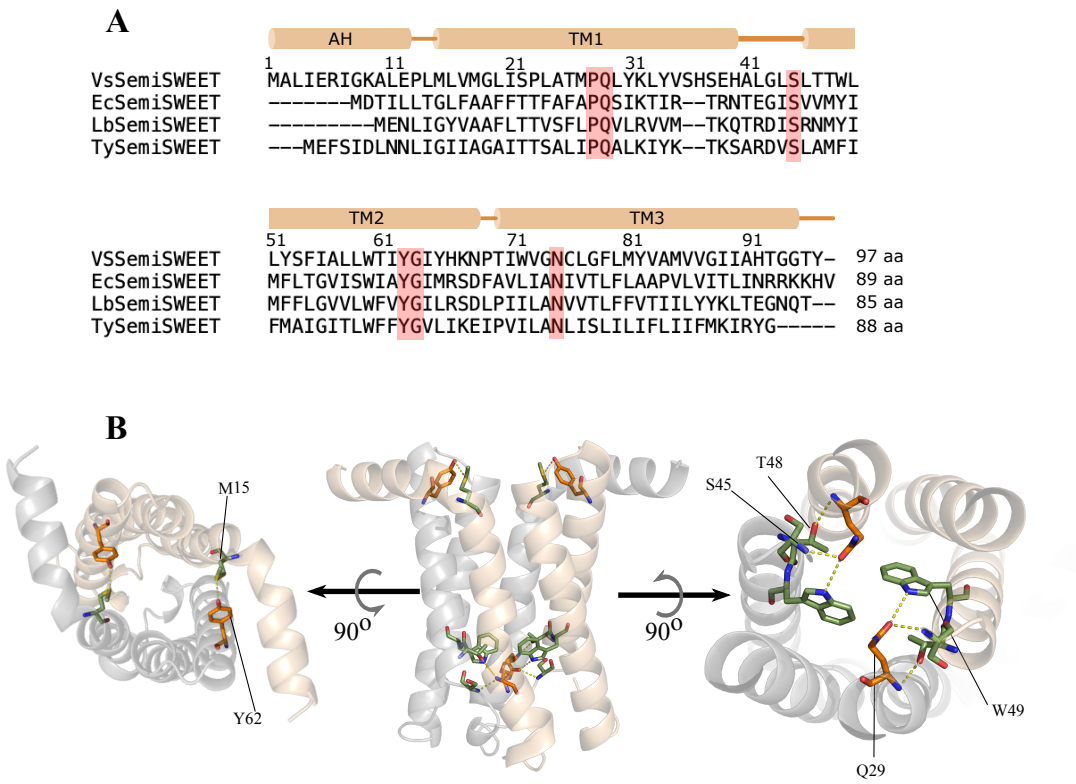

Figure 3.20. Sequence alignment of SemiSWEET from different species and ontacts of conserved residues in VsSemiSWEET.

A *Polar contacts of the the two residues selected for mutation (Q29 and Y62, orange). Q29 exhibits a hydrogen bridge to M15 (green). Q29 has polar contacts to S45, T48 and W49 (all green).* ***B*** *Alignment of SemiSWEET sequences. Conserved residues are shown in red.*

Therefore, I decided to use them for site-directed mutagenesis and 'bulk out' the interactions by changing these residues to tryptophans leading to two mutants Q29W and Y62W. Tryptophan was chosen over an alanine mutation to increase the disruption of the interface. The hypothesis was, that the Y62W mutation disrupts the interface mildly while the Q29W mutation has a stronger disruptive effect. These effects are probed through the same activation techniques which were used to investigate the stabilising effect of lipids: CIU and SID.

3.3.1 Mass Spectrometry of the Mutants

Both mutants were expressed and purified using the wild type protocol (section 3.5.1). Next mass spectrometry experiments were performed on a Synapt G2 (figure 3.21). These experiments showed that the Q29W mutant disrupts the interface so much it is present as a monomer but no dimer has been observed. In addition, an apo mass of 10804.96 ± 0.99 Da was observed. This is a cleavage of 800 Da from the expected mass of 11604 Da was observed corresponding to a loss of a sequence of LEVLFQ at the C-terminal side of the sequence. For the Y62W mutant this is not the case. Here, a suitable dimer population was present and no cleavage was observed. Therefore, CIU and SID experiments could be carried out only on the Y62W mutant.

To start CIU experiments, a cone energy ramp of the mutant with and without CDL bound has been carried out on the 7+ charge state to match WT experiments performed by Denis Shutin. The peaks used for dissociation are indicated in figure 3.21 (apo in purple, CDL-bound state in green). The results are shown in figure 3.22 left. A clear destabilisation compared to WT SemiSWEET (stable up to 180 V) is visible (see figure 3.4 with WT data provided by Denis Shutin). From this data a cone voltage of 130 V has been chosen to ensure SemiSWEET is not unfolded while maximising removal of detergent molecules. Next, CIU is carried out through a trap ramp to probe the stability of the 'cleaned-up' dimer. Here, the voltages have been stepwise increased by 2 V from 20 V to 80 V. The results of the subsequent IM analysis shown in figure 3.22. A stabilisation of 10.07 ± 0.70 V has been observed,

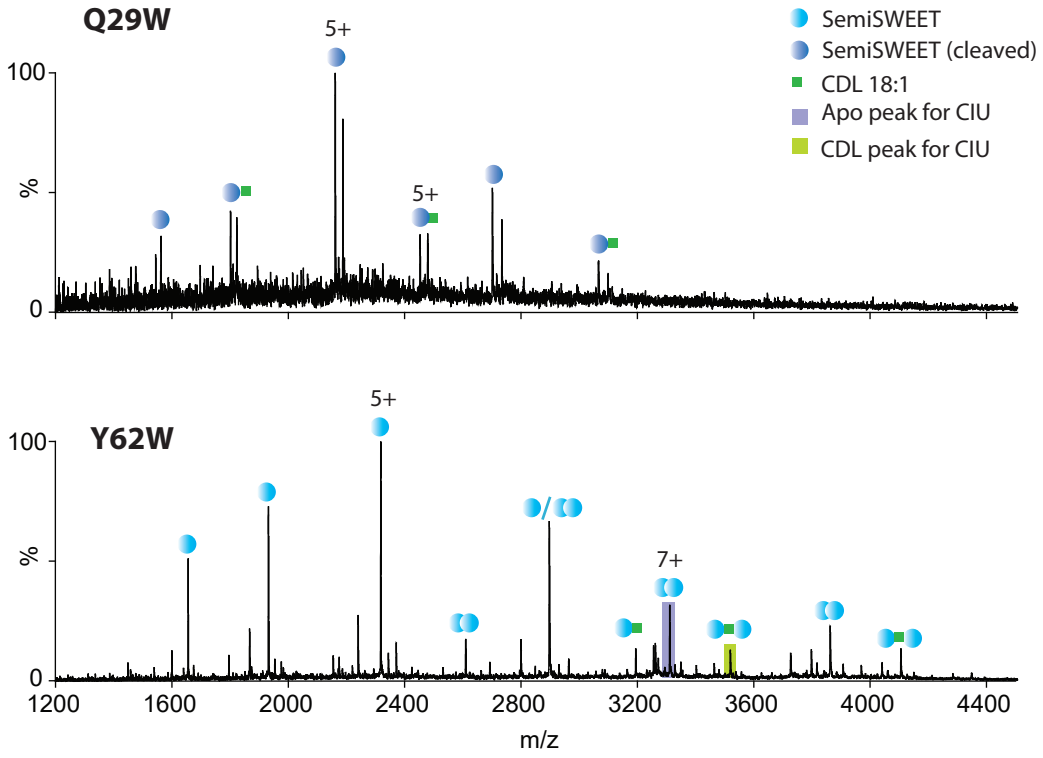

Figure 3.21. Mass spectra of SemiSWEET mutants upon CDL addition collected on a Waters Synapt G2.

Top *Spectrum of the Q29W mutant. Only monomer was visible for this mutant but CDL binding has been observed, too. A cleavage of the monomer has been observed.* ***Bottom*** *Spectrum of the Y62W mutant including an equilibrium of monomer and dimer.*

compared to a stabilisation of 9.24 ± 2.19 V for the WT SemiSWEET. Generally, the dimer is less stable than the WT SemiSWEET. This matches the expectation although CDL binding increases the stability slightly less than the WT stability. The details of this difference need further studies but could indicate a conformational change due to the steric hindrance caused by tryptophan.

In addition, SID experiments have been carried out by Denis Shutin (figure 3.23). Here, we unexpectedly observed a stabilisation of the dimer population through the mutant. A possible explanation might be, that a larger distance at the extracellular site might push the intracellular side closer together and increase the strength of

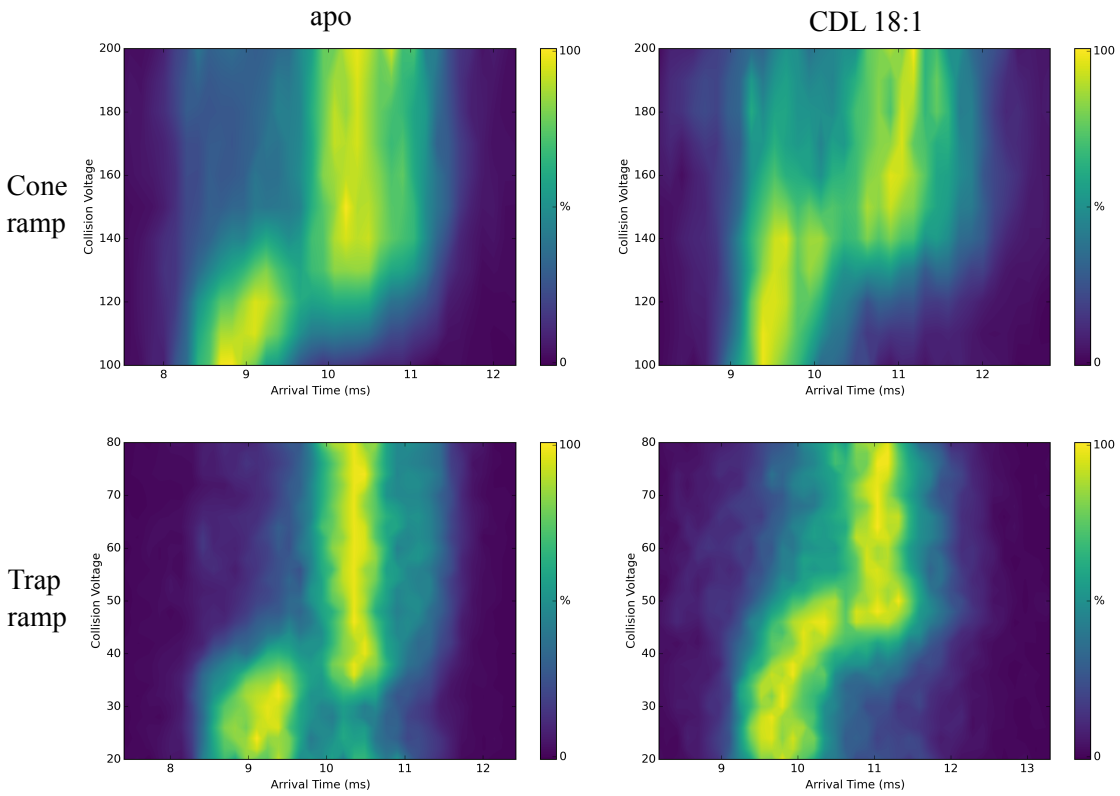

Figure 3.22. CIU data of the Y62W mutant.

Top *Cone ramp of Y62W. A clear stabilisation upon CDL binding is visible* ***Bottom*** *Trap ramp of the Y62W mutant. Like in the cone ramp, a stabilisation occurs.*

the network of interactions centred around the PQ-loop which form more contacts (figure 3.19). However, this is only a hypothesis. Further investigation of this finding is required. As the bottom line, the work on mutants shows that dimer interfaces can be modulated through carefully selected residues that can be predicted by the PSN-ensemble (one mutant shows a modulation of the interface strength while the other disrupt the interface enough that dissociation experiments are not necessary because there is no dimer left).

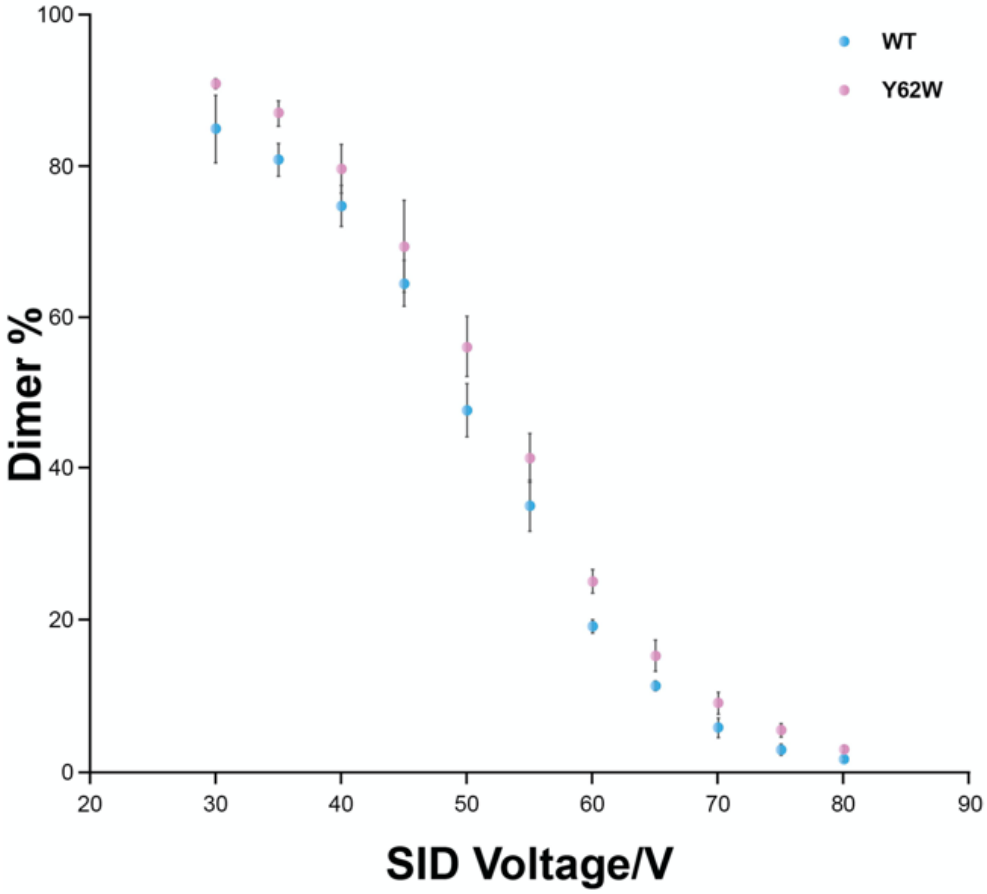

Figure 3.23. SID data of the Y62W mutant.
Dissociation experiments show a clear difference of the WT and the Y62W. Surprisingly, the WT showed a more stable interface. The values stem from three repeats and the error bars represent one standard deviation.

3.4 Conclusions and Future Directions

The SID data collected in this project shows that this technique is able to pick up the stabilisation effect of different lipids on the SemiSWEET interface. In addition, these experiments are able to distinguish even the subtle contributions on stability of different tail length and saturation. A difference in binding location has been observed for different classes of lipids in CGMD. The fact, that these subtle differences can be observed with the Martini model shows the power of CGMD and is somewhat surprising given that the hydrogen bonds are only included in the parametrisation

of beads. However, previous work by Parker *et al.* was able to study similar effects for a eukaryotic sugar transporter and PC lipids[352]. While the initial hypothesis of binding at the interface for longer tails and away from it for shorter tails could neither be falsified nor confirmed, a clear trend and statistically significant differences of binding preferences to a specific binding site (site IV from Deng *et al.*) was observed[224]. As this site has been found to be crucial to coordinate the dimerisation of SemiSWEET, binding to this site is a possible and likely explanation for the experimental results from SID.

Additional confidence in this effect can be gained through multiple routes. Experimentally, site-directed mutagenesis would be a suitable approach. To define which residues are the most suitable candidates for mutation I suggest longer atomistic simulations. While 100 ns of atomistic simulations have been carried out, they were not sufficient to see a difference for CDL 14:0 and CDL 18:1 in terms of stabilisation. These simulations are a promising approach despite being computationally expensive. Latoracca *et al.* performed more than 100 μ s of atomistic simulations and were able to see a conformational change of SemiSWEET from outward-open to inward-open in line with crystal structures inlcuding sugar transport[349]. Unfortunately, in that study only POPC was used for the lipid bilayer. Therefore, a bilayer mimicing the bacterial membrane could solve multiple questions at once. First, it could show whether CDL binding affects the conformation and by that the effect onto sugar transport. Also, it might be possible to distinguish a preference for a particular state for CDL binding as well as the effect on the equilibrium. Given that SID experiments noticed a stabilisation of the dimer, I consider this a promising

approach for future work despite the high computational costs.

CGMD can also contribute to extend this work in two ways. The question of state dependence (occluded, inward- and outward-open) can be addressed by perfoming systemic CGMD simulations with all CDL lipids created in this work. This might reveal a state-dependent binding similar as it has been shown to GPCRs for PIP2[169]. In addition, free energy calculations can inform on the energetics of different CDL binding as well as the effect of CDL binding on dimer stability (although the latter has the caveat that secondary structures are fixed in CGMD). Therefore, the more promising way might be to start with PMF/FEP of CDL binding although reaction coordinates would need to be decided and CGMD PMF might not notice significant differences for different CDL because they might be to similar. In addition, probing the dimer stability through PMFs might, however, require very long simulation times because of the degrees of freedom of the monomer. These simulations have briefly been attempted in this work, but complications such the intracellular end of the monomers sticking arose which led me abandoning this idea. PMFs of dimers would require too many restrictions of the monomer movement or require excessive sampling times.

For the second part of the project, it can be noted that the PSN-ensemble is able to predict residues that influence the interface stability. However, this is not without any caveats. The PSN-ensemble used distance as a metric for contacts and clusters. Of course, interacting residues are in contact. Combining a sequence analysis to find conserved residues and checking their contacts in structures do yield similar results.

It should be noted however, that it is not a trivial observation that a mutation of glutamine in the PQ-loop leads to a weakened interface. Despite being crucial for the conformational flexibility, the glutamine is not conserved in the eukaryotic SWEET transporters such as rice SWEET2b transporter[340,353] while the proline is conserved. This might indicate a crucial role of this glutamine residue in the interface stabilisation of SemiSWEET while SWEET transporters have other means for stabilising the functional conformations due to their monomeric 7-TM architecture. This needs further investigation. Also, it would be interesting to know what exactly causes the destabilisation: the 'bulking out' of interactions due to tryptophan or the lack of polar contacts. This could be investigated through alanine mutations instead of tryptophan mutations. In addition, MD simulations can help to investigate the stability question further through the same methods suggested above: PMF calculations with the same potential difficulties and the question of the conformation due to the mutation. In my opinion a closer investigation would be worthwhile because the PQ-loop is conserved for many transporters and even receptors such as the KDEL receptor which shares similarity to transporters[340].

The Y62W mutation was able to modulate the interface strength of SemiSWEET compared to the wildtype. This enabled me to collect CIU data and my collaborator to collect SID data. While the CIU data confirms the expected weakening, the SID data points the other direction, i.e. stabilising. Similar to the Q29W mutation, a mutation of Y62W to Y62A might be a suitable way forward as alanine does not induce a steric hindrance of tryptophans large side chain. This potential future work could also be complemented by MD simulations.

3.5 Materials & Methods

3.5.1 Expression and Purification of SemiSWEET

The expression and purification of SemiSWEET followed established protocol described previously[168]. The plasmid of VsSemiSWEET has been used to transform BL21 λ DE3 *E.coli* cells (Novagen). Colonies of these cells were used to inoculate flasks containing 100 ml of lysogeny Broth (LB) and 50 mg/ml kanamycin overnight at 37 °C. Afterwards, 15 ml of cells in LB were transferred 1 L flasks of the same mixture of LB and kanamycin to grow at 37 °C for three hours until an optical density of 0.6-0.8. To trigger expression 1 ml of 0.2 M of Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and expression took place overnight at 22 °C.

Cells were harvested by 10 minutes of centrifugation at 5,000 g at 4 °C and lysed through in a M-110 PS Microfluidiser (Microfluidics) with added protease inhibitor tablets (Roche) into the lysis buffer (300 mM sodium chloride, 20 mM Tris, pH 7.4). To separate the membranes cell debris was pelleted by 20 minutes of centrifugation at 20,000 g followed by ultracentrifugation at 100,000 g for 2 hours. After homogenisation membrane proteins were extracted with 2 % w/v of powdered Dodecyl β-D-maltopyranoside (DDM) at 4 °C overnight in resuspension buffer (100 mM sodium chloride, 20 mM Tris, 20%w/v glycerol; pH 8.0). The protein possesses a His-tag which allowed purification by immobilised metal affinity chromatography (IMAC) through a 5-ml HisTrap-HP column (GE Healthcare) with an equilibration buffer 150 mM sodium chloride, 20 mM Tris, 20 mM imidazole, 10% w/v glycerol, 1

mM DTT, 0.02 % w/v DDM; pH 8.0 and an elution buffer 50 mM sodium chloride, 20 mM Tris, 500 mM imidazole, 10% w/v glycerol, 1 mM DTT, 0.02 % w/v DDM; pH 8.0. After dialysis overnight with HRV 3C Protease to cleave the His-tag, the same column has been used for reverse IMAC followed by size exclusion chromatography (SEC) using a 24 ml Superdex 200 Increase 10/300 GL SEC column (GE Healthcare). At this stage, a buffer exchange into a Tetraethylene glycol monooctyl ether (C8E4) of 150 mM sodium chloride, 20 mM Tris, 10 % w/v glycerol, 1 mM DTT, 0.5 % w/v C8E4 at pH 8.0. was conducted. The sample was concentrated using a 10 kDa molecular weight cut-off MWCO concentrator and flash-frozen to be stored at $-80\,^{\circ}\mathrm{C}$.

The Q29W and Y62Q mutants have been expressed and purified in the same way.

3.5.2 Native Mass Spectrometry Experiments

3.5.2.1 Sample preparation

For nMS samples were exchanged into 200 mM ammonium acetate (AA) with double the critical micelle concentration (CMC) of detergent (here this means C8E4 at 0.5 % w/v) through a Biospin column (Micro Bio-Spin 6, Bio-Rad) for a final concentration of 10-20 μM. CDL was purchased from Avanti. It was diluted into a 200 μM stock solution and mixed with protein samples at a 4:1 volume ratio just before the nMS experiments were conducted.

3.5.2.2 CIU on a Synapt G2

The following parameters were used for the unfolding experiments of SemiSWEET mutants (although cone and trap voltage have been varied for the respective ramps). Data analysis was carried out using Waters MassLynx software in combination with UniDEC and NaViA[191].

Parameter	Value
Capillary voltage	1.2 kV
Sampling cone	130 V
Extraction cone	4 V
Transfer CE	2.0 V
Trap gas flow (Argon)	4 ml/min
IMS gas flow	40 ml/min
Helium cell gas flow	120 ml/min
Trap WV	250
Trap WH	3.0
IM WV	950
IM WH	40
Transfer WV	100
Transfer WH	1.0
Backing Pressure	5 mbar

Table 3.1. Parameters used in CIU experiments.

3.5.3 Homology Modelling and Simulations

3.5.3.1 Parameters for CGMD Simulations

The simulations of SemiSWEET have been carried out in Gromacs 2018 using the Martini forcefield version 2.2, which merges around four heavy atoms into a bead[254,257,258,354]. The atomistic structures of both *Vibrio sp.* SemiSWEET(PDB: 4QND) and the homology model have been coarse-grained and inserted into lipid bilayers using the MemProtMD pipeline[271,338]. The lipid bilayers for the simulations were composed of 70 % POPG, 20% POPE and 10% CDL of different chainlength (see figure 3.8). The shorter versions of CDL were created by shortening a 18:1 version of CDL with 5 Martini-beads and adjusting the parameters for the unsaturated bonds. For the unsaturated species, the unsaturated hydrogen bond was kept central. The system was neutralised through 150 mM sodium chloride and the simulation temperature was set to 310 K and coupled via v-rescale[267]. Pressure was kept at 1 bar though a semi-isotropic Berendsen-barostat with a compressibility of 5×10^{-6}. Electrostatics were computed with a reaction field[355]. Lennard-Jones interactions were determined using a velocity Verlet scheme with a cut-off to enable GPU calculations[356].All bond constrains were calculated by the P-LINCS algorithm[357].

The simulations were carried out for five microseconds and five replica for each lipid composition with a timestep of 20 fs. Contacts of residues was defined by a distance of less than six angstroms between a residue bead and a lipid headgroup bead. The interfacial residues have been defined visually using PyMol[358]. Contacts and binding sites were calculated using Gromacs, MDAnalysis and PyLipID[354,359,360].

3.5.3.2 Parameters for Atomistic Simulations

Starting points for atomistic simuations were created using the CG2AT script[361] from the final snapshot of CG simulations. The system was neutralised through 150 mM sodium chloride and the simulation temperature was set to 310 K and coupled via v-rescale[267]. Pressure was kept at 1 bar though a semi-isotropic Parinello-Rahman-barostat with a compressibility of 5×10^{-5}. Electrostatics were computed with a reaction field[355]. Lennard-Jones interactions were determined using a velocity Verlet scheme with a cut-off to enable GPU calculations[356].All bond constrains were calculated by the P-LINCS algorithm[357].

The simulations were carried out for 100 nanoseconds and four replica for each lipid composition with a timestep of 2 fs. Contacts of residues was defined by a distance of less than six angstroms between a residue bead and a lipid headgroup bead. The interfacial residues have been defined visually. Contacts and binding sites were calculated using MDAnalysis, PyLipID and custom scripts as well as Gromacs tools for root mean square deviavations[354,359].

3.5.3.3 Homology Modelling

Homology modelling has been carried out based CCS using Swiss Model[362] with the sequence of VsSemiSWEET based on the inward-open structure.

3.5.3.4 ANOVA Tests

ANOVA tests including post-hoc Tukey tests have been carried through Online Web Statistical Calculators developed by Navendu Vasavada[363].

3.5.4 Contributions

- Denis Shutin performed all SID work in this chapter
- Kallol Gupta helped designing the mutants for the second part of the chapter
- Jani Bolla helped designing the geneblock for the mutants
- Denis Shutin helped co-purifying the Q29W and Y62W mutants and performed SID experiments on the latter
- CCS measurement and IMPACT calculation for structures were done by Denis Shutin

Chapter 4

Lipid Interactions of a Bacterial Potassium Channel

Abstract

Calcium-activated potassium channels are crucial regulators of neural excitability. Their malfunction is associated with a wide range of conditions and diseases. Therefore, a better understanding of the regulation of these ion channels could contribute to drug discovery of these medically relevant proteins. Calcium-activated potassium channels such as the BK channel are regulated through many stimuli including lipids. Among lipids PIP2 has been identified as a key player for the activation of BK channels. The bacterial homologue of the BK channel, MthK, is often used as a proxy to study the more complex eukaryotic channel and contributed to the understanding of this class of ion channel. Here, we investigate lipid binding of MthK through native mass spectrometry experiments and molecular dynamics simulations. Cardiolipin is identified to bind to MthK in a specific manner. In addition, lipid binding is compared to the PIP2 binding to the BK channel.

4.1 Introduction

4.1.1 BK Channel

The calcium and voltage-activated potassium channel (BK channel, also known as slo1 or Maxi-K) is a key player in allowing potassium ions to cross cell membranes. The BK channel has the highest single channel conductance among all potassium channels at 100–270 pS[364] and are ubiquitously expressed throughout tissues of the human body. It is involved in regulating various physiological processes including muscle contraction, hearing, regulating blood pressure and the regulation of neurotransmitter release[365,366]. Details of these involvement can be found an a review by Bentzen *et al.*[366]. The main role of the BK channel is in the central nervous system, where it is co-located with voltage-gated calicum channels and balances membrane polarisation in an event of hyperpolarisation (i.e. excessive calcium influx)[366]. This diverse role also causes malfunction of the BK channel to be associated with diseases such as hypertension through knockout of the $\beta 1$ subunit or epilepsy through mutations in the RCK domain[367–370]. The association with a variety of conditions makes it a relevant drug target although so far drug discovery has largely been unsuccessful[366]. Therefore, a better understanding of the channel and its regulation could inform future drug discovery.

The BK channel shares the canonical tetrameric architecture of most K^+ channels with the highly conserved selectivity filter formed by transmembrane helix S5 and S6[371]. In addition, this channel possesses an N-terminal voltage-sensing domain

consisting of transmembrane helices S0 to S4 as illustrated in figure 4.1 below. The cytosolic domain on the C-terminus consists of two Regulation of Conductance of potassium ions domains (RCK)[40]. These domains bind Ca^{2+}, the ligand of the BK channel. The RCK domains are not identical and consist of 270 and 415 residues for the first and second RCK domain, respectively (starting from position 340 and 640 in the human channel)[364]. This unit is called α subunit. Furthermore, the BK channel binds different auxiliary subunits, the β and γ subunits. The β subunits consist of two transmembrane helices. There are four types of β subunits $\beta 1 - \beta 4$ and the expression of these subunits is tissue dependent. They all bind to the tetramer of the in a one to one stoichiometry, i.e. a tetramer α subunits is complemented by a homotetramer of a type of β subunit[364].

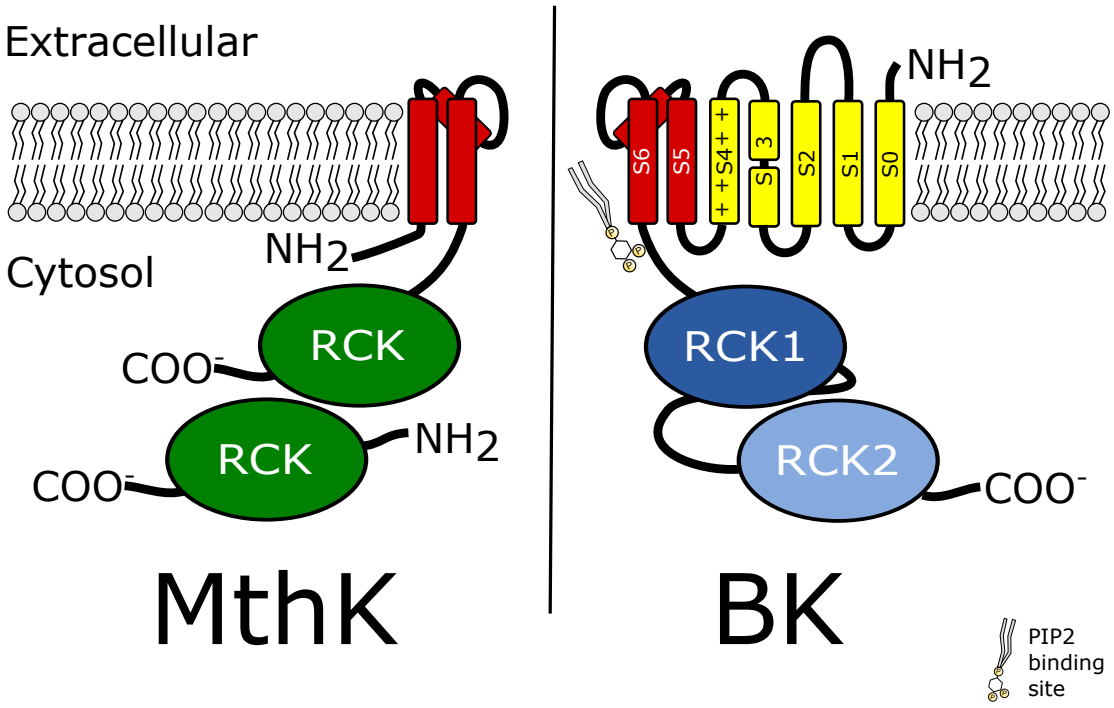

Figure 4.1. Architecture of the the BK channel and MthK.
The helices forming the selectivity filter are shown in red. The voltage-sensing domain including charges on S4 of BK are coloured yellow. The TMD of BK consists of the VSD and the SF. The RCK domains of MthK and BK are shown in blue and green respectively.

4.1.2 Regulation of the BK channel

BK is regulated through various mechanisms. As a ligand-gated ion channel, it opens upon binding Ca^{2+} in the cytosolic RCK domains with an affinity in a 1-10 micromolar range[372]. In addition, voltage activates this channel due to the VSD at the N-terminal domain[371]. In addition to other lipids, PIP_2 was shown to facilitate activation upon calcium binding[82,84] and a binding site for PIP_2 was located at the linker region between transmembrane domain and the first RCK domain[83].

4.1.3 The bacterial homologue of BK:MthK

Due to the eukaryotic nature of the BK channel, structural studies are challenging with the first structure of this channel published in 2017[373]. Therefore, its bacterial homologue MthK is often studied as a suitable proxy due to structural similarities[238]. This approach was adopted as part of the work discussed in this chapter.

MthK was named for the organism where it was first identified in 1997, *Methanobacterium thermoautotrophicum*[374]. In 2002, the crystal structure of MthK was obtained and sequence similarities to BK made it a model for understanding channels with RCK domains ever since[40,375,376]. The structure of the channel is tetrameric like most K^+ channels but showed an octameric gating ring of the RCK domains[375]. This was confirmed in later studies expressing only the RCK domain[377] (see figure 4.1). Unlike the BK channel, both RCK domains are identical and the second RCK domain is bound non-covalently while both domains are linked in the BK channel. In addition, both channels share the pore domain common to K channels[40] while the

VSD of the BK channel is absent in MthK. Both channels perform similar motions upon calcium binding[40].

In line with calcium activation of BK, MthK is also activated by calcium with the affinity in in $1 - 10mM$ range, around three orders larger than the BK channel while showing a steep Hill-coefficient of around eight[378]. This indicates a cooperative binding. However, it is not known whether calcium is the physiological ligand for multiple reasons. Firstly, the concentration of calcium in prokaryotes is an order of magnitude lower than the concentrations measured for activation (although this can be lower by at least a factor of 100 through increasing the temperature to $37°C$)[378,379]. Secondly, the role of calcium in bacteria remains elusive, while it is known to be an important secondary messenger in eukaryotic cells[380–382].

Most work on MthK until recently has focussed on electrophysiology and structural implications of calcium binding to MthK[238,375,378,383]. Other directions such as temperature regulations have been pursued as well[379,384]. The regulation of MthK by lipids by a computational study. However, in this publication the binding of eukaryotic lipids were assessed with MthK; a prokaryotic protein[385]. Therefore, in this chapter I investigate lipid interactions with MthK and compare it to PIP_2 interactions of its eukaryotic homologue, the BK channel.

4.1.4 Hypothesis on Expression and Stoichiometry

Hypotheses of MthK expression and stoichiometry were first described in 2002 with the crystal structure of MthK being solved by the MacKinnon group[375]. In this study bands of 26 kDa and 200 kDa were observed when expressing the wild-type of MthK as outlined in figure 4.2 below. This was rationalised by a single gene giving rise to two protein products: a soluble RCK domain is expressed from the internal starting point at position 107 which causes the 26 kDa band while having a tetramer of the full-length MthK expressed from position 1 of the gene for the 200 kDa band. Additional evidence was provided by mutating a methionine to isoleucine at position 107, at the beginning of the RCK domain, a small shift of the absorption peak in the (SEC) was observed and the 26 kDa band disappeared in line with the explanation. However, only the M107I mutant was crystallised and deposited in the protein data bank (PDB).

This rationale neglects the mechanisms of protein expression in *E.coli*. When ribosomes bind to the ribosome binding site upstream of the start codon, they start protein expression at the next start codon up to the next stop codon instead of skipping more than 300 bases pairs before starting the protein expression at another start codon. In addition, there are two start codons located upstream of position 107 and five more downstream; yet other shortened variants of MthK were observed by the MacKinnon group or by me.

Starting from the hypothesis, multiple experimental routes had to be pursued until

in vivo stoichiometry was achieved by expressing the RCK domain in separately in a Duet-vector. These approaches will be outlined in the following section.

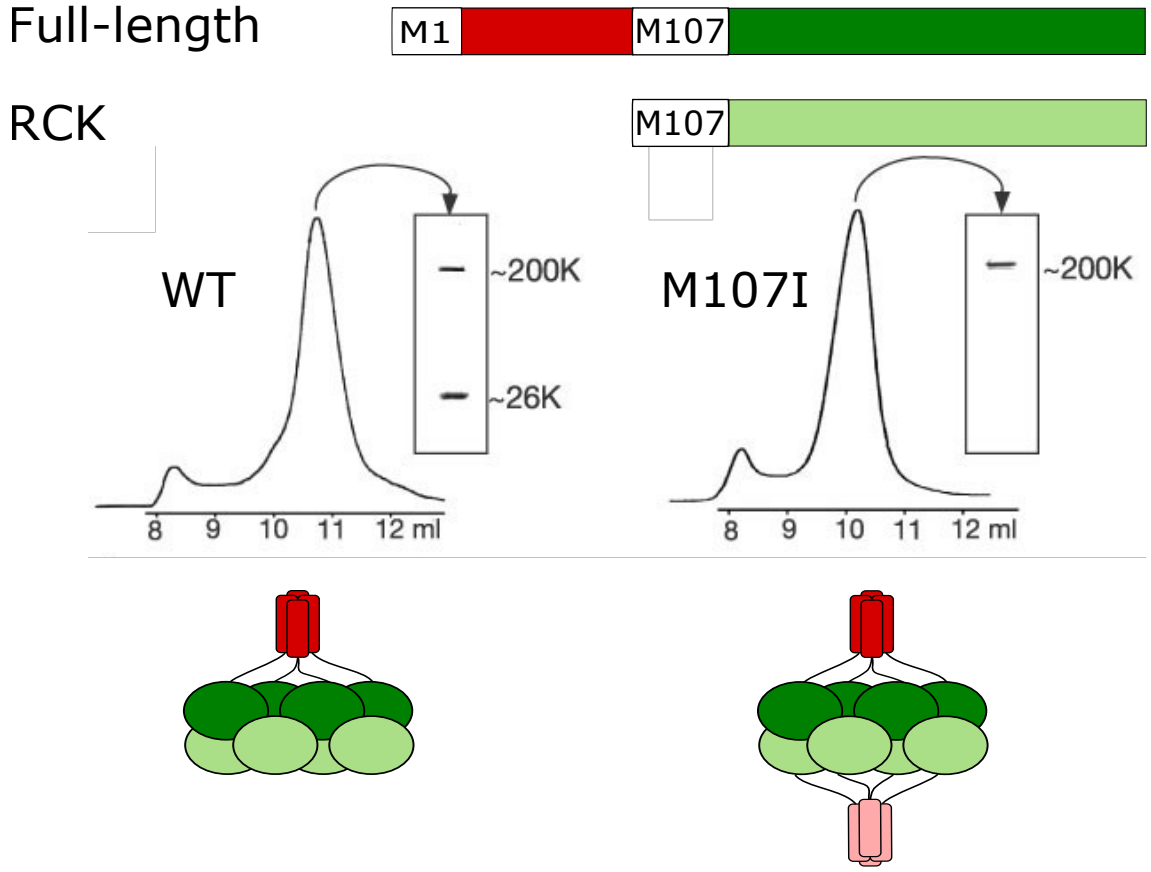

Figure 4.2. Hypothesis on MthK expression and stoichiometry on the first crystal structure of MthK according to the MacKinnon paper.
When expressing the wild-type protein both the full-length protein (top) and the RCK domain (top, below) on its own is thought of to be expressed. These are assumed to form the octameric gating ring. When mutating M107I, it is thought of forming an octamer, as shown on the right as the RCK domain is not expressed. The octamer can form because the RCK domain is also part of the full-length protein. So all contacts are able to form. adopted from reference[375]

4.2 Obtaining the *in vivo* stoichiometry

To study a membrane protein, the expression and purification needs to be optimised. In the case of MthK, purification consists of IMAC through a 6 His-tag, tag removal with the appropriate protease, reverse IMAC followed by a SEC column. More

details on the exact buffers used can be found in the methods section of this chapter. Stoichiometry plays a key role in this section, therefore the molecular weights of different stoichiometries are given in figure 4.3.

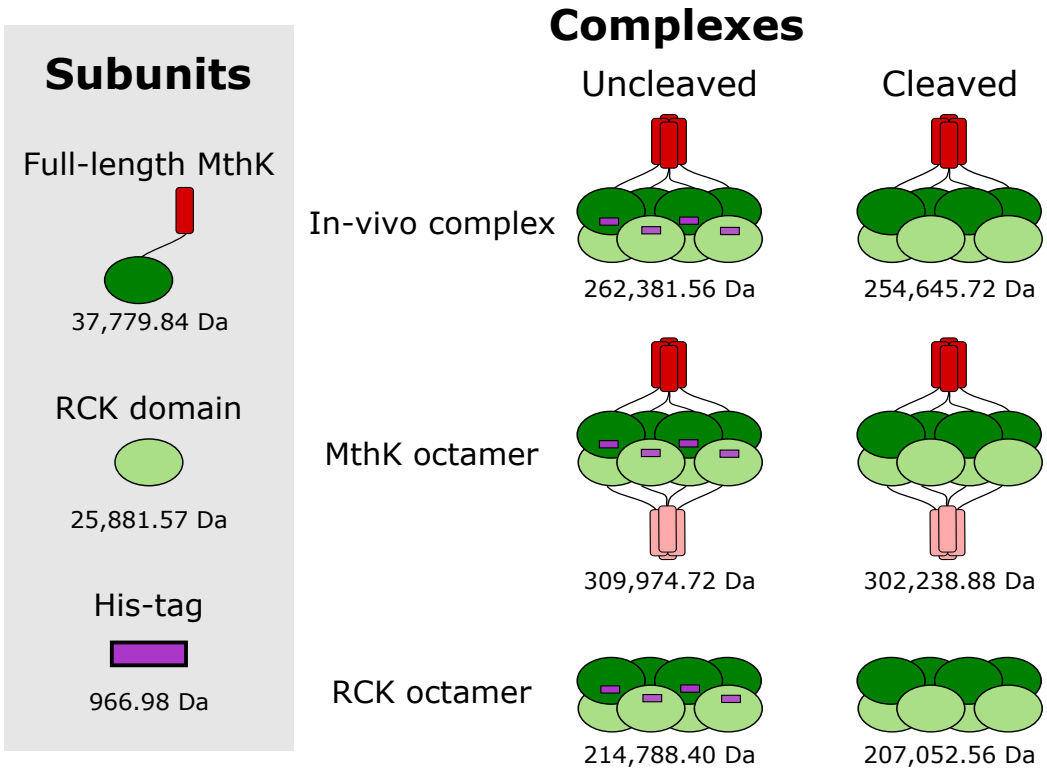

Figure 4.3. Overview of possible stoichiometries for MthK.
Different subunits of MthK are shown in the grey box. Stable complexes with octameric gating rings are shown on the right. Full-length MthK and RCK domains include a C-terminal sequence of leucine, valine, proline and arginine because a thrombin cleavage site. The his-tag has an additional glycine and serine for the same reason.

4.2.1 Different Cells and Vectors

In order to optimise MthK purification, the first step consisted of choosing the appropriate detergent (which is also crucial for nMS)[171]. Therefore, a detergent screen has been performed (figure 4.4A) indicating that many detergents are suitable to purify MthK. Therefore, I finally chose DM since it was previously used to purify MthK[375,378]. Initial constructs made use of a cleavage site for Tobacco Etch Virus

nuclear-inclusion-a endopeptidase (TEV) and TEV as protease coupled with a green fluorescent protein GFP. However, incubation with TEV protease led to protein precipitation (figure 4.4B). Therefore, I switched to a construct containing a thrombin cleavage site as it was used before and omitted a GFP-tag[375].

This construct was used for the expression in BL21 (DE3) cells. The results are shown in figure 4.5 including SEC profile and electrophoresis gel (panels A and B, respectively). The observations are in line with the purification results in the literature[375]. Spectra of MthK are shown in panels C for DM and D for C8E4. While DM appears less delipidating, in both cases the observed mass corresponds to an octamer, along with a dissociated monomer in the case of DM due to higher HCD voltage of 300 V in contrast to the 200 V for C8E4. This is because of the high voltage applied to the DM sample. Both used detergents allow a sable octamer to be observed.

Proteomics analysis of the 26 kDa band (figure 4.5E) indicates the presence of fragments of residues from the TMD. This directly contradicts the hypothesis that that band contains only the RCK domain. This leads to the question whether the RCK domain is not expressed or lost during the purification process. While the latter might be possible, I consider the former more likely. If the RCK domain was expressed, it should copurify because the His-tag is on the C-terminal side of the protein and if co-purified, they would be visible in spectra as sections below of mixing purified RCK and MthK indicate (see section 4.2.4).

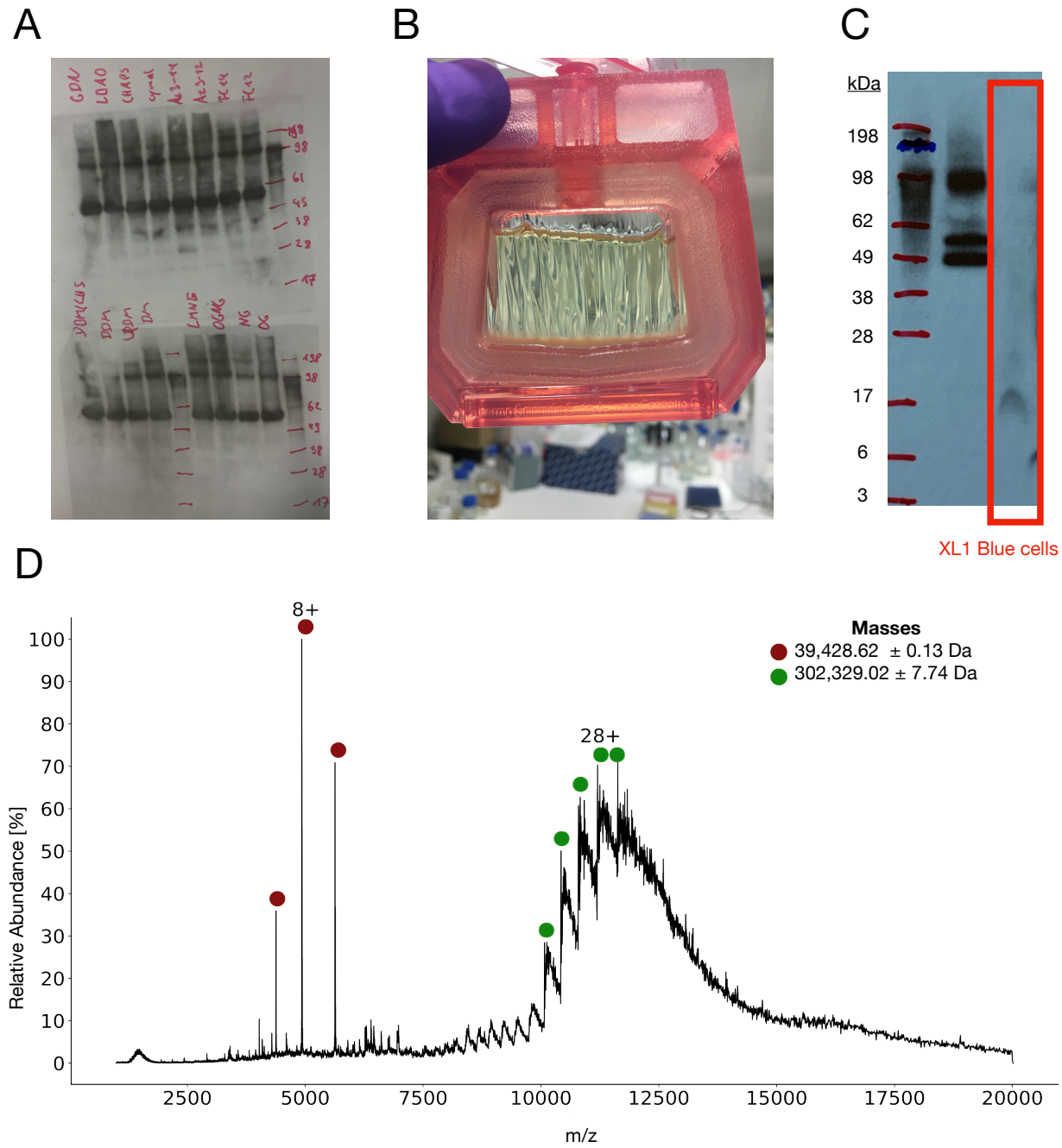

Figure 4.4. Detergent screen, photo of precipitated MthK, western blot of expression in XL1 cells.

A *Detergent screen of MthK.* ***B*** *Precipitated MthK with a TEV cleaving site.* ***C*** *Western blot of MthK expression in XL1 Blue Cells.* ***D*** *Spectrum of MthK expressed in C43 cells. Collected at a capillary voltage of 1.4 kV and a capillary temperature of 100 °C with an HCD voltage of 200 V and IST voltage of 200 V.*

The absence of the in vivo stoichiometry stimulated further optimisation through use of different *E.coli* expression strains. I initially employed C43 (DE3) cells, a strain derived from BL21 (DE3) cells optimised for membrane protein expression

(figure 4.4D)[386]. The resulting spectra are similar to the ones observed following expression in BL21(DE3) cells, showing an octameric stoichiometry but no *in vivo* complex. Therefore, as a final attempt I tried to mimic the approach taken by Zadek and Nimigean which used XL1 Blue cells[378]. These cells, which are more commonly used for cloning than for protein expression, were unable to express the protein as shown in a Western blot (figure 4.4C)[387]. Since there is little difference between cell lines, I continued experiments with BL21 (DE3) cell lines.

4.2.2 Sonication of Lipid Vesicles Mass Spectrometry

As a next step, SoLVe-MS was employed to investigate whether the RCK domain was lost during the purification process. This technique omits any purification process (see section 1.3.1). The results of following the SoLVe-MS protocol on membranes BL21 (DE3) cell overexpressing MthK are shown in figure 4.6. MthK is certainly present, since it has been purified from the same batch of cells for lipid binding experiments (see section 4.3).

While there are two series with masses of $211,733.70 \pm 128.64\,\mathrm{Da}$ (dark green) and $208,639.53 \pm 19.48\,\mathrm{Da}$ (light blue) close to the octamer of RCK, they are too light because in this expression the His-tag has not been cleaved. Similarly, the $102,741.22 \pm 5.9\,\mathrm{Da}$ (grey) and $141,893.97 \pm 16.16\,\mathrm{Da}$ (orange) do not match any plausible MthK stoichiometry. All observed series are most likely other proteins than MthK and SoLVe-MS has not been pursued further in this work. Otherwise, because it is overexpressed and soluble, it should be present if is was dissociating from the complex to a degree. As soluble protein, it would be transmitting well and

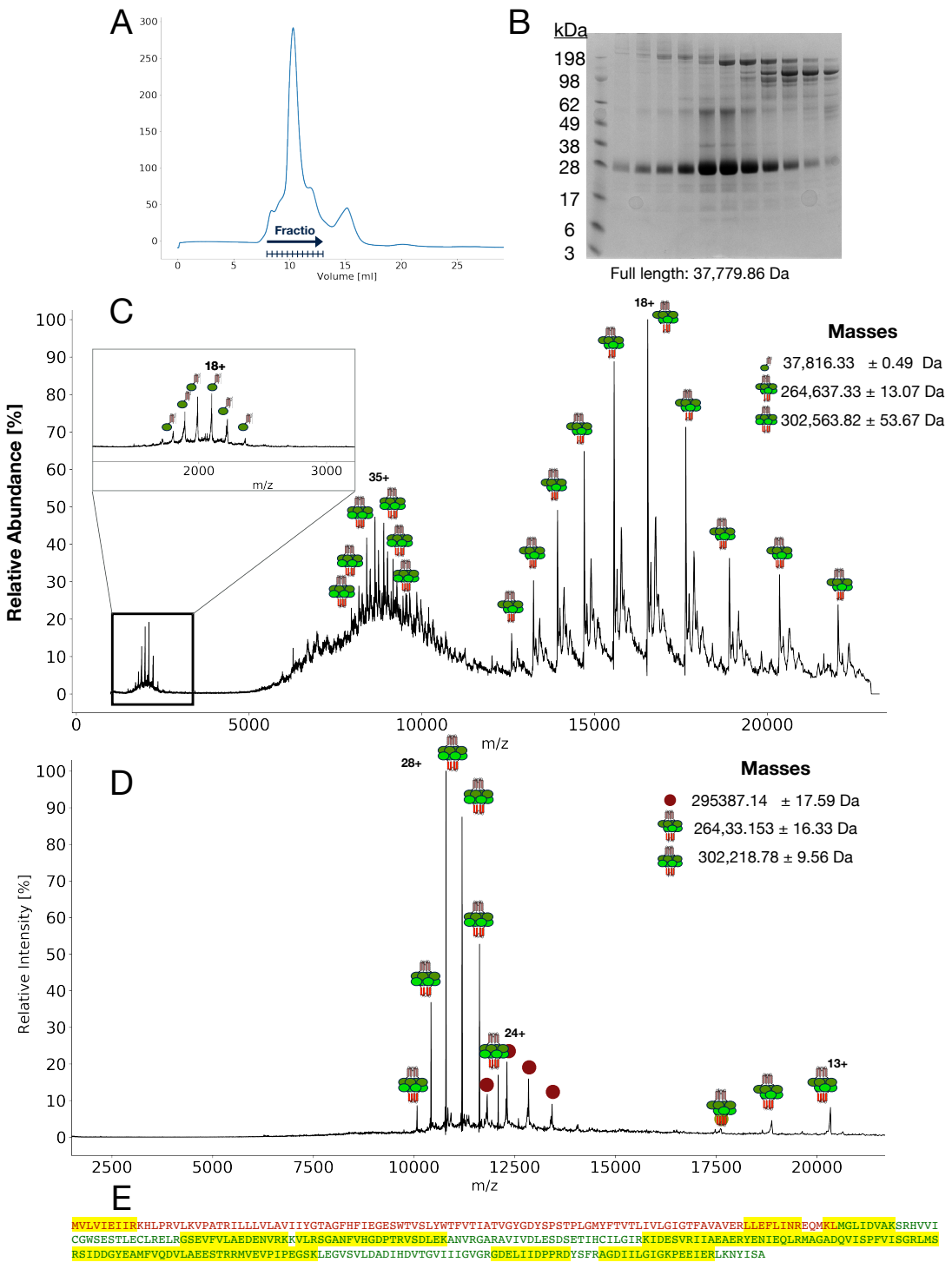

Figure 4.5. Purification and mass spectrometry of MthK purified from BL21 cells.

A *SEC profile and* ***B*** *electrophoresis gels of MthK.* ***C*** *Spectrum of MthK in DM and* ***D*** *C8E4. Collected at a capillary voltage of 1.4 kV and a capillary temperature of 300 °C with an HCD voltage of 300 V and IST voltage of 300 V.* ***E*** *Fragments found (yellow) in proteomics of the 26 kDa band. TMD sequence in red and RCK domain in green. This indicates that the full-length protein also shows in this band.*

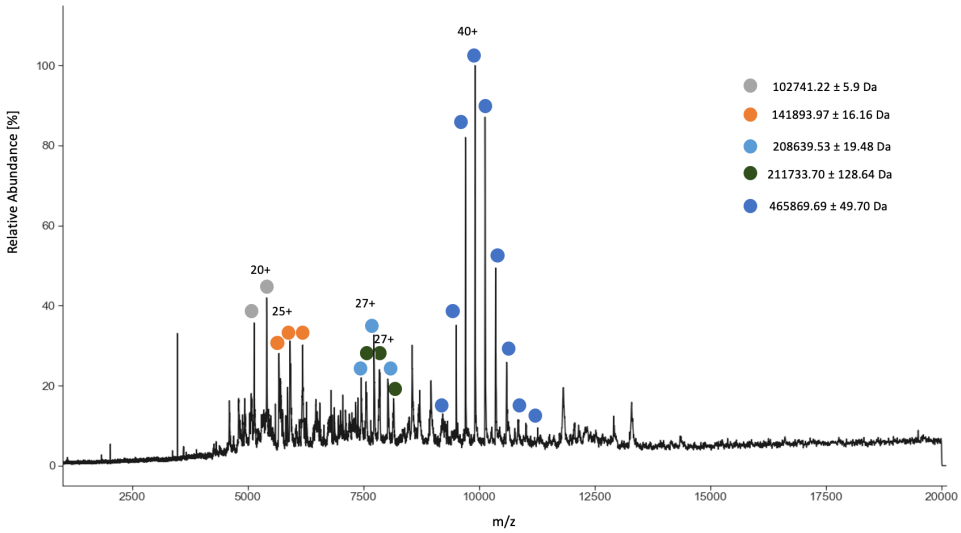

Figure 4.6. SoLVe-MS spectra of BL21 sonicated membranes with overexpressed MthK.
Collected at a capillary voltage of 1.4 kV and a capillary temperature of 300 °C with an HCD voltage of 300 V and IST voltage of 300 V.

therefore be observable in mass spectra.

While SoLVe-MS was unable to solve the riddle of the missing RCK domain conclusively, the complexity of the spectra lead to a search for better analysis tools. Due to a lack of alternatives, I started working on NaViA, which is presented in chapter 5.4 of this thesis. The absence of RCK lead me to consider approaches to add the RCK domain to MthK to form the *in vivo* complex.

4.2.3 Expressing TMD and RCK domain separately

To do so, RCK has been expressed and purified separately in a similar construct to MthK, i.e. with the same cleavage site and 6xHis-tag. The purification protocol is the same as for MthK without the membrane purification. The results of the

purification for the RCK domain are shown in figure 4.7. Panel A and B show similar profiles as the full-length MthK expression. Most of the protein is present in the octameric state which is known in the literature to be highly stable in solution[388].

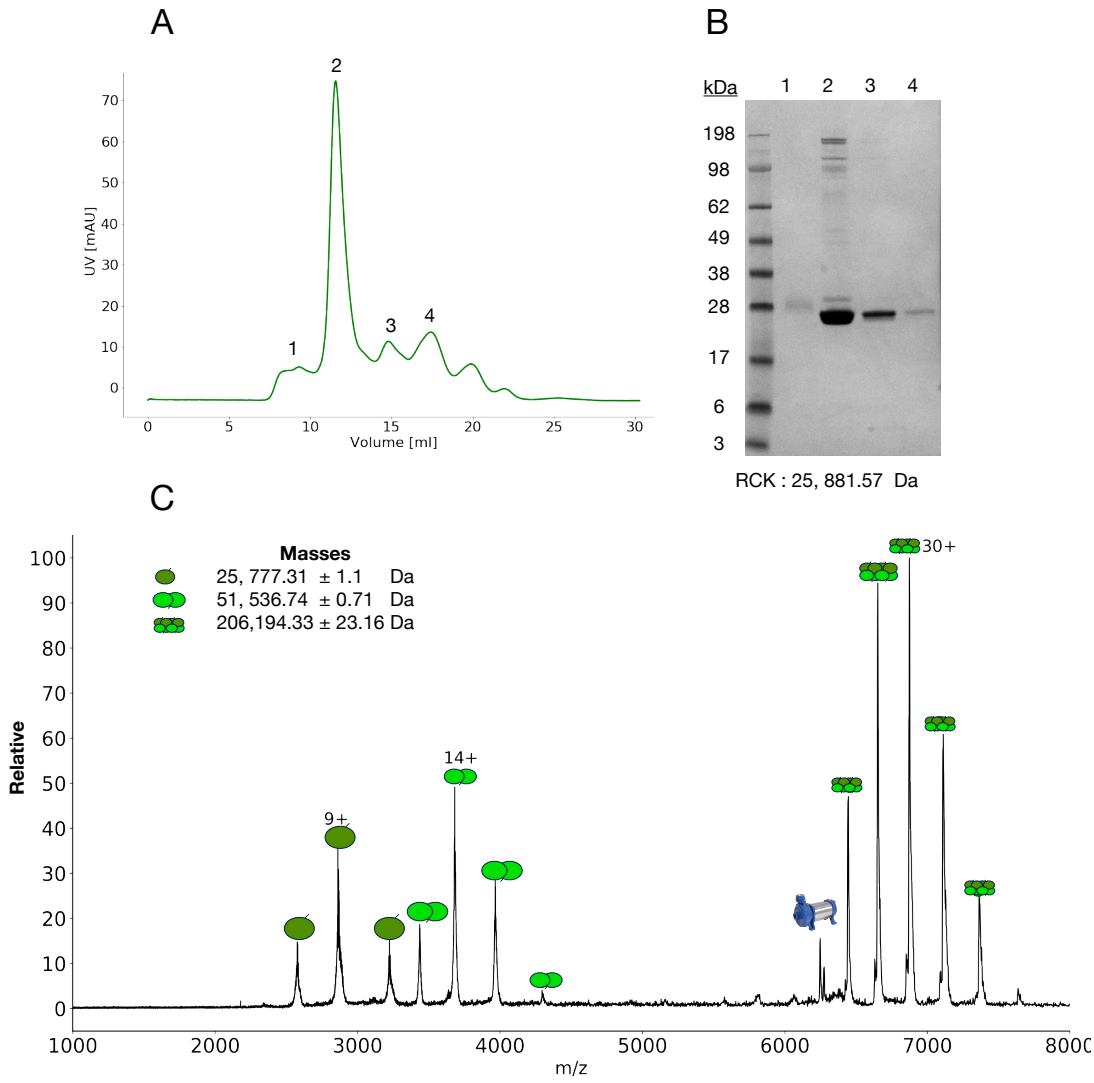

Figure 4.7. Purification and mass spectrometry of TMD of MthK.
A *SEC profile of RCK domain.* ***B*** *Electrophoresis gel of RCK domain.* ***C*** *Spectrum of RCK domain with monomers, dimers and octamers present.Collected at a capillary voltage of 1.4 kV and a capillary temperature of 100 °C with an HCD voltage of 200 V and IST voltage of 200 V. The blue vacuum pump indicates pump noise (around 6,100).*

In addition, the TMD has been expressed and purified separately (figure 4.8). The SEC profile showed an elution of the main peak used for mass spectra collection at around 15 ml (figure 4.8A) with the expected bands matching the molecular weight

of the 12.381 Da. Acquisition of resolved mass spectra required the application of high voltages and lead to a dissociation of the complex. This is not unsurprising because that the TMD is mostly located in the detergent micelle and requires strong collision due to the lack of a soluble domain. The dissociation of the complex shows that MthK is mostly stabilised by the non-covalent bonds in the octameric gating ring.

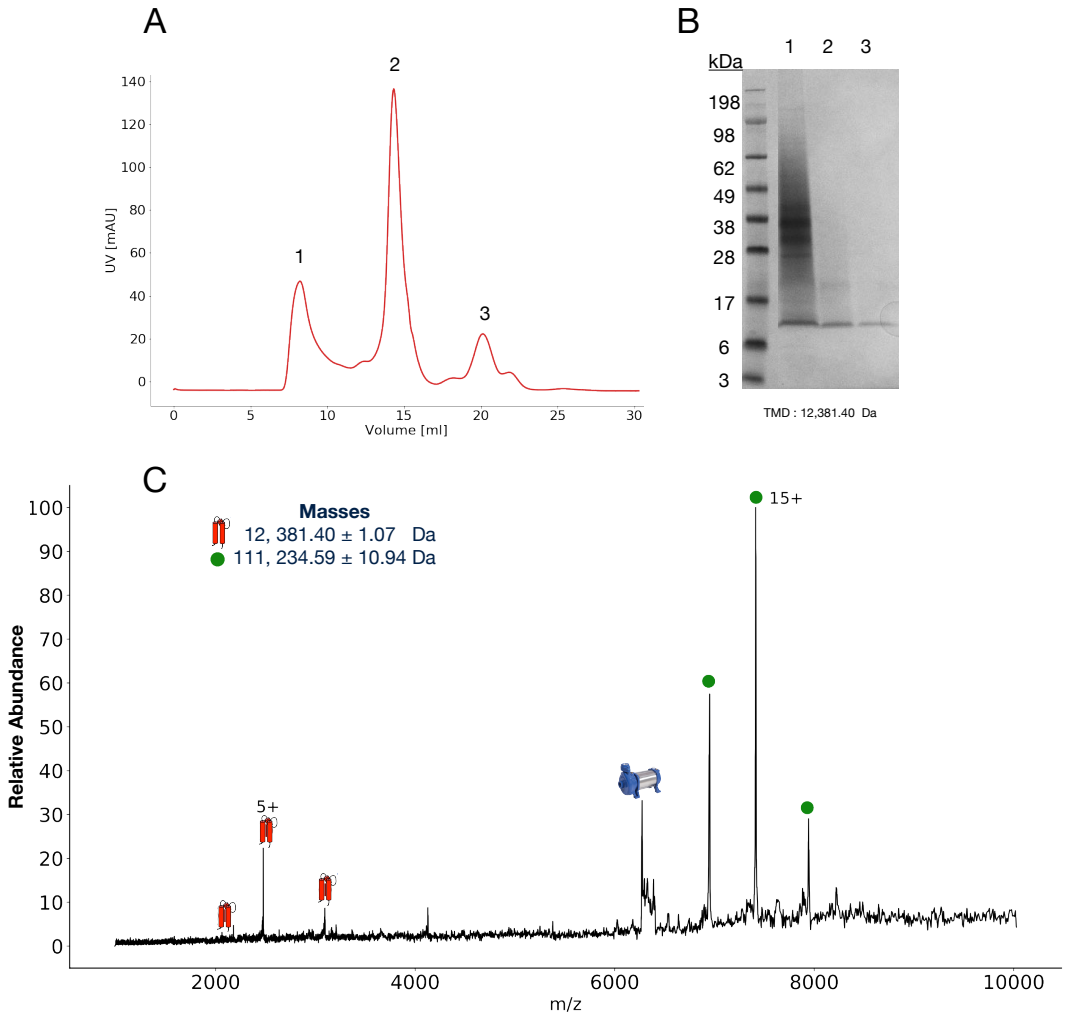

Figure 4.8. Purification and mass spectrometry of TMD of MthK.

A *SEC profile of TMD.* ***B*** *Electrophoresis gel of TMD.* ***C*** *Spectrum of TMD with monomers (red TMD) and impurity (green dot) present. Collected at a capillary voltage of 1.2 kV and a capillary temperature of 300 °C with an HCD voltage of 100 V and IST voltage of 300 V. The blue vacuum pump indicates pump noise (around 6,100).*

4.2.4 Mixing purified proteins

The next step consisted of mixing purified RCK with purified MthK and yielded the results are shown in figure 4.9 for a 1:1 mixture of RCK:MthK in panel A and a 1:5 mixture in panel B. This allows to estimate the relative transmission efficiency into the mass spectrometer. In the 1:1 mixture, only RCK is visible while a ratio of 1:5 is required to make MthK visible, too. This matches the expectation because RCK is a soluble protein which can be desolvated and transmitted more easily. Also, it confirms the hypothesis that RCK should be visible if it was present in the complex. Therefore, I conclude that it is necessary to add RCK back into cells for forming the *in vivo* complex. Mixing readily expressed proteins at the mass spectrometry stage is impossible because of the stability of the octameric gating ring[40]. Therefore, it was co-purified.

4.2.5 Co-purifying MthK and RCK domain

When MthK is expressed in cells it is unlikely to directly form an octamer given the position of the second TMD (see figure 4.2 - bottom right). Thus, it most probably forms tetramers as this matches its *in vivo* stoichiometry. Due to the C-terminal RCK domain in the full-length complex, it binds to another exposed tetramer during the purification process. This matches the expectation in the absence of pure RCK domains because the RCK domain is a part of the full-length complex. This formation of the octamer is implausible to take place in cells or in purified membranes due to the large size of these patches around tens of microns[389] which is orders of magnitude larger than proteins. Therefore, the most plausible point in time would

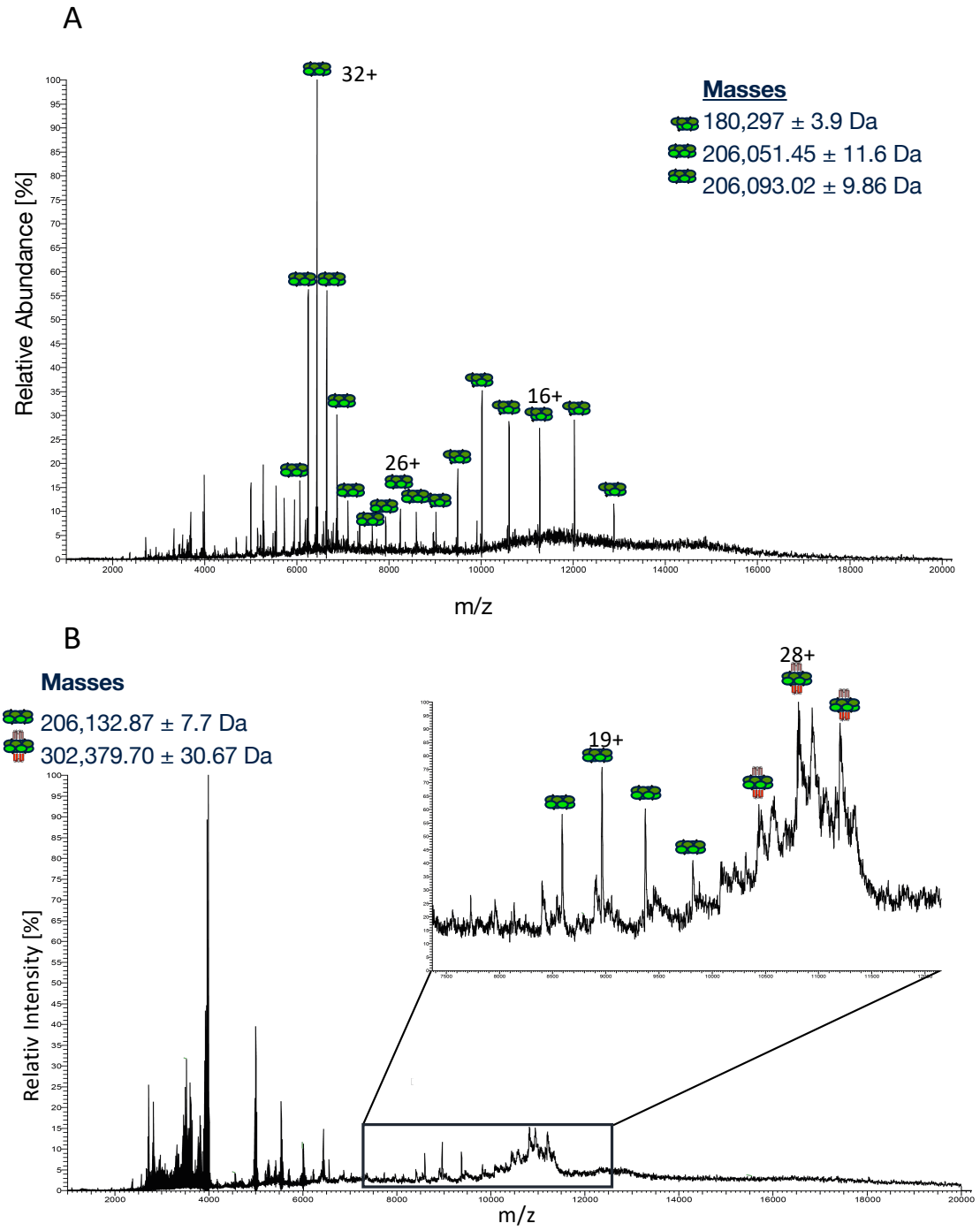

Figure 4.9. Mixing of purified MthK and RCK for two ratios.

A *RCK:MthK 1:1 mixture.* ***B*** *RCK:MthK 1:5 mixture. Collected at a capillary voltage of 1.2 kV and a capillary temperature of 100 °C with an HCD voltage of 200 V and IST voltage of 250 V.*

be the membrane solubilisation stage.

To prevent this from happening, I attempted to mix purified RCK domains with cleaved His-tags with MthK-containing membranes. The 6 His-tag was cleaved beforehand to avoid the formation of his-tagged RCK octamers being purified instead of MthK-bound RCK domains. Since RCK is soluble, the his-tagged RCK octamer signal would overlay MthK as soluble proteins have substantially higher yields. The results of this expression including gel and SEC profile are shown in figure 4.10.

While SEC profile (panel A) and gel (panel B) look similar to a standard purification in BL21 (DE3) cells, the mass spectra show that some RCK could be bound although it seems to form dimer and trimers. In addition, the octamer is visible at 300 V in the HCD cell shows that the octamer is present (blue). There is also a peak series close to the expected mass of the *in vivo* complex although the mass is around 8 kDa lower than expected. However, this is not conclusive and the *in vivo* complex could not be considered stable. To address the issue of stability which might be caused by stable RCK octamers not dissociating enough to form in *in vivo* complexes, I attempted to go a step further and express both subunits together in the same cells.

4.2.6 Duet Vector for Expressing MthK and RCK

Many protein form complexes of different subunits. To overexpress these types of protein, specific vectors with multiple expression sites have been developed. Since previous attempts of obtaining the *in vivo* stoichiometry were mostly unsuccessful, I set out to employ pETDuet™-1 vectors (Merck). Here, His-tagged full length MthK has been inserted in site 1 and Flag-tagged (sequence DYKDDDDK) RCK domain was inserted in site 2[390,391]. Western Blot targeting His and Flag indicated that both

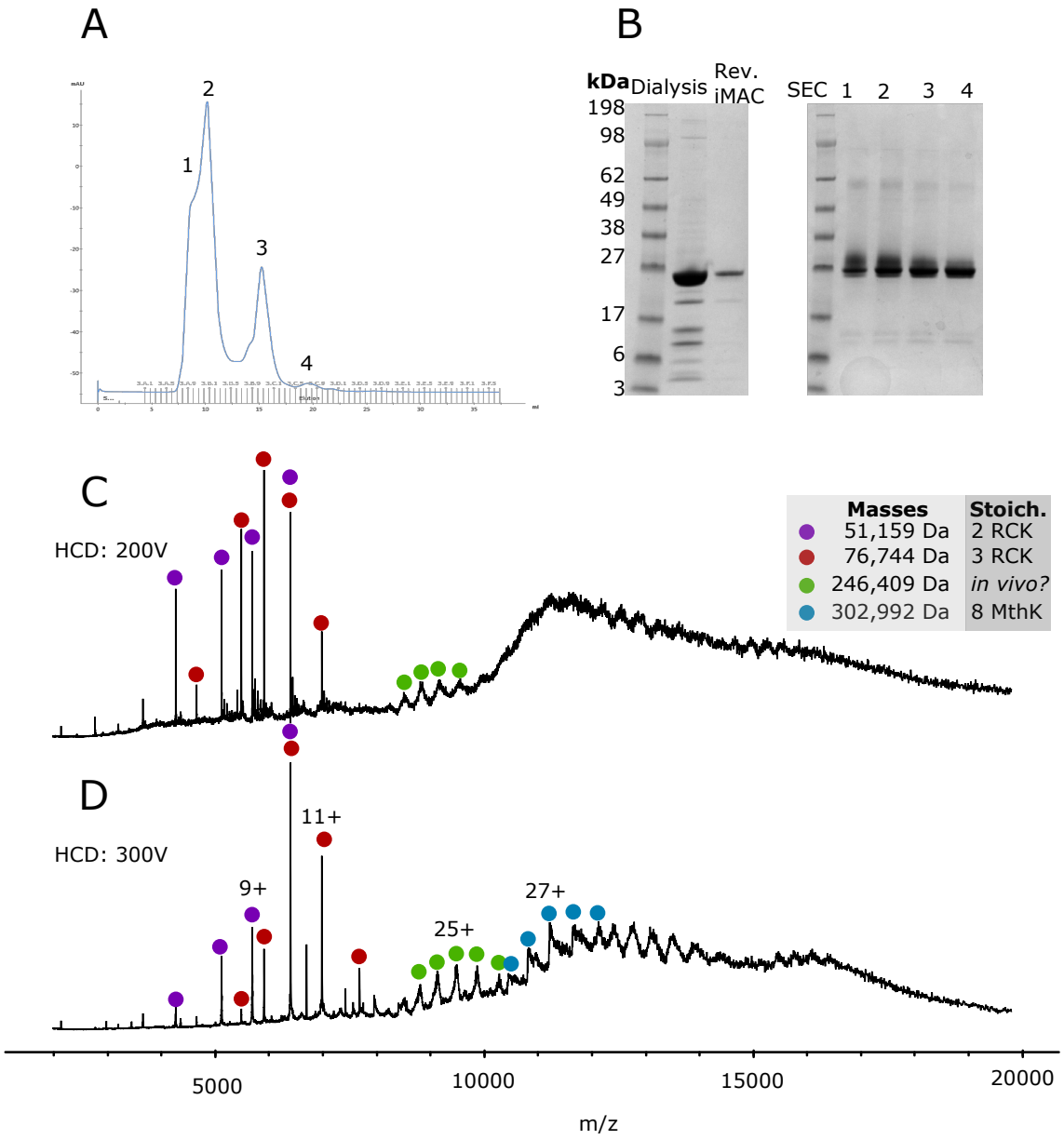

Figure 4.10. Purification and mass spectrometry of MthK copurified with RCK.

A *SEC profile of the co-purification.* ***B*** *Electrophoresis gel of the purification.* ***C*** *Mass spectrum with a HCD voltage of 200 V and* ***D*** *300 V. Collected at a capillary voltage of 1.2 kV and a capillary temperature of 100 °C and IST voltage of 300 V.*

proteins were expressed (figure 4.11A). Thus, I set out to purify membranes using DM detergent as before but without cleavage of the tags. Following purification, native mass spectra have been collected in C8E4 as detergent. Native mass spectra have been collected in C8E4 as detergent.

The co-expression of both subunits allows to observe the *in vivo* stoichiometry for

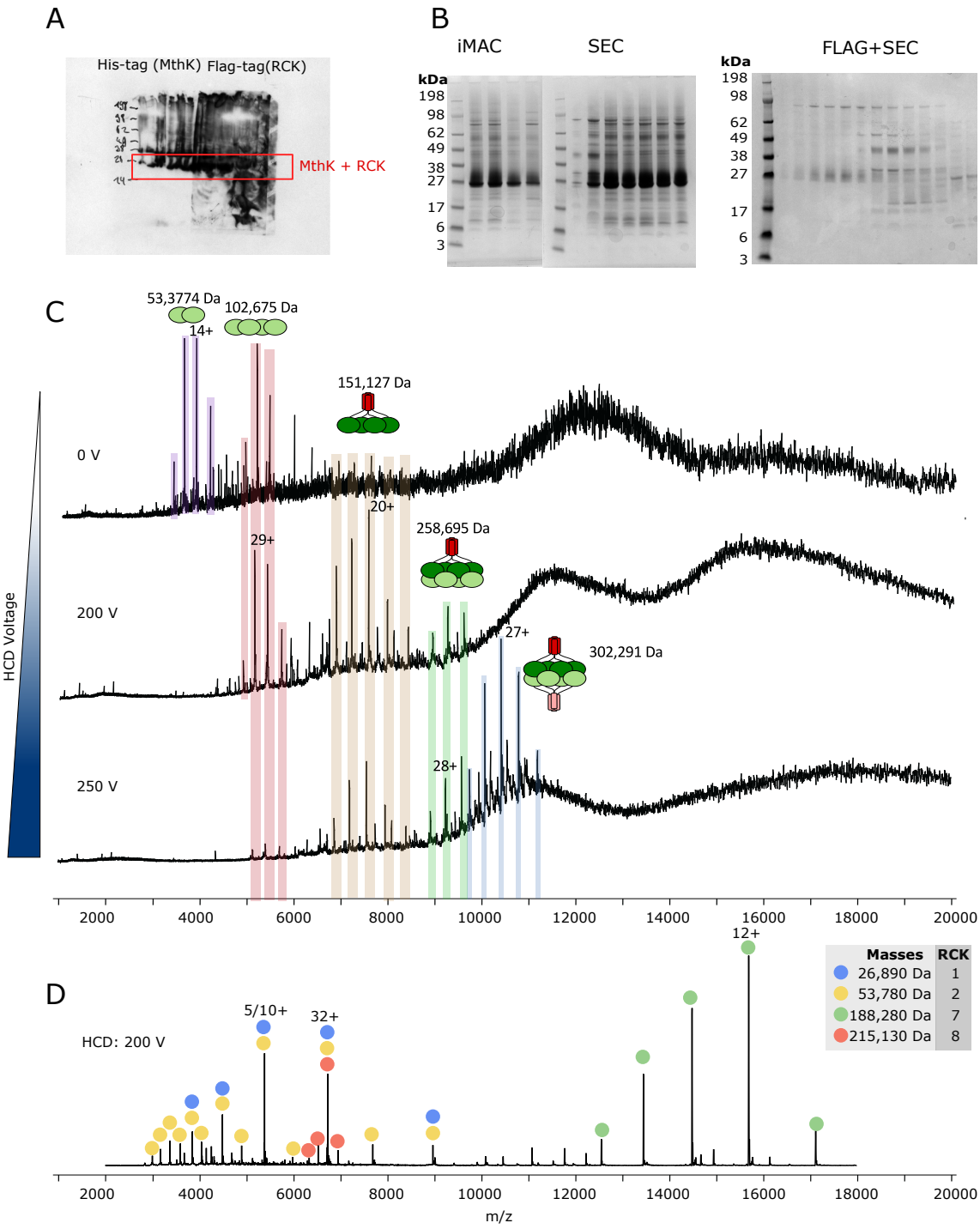

Figure 4.11. Western blot, purification and mass spectra of MthK and RCK expressed in Duet vectors.

A *Western blot of cell lysate for His- and Flag-tag detection.* ***B*** *IMAC and SEC gels of the co-expressed proteins (left) and a SEC gel after a Flag-affinity purification.* ***C*** *Mass spectra of co-expressed MthK and RCK at different HCD voltages.* ***D*** *Spectra of MthK and RCK after Flag purification and SEC column at different HCD voltages. Collected at a capillary voltage of 1.2 kV and a capillary temperature of 100 °C with different HCD voltage and IST voltage of 300 V.*

similar gels profiles as previous purifications of MthK alone (panel B). While it was possible to observe the *in vivo* stoichiometry of MthK in nMS experiments (panel C), it was not possible to achieve a state in which only the *in vivo* stoichiometry present despite optimising voltages at different stages. Attempts to use a Flag-affinity purification resulted in only the RCK domain being visible (panel B on the right and panel D). In addition, many previous studies on MthK uses at its octameric stoichiometry as a basis[376,385]. Therefore, I used the expression in BL21 (DE3) cells with the purification described in the methods section of this chapter for further investigations.

4.3 Lipid-binding of the MthK channel

MthK is an extensively studied system for calcium binding and the gating mechanism as indicated in section 4.1.3. The lipid-composition of membranes has not attracted much attention except from a computational study by Gu and de Groot which probed the effect of different lipid classes[385]. Unfortunately, in the lipids employed in that study (phosphocholine and cholesterol) are all present in eukaryotic membranes but not in bacterial membranes. Despite the non-biologically relevant combination of lipids and protein, Gu and de Groot find that lipids tails modulate the equilibrium of closed and open state of MthK. This suggests a tight lipid regulation of MthK. Therefore, I an investigated binding features of bacterial lipids bind to MthK.

4.3.1 Endogenous Lipids and Lipidomics

The first indication of tightly bound lipids can be derived from a partially delipidated spectrum (figure 4.12A). The peak series indicated endogenous lipids bound to MthK. While a mass of 3,680 Da corresponds to LPS, which is often copurified to bacterial membrane proteins of the inner membrane, the lipids of the masses of 658.24 Da to 835.24 Da and 1,389.05 Da to 1,407.87 Da are likely to be biologically relevant. The former corresponds to either PE or PG lipid bound, the latter could either be CDL or a combination of two phospholipids.

For a closer investigation, lipidomics of the same sample has been carried out. The elution at minutes 18-22 minutes corresponds to phospholipids and is shown in figure 4.12B while the elution at minutes 26-32 (figure 4.12C) corresponds to CDL. It is visible that both types of lipids are present in the sample and are therefore candidates for lipid binding to MthK. However, lipidomics probes the whole content of lipids present in the sample not distinguishing between types of lipid-binding. Nonetheless, the spectrum in figure 4.12A shows a higher abundance of the 1.4 kDa species bound to MthK, compared to lower MW species. CDL occurs in lower concentration in bacterial membranes, yet the 1.4 kDa peak in 4.12A is more intense, hence suggesting that CDL may be bound to MthK.

4.3.2 Lipid Titration Experiments

To probe this more closely, a titration experiment of CDL to MthK has been conducted. The concentration of MthK was kept at 5μM while the concentration of

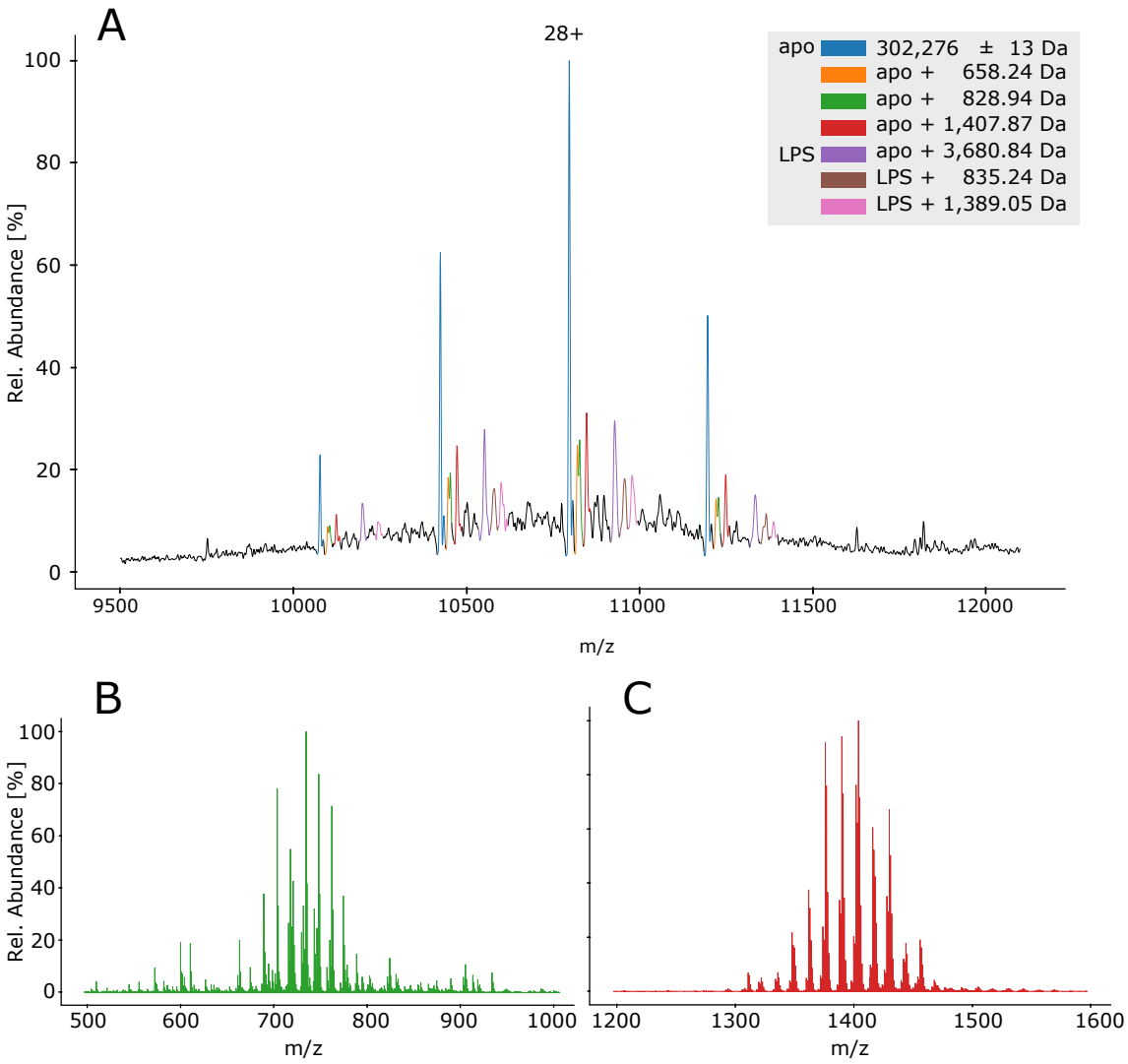

Figure 4.12. Apo spectrum of partially delipidated MthK and lipidomics

A *Spectrum of partially delipidated MthK octamer. Multiple series possibly corresponding to lipids are visible.* ***B*** *Spectrum of lipidomics data eluded minutes 18-22.* ***C*** *Lipidomics data eluded minutes 26-32.*

CDL was varied from $1.25 - 40\mu$M.

The resulting spectra are shown in figure 4.13. Here, up to 8 CDL molecules bound to MthK can be observed in the highest CDL concentration. This matches the expectation, since an octameric stoichiometry would yield 8 potential binding sites. The fact that the maximum number of lipids bound has not been reached below 40μM makes it impossible to determine K_D. However, I then compared CDL binding

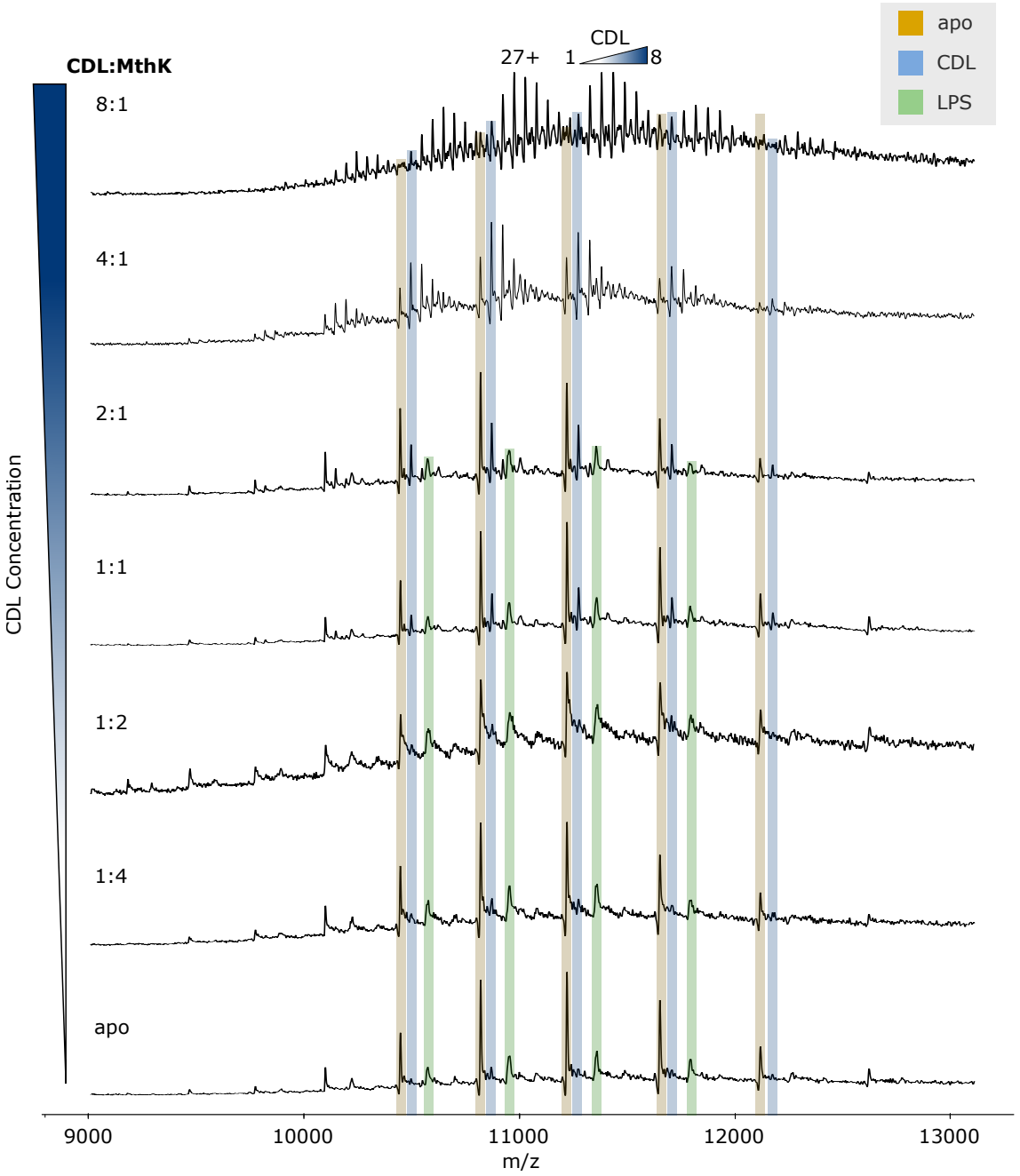

Figure 4.13. Titration of CDL binding to MthK

Spectra of different concentrations of CDL for 5μM MthK. LPS and CDL binding is visibleSpectra have been collected at a in-source trapping energy of 200 V and an HCD cell voltage of 200 V.

to the negatively charged PG at 40μM concentration with 5μM protein concentration (figure 4.14).

While up to 8 CDL appear bound to MthK (figure 4.14A), only 3 PG molecules could be observed (figure 4.14B). This observation is confirmed in a quantitative

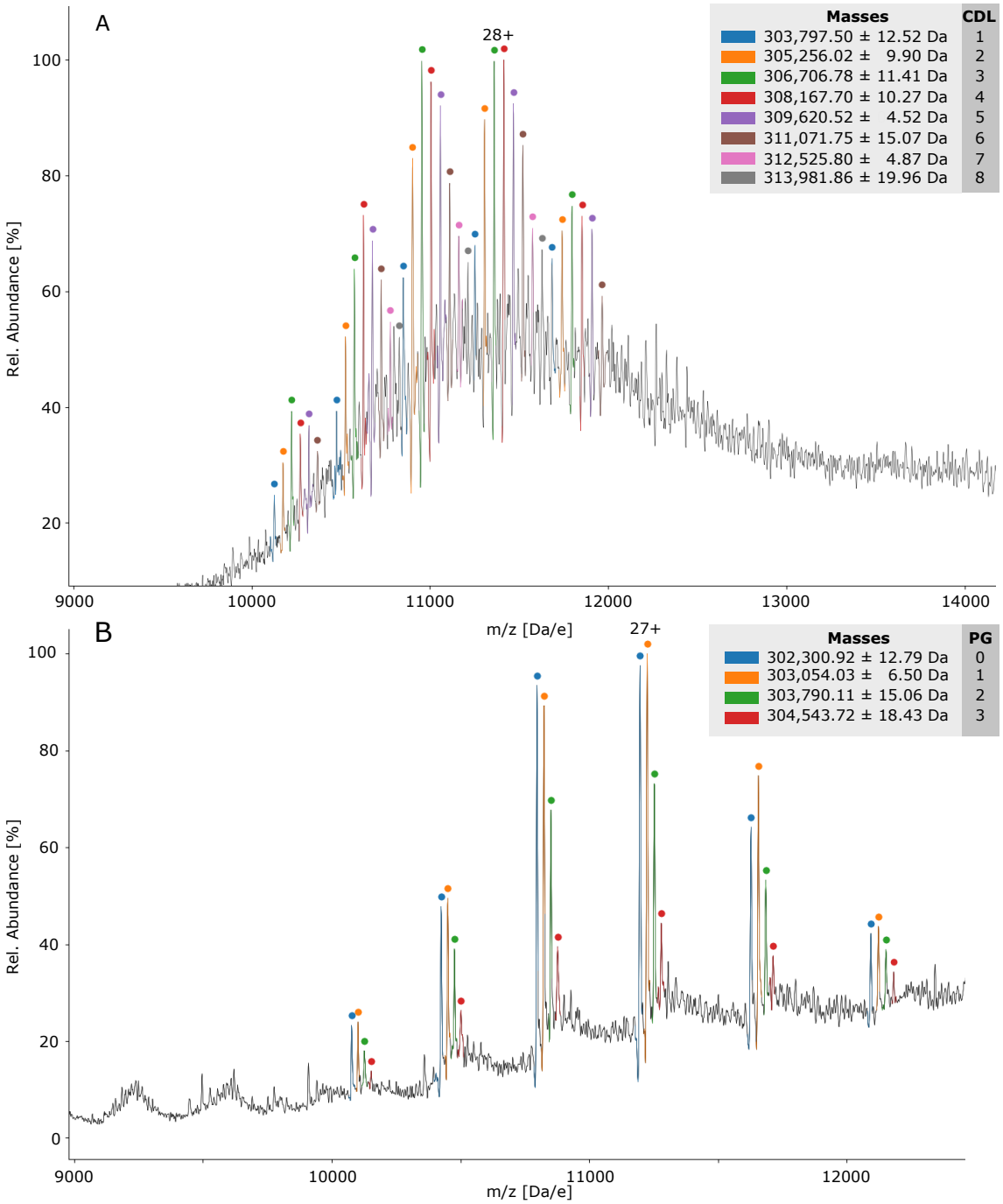

Figure 4.14. Spectra of CDL and PG binding to MthK

A *Spectrum of CDL binding to MthK.* ***B*** *Spectrum of PG binding to MthK. Both spectra have been collected at a in-source trapping energy of 200 V and an HCD cell voltage of 200 V.*

comparison of the weighted average number of lipids bound to MthK from three repeats at the same conditions yields

$$CDL: \quad 2.45 \pm 1.12$$

$$PG: \quad 1.33 \pm 0.04.$$

This data shows there is a clear difference in binding affinity with CDL binding stronger than PG. However, given the number of binding sites on MthK due to the octameric stoichiometry, , we cannot exclude a contribution from non-specific binding. In addition, the binding site is unknown. These questions are addressed in the next section through CGMD.

4.4 Computational Studies of Lipid-binding to MthK

The binding location and specificity of lipid binding to MthK have been addressed by CGMD. The starting point of such simulations is the selection of a structure for simulations. I based my model on the structure published by Jiang *et al.* from the MacKinnon group[375]. Here, the second transmembrane of the octamer domain is removed to mimic the *in vivo* stoichiometry with four soluble RCK domains bound and missing regions are added though Modeller. This corresponds to move from the crystallised M107I structure (see figure 4.2 - right) to the WT structure (see figure 4.2 - left). The resulting structure is shown in figure 4.15A-B. With a setup similar of a POPE:POPG:CDL 70:20:10 bilayer and a simulation time of $5\mu s$ this setup is similar to the one used for SemiSWEET in the previous chapter. Although the simulation box is larger to accommodate full MthK, which is around 15 times larger

than SemiSWEET (with regard to mass).

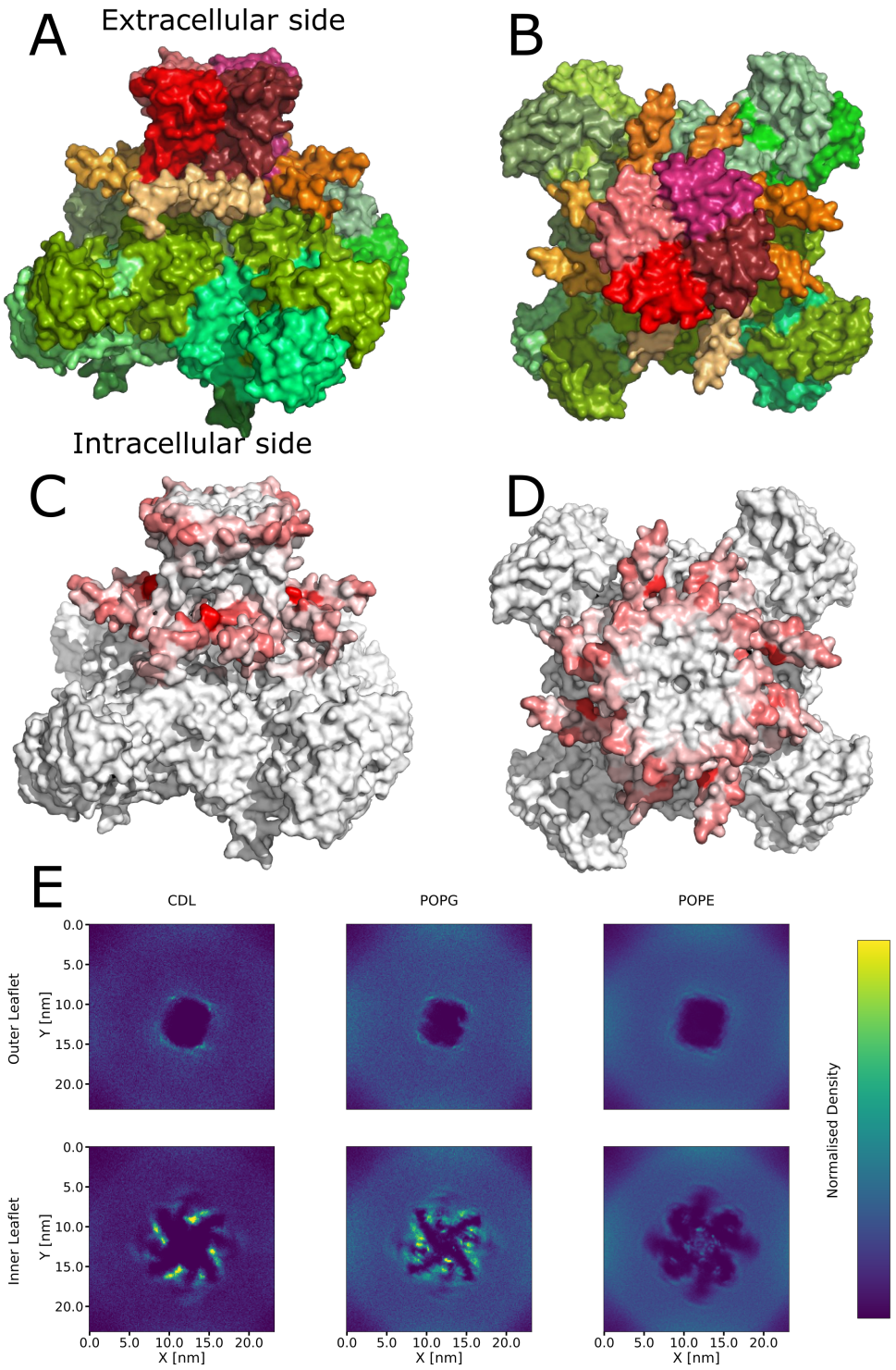

Figure 4.15. Structure and lipid contacts of MthK

A *Side-view of structure of MthK with the TMD in red shades and the RCK domains in green shades.* ***B*** *View from the extracellular side. Sections added by Modeller are shown in orange.* ***C*** *CDL contacts mapped onto MthK (side-view) and* ***D*** *view from extracellular side.* ***E*** *The location of CDL, POPE and POPG is shown for extracellular (outer) and intracellular (inner) leaflets of BK. Colours are rescaled by the concentration of lipids.*

To see whether there are specific binding sites, a heatmap of the location of different lipid headgroups has been created on a 200x200 grid for binning around centered and rotated trajectories of MthK in a bilayer (figure 4.15)C-E. In order to account for different lipid concentrations, the density in the bins is divided by the concentration of the lipid.

The heatmap of different bilayers shows that POPE is not preferentially found near MthK, while both POPG and CDL are found near MthK. POPG shows a spread out density around MthK in the lower leaflet, while CDL appears more localised. This indicates specific binding of CDL while POPG is less specific. Both lipids bind only in the intracellular leaflet while the extracellular leaflet does not show any specific or unspecific preference. Taken together, these observations are in line with the experimental findings that CDL binds stronger and specific while POPG is not binding strongly because it an annular lipid.

The binding site of CDL is located between the amphiphilic helix and the linker region between TMD and RCK domain. A plot of contacts mapped onto the sequence (figure 4.16) reveals that residues from different regions are involved in binding (H11, R101). Following from this observation is the question what drives the CDL binding. One possibility could be the amphiphilic helix at the N-terminal side. This helix has been implied in desensitivation of MthK through a ball-and-chain mechanism[238]. This lead to the hypothesis that the binding is coordinated by the amphiphilic helix.Therefore, a new set of simulations has been performed of MthK without the positions 1-18. The structure is this construct, $\Delta 1-18$ MthK shown in figure 4.17A and B.

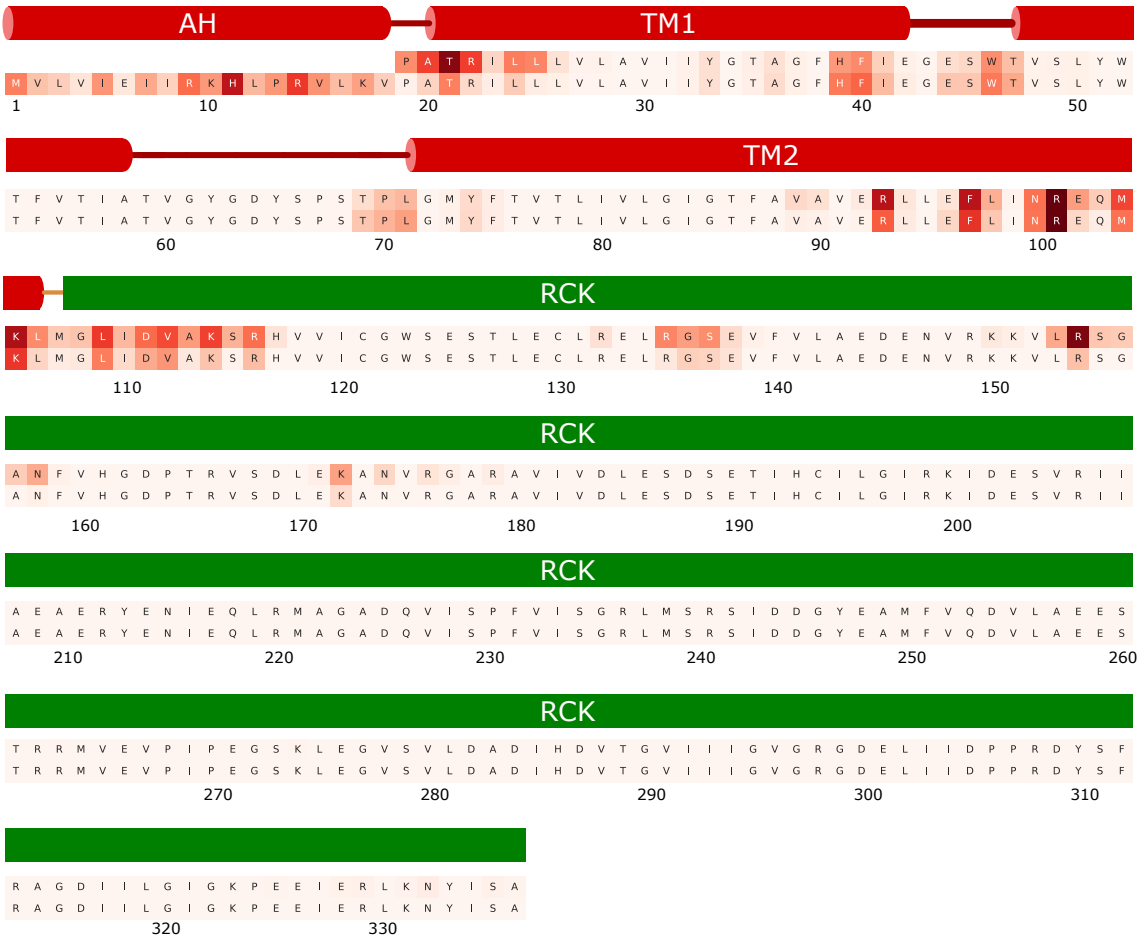

Figure 4.16. Normalised contacts of CDL to residues of MthK

Contacts are shown for the $\Delta 1-18$ *MthK in in side-view and for full-length MthK in the lower row.*

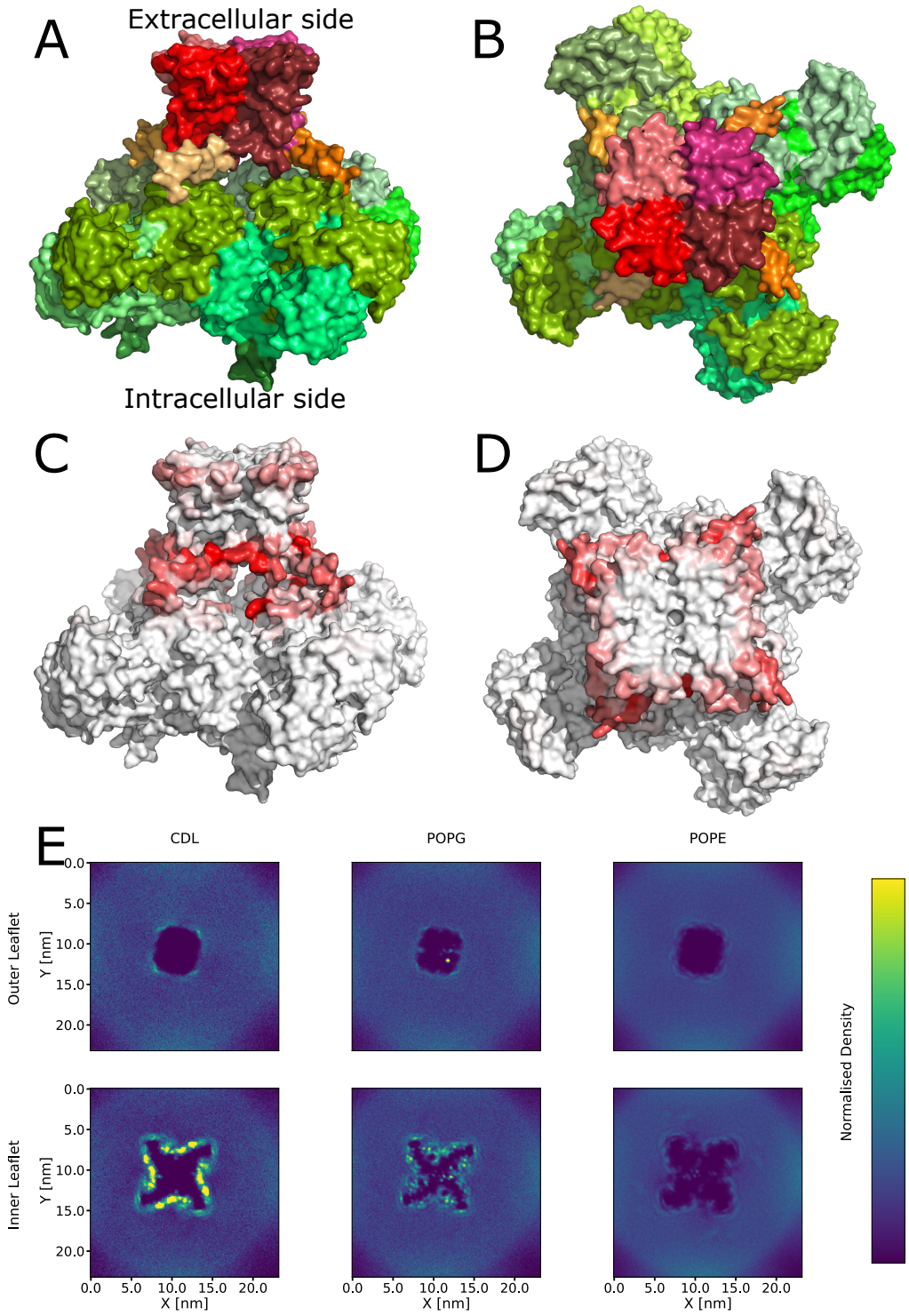

Figure 4.17. Structure of* $\Delta 1-18$ *MthK used in simulations

A *Structure of MthK with the TMD in red shades and the RCK domains in green shades in side-view and* ***B*** *from the extracellular side. Sections added by Modeller are shown in orange.* ***C*** *Contacts mapped onto* $\Delta 1-18$ *MthK in side-view and* ***D*** *from extracellular side.* ***E*** *The location of CDL, POPE and POPG is shown for extracellular (outer) and intracellular (inner) leaflets of BK. Colours are rescaled by the concentration of lipids.*

These simulations are performed for five microseconds, same as the full MthK simulations. To investigate lipid binding location, a heatmap of the location of the headgroups and a structure with contacts mapped are shown in figure 4.17C-E. Both indicate that the binding site is preserved without the amphiphilic helix. This is confirmed in the sequence plot in figure 4.16. Most residues forming the contact to CDL are similar for both versions of MthK. A noticeable difference is the involvement of residues from the RCK domain in Δ1-18 MthK. The reason for this is, that they become exposed because of the missing helix.

4.4.1 Comparing CDL Binding to MthK to Eukaryotic Homologues

CDL binds to MthK in the intracellular leaflet of the membrane and shows binding to the amphiphilic helix at the N-terminal side of MthK and the linker region to the RCK domain. The former invites the comparison to PIP_2 the BK channel since PIP_2 is known to activate the BK channel[82]. To investigate the homology, PIP_2 binding to the BK channel (figure 4.18A-B) has been simulated in a POPC bilayer in a 10:90 ratio. Notably, PIP_2 has also been added in the extracellular leaflet despite being only present in the intracellular leaflet in nature. The reason is to mimic the situation in MthK with CDL. Like MthK, five repeats of five microseconds have been performed.

Heatmaps for lipid binding locations are shown in figure 4.18B-E. Unlike the MthK, the are various different binding location for PIP_2 in both bilayers. While the binding in the extracellular leaflet has little biological relevance, the contacts at

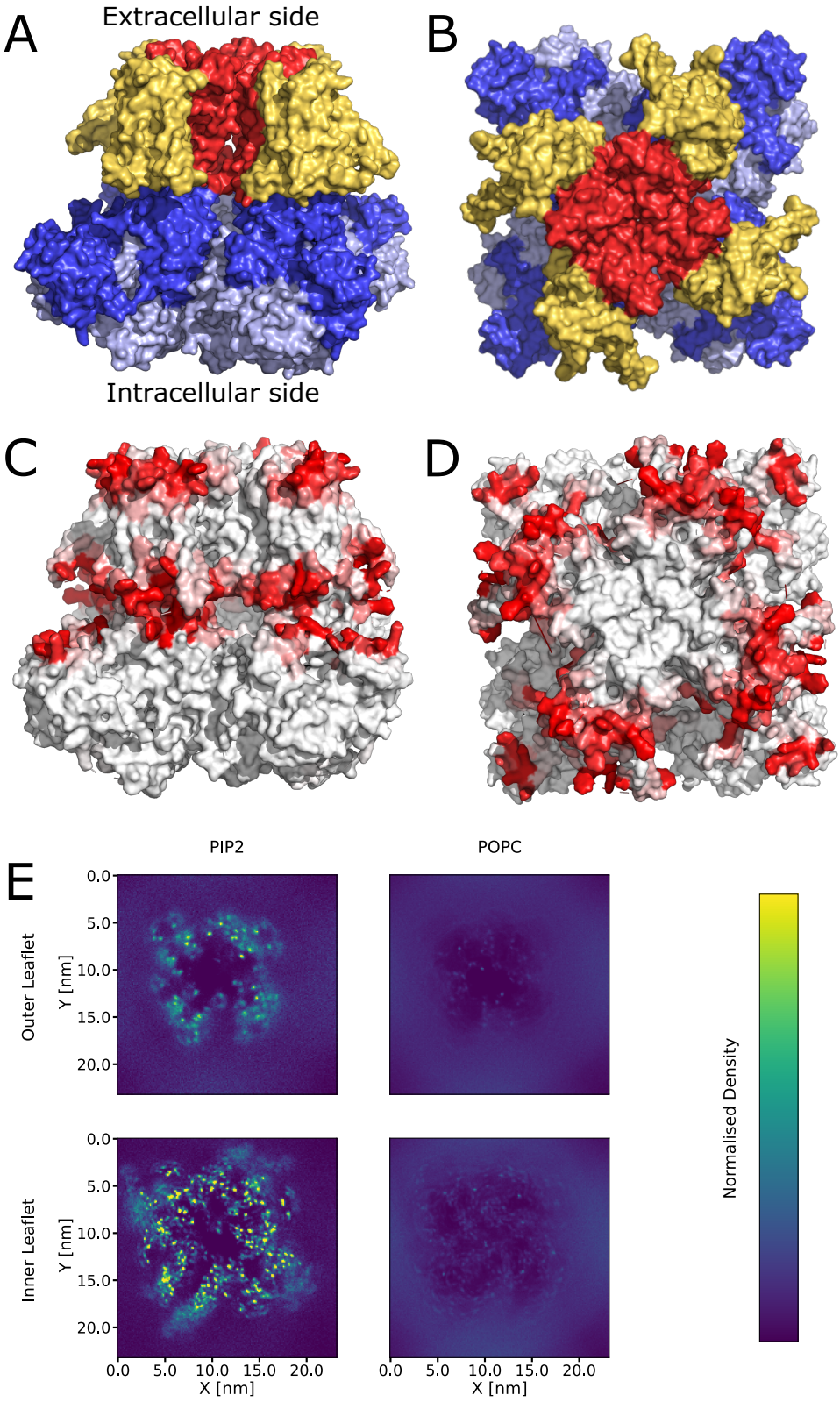

Figure 4.18. Structure of BK used in simulations

A *Structure of BK with the TMD in red shades, the VSD in yellow and the RCK domains in blue shades in side-view and* ***B*** *from the extracellular side.* ***C*** *Contacts mapped onto BK in side-view and* ***D*** *extracellular view.* ***E*** *The location of PIP_2 and POPC is shown for extracellular (outer) and intracellular (inner) leaflets of BK. Colours are rescaled by the concentration of lipids.*

the intracellular leaflet show that most binding occurs near the VSD of the BK channel. This behaviour is not unexpected given that PIP_2 binding sites have been identified in VSD of K_V channels[77–79]. Given the diversity of K_V channels and in PIP_2 interaction (see section 1.2) a comparison to these channels would be an interesting project for future research.

While the VSD binds PIP_2, it also forms a steric hindrance for PIP_2 reaching the pore domain which is also present in MthK. Therefore, I performed simulations of $\Delta 1 - 115$ BK, the BK channel without the VSD (position 1-115) with a structure shown in figure 4.19A-B to establish a closer connection between the homologues.

The results of these simulations are shown in figure 4.19C-E mapped onto the structure and as a heatmap. Both show no specific binding site with PIP_2 binding to almost all locations. Possibly the concentration of 10% could be a reason. Also, PIP_2 tends to bind strongly to proteins in CGMD due to its high negative charge. Therefore, these simulations were unable to resolve whether the CDL binding site is conserved in eukaryotic homologues. However, recent structures of the BK channel with a $\beta 4$ show that these domains would block the space between the VSD[392]. So the binding site might not be accessible for lipids *in vivo*.

4.5 Conclusions and Future Directions

In this chapter the integration of simulations and mass spectrometry showed that CDL binds specifically to MthK. This finding could be pursued further in multiple directions: atomistic simulations could be employed to narrow down the residues involved in the binding process for site-directed mutagenesis experiments in mass

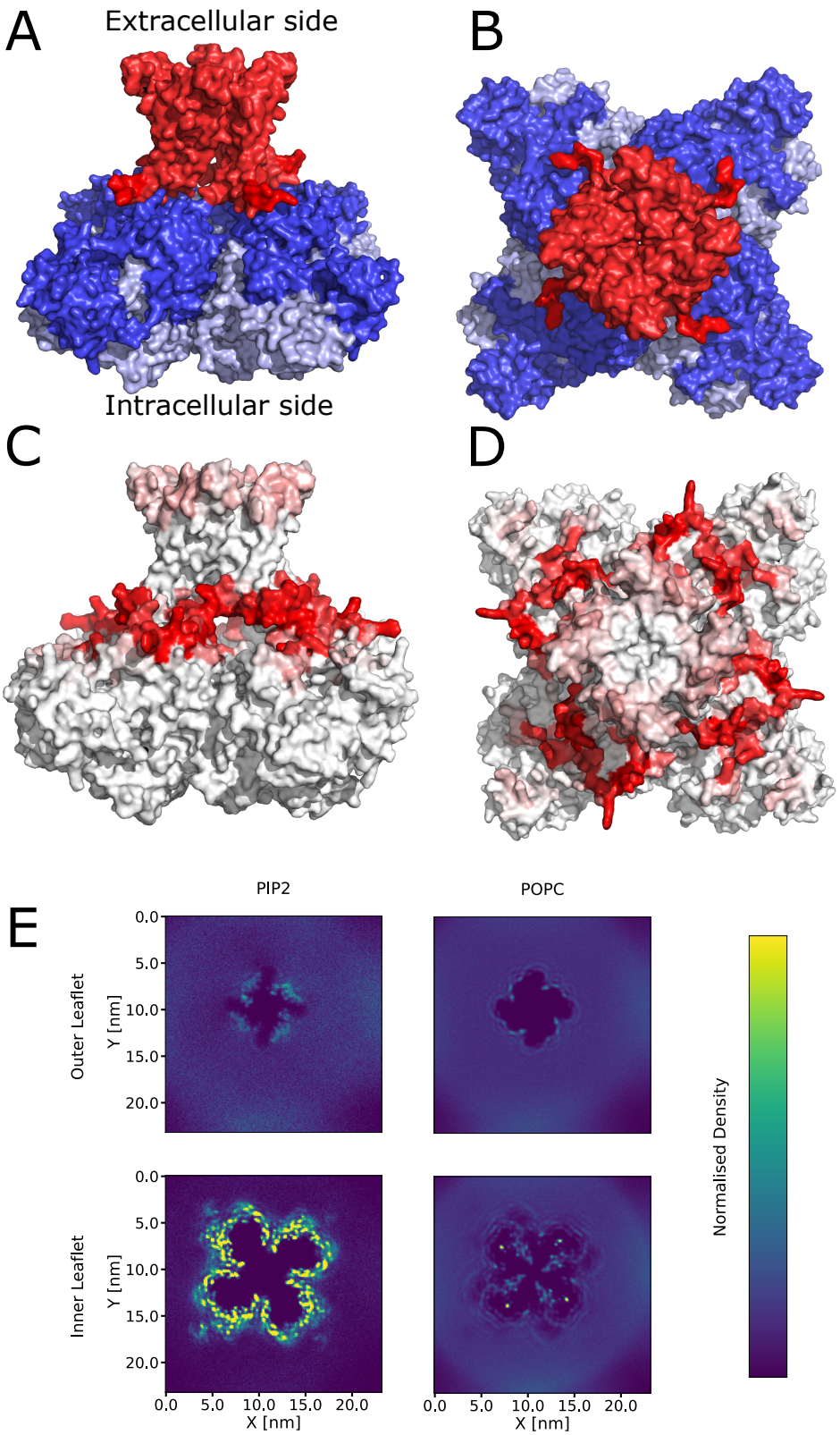

Figure 4.19. Structure of $\Delta 1-115$ BK used in simulations

A *Structure of BK with the TMD in red shades and the RCK domains in blue shades in side-view and* ***B*** *from extracellular side.* ***C*** *Contacts mapped onto $\Delta 1-115$ BK in side-view and* ***D*** *from extracellular side.* ***E****The location of PIP_2 and POPC is shown for extracellular (outer) and intracellular (inner) leaflets of BK. Colours are rescaled by the concentration of lipids.*

spectrometry. In addition and complementarily, this work could be extended towards electrophysiology. Such experiments could investigate the functional relevance of CDL binding similar to the BK channel. The comparison to the BK channel also invites to study the combination of calcium and CDL binding because PIP_2 only opens BK in the presence of Ca^{2+}. Probably, electrophysiology is best suited for such experiments, since mass spectrometry with salts is challenging. I attempted to use calcium with MthK but only received well resolved spectra up to the high micromolar range which is below concentration needed to open MthK at room temperature[379].

The study of lipid interactions of the BK channel could also be extended to include different classes of lipids in a systematic. While experimental approaches often add a single species of lipids, CGMD is able to embrace the complexity more realistic lipid bilayers[393]. It would be interesting to see whether different lipid compositions have an influence on the BK channel possibly mimicking different tissues where the BK channel is present. This could also include β-subunits since a structure with $\beta 4$ has recently been published[392].

4.6 Materials & Methods

4.6.1 Expression and Purification of MthK

The expression and purification of MthK followed established protocol described previously[375]. The plasmid of MthK has been used to transform BL21 λ DE3 *E.coli* cells (New England Biolabs). Colonies of these cells were used to inoculate flasks containing 100 ml of lysogeny Broth (LB) and 100 mg/ml ampicilin overnight at 37 °C. Afterwards, 15 ml of cells in LB were transferred 1 L flasks of the same mixture of LB and ampicillin to grow at 37 °C for three hours until an optical density of 0.6-0.8. To trigger expression 1 ml of 0.5 M of Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and expression took place for three hours at 37 °C.

Cells were harvested by 10 minutes of centrifugation at 5,000 g at 4 °C and lysed through in a M-110 PS Microfluidiser (Microfluidics) with added protease inhibitor tablets (Roche) into the lysis buffer (100 mM potassium chloride, 20 mM Tris, pH 7.4). To separate the membranes cell debris was pelleted by 20 minutes of centrifugation at 20,000 g followed by ultracentrifugation at 100,000 g for 1 hours 15 minutes. After homogenisation membrane proteins were extracted with 2% w/v of powdered Decyl β-D-maltopyranoside (DM) at 4 °C for three hours in resuspension buffer (100 mM potassium chloride, 20 mM Tris, pH 7.4). The protein possesses a His-tag which allowed purification by immobilised metal affinity chromatography (IMAC) through a 5-ml HisTrap-HP column (GE Healthcare) with an equilibration buffer 100 mM potassium chloride, 20 mM Tris, 20 mM imidazole, 5 mM DM; pH

8.0 and an elution buffer 100 mM potassium chloride, 20 mM Tris, 300 mM imidazole, 5 mM DM; pH 7.4. After dialysis at room temperature for three hours with thrombin to cleave the His-tag, the same column has been used for reverse IMAC followed by size exclusion chromatography (SEC) using a 24 ml Superdex 200 Increase 10/300 GL SEC column (GE Healthcare) with a buffer of 100 mM potassium chloride, 20 mM Tris, 5 mM DM at pH 7.4. The sample was concentrated using a 30 kDa molecular weight cut-off MWCO concentrator and flash-frozen to be stored at $-80\,^{\circ}\mathrm{C}$.

4.6.1.1 Western Blots

Western blots have been carried out following standard protocols[394]. The antibodies used were Anti-6X His tag antibody from mouse (Abcam) and the respective anti-mouse antibody. For the Flag-tag an Anti-DDDDK tag from rabbit (Abcam) was used with the matching anti-rabbit antibody for labelling.

4.6.2 Native Mass Spectrometry Experiments

4.6.2.1 Sample preparation

For nMS samples were exchanged into 200 mM ammonium acetate (AA) with double the critical micelle concentration (CMC) of detergent (here this means C8E4 at 0.5 % w/v or DM at 5 mM) through a Biospin column (Micro Bio-Spin 6, Bio-Rad) for a final concentration of 10-20 μM. All lipids was purchased from Avanti and diluted into a 200 μM stock solution and mixed with protein samples at a 4:1 volume ratio

just before the nMS experiments were conducted.

4.6.3 Mass Spectra Collection

The nMS experiments were carried out on an Q Exactive UHMR (see section 2.5.2) using gold coated borosilicate needles created in-house[395]. For data collection, the following parameters were used. The following parameters were used: HCD cell pressure of Argon as collision gas, 1×10^{-9} mbar, S-lens RF, 100. Other parameters have been varied depending on experiments: capillary voltage, 1.2-1.4 kV; source temperature, $100 - 300°C$; in-source trapping and HCD voltage were adjusted to the protein.

4.6.4 Data Analysis

Data analysis was carried out using the Thermo Fisher Scientific Excalibur suite in conjunction with UniDEC, MacSED or NaViA[191,285].

4.6.5 Homology Modelling and Simulations

The simulations of MthK have been carried out in Gromacs 2018 using the Martini forcefield version 2.2 and 3[254,257,258,261,354]. Version 3 has been used for bacterial membranes while version 2.2 has been employed for the PIP_2 simulations. The atomistic structures of MthK (PDB: 1LNQ) has been coarse-grained and inserted into lipid bilayers using the MemProtMD pipeline or *insane*[271,272,375]. The amphiphilic helix parallel to the membrane was added via Swiss Model[362]. For the BK channel

the calcium-bound state has been used (PDB ID:5TJ6)[373].

The lipid bilayers for the simulations were composed of 70 % POPG 20% POPE and 10% CDL for MthK and 90 % POPC with 10 % PIP_2 for the BK channel. The system was neutralised through 150 mM sodium chloride at the simulation temperature was set to 310 K and coupled via v-rescale[267]. Pressure was kept at 1 bar though a semi-isotropic Berendsen-barostat with a compressibility of 5×10^{-6}. Electrostatics were computed with a reaction field[355]. Lennard-Jones interactions were determined using a velocity Verlet scheme with a cut-off to enable GPU calculations[356].All bond constrains were implemented by the P-LINCS algorithm[357].

The simulations were carried out for five microseconds and five replica for each lipid composition with a timestep of 20 fs. Contacts of residues was defined by a distance of less than six angstroms between a residue bead and a lipid headgroup bead. The interfacial residues have been defined visually. Contacts and binding sites were calculated using MDAnalysis, PyLipID and Gromacs[354,359,360].

4.6.6 Contributions

- Mark Agasid helped design constructs
- Di Wu carried out proteomics experiments
- Robin Corey helped setting up CGMD simulations

Chapter 5

A Web-based Framework for the Analysis of Native Mass Spectra

Abstract

Native mass spectrometry became an established technique in investigating the stoichiometry and interactions of proteins. Recent advances in sample preparation lead to an increasing complexity of the resulting mass spectra of many overlapping molecular species. These are tedious to analyse manually but almost impossible to track by automated methods because they require prior knowledge from various sources. NaViA was designed to bridge that gap as a visual easy-to-use online tool to augment the manual assignment of native mass spectra. It is freely available on the web and does not require installation. In addition to analytic capabilities it offers to export data in different image formats for publication and to store sessions of analysis for sharing with collaborators or deposition for publication. This in combination with the integration with standard tools such as UniDec makes NaViA a useful addition to the suite of native mass spectrometry software.

5.1 Introduction

Native mass spectrometry is an excellent tool to investigate the stoichiometry of protein complexes as well as study the interactions of proteins with one another or ligands. Depending on the design of the experiment, these interactions can lead to very complex spectra[156,176]. This is especially true for spectra from SoLVe-MS, such as those obtained for MthK in section 4.2.2 in the previous chapter.

Native mass spectra are analysed manually or by using automated approaches such as UniDec[191]. Interpretation of these spectra requires a priori knowledge of the molecules involved in the interaction. Therefore, utilising knowledge from other biophysical methods, such as gel electrophoresis, western blot, or 'omics techniques is essential. In addition, structural information from literature or structural biology can assist in the interpretation of nMS results. The plethora of complementary knowledge used in nMS lead to challenges for automated approaches in the case of SoLVe-MS because it consists of many overlapping charge state series produced by various proteins that often have unknown identities prior to the experiment. The alternative of manual assignment is equally hard, because a correct manual assignment of many charge states can be incredibly time consuming.

Current software for native mass spectrometry also falls short in another area: the ability to save, load and share raw and processed data. To date, most software such as XCalibur and MassLynx requires the user to export raw data which is then imported into graphics illustration programs for annotation. This leads to several problems. It is hard to continue an analysis of a complex spectrum or resume work after a long interruption. Furthermore, it is difficult to share spectra with

peak assignment in a verifiable way. It is possible to share raw data where anybody checking a peak assignment (given that person has access to the software) would need to redo the analysis; or images of the data are shared which renders a verification of the peak assignment impossible.

These necessities motivated the creation of Native Visual Analyser (NaViA) shown in figure 5.1. The aim of NaViA is to occupy the middle ground between manual and automated analysis. A manual analysis requires copying the m/z values of peaks from instrument software manually and subsequently calculating the mass though e.g. Microsoft Excel. In contrast, UniDec imports spectra in proprietary format or txt and then performs preprocessing and peak assignment automatically. NaViA is an online service which allows a semi-automated analysis of native mass spectra through an easy-to-use interface of a web page (`https://navia.ms`). Charge states corresponding to a particular complex are selected visually and masses are assigned automatically following the MacSED algorithms (see section 5.3.2). The focus on augmenting a manual assignment is also instructive for scientists entering the field of nMS, since it provides key information on where the peaks are to be expected for a certain complex and highlights the peak series for this complex. NaViA allows to share 'sessions' in a similar way to PyMol or VMD[358,396]. They are loaded using common internet browsers. Given the push of in the scientific community for open science and open data, NaViA can contribute to this effort by deposition of NaViA sessions for publications as well as making collaboration easier in the field of native mass spectrometry and support the credibility of the field.

5.2 Implementation

NaViA is a standalone web page. I decided for a standalone website for two reasons. The primary reason is accessibility. Users should not be required to install additional software because it puts a hurdle for new users such as collaborators. A further reason concerned data security. Once the website is loaded no further connection needs to be made to any server and all of the data is processed locally on the user's computer. Thus, the security and ownership of the mass spectra are protected. This is particularly important for industrial users. Another advantages of a standalone website is that it also possible to download NaViA and run it in a browser of a computer without internet access, e.g. an instrument computers. As a side effect, basing the program on a web browser which is available on most computer simplifies the maintenance of the program and avoids incompatibilities with different operating systems.

The program is implemented in Python 3 using the Bokeh visualisation library[397,398]. It allows an easy implementation of basic features in Python. In addition, the licence of Bokeh guarantees free usage without constrains for e.g. industrial users in contrast to other libraries such as Plotly[399]. A final reason is the possibility to include JavaScript for functions, which is essential for making NaViA a standalone website. Without JavaScript inserts the realisation of the crucial charge state selection would have been impossible.

The project is hosted on GitHub with a custom domain linking to the program.

The project is hosted under `https://github.com/d-que/navia` and the custom domain is `https://navia.ms`. GitHub was chosen for the simplicity of hosting with the opportunity to push new versions easily. A custom domain was deemed handy so users do not have to remember a lengthy subdomain to access a tool made for easy and intuitive use. The top-level domain of Montserrat (.ms) was chosen because ms is also frequently used to abbreviate mass spectrometry, see nMS for native mass spectrometry.

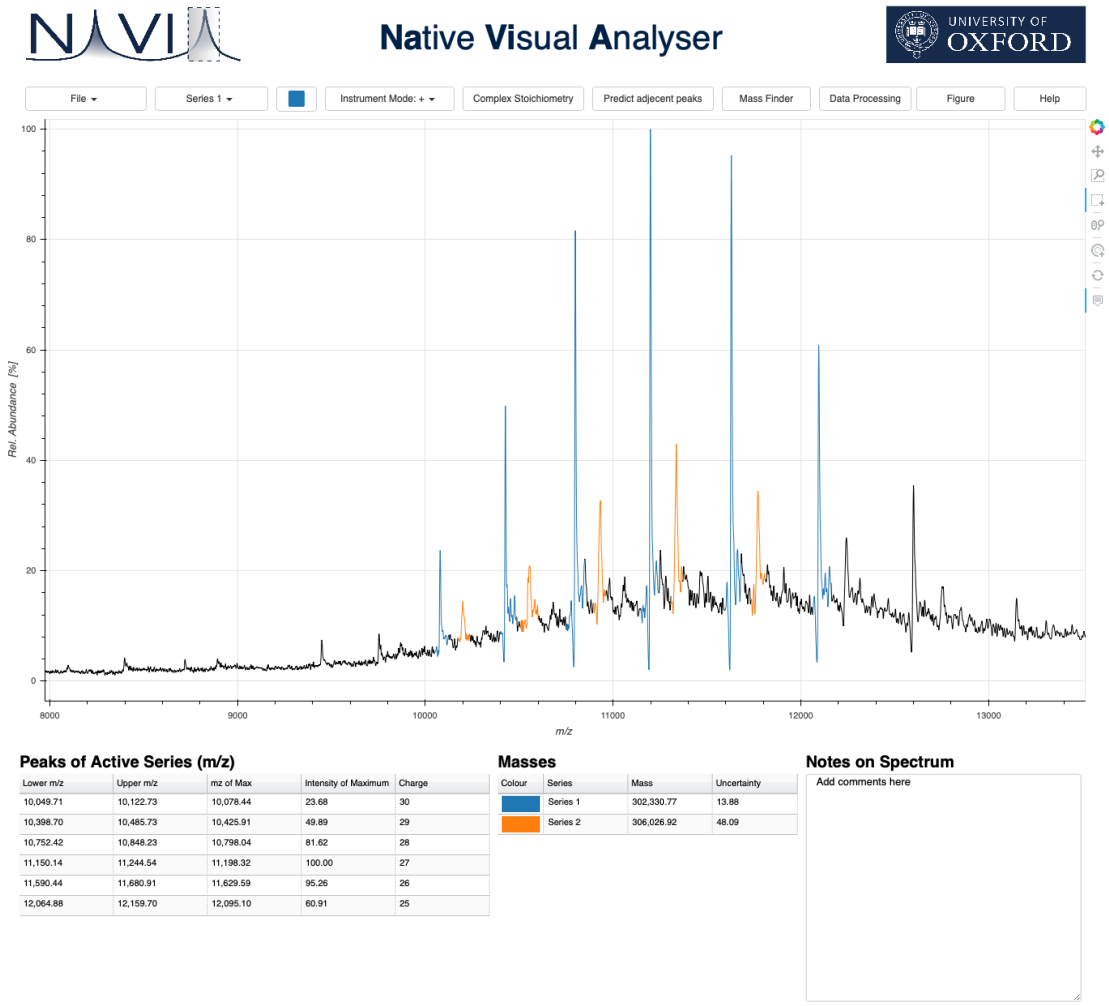

Figure 5.1. Screenshot of the interface of NaViA

NaViA with an MthK example data loaded. The interface consists of a top row with buttons for most functions. The central window allows to visualise the spectrum and select charge states. Tools for selection and correction of the charge states as well as general tools to move the window are located on the right of the central window. Below, there is a list of the active peaks and a colour coded list of masses of the series as well as a comment window.

5.3 NaViA features

5.3.1 Charge State Selection

Peaks are selected through a range on the m/z-axis using the mouse cursor. The selected regions are then highlighted in the colour of the current series of charge states to tell them apart from the rest of the spectrum. This visual selection is shown in figure 5.2 and is one of the main features of NaViA. The colour of a series by default chosen as Tableau colours[400] to create a visually appealing colour palette. Colours of series can be changed and saved manually to e.g. illustrate the binding of multiple ligands with similar colours.

When a range is marked for as a selection there is the possibility for overlap which needs to be accounted for. It follows the diagram shown below in figure 5.3. The algorithm is chosen to create the most natural merging and overwriting experience for the user.

Once multiple charge states are selected, the corresponding mass is calculated by the MacSED algorithm.

5.3.2 MacSED Algorithm for Mass Assignment

The mass of a molecular species is calculated using the Mass and Charge State Evaluation and Determination (MacSED) algorithm. It was employed by McKay *et al.* in 2006[285]. This algorithm makes use of the fact that all of the charge states measure represent the same mass of a complex and that there are a limited number

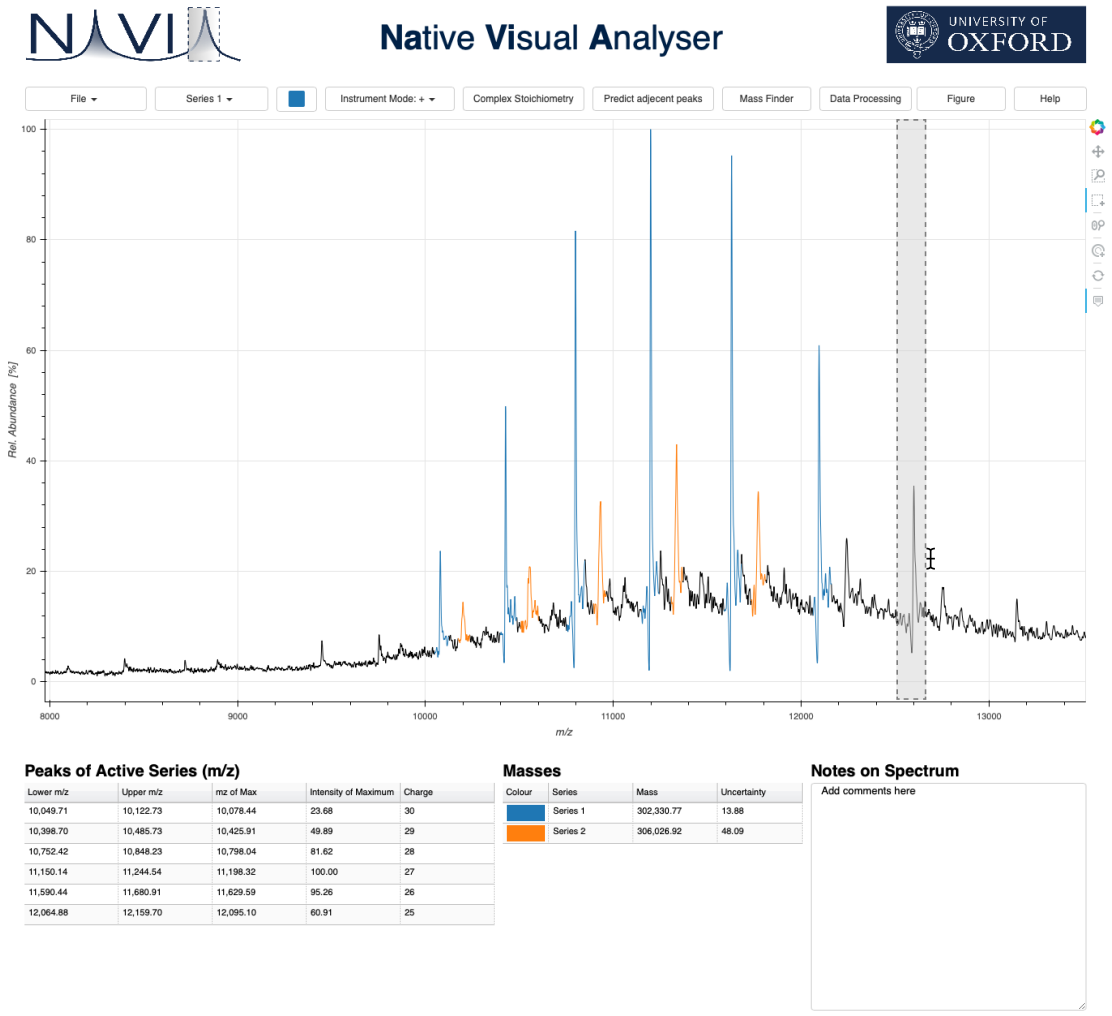

Figure 5.2. Selection of Charge States in NaViA

Charge states are selected visually by the user through a mouse click. A grey transparent range window will appear and mark the area that the user assesses to be a part of a series. The maximum intensity of the window is then used as input for the MacSED algorithm.

of possible charges as the mass of a complex is given by

$$\left(\frac{m}{z}\right)_n = \frac{m_{com} + n \times m_p}{n \times e}, \tag{5.1}$$

where m_{com} is the complex mass, m_p the mass of a proton and e the elementary charge and $n \, \epsilon \, \mathbb{N}$ the number of charges. Therefore the m_{com} is given by

$$\Rightarrow m_{com} = n \times \left(\left(\frac{m}{z}\right)_n \times e - m_p\right). \tag{5.2}$$

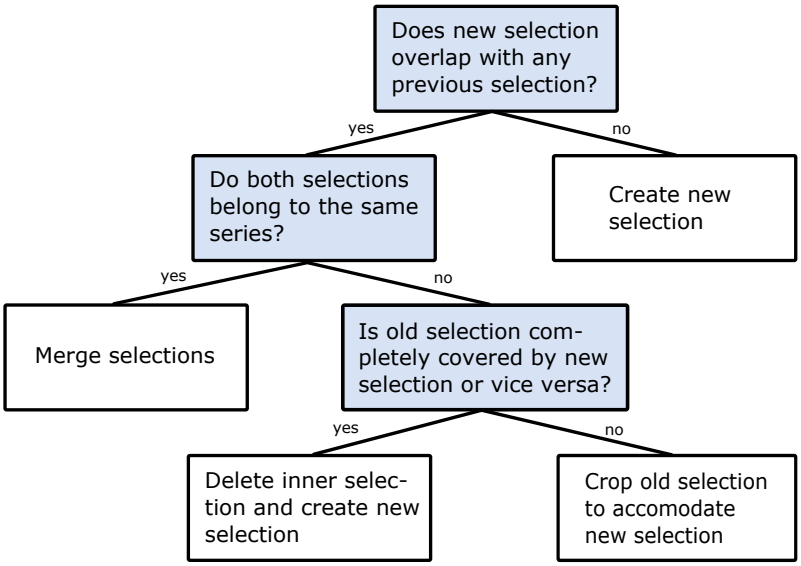

Figure 5.3. Algorithm for Selection Overlap

In case of an overlap between selections, the algorithm needs to determine whether the new and old selection belong to the same series. Afterwards, depending on partial or complete overlap, the algorithm decides on whether to crop the old selection or delete the surrounded selection.

With this it is possible to try out different charge values for all of the peaks and determine determine the correct charge states by minimising the standard deviation for the m_{com} values calculated at each of those peaks.

This algorithm is not described anywhere in great details despite being the foundation of an excel spreadsheet commonly used by multiple research groups. Therefore, it is described here. The algorithm starts with extracting a series of m/z values from a mass spectrum. This is not directly part of the algorithm (in NaViA this is done as described in section 5.3.1 above). The intensities of these charge states are not used. The charge states are inserted into a vector $\overrightarrow{m/z} = (m/z_1 \dots m/z_n)^T$ in descending order since the highest m/z value has the lowest charge of the charge states belong to one complex. Next, a range of possible elementary charges is needed $1 \dots m+n$. While relatively low charge states are usually produced by a nMS experiment, in

some cases charge states well above 100 can be observed, requiring a high charge state limit for the algorithm (in NaViA it is set to 1000). From this a matrix of possible charges $\mathbf{C}$ is determined

$$\mathbf{C} = \begin{pmatrix} 1 & \cdots & n+1 \\ \vdots & \ddots & \vdots \\ m & \cdots & m+n \end{pmatrix},$$

which is then used to obtain a mass matrix $\mathbf{M}$ of the same dimensions by

$$\begin{aligned} \mathbf{M} &= \mathbf{C} \times \mathbf{I_m} \times \overrightarrow{m/z}, \\ \Rightarrow \mathbf{M} &= \begin{pmatrix} \overrightarrow{m/z}_1 \times 1 & \cdots & \overrightarrow{m/z}_n \times n \\ \vdots & \ddots & \vdots \\ \overrightarrow{m/z}_1 \times m & \cdots & \overrightarrow{m/z}_n \times (m+n) \end{pmatrix}, \end{aligned}$$

where $\mathbf{I_m}$ is the identity matrix of size m. A row of this matrix has a mass measurement for each of the charge states of $\overrightarrow{m/z}$. Hence, the mean and the standard deviation σ are calculated using the elements of a row in M yielding a vector of standard deviations and a vector of means. Calculating the mean and the standard deviation gives an easy measure to determine the true charge state. Since m_{com} is linear in n, the standard deviation is also linear in the difference to the true charge state. The row with the lowest standard deviation therefore represents the true charge state. An example of the algorithm is illustrated in figure 5.4.

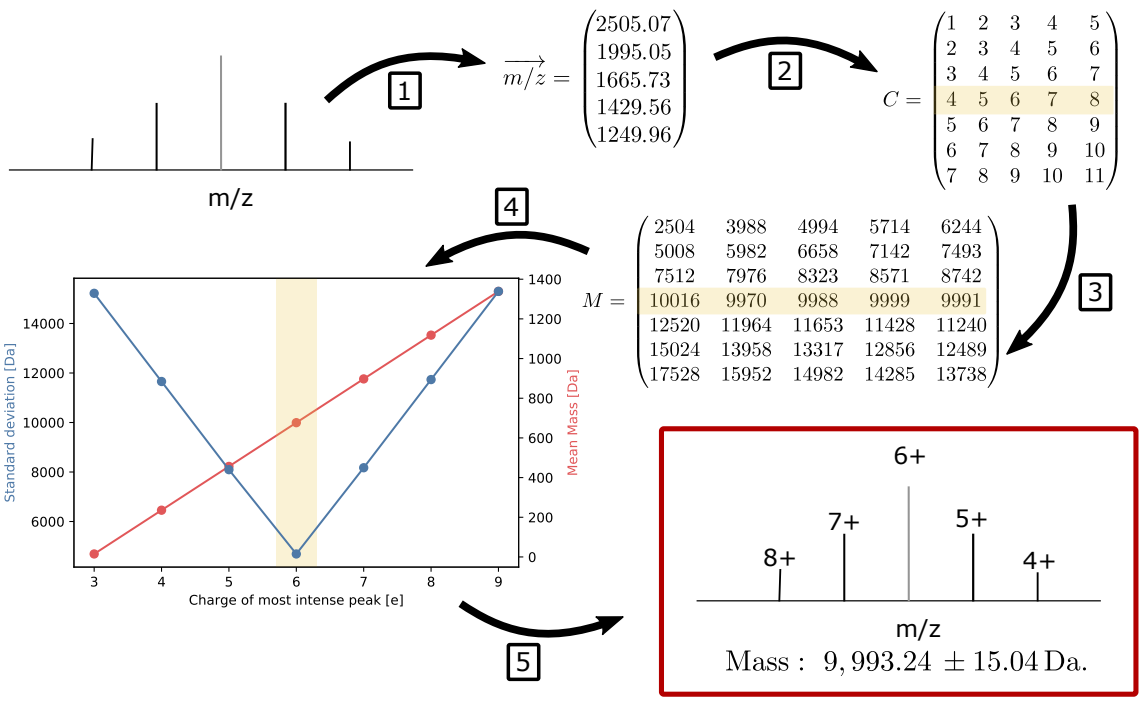

Figure 5.4. Example of the MacSED algorithm

In this example the MacSED algorithm was applied to a complex of mass 10,000 Da in charge states 4+ to 8+ has been given. It follows the several steps: 1. Extract the m/z values of the peaks (they have a bit of noise in this example). 2. Generate the charge matrix ***C*** *for a range of possible charges. Here, 3+ to 9+ was chosen for the charge of the most intense peak (grey) for visualisation purposes. 3. Calculate the potential masses for the mass matrix* ***M****. 4. Do a row-wise calculation of the mean and the standard deviation of the mass. The results of the mean mass are shown in red and standard deviation in blue. Choose the a charge of 6+ as the correct charge because it has the lowest standard deviation (yellow background). 5. Use the mean and the standard deviation as the true mass and its uncertainty, respectively, and assign the charge states to all the peaks.*

5.3.3 Comment Window

Interpreting native mass spectra can benefit greatly from additional information about the sample. Therefore, a comment window allowing to enter text was added to NaViA. This allows to store all the information in one NaViA file or at least to record where additional files (such as images of electrophoresis) are saved. Therefore, it can be helpful in collaborations as well. The comment is saved in `.navia` files as described in section 5.3.7.

5.3.4 Data Processing

NaViA enables data processing of native mass spectra using the same algorithm as the standard software e.g. UniDec. This choice was made for consistency and based on the experience of what is needed for the analysis of native mass spectra. Processing of spectra is mostly needed when relatively poor signal-to-noise ratios are observed, which is common when older instruments are used or when challenging samples are investigated. The processing options available in NaViA are described below.

Range mz In this function, an m/z range can be defined to crop the spectrum to an area of interest for visualisation.

Intensity Threshold Allows to define a threshold manually. All intensities below this value are set to zero.

Gaussian Smoothing Gaussian smoothing allows to correct for high frequency noise which is routinely encountered when collecting native mass spectra. For each intensity data point a Gaussian centred around its m/z is calculated with a given σ. Then data around the data point are averaged with weights according to this Gaussian and a cut-off of equal or larger than 5σ. On the boundaries, the distribution is mirrored to achieve correct normalisation.

5.3.5 Background Subtraction

NaViA offers three different modes of background subtraction: minimum, linear and curved, like UniDec. The details of these modes are explained below and illustrated in figure 5.5. They allow to subtracting a potential offset in different ways.

1. Subtract Minimum In this option, the minimum intensity of the spectrum is determined and then subtracted from all intensities.

2. Linear Subtraction For a linear subtraction a buffer variable b is needed. Then the intensities for the lowest b m/z values are averaged. The same procedure is applied to the intensities of the highest b m/z intensities. Afterwards an interpolation between these two point from their m/z region is applied to the spectrum in the middle. In order to avoid negative values, they are set to zero.

3. Curved Subtraction The curved subtraction was first introduced in Massign[199]. The curved subtraction performs a two step process. First, the buffer b is used to define a window of $b/2$ around a m/z value and determine the minimum intensity value in this window. This is done for the complete m/z range to define an array of minimum intensities for the whole m/z range. Afterwards, this minimum intensity array is subtracted from the actual intensities. As a final step, the spectrum is smoothed through a Gaussian filter.

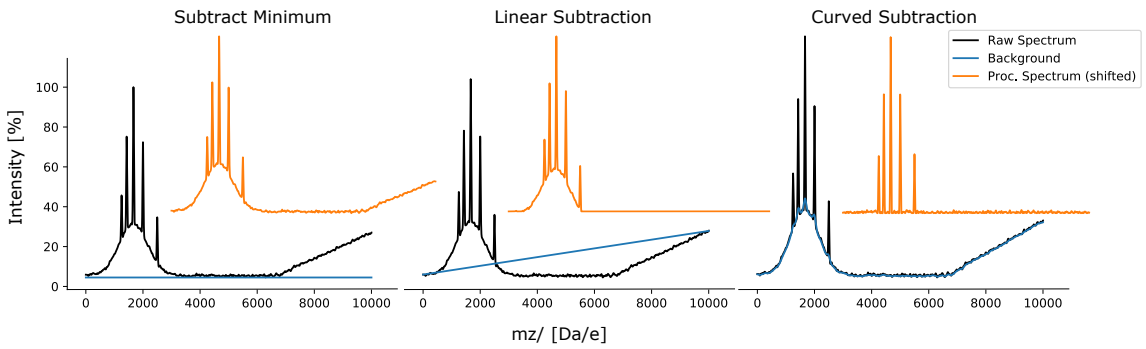

Figure 5.5. Illustration of the options to deduct the background of a spectrum. *In all three options the detected background is shown in blue, the original spectrum in black and the processed spectrum in orange but shifted. On the left the minimum value of a spectrum is deducted. This levels off any offset, but falls short in deducting any background which does not occur evenly across the m/z range. In contrast the linear subtraction manages to rectify the larger background in the high m/z region but does not straighten the bump around the peaks. This is only possible with the curved subtraction mode on the right.*

5.3.6 Input Files

NaViA uses simple comma separated values files (CSV) with a header of up to ten lines and two columns denoting m/z and intensity. These can be readily imported by instrument software. In addition, it allows to load sessions as well as subunit tables for assigning complex stoichiometry as described in section 5.3.9.

5.3.7 NaViA file format

NaViA allows sessions to be saved, including raw data, data processing and peak assignment, in a `.navia`-file in JSON format. This feature allows users to share either assigned spectra or raw data. In addition, JSON formats are easily readable by various programming languages and software packages e.g. Python or Origin. Therefore, `.navia`-files are well suited for consecutive data analysis. The format works as a dictionary and the key words and content are given below.

'raw_data': a dictionary with 'm/z' and raw 'Intensity' as keys, each comprising a list of floats.

'proc_data': adictionary with 'm/z' and processed 'Intensity' as keys, each comprising a list of floats.

'ser_dict': a dictionary for names of the series (naming planned for future release)

'masses': a dictionary with 'Series', 'Mass', 'Uncertainty' and 'Colour'(line colour) as keys.

'sele': a nested dictionary for all 20 possible series containing lists of integers for lower, higher and maximum intensity m/z indices ('i_low', 'i_upp' and 'i_max' respectively)

'ser_data': a similar to 'sele'. containing ('x_low', 'x_upp', 'max_int' and 'x_max') mz-values and the intensity of the maxima or assigned peak if corrected.

'charge_sign': either 'positive' or 'negative' depending on instrument mode.

'subunits': contains lists of properties of subunits of complex ('mass', 'min', 'max', 'stride', 'name').

'stoichiometry': lists of properties of complex stoichiometry to measured masses for NaViA.

'dtp': lists of properties of data processing.

'comment': string of the comment field.

5.3.8 Figure

The figure menu contains all the options related to exporting figures. Many elements of the figure can be shown and hidden according to user needs. These are grid, axes, axes labels and m/z of peaks for correcting them. In addition, linewidth and the x- and y-ranges can be specified. These tools can be useful when evaluating a series of spectra and exact alignment is desirable.

NaViA provides the possibility to export images in 4K quality sufficient for A4 pages in publications as well as vector graphics in the SVG format. In order to keep NaViA running smoothly in the browser, these figures need to be created on demand. The export of these images allows using them for publications or presentations and in standard software such as Inkscape or Adobe Illustrator.

All functions are working in major web browsers. Google Chrome, Mozilla Firefox and Apple Safari have been tested. However, in tests Apple's Safari browser on macOS showed to be very slow rendering 4K and SVG files. This improved in Safari Version 14.0 but was still slower than Chrome. Therefore, picture export can be guaranteed in all of the major browsers except Safari. This does not affect the use of Safari per se, so it is still possible to use Safari for all analysis purposes. While Safari

has its limits on macOS, the iOS version did not show similar problems. This indicates that it is related only to the maOS version of Safari. Since other browsers are available on macOS, this issue does not harm cross-platform availability of NaViA.

5.3.9 Complex Stoichiometry

The "Complex Stoichiometry" function makes it possible to see whether a complex with a known mass is present. This can be very helpful for spectra where many molecular species are present. It allows to assign the subunits with name, mass as well as minimum, maximum and stride of appearance in a complex. With these subunits, a complex stoichiometry can be calculated for all series. It is also possible to save and load subunit tables in an CSV format in order to facilitate analysis of how an interaction changes in e.g. titration experiments. Complex Stoichiometry needs to be activated by clicking the Complex Stoichiometry button in the top row. Afterwards the Subunit and Stoichiometry boxes allow adding subunits and fitting these units to a series with a given delta respectively as shown in figure 5.6.

5.3.10 Predict Adjacent Peaks

The prediction of adjacent charge states is a tool to find additional charge states. It requires the user to have identified at least two charge states of a complex. For such a series, equation 5.1 allows calculating m/z values for the charges $n_{min} - 1$ and $n_{max} + 1$. These states are marked with full vertical lines. To estimate a range of possible m/z values, dashed lines are showing a range of $5 \times \sigma//(n_{max/min} \pm 1)$

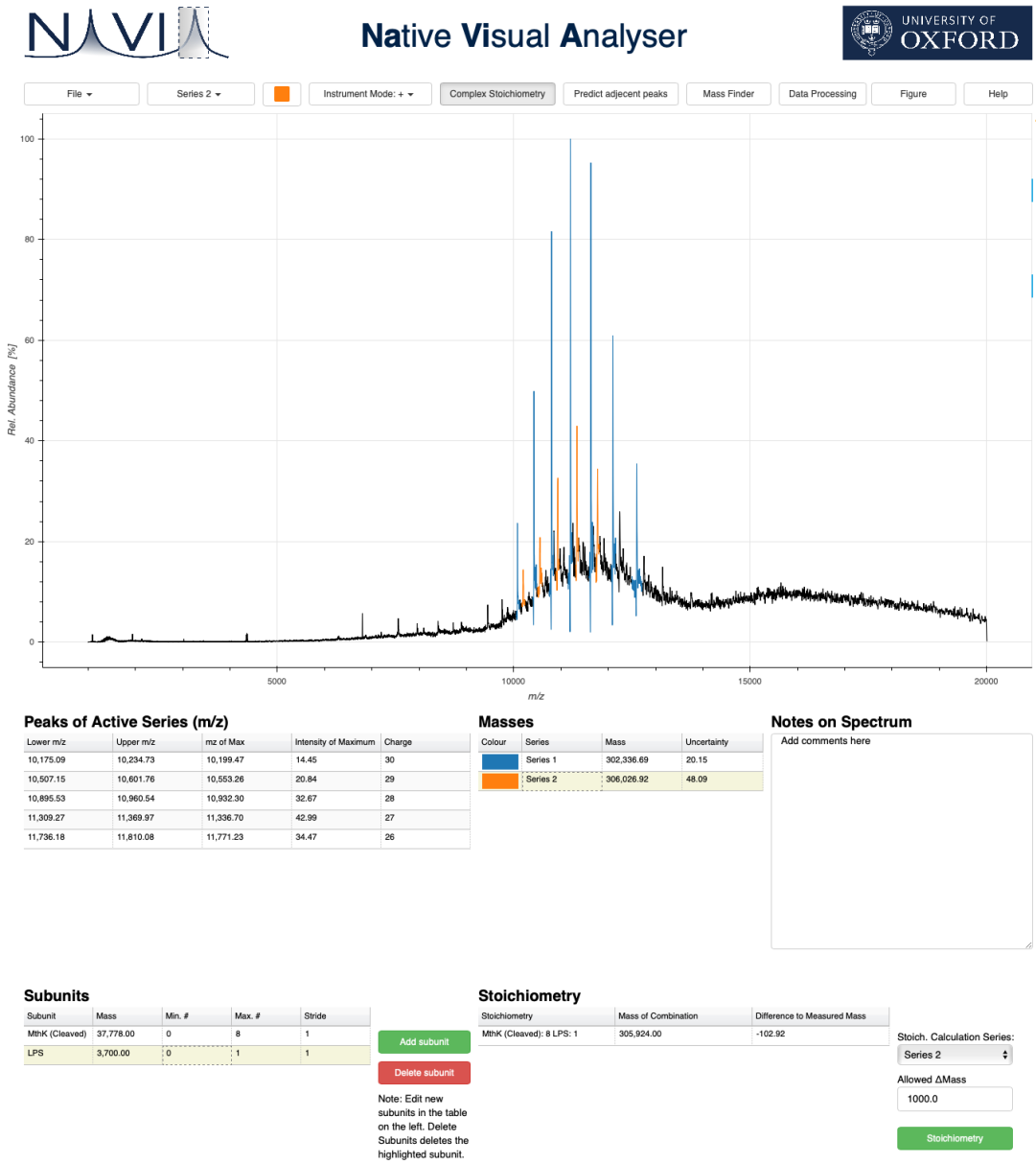

Figure 5.6. Complex Stoichiometry in NaViA

The Complex Stoichiometry function is shown below the standard interface upon activation in the top row. It consists of the Subunits list on the left and the Stoichiometry list on the right. In Subunits, additional units can be added. In Stoichiometry, these units are fitted against the mass of a selected series within the allowed mass difference $\Delta Mass$.

around the estimated m/z values. Assuming that σ is a good estimator for the uncertainty and a Gaussian distribution of noise, this indicates a likelihood of less than 5.733×10^{-9} % for an additional peak to be outside this window. An example for of peak predictions is given in figure 5.7.

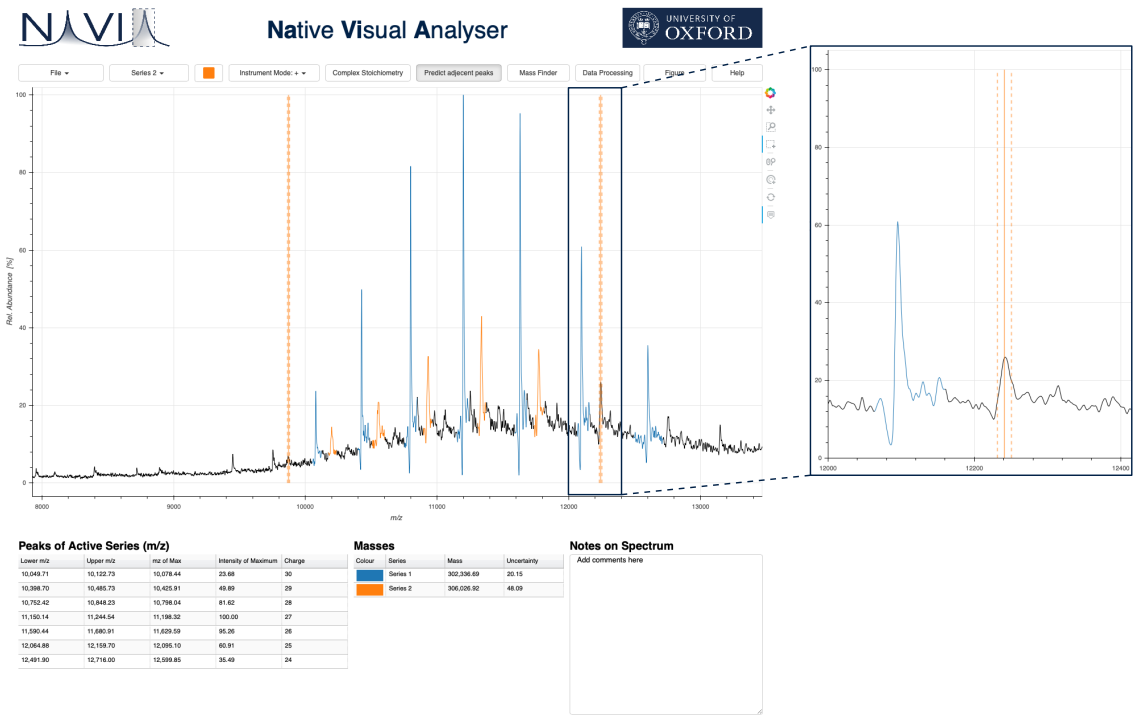

Figure 5.7. Adjacent Peak Prediction in NaViA

Adjacent Peak Prediction can be activated in the top row. It shows where the next peaks for a series are to be expected for the active series. The exact m/z value is shown with a solid line and a range is given by a dashed line in the distance of five standard deviations. This is shown for the MthK + LPS example including an expanded window on the right.

5.3.11 Mass Finder

Mass Finder is a function to augment the search for charge states corresponding to a specific complex of known mass, e.g. a specific stoichiometry for known subunits. The reason for including this function is, that in SoLVe spectra, some complexes are hard to identify in the many overlapping series of peaks. For the function the mass of the complex as well as the range of charges need to be specified. Then equation 5.1 is solved to create possible m/z values. This is shown in figure 5.8 below. Vertical lines indicate where the peaks for the charge states are expected. Mass of the complex as well as the range of charges can be updated on the fly.

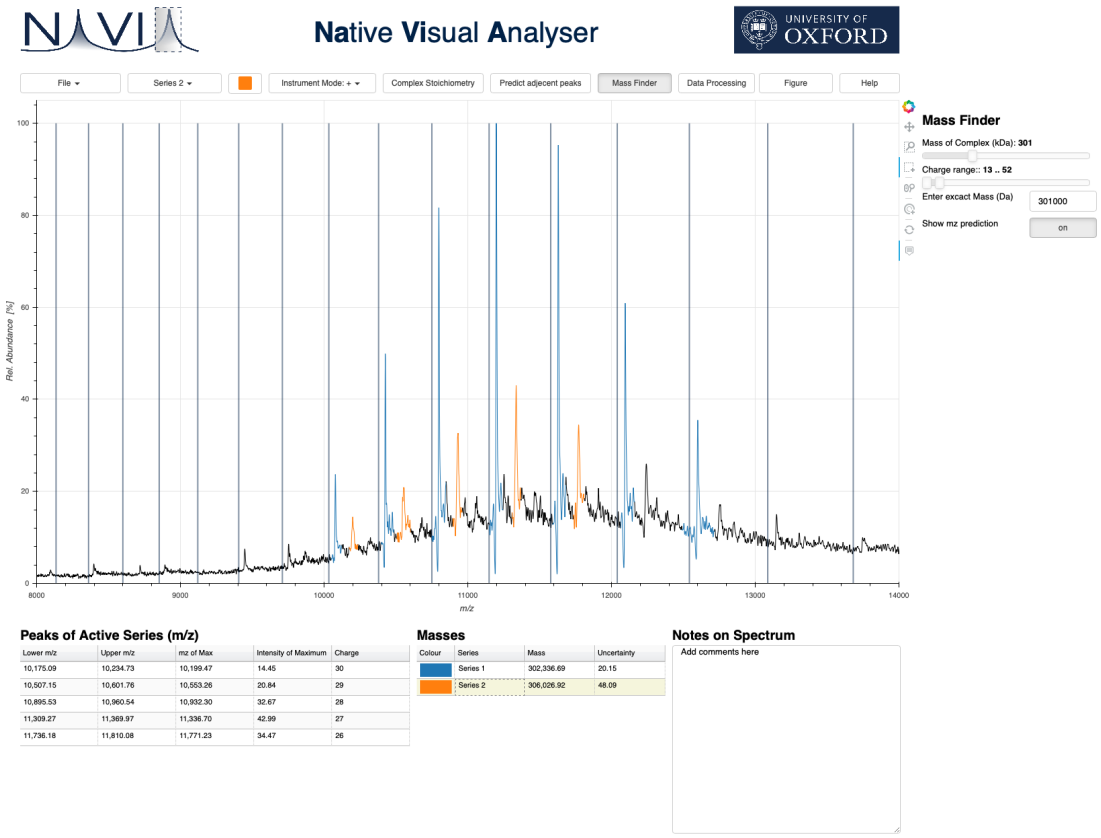

Figure 5.8. Mass Finder in NaViA

Mass Finder allows calculating m/z values where a peak for a certain range of charge states for a complex of a given mass are to be expected. They are shown as vertical lines for a complex of $300,000$ *Da in this example.*

5.3.12 Help

The help window allows the user to reach websites with further instructions. These are the GitHub project where the source code and additional information such as the `navia` file format can be found. In addition, there is a link to the YouTube channel explaining how to use the program with a video (https://youtu.be/BfC-9olFkaY).

5.4 Conclusions & Future directions

NaViA was implemented to have a easy-to-use tool to augment the analysis native mass spectra. Feedback from early users (in the Robinson and Benesch groups) indi-

cates that it serves this purpose well. Sessions of native mass spectra can be shared and deposited with ease. Therefore, I am hopeful for the program to be appreciated by the wider community.

The current implementation of NaViA leaves a lot of opportunities for extension. While it is possible to implement a lot of functions for niche applications, it would clutter the interface and oppose the idea of a simple tool. However, a few ideas might be worth considering. First, the peak selection could be made smarter. For example, peaks could be selected such that only a concave region around the highest intensity is selected. However, this would require some smoothing for noisy data. In addition, peaks belonging to a series could be automatically selected given a few peaks of that series are known. This feature has not yet been implemented, because the prime application I had in mind as a developer were SoLVe-spectra which are naturally complex and require additional knowledge. Also, for automatic assignment there are tools such as UniDec which integrate with NaViA. A closer integration with UniDec will be a more promising route to ensure sustainability of the development. An early version of NaViA was implemented in wxPython (a common framework for developing graphical user interfaces for Python programs), the same framework that UniDec uses. Thus, NaViA could be translated into a module for the UniDec suite and be integrated. In addition, loading UniDec sessions into NaViA would be useful and is at the top of the priority list. Recently, UniScore was added to UniDec[190]. This score rates the quality of a deconvolution of native mass spectra through the multiple criteria including consistency of peak shape and a smooth charge distribution. Adding this feature to NaViA would also be helpful

to assess the assignment of peaks by the user.

An additional route for extending NaViA would be to make NaViA a python package. This would help to integrate it into modern workflows for data analysis such as Jupyter Notebooks. Doing so, could also help integrating NaViA into example notebooks for typical nMS applications, such as ligand binding curves or ligand affinities. Providing a suite of Jupyter notebooks could also be helpful to illustrate how to integrate NaViA into Python work flows for the nMS community.

Chapter 6

Concluding Remarks

In this thesis I used native mass spectrometry and molecular dynamics simulations to investigate lipid binding to membrane proteins and developed a website to analyse native mass spectra visually with the help of auxiliary tools. It consists of three projects which each have a distinct character.

In the first results chapter (chapter 3), a collaborative project on using SID to probe the effect of lipid binding on the stability of dimer interface of a bacterial transporter, SemiSWEET, is described. It observed clear stability differences between classes of lipids and even noticed the subtle differences of a chain length dependence for the stabilising effect of different CDL. I performed CGMD simulations to shed light on the molecular causes of these observations. My simulations offer a rationale for the observed differences between classes as well as for the stability differences between different CDL species. Different preferences to bind to a site involved in coordinating dimerisation is indicated to be the underlying cause. These simulations can be extended by atomistic molecular dynamics simulations with longer simulation times to see whether there is any state dependence. PMF calculations on the dimer

stability would be interesting, but almost impossible to perform in practise due to the high number of degrees of freedom. Therefore, I consider long atomistic simulations the most feasible way.

For the second part of this chapter, I performed CIU experiments on mutants of SemiSWEET. These mutants were designed to selectively weaken the dimer interface through site-directed mutagenesis. While my experiments showed a lower stability, SID experiments by my collaborator curiously pointed towards a higher stability compared to the wild-type for the Y62W mutant. The Q29W mutant destabilised the interface enough that only monomer was visible. This work could be extended by choosing mutations to alanine which also disrupts the interactions but does not "bulk out" the interface. If some effect is found for mutants, it would also be promising to start MD studies on it with a similar approach to the one used in the first part of this chapter. Extending the work on mutants of the dimer interface would be the approach I would most likely pursue if given more time.

Chapter 4 describes my investigation on lipid binding to a bacterial potassium channel, MthK. While the previous chapter mostly consisted of simulations, this chapter comprised an extensive amount of experimental work conducted by me. A combination of mass spectrometry and molecular dynamics simulations is employed to characterise lipid binding after initial experiments indicated endogenous lipids being bound to MthK. Cardiolipin has been shown specifically bound to the linker region of MthK through a combination of experiments and simulations. These results are compared to the homologous situation in eukaryotes (PIP_2 binding to the BK channel) by further simulations without concrete results. This project could be extended by performing atomistic simulations to obtain a better understanding of

the residues involved in CDL interactions. However, the most promising route would be to perform electrophysiology experiments to address the biological relevance of CDL binding.

I am proud to have achieved the in-vivo stoichiometry, but it did not prove useful for lipid-binding experiments. The main difficulty of the MthK project was that it was incredibly challenging to obtain this stoichiometry and to express the channel in sufficient quantities. While trying different techniques to achieve the in-vivo stoichiometry allowed me to learn a lot about nMS, I would do it differently now and challenge earlier whether it is necessary to spend further time and effort on obtaining this stoichiometry. Eventually, the octamer was used to study lipid binding. Therefore, I would challenge the necessity of experiments more if I were to start a different project. It could have saved a lot of time and effort.

MthK is not an ideal protein to work on with nMS. With the benefit of hindsight, I would have chosen a different protein. An ion-gated ion channel is not an ideal choice to investigate with mass spectrometry because the presence of many adducts to protein complexes renders the analysis of spectra almost impossible and small ions in the buffer will impede the effective transmission of protein complexes. A different issue is that while it is possible to investigate lipid-binding through nMS, it is not possible to draw conclusions of the biological relevance from these experiments alone. This would require techniques like electrophysiology. Therefore, other proteins are better suited for these techniques. My conclusion would be to choose a protein, where the biological relevance of the experiments can be addressed with nMS and MD.

The first part of this chapter outlines a number of different approaches to obtain the *in vivo* stoichiometry which was eventually observed (albeit not stable enough for lipid binding experiments). Among the different experimental routes, I employed SoLVe-MS, which results in very complex mass spectra. Frustrated by the available software solutions for the analysis of these spectra, I decided to create a program of my own. Thus, these experiments sparked the development of software described in chapter 5.

In the final results chapter Id describe a web-based program for the analysis of native mass spectra, NaViA. This program allows the user to analyse native mass spectra through a visual user interface directly in a web browser. NaViA also offers a suite of auxiliary tools to facilitate the analysis such as predicting adjacent charge states or finding the stoichiometry from known subunits. In addition, the ability to save and load sessions for sharing spectra including analysis or depositing them for publication is implemented. These abilities will prove useful for the nMS community as they enable easier collaboration and they provide a way to check the claims of results directly. Therefore, they will improve accountability of nMS.

The ease of use and the instructive auxiliary tools make it a valuable addition to the suite of software available for the analysis of native mass spectrometry. I hope it to be helpful for new scientists in the field of nMS and experienced mass spectrometrists alike. In the future, NaViA should be integrated into UniDEC. Since I am about to graduate with this thesis, the chances of a successor maintaining NaViA are slim. The best option is to integrate NaViA into UniDEC which is doable in the Python framework of UniDEC. This ensures sustainable development and integration of the software with the one that is commonly used in nMS community. In addition, I

believe some of the features of UniDEC would also benefit NaViA. Some examples include an enhancement of the user interface by allowing the user to rename series and the integration of UniScore into NaViA to have a gauge of the quality of the fit of a series. However, I keep NaViA as a rather simple to use tool and avoid duplicating the majority of UniDEC features. NaViA is meant to be a lightweight in-browser supplement to the more extensive UniDEC suite.

References

1. Clare M O'Connor, Jill U Adams, and Jennifer Fairman. Essentials of cell biology. *Cambridge, MA: NPG Education*, 1:54, 2010.

2. S. Macbeth. Cellular gatekeepers. *Nature Structural & Molecular Biology*, 23, 06 2016. doi: 10.1038/nsmb.3246.

3. Bruce Alberts. Molecular biology of the cell, 2008.

4. Jonathan Lombard. Once upon a time the cell membranes: 175 years of cell boundary research, dec 2014. ISSN 17456150.

5. Harvey Lodish, Arnold Berk, Chris A Kaiser, Chris Kaiser, Monty Krieger, Matthew P Scott, Anthony Bretscher, Hidde Ploegh, Paul Matsudaira, et al. *Molecular cell biology (Ch. 7)*. Macmillan, 2008.

6. Gerrit Van Meer, Dennis R Voelker, and Gerald W Feigenson. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2):112–124, 2008.

7. Markus R Wenk. Lipidomics: new tools and applications. *Cell*, 143(6):888–895, 2010.

8. Takeshi Harayama and Howard Riezman. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology*, 19(5):281, 2018.

9. Christian Sohlenkamp and Otto Geiger. Bacterial membrane lipids: Diversity in structures and pathways. *FEMS Microbiology Reviews*, 40(1):133–159, 2015. ISSN 15746976. doi: 10.1093/femsre/fuv008.

10. CR Raetz and William Dowhan. Biosynthesis and function of phospholipids in escherichia coli. *Journal of Biological Chemistry*, 265(3):1235–1238, 1990.

11. Doralicia Casares, Pablo V Escribá, and Catalina Ana Rosselló. Membrane lipid composition: effect on membrane and organelle structure, function and compartmentalization and therapeutic avenues. *International journal of molecular sciences*, 20(9):2167, 2019.

12. Byung-Chang Suh and Bertil Hille. Pip2 is a necessary cofactor for ion channel function: how and why? *Annu. Rev. Biophys.*, 37:175–195, 2008.

13. Giuseppe Zaccai. Proteins as nano-machines: dynamics–function relations studied by neutron scattering. *Journal of Physics: Condensed Matter*, 15(18): S1673, 2003.

14. Linn Fagerberg, Kalle Jonasson, Gunnar von Heijne, Mathias Uhlén, and Lisa Berglund. Prediction of the human membrane proteome. *Proteomics*, 10(6): 1141–1149, 2010.

15. Hang Yin and Aaron D. Flynn. Drugging Membrane Protein Interactions, jul 2016. ISSN 15454274.

16. John P Overington, Bissan Al-Lazikani, and Andrew L Hopkins. How many drug targets are there? *Nature Reviews Drug Discovery*, 5(12):993–996, 2006.

17. Petr Klein, Minoru Kanehisa, and Charles DeLisi. The detection and classification of membrane-spanning proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 815(3):468–476, 1985.

18. Markus Sällman Almén, Karl JV Nordström, Robert Fredriksson, and Helgi B Schiöth. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC biology*, 7(1):1–14, 2009.

19. MICHAEL G Wolfersberger. Uniporters, symporters and antiporters. *Journal of Experimental Biology*, 196(1):5–6, 1994.

20. PM Jones and AM George. The abc transporter structure and mechanism: perspectives on recent research. *Cellular and Molecular Life Sciences CMLS*, 61(6):682–699, 2004.

21. Joseph J Feher. *Quantitative human physiology: an introduction (Ch. 2.6)*. Academic press, 2017.

22. Francesco Fiorentino, Joshua B Sauer, Xingyu Qiu, Robin A Corey, C Keith Cassidy, Benjamin Mynors-wallis, Shahid Mehmood, Jani R Bolla, Phillip J Stansfeld, and Carol V Robinson. substrate and an antimicrobial peptide. *Nature Chemical Biology*. ISSN 1552-4469. doi: 10.1038/s41589-020-00694-2.

23. Helen M Knight, Benjamin S Pickard, Alan Maclean, Mary P Malloy, Dinesh C Soares, Allan F McRae, Alison Condie, Angela White, William Hawkins, Kevin McGhee, et al. A cytogenetic abnormality and rare coding variants identify abca13 as a candidate gene in schizophrenia, bipolar disorder, and depression. *The American Journal of Human Genetics*, 85(6):833–846, 2009.

24. Mitsuhiro Nakato, Naoko Shiranaga, Maiko Tomioka, Hitomi Watanabe, Junko Kurisu, Mineko Kengaku, Naoko Komura, Hiromune Ando, Yasuhisa Kimura, Noriyuki Kioka, et al. Abca13 dysfunction associated with psychiatric disorders causes impaired cholesterol trafficking. *Journal of Biological Chemistry*, 296: 100166, 2021.

25. Frances M Ashcroft. From molecule to malady. *Nature*, 440(7083):440–447, 2006.

26. Candice Contet, Scott P Goulding, Dika A Kuljis, and AL Barth. Bk channels in the central nervous system. *International review of neurobiology*, 128:281–342, 2016.

27. Jeet Kalia, Mirela Milescu, Juan Salvatierra, Jordan Wagner, Julie K Klint, Glenn F King, Baldomero M Olivera, and Frank Bosmans. From foe to friend: using animal toxins to investigate ion channel function. *Journal of molecular biology*, 427(1):158–175, 2015.

28. Gregory J. Kaczorowski, Owen B. McManus, Birgit T. Priest, and Maria L. Garcia. Ion channels as drug targets: The next GPCRs, may 2008. ISSN 00221295.

29. A. M. Waszkielewicz, A. Gunia, N. Szkaradek, K. Sloczynska, S. Krupinska, and H. Marona. Ion Channels as Drug Targets in Central Nervous System Disorders. *Current Medicinal Chemistry*, 20(10):1241–1285, mar 2013. ISSN 09298673. doi: 10.2174/0929867311320100005.

30. Jie Zheng and Matthew C. Trudeau. *HANDBOOK OF ION CHANNELS*. 2015. ISBN 9781466551428.

31. Sharan K. Bagal, Alan D. Brown, Peter J. Cox, Kiyoyuki Omoto, Robert M. Owen, David C. Pryde, Benjamin Sidders, Sarah E. Skerratt, Edward B. Stevens, R. Ian Storer, and Nigel A. Swain. Ion channels as therapeutic targets: A drug discovery perspective, feb 2013. ISSN 00222623.

32. Declan A Doyle, Joao Morais Cabral, Richard A Pfuetzner, Anling Kuo, Jacqueline M Gulbis, Steven L Cohen, Brian T Chait, and Roderick MacKinnon. The structure of the potassium channel: molecular basis of k+ conduction and selectivity. *science*, 280(5360):69–77, 1998.

33. Qie Kuang, Pasi Purhonen, and Hans Hebert. Structure of potassium channels. *Cellular and molecular life sciences*, 72(19):3677–3693, 2015.

34. Steve W Lockless, Ming Zhou, and Roderick MacKinnon. Structural and Thermodynamic Properties of Selective Ion Binding in a K+ Channel. *PLoS Biology*, 5(5):e121, may 2007. ISSN 1545-7885. doi: 10.1371/journal.pbio.0050121.

35. Gary Yellen. The voltage-gated potassium channels and their relatives. *nature*, 419(6902):35–42, 2002.

36. Stephen B Long, Ernest B Campbell, and Roderick MacKinnon. Crystal structure of a mammalian voltage-dependent shaker family k+ channel. *Science*, 309(5736):897–903, 2005.

37. Sanjay Kumar Aggarwal and Roderick MacKinnon. Contribution of the s4 segment to gating charge in the shaker k+ channel. *Neuron*, 16(6):1169–1177, 1996.

38. Sang-Ah Seoh, Daniel Sigg, Diane M Papazian, and Francisco Bezanilla. Voltage-sensing residues in the s2 and s4 segments of the shaker k+ channel. *neuron*, 16(6):1159–1167, 1996.

39. Ravi C. Balijepalli and Timothy J. Kamp. Caveolae, ion channels and cardiac arrhythmias, oct 2008. ISSN 00796107.

40. Teresa Giraldez and Brad S Rothberg. Understanding the conformational motions of RCK gating rings. *The Journal of General Physiology*, 149(4):431–441, 2017. ISSN 0022-1295. doi: 10.1085/jgp.201611726.

41. Fabrice Duprat, Florian Lesage, Michel Fink, Roberto Reyes, Catherine Heurteaux, and Michel Lazdunski. Task, a human background k+ channel to sense external ph variations near physiological ph. *The EMBO journal*, 16 (17):5464–5471, 1997.

42. Stephen G Brohawn, Zhenwei Su, and Roderick MacKinnon. Mechanosensitivity is mediated directly by the lipid membrane in traak and trek1 k+ channels. *Proceedings of the National Academy of Sciences*, 111(9):3614–3619, 2014.

43. Prafulla Aryal, Viwan Jarerattanachat, Michael V Clausen, Marcus Schewe, Conor McClenaghan, Liam Argent, Linus J Conrad, Yin Y Dong, Ashley CW Pike, Elisabeth P Carpenter, et al. Bilayer-mediated structural transitions control mechanosensitivity of the trek-2 k2p channel. *Structure*, 25(5):708–718, 2017.

44. Hiroshi Hibino, Atsushi Inanobe, Kazuharu Furutani, Shingo Murakami, IAN Findlay, and Yoshihisa Kurachi. Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiological reviews*, 90(1):291–366, 2010.

45. H Sakura, N Wat, V Horton, H Millns, RC Turner, and FM Ashcroft. Sequence variations in the human kir6. 2 gene, a subunit of the beta-cell atp-sensitive k-channel: no association with niddm in white caucasian subjects or evidence of abnormal function when expressed in vitro. *Diabetologia*, 39(10):1233–1236, 1996.

46. Donald W Hilgemann and Rebecca Ball. Regulation of cardiac na+, ca2+ exchange and katp potassium channels by pip2. *Science*, 273(5277):956–959, 1996.

47. Valentina Corradi, Besian I. Sejdiu, Haydee Mesa-Galloso, Haleh Abdizadeh, Sergei Yu Noskov, Siewert J. Marrink, and D. Peter Tieleman. Emerging Diversity in Lipid-Protein Interactions, may 2019. ISSN 15206890.

48. Heidi Koldsø and Mark S.P. Sansom. Local lipid reorganization by a transmembrane protein domain. *Journal of Physical Chemistry Letters*, 3(23):3498–3502, dec 2012. ISSN 19487185. doi: 10.1021/jz301570w.

49. Valentina Corradi, Eduardo Mendez-Villuendas, Helgi I. Ingólfsson, Ruo-xu Gu, Iwona Siuda, Manuel N Melo, Anastassiia Moussatova, Lucien J. DeGagné, Besian I. Sejdiu, Gurpreet Singh, Tsjerk A. Wassenaar, Karelia Delgado Magnero, Siewert J. Marrink, and D. Peter Tieleman. Lipid–Protein Interactions Are Unique Fingerprints for Membrane Proteins. *ACS Central Science*, page acscentsci.8b00143, 2018. ISSN 2374-7943. doi: 10.1021/acscentsci.8b00143.

50. George Hedger and Mark SP Sansom. Lipid interaction sites on channels, transporters and receptors: recent insights from molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1858(10):2390–2400, 2016.

51. Anna L Duncan, Wanling Song, and Mark SP Sansom. Lipid-dependent regulation of ion channels and g protein–coupled receptors: insights from structures and simulations. *Annual review of pharmacology and toxicology*, 60:31–50, 2020.

52. Jani Reddy Bolla, Mark T. Agasid, Shahid Mehmood, and Carol V. Robinson. Membrane Protein–Lipid Interactions Probed Using Mass Spectrometry. *Annual Review of Biochemistry*, 88(1):85–111, jun 2019. ISSN 0066-4154. doi: 10.1146/annurev-biochem-013118-111508.

53. Ünal Coskun, Michał Grzybek, David Drechsel, and Kai Simons. Regulation of human EGF receptor by lipids. *Proceedings of the National Academy of Sciences of the United States of America*, 108(22):9044–9048, may 2011. ISSN 00278424. doi: 10.1073/pnas.1105666108.

54. M Gertrude Gutierrez, Kylee S Mansfield, and Noah Malmstadt. The functional activity of the human serotonin 5-HT1A receptor is controlled by lipid bilayer composition. *Biophysical Journal*, 110(11):2486–2495, 2016. ISSN 1542-0086. doi: 10.1016/j.bpj.2016.04.042. URL `http://www.ncbi.nlm.nih.gov/pubmed/27276266`.

55. RJ Froud, CRA Earl, JM East, and AG Lee. Effects of lipid fatty acyl chain structure on the activity of the (ca2++ mg2+)-atpase. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 860(2):354–360, 1986.

56. RJ Froud, JM East, OT Jones, AG Lee, and AG Lee. Effects of lipids and long-chain alkyl derivatives on the activity of calcium-magnesium atpase. *Biochemistry*, 25(23):7544–7552, 1986.

57. AP Starling, JM East, and AG Lee. Effects of phosphatidylcholine fatty acyl chain length on calcium binding and other functions of the calcium-magnesium-atpase. *Biochemistry*, 32(6):1593–1600, 1993.

58. Anthony G. Lee. Lipid-protein interactions. In *Biochemical Society Transactions*, volume 39, pages 761–766. Portland Press, jun 2011. doi: 10.1042/BST0390761.

59. Tim Rasmussen. How do mechanosensitive channels sense membrane tension? *Biochemical Society Transactions*, 44(4):1019–1025, aug 2016. ISSN 14708752. doi: 10.1042/BST20160018.

60. Charles D. Cox, Navid Bavi, and Boris Martinac. Bacterial Mechanosensors, feb 2018. ISSN 15451585.

61. Linda Volkers, Yasmine Mechioukhi, and Bertrand Coste. Piezo channels: from structure to function, jan 2015. ISSN 14322013.

62. Chérine Bechara, Anne Nöll, Nina Morgner, Matteo T. Degiacomi, Robert Tampé, and Carol V. Robinson. A subset of annular lipids is linked to the flippase activity of an ABC transporter. *Nature Chemistry*, 7(3):255–262, feb 2015. ISSN 17554349. doi: 10.1038/nchem.2172.

63. Richard K Hite, Zongli Li, and Thomas Walz. Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. *The EMBO Journal*, 29(10):1652–1658, may 2010. ISSN 0261-4189. doi: 10.1038/emboj.2010.68.

64. Jacques Fantini and Francisco Jose Barrantes. How membrane lipids control the 3d structure and function of receptors. 2018.

65. Kallol Gupta, Joseph A. C. Donlan, Jonathan T. S. Hopper, Povilas Uzdavinys, Michael Landreh, Weston B. Struwe, David Drew, Andrew J. Baldwin, Phillip J. Stansfeld, and Carol V. Robinson. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature*, 541(7637):1–4, 2017. ISSN 0028-0836. doi: 10.1038/nature20820.

66. Coeli M.B. Lopes, Hailin Zhang, Tibor Rohacs, Taihao Jin, Jian Yang, and Diomedes E. Logothetis. Alterations in conserved Kir channel-PIP2 interactions underlie channelopathies. *Neuron*, 34(6):933–944, jun 2002. ISSN 08966273. doi: 10.1016/S0896-6273(02)00725-0.

67. Shozeb Haider, Andrei I Tarasov, Tim J Craig, Mark S P Sansom, and Frances M Ashcroft. Identification of the PIP2-binding site on Kir6.2 by molecular modelling and functional analysis. *The EMBO Journal*, 26(16):3749–3759, aug 2007. ISSN 0261-4189. doi: 10.1038/sj.emboj.7601809.

68. Emre Lacin, Prafulla Aryal, Ian W. Glaaser, Karthik Bodhinathan, Eric Tsai, Nidaa Marsh, Stephen J. Tucker, Mark S.P. Sansom, and Paul A. Slesinger. Dynamic role of the tether helix in PIP2-dependent gating of a G protein-gated potassium channel. *Journal of General Physiology*, 149(8):799–811, aug 2017. ISSN 15407748. doi: 10.1085/jgp.201711801.

69. Decha Enkvetchakul, Iana Jeliazkova, and Colin G Nichols. Direct modulation of kir channel gating by membrane phosphatidylinositol 4, 5-bisphosphate. *Journal of Biological Chemistry*, 280(43):35785–35788, 2005.

70. Jean Chemin, Amanda Jane Patel, Fabrice Duprat, Inger Lauritzen, Michel Lazdunski, and Eric Honoré. A phospholipid sensor controls mechanogating of the k+ channel trek-1. *The EMBO journal*, 24(1):44–53, 2005.

71. Coeli MB Lopes, Tibor Rohács, Gábor Czirják, Tamás Balla, Péter Enyedi, and Diomedes E Logothetis. Pip2 hydrolysis underlies agonist-induced inhibition and regulates voltage gating of two-pore domain k+ channels. *The Journal of physiology*, 564(1):117–129, 2005.

72. Jean Chemin, Amanda Jane Patel, Fabrice Duprat, Frederick Sachs, Michel Lazdunski, and Eric Honore. Up- and down-regulation of the mechano-gated K2P channel TREK-1 by PIP2 and other membrane phospholipids, oct 2007. ISSN 00316768.

73. Mayra Delgado-Ramírez, José J. De Jesús-Pérez, Iván A. Aréchiga-Figueroa, Jorge Arreola, Scott K. Adney, Carlos A. Villalba-Galea, Diomedes E. Logothetis, and Aldo A. Rodríguez-Menchaca. Regulation of Kv2.1 channel inactivation by phosphatidylinositol 4,5-bisphosphate. *Scientific Reports*, 8(1):1–13, dec 2018. ISSN 20452322. doi: 10.1038/s41598-018-20280-w.

74. Mark A. Zaydman and Jianmin Cui. PIP2 regulation of KCNQ channels: Biophysical and molecular mechanisms for lipid modulation of voltage-dependent gating, may 2014. ISSN 1664042X.

75. Jin Song Bian, Anna Kagan, and Thomas V. McDonald. Molecular analysis of PIP2 regulation of HERG and IKr. *American Journal of Physiology - Heart and Circulatory Physiology*, 287(5 56-5):2154–2163, nov 2004. ISSN 03636135. doi: 10.1152/ajpheart.00120.2004.

76. Fayal Abderemane-Ali, Zeineb Es-Salah-Lamoureux, Lucie Delemotte, Marina A. Kasimova, Alain J. Labro, Dirk J. Snyders, David Fedida, Mounir Tarek, Isabelle Baró, and Gildas Loussouarn. Dual effect of phosphatidyl (4,5)-bisphosphate PIP2on shaker K+ channels. *Journal of Biological Chemistry*, 287 (43):36158–36167, oct 2012. ISSN 00219258. doi: 10.1074/jbc.M112.382085.

77. Mark A Zaydman, Jonathan R Silva, Kelli Delaloye, Yang Li, Hongwu Liang, H Peter Larsson, Jingyi Shi, and Jianmin Cui. Kv7. 1 ion channels require a lipid to couple voltage sensing to pore opening. *Proceedings of the National Academy of Sciences*, 110(32):13180–13185, 2013.

78. Ji Sun and Roderick MacKinnon. Structural basis of human kcnq1 modulation and gating. *Cell*, 180(2):340–347, 2020.

79. Yongfeng Liu, Xianjin Xu, Junyuan Gao, Moawiah M Naffaa, Hongwu Liang, Jingyi Shi, Hong Zhan Wang, Nien-Du Yang, Panpan Hou, Wenshan Zhao, et al. A pip 2 substitute mediates voltage sensor-pore coupling in kcnq activation. *Communications biology*, 3(1):1–12, 2020.

80. Marina A Kasimova, Mark A Zaydman, Jianmin Cui, and Mounir Tarek. Pip 2-dependent coupling is prominent in kv7. 1 due to weakened interactions between s4-s5 and s6. *Scientific reports*, 5(1):1–9, 2015.

81. Aldo A Rodriguez-Menchaca, Scott K Adney, Qiong-Yao Tang, Xuan-Yu Meng, Avia Rosenhouse-Dantsker, Meng Cui, and Diomedes E Logothetis. Pip2 controls voltage-sensor movement and pore opening of kv channels through the s4–s5 linker. *Proceedings of the National Academy of Sciences*, 109(36):E2399–E2408, 2012.

82. Thirumalini Vaithianathan, Anna Bukiya, Jianxi Liu, Penchong Liu, Maria Asuncion-Chin, Zheng Fan, and Alejandro Dopico. Direct regulation of BK channels by phosphatidylinositol 4,5-bisphosphate as a novel signaling pathway. *Journal of General Physiology*, 132(1):13–28, jul 2008. ISSN 00221295. doi: 10.1085/jgp.200709913.

83. Qiong Yao Tang, Zhe Zhang, Xuan Yu Meng, Meng Cui, and Diomedes E. Logothetis. Structural determinants of phosphatidylinositol 4,5-bisphosphate (PIP2) regulation of BK channel activity through the RCK1 Ca2+ coordination site. *Journal of Biological Chemistry*, 289(27):18860–18872, jul 2014. ISSN 1083351X. doi: 10.1074/jbc.M113.538033.

84. Alex M. Dopico and Anna N. Bukiya. Lipid regulation of BK channel function. *Frontiers in Physiology*, 5 AUG(August):1–7, 2014. ISSN 1664042X. doi: 10.3389/fphys.2014.00312.

85. Gang Yue, Bela Malik, Guichin Yue, and Douglas C. Eaton. Phosphatidylinositol 4,5-bisphosphate (PIP2) stimulates epithelial sodium channel activity in A6 cells. *Journal of Biological Chemistry*, 277(14):11965–11969, apr 2002. ISSN 00219258. doi: 10.1074/jbc.M108951200.

86. Crystal R. Archer, Benjamin T. Enslow, Chase M. Carver, and James D. Stockand. Phosphatidylinositol 4,5-bisphosphate directly interacts with the β and γ subunits of the sodium channel ENaC. *Journal of Biological Chemistry*, 295(23):7958–7969, apr 2020. ISSN 1083351X. doi: 10.1074/JBC.RA120.012606.

87. Qiusheng Tong, Nikita Gamper, Jorge L. Medina, Mark S. Shapiro, and James D. Stockand. Direct activation of the epithelial Na+ channel by phosphatidylinositol 3,4,5-trisphosphate and phosphatidylinositol 3,4-bisphosphate produced by phosphoinositide 3-OH kinase. *Journal of Biological Chemistry*, 279(21):22654–22663, may 2004. ISSN 00219258. doi: 10.1074/jbc.M401004200.

88. Nazzareno D'Avanzo, Emily C. McCusker, Andrew M. Powl, Andrew J. Miles, Colin G. Nichols, and B. A. Wallace. Differential Lipid Dependence of the Function of Bacterial Sodium Channels. *PLoS ONE*, 8(4):e61216, apr 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0061216.

89. Victor G Romanenko, Yun Fang, Fitzroy Byfield, Alexander J Travis, Carol A Vandenberg, George H Rothblat, and Irena Levitan. Cholesterol sensitivity and lipid raft targeting of kir2. 1 channels. *Biophysical journal*, 87(6):3850–3861, 2004.

90. Nazzareno D'Avanzo, Krzysztof Hyrc, Decha Enkvetchakul, Douglas F Covey, and Colin G Nichols. Enantioselective protein-sterol interactions mediate regulation of both prokaryotic and eukaryotic inward rectifier k+ channels by cholesterol. *PLoS One*, 6(4):e19393, 2011.

91. Valentin K Gribkoff, John E Starrett Jr, and Steven I Dworetzky. Maxi-k potassium channels: form, function, and modulation of a class of endogenous regulators of intracellular calcium. *The Neuroscientist*, 7(2):166–177, 2001.

92. Florina Zakany, Tamas Kovacs, Gyorgy Panyi, and Zoltan Varga. Direct and indirect cholesterol effects on membrane proteins with special focus on potassium channels. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1865(8):158706, 2020.

93. Florina Zakany, Pal Pap, Ferenc Papp, Tamas Kovacs, Peter Nagy, Maria Peter, Lajos Szente, Gyorgy Panyi, and Zoltan Varga. Determining the target of membrane sterols on voltage-gated potassium channels. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1864(3):312–325, 2019.

94. Péter Hajdú, Zoltán Varga, Carlo Pieri, György Panyi, and Rezső Gáspár. Cholesterol modifies the gating of kv1. 3 in human t lymphocytes. *Pflügers Archiv*, 445(6):674–682, 2003.

95. V Bolotina, V Omelyanenko, B Heyes, U Ryan, and P Bregestovski. Variations of membrane cholesterol alter the kinetics of ca 2+-dependent k+ channels and membrane fluidity in vascular smooth muscle cells. *Pflügers Archiv*, 415(3): 262–268, 1989.

96. HM Chang, R Reitstetter, RP Mason, and R Gruener. Attenuation of channel kinetics and conductance by cholesterol: an interpretation using structural stress as a unifying concept. *The Journal of membrane biology*, 143(1):51–63, 1995.

97. Anna N Bukiya, Jitendra D Belani, Scott Rychnovsky, and Alex M Dopico. Specificity of cholesterol and analogs to modulate bk channels points to direct sterol–channel protein interactions. *Journal of General Physiology*, 137(1):93–110, 2011.

98. Xiaolian Xu and Erwin London. The effect of sterol structure on membrane lipid domains reveals how cholesterol can induce lipid domain formation. *Biochemistry*, 39(5):843–849, 2000.

99. Gerald Gimpl, Katja Burger, and Falk Fahrenholz. Cholesterol as modulator of receptor function. *Biochemistry*, 36(36):10959–10974, 1997.

100. Çagla Eroglu, Britta Brügger, Felix Wieland, and Irmgard Sinning. Glutamate-binding affinity of drosophila metabotropic glutamate receptor is modulated by association with lipid rafts. *Proceedings of the National Academy of Sciences*, 100(18):10219–10224, 2003.

101. Sandeep Shrivastava, Thomas J Pucadyil, Yamuna Devi Paila, Sourav Ganguly, and Amitabha Chattopadhyay. Chronic cholesterol depletion using statin impairs the function and dynamics of human serotonin1a receptors. *Biochemistry*, 49(26):5426–5435, 2010.

102. Huixian Wu, Chong Wang, Karen J Gregory, Gye Won Han, Hyekyung P Cho, Yan Xia, Colleen M Niswender, Vsevolod Katritch, Jens Meiler, Vadim Cherezov, et al. Structure of a class c gpcr metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science*, 344(6179):58–64, 2014.

103. Wei Liu, Daniel Wacker, Cornelius Gati, Gye Won Han, Daniel James, Dingjie Wang, Garrett Nelson, Uwe Weierstall, Vsevolod Katritch, Anton Barty, et al. Serial femtosecond crystallography of g protein–coupled receptors. *Science*, 342 (6165):1521–1524, 2013.

104. Antoine Koehl, Hongli Hu, Dan Feng, Bingfa Sun, Yan Zhang, Michael J Robertson, Matthew Chu, Tong Sun Kobilka, Toon Laeremans, Jan Steyaert, et al. Structural insights into the activation of metabotropic glutamate receptors. *Nature*, 566(7742):79–84, 2019.

105. Xavier Prasanna, Durba Sengupta, and Amitabha Chattopadhyay. Cholesterol-dependent conformational plasticity in gpcr dimers. *Scientific reports*, 6(1): 1–12, 2016.

106. Yasuhisa Kimura, Noriyuki Kioka, Hiroaki Kato, Michinori Matsuo, and Kazumitsu Ueda. Modulation of drug-stimulated atpase activity of human mdr1/p-glycoprotein by cholesterol. *Biochemical Journal*, 401(2):597–605, 2007.

107. Zsuzsanna Gutay-Tóth, Ferenc Fenyvesi, Orsolya Bársony, Lajos Szente, Katalin Goda, Gábor Szabó, and Zsolt Bacsó. Cholesterol-dependent conformational changes of p-glycoprotein are detected by the 15d3 monoclonal antibody. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1861(3):188–195, 2016.

108. Tanadet Pipatpolkai, Samuel Usher, Phillip J Stansfeld, and Frances M Ashcroft. New insights into k atp channel gene mutations and neonatal diabetes mellitus. *Nature Reviews Endocrinology*, 16(7):378–393, 2020.

109. Jie Wu, Wei-Guang Ding, and Minoru Horie. Molecular pathogenesis of long qt syndrome type 1. *Journal of arrhythmia*, 32(5):381–388, 2016.

110. Takeshi Murata, Ichiro Yamato, Yoshimi Kakinuma, Andrew GW Leslie, and John E Walker. Structure of the rotor of the v-type na+-atpase from enterococcus hirae. *Science*, 308(5722):654–659, 2005.

111. Min Zhou, Nina Morgner, Nelson P Barrera, Argyris Politis, Shoshanna C Isaacson, Dijana Matak-Vinković, Takeshi Murata, Ricardo A Bernal, Daniela Stock, and Carol V Robinson. Mass spectrometry of intact v-type atpases reveals bound lipids and the effects of nucleotide binding. *Science*, 334(6054): 380–385, 2011.

112. Caroline Koshy and Christine Ziegler. Structural insights into functional lipid–protein interactions in secondary transporters. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850(3):476–487, 2015.

113. Hartmut Luecke, Brigitte Schobert, Hans-Thomas Richter, Jean-Philippe Cartailler, and Janos K Lanyi. Structure of bacteriorhodopsin at 1.55 å resolution. *Journal of molecular biology*, 291(4):899–911, 1999.

114. Stephen B Long, Xiao Tao, Ernest B Campbell, and Roderick MacKinnon. Atomic structure of a voltage-dependent k+ channel in a lipid membrane-like environment. *Nature*, 450(7168):376–382, 2007.

115. Olivier Soubias and Klaus Gawrisch. Probing specific lipid- protein interaction by saturation transfer difference nmr spectroscopy. *Journal of the American Chemical Society*, 127(38):13110–13111, 2005.

116. Binyong Liang and Lukas K Tamm. Nmr as a tool to investigate the structure, dynamics and function of membrane proteins. *Nature structural & molecular biology*, 23(6):468, 2016.

117. Gaddafi I Danmaliki and Peter M Hwang. Solution nmr spectroscopy of membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, page 183356, 2020.

118. Christophe Chipot, François Dehez, Jason R Schnell, Nicole Zitzmann, Eva Pebay-Peyroula, Laurent J Catoire, Bruno Miroux, Edmund RS Kunji, Gianluigi Veglia, Timothy A Cross, et al. Perturbations of native membrane protein structure in alkyl phosphocholine detergents: a critical assessment of nmr and biophysical studies. *Chemical reviews*, 118(7):3559–3607, 2018.

119. Stefan Bibow and Sebastian Hiller. A guide to quantifying membrane protein dynamics in lipids and other native-like environments by solution-state nmr spectroscopy. *The FEBS journal*, 286(9):1610–1623, 2019.

120. Yun Huang, Xiaoyu Wang, Guohua Lv, Asghar M Razavi, Gerard HM Huysmans, Harel Weinstein, Clay Bracken, David Eliezer, and Olga Boudker. Use of paramagnetic 19 f nmr to monitor domain movement in a glutamate transporter homolog. *Nature chemical biology*, 16(9):1006–1012, 2020.

121. Daniel Huster. Solid-state nmr spectroscopy to study protein–lipid interactions. *Biochimica Et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1841(8):1146–1160, 2014.

122. Xiao-Chen Bai, Greg McMullan, and Sjors HW Scheres. How cryo-em is revolutionizing structural biology. *Trends in biochemical sciences*, 40(1):49–57, 2015.

123. Daniel Cressey and Ewen Callaway. Cryo-electron microscopy wins chemistry nobel. *Nature News*, 550(7675):167, 2017.

124. Jean-Paul Renaud, Ashwin Chari, Claudio Ciferri, Wen-ti Liu, Hervé-William Rémigy, Holger Stark, and Christian Wiesmann. Cryo-em in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery*, 17 (7):471–492, 2018.

125. Yuan Gao, Erhu Cao, David Julius, and Yifan Cheng. Trpv1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature*, 534(7607): 347–351, 2016.

126. Ji She, Jiangtao Guo, Qingfeng Chen, Weizhong Zeng, Youxing Jiang, and Xiao-chen Bai. Structural insights into the voltage and phospholipid activation of the mammalian tpc1 channel. *Nature*, 556(7699):130–134, 2018.

127. Scott M Jackson, Ioannis Manolaridis, Julia Kowal, Melanie Zechner, Nicholas MI Taylor, Manuel Bause, Stefanie Bauer, Ruben Bartholomaeus, Guenther Bernhardt, Burkhard Koenig, et al. Structural basis of small-molecule inhibition of human multidrug transporter abcg2. *Nature structural & molecular biology*, 25(4):333–340, 2018.

128. Mark T. Agasid and Carol V. Robinson. Probing membrane protein–lipid interactions. *Current Opinion in Structural Biology*, 69:78–85, aug 2021. ISSN 0959440X. doi: 10.1016/j.sbi.2021.03.010.

129. Chloe Martens and Argyris Politis. A glimpse into the molecular mechanism of integral membrane proteins through hydrogen–deuterium exchange mass spectrometry. *Protein Science*, 29(6):1285–1301, 2020. ISSN 1469896X. doi: 10.1002/pro.3853.

130. Michael T Marty. Nanodiscs and mass spectrometry: Making membranes fly. *International Journal of Mass Spectrometry*, 458:116436, 2020.

131. Chloe Martens, Mrinal Shekhar, Antoni J. Borysik, Andy M. Lau, Eamonn Reading, Emad Tajkhorshid, Paula J. Booth, and Argyris Politis. Direct protein-lipid interactions shape the conformational landscape of secondary transporters. *Nature Communications*, 9(1):4151, 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06704-1. URL `http://www.nature.com/articles/s41467-018-06704-1`.

132. Chloe Martens, Richard A Stein, Matthieu Masureel, Aurelie Roth, Smriti Mishra, Rosie Dawaliby, Albert Konijnenberg, Frank Sobott, Cedric Govaerts, and Hassane S Mchaourab. Lipids modulate the conformational dynamics of a secondary multidrug transporter. *Nature structural & molecular biology*, 23 (8):744–751, 2016.

133. Eamonn Reading, Zoe Hall, Chloe Martens, Tabasom Haghighi, Heather Findlay, Zainab Ahdash, Argyris Politis, and Paula J Booth. Interrogating membrane protein conformational dynamics within native lipid compositions. *Angewandte Chemie International Edition*, 56(49):15654–15657, 2017.

134. S-L Shyng and CG Nichols. Membrane phospholipid control of nucleotide sensitivity of katp channels. *Science*, 282(5391):1138–1141, 1998.

135. Thomas Baukrowitz, Uwe Schulte, Dominik Oliver, Stefan Herlitze, Tobias Krauter, Stephen J Tucker, J Peter Ruppersberg, and Bernd Fakler. Pip2 and pip as determinants for atp inhibition of katp channels. *Science*, 282(5391): 1141–1144, 1998.

136. Tibor Rohács, Coeli Lopes, Tooraj Mirshahi, Taihao Jin, Hailin Zhang, and Diomedes E Logothetis. Assaying phosphatidylinositol bisphosphate regulation of potassium channels. *Methods in enzymology*, 345:71–92, 2002.

137. Eugene Gabev, John Kasianowicz, Tammy Abbott, and Stuart McLaughlin. Binding of neomycin to phosphatidylinositol 4, 5-biphosphate (pip2). *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 979(1):105–112, 1989.

138. David J Posson, Radda Rusinova, Olaf S Andersen, and Crina M Nimigean. Stopped-flow fluorometric ion flux assay for ligand-gated ion channel studies. In *Potassium Channels*, pages 223–235. Springer, 2018.

139. Hai-bo Yu, Min Li, Wei-ping Wang, and Xiao-liang Wang. High throughput screening technologies for ion channels. *Acta Pharmacologica Sinica*, 37(1): 34–43, 2016.

140. Jana Kusch and Giovanni Zifarelli. Patch-clamp fluorometry: electrophysiology meets fluorescence. *Biophysical journal*, 106(6):1250–1257, 2014.

141. Cerrone Cabanos, Miao Wang, Xianlin Han, and Scott B Hansen. A soluble fluorescent binding assay reveals pip2 antagonism of trek-1 channels. *Cell reports*, 20(6):1287–1294, 2017.

142. E Nicholas Petersen, Mahmud Arif Pavel, Hao Wang, and Scott B Hansen. Disruption of palmitate-mediated localization; a shared pathway of force and anesthetic activation of trek-1 channels. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1862(1):183091, 2020.

143. A. J. Dempster. A new method of positive ray analysis. *Physical Review*, 11 (4):316–325, apr 1918. ISSN 0031899X. doi: 10.1103/PhysRev.11.316.

144. F.W. Aston. LXXIV. A positive ray spectrograph . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 38(228):707–714, dec 1919. ISSN 1941-5982. doi: 10.1080/14786441208636004.

145. Nobel Foundation. Nobel prize in chemistry 1922, 1922. URL `https://www.nobelprize.org/prizes/chemistry/1922/summary/`.

146. William E. Parkins. The uranium bomb, the calutron, and the space-charge problem, may 2005. ISSN 00319228.

147. C. J. Porter, J. H. Beynon, and T. Ast. The modern mass spectrometer—a complete chemical laboratory. *Organic Mass Spectrometry*, 16(3):101–114, mar 1981. ISSN 0030-493X. doi: 10.1002/oms.1210160302. URL `http://doi.wiley.com/10.1002/oms.1210160302`.

148. Malcolm Dole, L. L. Mack, R. L. Hines, Department Of Chemistry, R. C. Mobley, L. D. Ferguson, and M. B. Alice. Molecular beams of macroions. *The Journal of Chemical Physics*, 49(5):2240–2249, sep 1968. ISSN 00219606. doi: 10.1063/1.1670391.

149. Masamichi Yamashita and John B. Fenn. Electrospray ion source. Another variation on the free-jet theme. *Journal of Physical Chemistry*, 88(20):4451–4459, 1984. ISSN 00223654. doi: 10.1021/j150664a002.

150. Viswanatham Katta and Brian T Chait. Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *Journal of the American Chemical Society*, 113(22):8534–8535, 1991.

151. Joseph A Loo. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass spectrometry reviews*, 16(1):1–23, 1997.

152. Adam A Rostom and Carol V Robinson. Detection of the intact groel chaperonin assembly by mass spectrometry. *Journal of the American Chemical Society*, 121(19), 1999.

153. Robert HH van den Heuvel and Albert JR Heck. Native protein mass spectrometry: from intact oligomers to functional machineries. *Current opinion in chemical biology*, 8(5):519–526, 2004.

154. Aneika C. Leney and Albert J.R. Heck. Native Mass Spectrometry: What is in the Name? *Journal of the American Society for Mass Spectrometry*, 28(1): 5–13, jan 2017. ISSN 18791123. doi: 10.1007/s13361-016-1545-3.

155. Peter G Wolynes. Biomolecular folding in vacuo!!!(?). *Proceedings of the National Academy of Sciences of the United States of America*, 92(7):2426, 1995.

156. Justin L P Benesch, Brandon T Ruotolo, Douglas a Simmons, and Carol V Robinson. Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chemical Reviews (Washington, DC, United States)*, 107(Copyright (C) 2014 American Chemical Society (ACS). All Rights Reserved.):3544–3567, 2007. ISSN 0009-2665. doi: 10.1021/cr068289b.

157. Albert J.R. Heck. Native mass spectrometry: A bridge between interactomics and structural biology. *Nature Methods*, 5(11):927–933, 2008. ISSN 15487091. doi: 10.1038/nmeth.1265.

158. Carol V Robinson, Tibor Rohacs, and Scott B Hansen. Tools for understanding nanoscale lipid regulation of ion channels. *Trends in biochemical sciences*, 44 (9):795–806, 2019.

159. Jani R Bolla, Francesco Fiorentino, and Carol V Robinson. Mass spectrometry informs the structure and dynamics of membrane proteins involved in lipid and drug transport. *Current Opinion in Structural Biology*, 70:53–60, 2021.

160. Timothy M. Allison and Mark T. Agasid. Native Protein Mass Spectrometry. In *Methods in Molecular Biology*, volume 2073, pages 287–299. Humana Press Inc., 2020. doi: 10.1007/978-1-4939-9869-2_15.

161. Roslyn M. Bill, Peter J.F. Henderson, So Iwata, Edmund R.S. Kunji, Hartmut Michel, Richard Neutze, Simon Newstead, Bert Poolman, Christopher G. Tate, and Horst Vogel. Overcoming barriers to membrane protein structure determination, apr 2011. ISSN 10870156.

162. Nelson P. Barrera, Natalie Di Bartolo, Paula J. Booth, and Carol V. Robinson. Micelles protect membrane complexes from solution to vacuum. *Science*, 321 (5886):243–246, jul 2008. ISSN 00368075. doi: 10.1126/science.1159292.

163. Nicholas G. Housden, Jonathan T.S. Hopper, Natalya Lukoyanova, David Rodriguez-Larrea, Justyna A. Wojdyla, Alexander Klein, Renata Kaminska, Hagan Bayley, Helen R. Saibil, Carol V. Robinson, and Colin Kleanthous. Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science*, 340(6140):1570–1574, jun 2013. ISSN 10959203. doi: 10.1126/science.1237864.

164. Arthur Laganowsky, Eamonn Reading, Timothy M. Allison, Martin B. Ulmschneider, Matteo T. Degiacomi, Andrew J. Baldwin, and Carol V. Robinson. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature*, 510(7503):172–175, 2014. ISSN 0028-0836. doi: 10.1038/nature13419.

165. Arthur Laganowsky, Eamonn Reading, Jonathan T S Hopper, and Carol V Robinson. Mass Spectrometry of Intact Membrane Protein Complexes. *Nature protocols*, 8(4):639–651, 2014. ISSN 1750-2799. doi: 10.1038/nprot.2013.024. Mass. URL `http://www.ncbi.nlm.nih.gov/pubmed/23471109`.

166. Joseph Gault, Joseph A C Donlan, Idlir Liko, Jonathan T S Hopper, Kallol Gupta, Nicholas G Housden, Weston B Struwe, Michael T Marty, Todd Mize, Cherine Bechara, Ya Zhu, Beili Wu, Colin Kleanthous, Mikhail Belov, Eugen Damoc, Alexander Makarov, and Carol V Robinson. High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nature Methods*, 13(4):333–336, 2016. ISSN 1548-7091. doi: 10.1038/nmeth.3771. URL `http://www.nature.com/doifinder/10.1038/nmeth.3771`.

167. Kallol Gupta, Jingwen Li, Idlir Liko, Joseph Gault, Cherine Bechara, Di Wu, Jonathan T.S. Hopper, Kevin Giles, Justin L.P. Benesch, and Carol V. Robinson. Identifying key membrane protein lipid interactions using mass spectrometry. *Nature Protocols*, 13(5):1106–1120, 2018. ISSN 17502799. doi: 10.1038/nprot.2018.014.

168. Kallol Gupta, Joseph A. C. Donlan, Jonathan T. S. Hopper, Povilas Uzdavinys, Michael Landreh, Weston B. Struwe, David Drew, Andrew J. Baldwin, Phillip J. Stansfeld, and Carol V. Robinson. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature*, 541(7637):1–4, 2017. ISSN 0028-0836. doi: 10.1038/nature20820. URL `http://www.nature.com/doifinder/10.1038/nature20820`.

169. Hsin Yung Yen, Kin Kuan Hoi, Idlir Liko, George Hedger, Michael R. Horrell, Wanling Song, Di Wu, Philipp Heine, Tony Warne, Yang Lee, Byron Carpenter, Andreas Plückthun, Christopher G. Tate, Mark S.P. Sansom, and Carol V. Robinson. PtdIns(4,5)P2stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature*, 559(7714):423–427, 2018. ISSN 14764687. doi: 10.1038/s41586-018-0325-6.

170. Hsin-Yung Yen, Jonathan TS Hopper, Idlir Liko, Timothy M Allison, Ya Zhu, Dejian Wang, Monika Stegmann, Shabaz Mohammed, Beili Wu, and Carol V Robinson. Ligand binding to a g protein–coupled receptor captured in a mass spectrometer. *Science advances*, 3(6):e1701016, 2017.

171. Eamonn Reading, Troy A. Walton, Idlir Liko, Michael T. Marty, Arthur Laganowsky, Douglas C. Rees, and Carol V. Robinson. The effect of detergent, temperature, and lipid on the oligomeric state of MscL constructs: Insights from mass spectrometry. *Chemistry and Biology*, 22(5):593–603, 2015. ISSN 10745521. doi: 10.1016/j.chembiol.2015.04.016.

172. Michael Landreh, Joana Costeira-Paulo, Joseph Gault, Erik G Marklund, and Carol V Robinson. Effects of detergent micelles on lipid binding to proteins in electrospray ionization mass spectrometry. *Analytical chemistry*, 89(14): 7425–7430, 2017.

173. Youzhong Guo. Be cautious with crystal structures of membrane proteins or complexes prepared in detergents. *Crystals*, 10(2):86, 2020.

174. Leonhard H. Urner, Idlir Liko, Hsin Yung Yen, Kin Kuan Hoi, Jani Reddy Bolla, Joseph Gault, Fernando Gonçalves Almeida, Marc Philip Schweder, Denis Shutin, Svenja Ehrmann, Rainer Haag, Carol V. Robinson, and Kevin Pagel. Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors. *Nature Communications*, 11(1):1–10, dec 2020. ISSN 20411723. doi: 10.1038/s41467-020-14424-8.

175. James E Keener, Guozhi Zhang, and Michael T Marty. Native mass spectrometry of membrane proteins. *Analytical Chemistry*, 2020.

176. Dror S. Chorev, Lindsay A. Baker, Di Wu, Victoria Beilsten-Edmands, Sarah L. Rouse, Tzviya Zeev-Ben-Mordehai, Chimari Jiko, Firdaus Samsudin, Christoph Gerle, Syma Khalid, Alastair G. Stewart, Stephen J. Matthews, Kay Grünewald, and Carol V. Robinson. Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. *Science*, 362(6416): 829–834, 2018. ISSN 0036-8075. doi: 10.1126/science.aau0976.

177. Timothy H Bayburt, Yelena V Grinkova, and Stephen G Sligar. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano letters*, 2(8):853–856, 2002.

178. Michael T Marty, Kin Kuan Hoi, and Carol V Robinson. Interfacing membrane mimetics with mass spectrometry. *Accounts of chemical research*, 49(11):2459–2467, 2016.

179. Michael T Marty, Kin Kuan Hoi, Joseph Gault, and Carol V Robinson. Probing the lipid annular belt by gas-phase dissociation of membrane proteins in nanodiscs. *Angewandte Chemie International Edition*, 55(2):550–554, 2016.

180. Thirupathi Ravula, Nathaniel Z Hardin, and Ayyalusamy Ramamoorthy. Polymer nanodiscs: Advantages and limitations. *Chemistry and physics of lipids*, 219:45–49, 2019.

181. Michael Overduin and Mansoore Esmaili. Native nanodiscs and the convergence of lipidomics, metabolomics, interactomics and proteomics. *Applied Sciences*, 9(6):1230, 2019.

182. Nils Hellwig, Oliver Peetz, Zainab Ahdash, Igor Tascón, Paula J Booth, Vedrana Mikusevic, Marina Diskowski, Argyris Politis, Yvonne Hellmich, Inga Hänelt, et al. Native mass spectrometry goes more native: investigation of membrane protein complexes directly from smalps. *Chemical Communications*, 54(97):13702–13705, 2018.

183. Dror S. Chorev, Haiping Tang, Sarah L. Rouse, Jani Reddy Bolla, Andriko von Kügelgen, Lindsay A. Baker, Di Wu, Joseph Gault, Kay Grünewald, Tanmay A.M. Bharat, Stephen J. Matthews, and Carol V. Robinson. The use of sonicated lipid vesicles for mass spectrometry of membrane protein complexes. *Nature Protocols*, 15(5):1690–1706, may 2020. ISSN 17502799. doi: 10.1038/s41596-020-0303-y.

184. Judy Hirst, Edmund R.S. Kunji, and John E. Walker. Comment on "Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry", nov 2019. ISSN 10959203. URL `http://science.sciencemag.org/`.

185. Dror S. Chorev and Carol V. Robinson. Response to Comment on "Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry", nov 2019. ISSN 10959203. URL `http://science.sciencemag.org/`.

186. Tom N Durrant and Ingeborg Hers. Pi3k inhibitors in thrombosis and cardiovascular disease. *Clinical and translational medicine*, 9(1):1–21, 2020.

187. Marshall Bern, Tomislav Caval, Yong J. Kil, Wilfred Tang, Christopher Becker, Eric Carlson, Doron Kletter, K. Ilker Sen, Nicolas Galy, Dominique Hagemans, Vojtech Franc, and Albert J.R. Heck. Parsimonious Charge Deconvolution for Native Mass Spectrometry. *Journal of Proteome Research*, 17(3):1216–1226, 2018. ISSN 15353907. doi: 10.1021/acs.jproteome.7b00839.

188. A. G. Ferrige, M. J. Seddon, S. Jarvis, John Skilling, and Robert Aplin. Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry*, 5(8):374–377, aug 1991. ISSN 0951-4198. doi: 10.1002/rcm.1290050810.

189. A. G. Ferrige, M. J. Seddon, B. N. Green, S. A. Jarvis, J. Skilling, and J. Staunton. Disentangling electrospray spectra with maximum entropy. *Rapid Communications in Mass Spectrometry*, 6(11):707–711, nov 1992. ISSN 0951-4198. doi: 10.1002/rcm.1290061115.

190. Michael T. Marty. A Universal Score for Deconvolution of Intact Protein and Native Electrospray Mass Spectra. *Analytical Chemistry*, 92(6):4395–4401, mar 2020. ISSN 15206882. doi: 10.1021/acs.analchem.9b05272.

191. Michael T. Marty, Andrew J. Baldwin, Erik G. Marklund, Georg K.A. Hochberg, Justin L.P. Benesch, and Carol V. Robinson. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Analytical Chemistry*, 87(8):4370–4376, 2015. ISSN 15206882. doi: 10.1021/acs.analchem.5b00140.

192. Deseree J. Reid, Jessica M. Diesing, Matthew A. Miller, Scott M. Perry, Jessica A. Wales, William R. Montfort, and Michael T. Marty. MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra. *Journal of the American Society for Mass Spectrometry*, 30(1):118–127, jan 2019. ISSN 18791123. doi: 10.1007/s13361-018-1951-9.

193. Michael T. Marty. Eliminating Artifacts in Electrospray Deconvolution with a SoftMax Function. *Journal of the American Society for Mass Spectrometry*, 30 (10):2174–2177, oct 2019. ISSN 18791123. doi: 10.1007/s13361-019-02286-4.

194. Iain D.G. Campuzano, John H. Robinson, John O. Hui, Stone D.H. Shi, Chawita Netirojjanakul, Michael Nshanian, Pascal F. Egea, Jennifer L. Lippens, Dhanashri Bagal, Joseph A. Loo, and Marshall Bern. Native and

denaturing ms protein deconvolution for biopharma: monoclonal antibodies and antibody-drug conjugates to polydisperse membrane proteins and beyond. *Analytical Chemistry*, 91(15):9472–9480, aug 2019. ISSN 15206882. doi: 10.1021/acs.analchem.9b00062.

195. Sean P Cleary, Avery M Thompson, and James S Prell. Fourier analysis method for analyzing highly congested mass spectra of ion populations with repeated subunits. *Analytical chemistry*, 88(12):6205–6213, 2016.

196. Sean P Cleary, Huilin Li, Dhanashri Bagal, Joseph A Loo, Iain DG Campuzano, and James S Prell. Extracting charge and mass information from highly congested mass spectra using fourier-domain harmonics. *Journal of the American Society for Mass Spectrometry*, 29(10):2067–2080, 2018.

197. Sean P. Cleary and James S. Prell. Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gábor Transform. *ChemPhysChem*, 20(4):519–523, feb 2019. ISSN 14394235. doi: 10.1002/cphc.201900022.

198. Jonathan Lu, Michael J. Trnka, Soung Hun Roh, Philip J.J. Robinson, Carrie Shiau, Danica Galonic Fujimori, Wah Chiu, Alma L. Burlingame, and Shenheng Guan. Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra from Large Protein Complexes. *Journal of the American Society for Mass Spectrometry*, 26(12):2141–2151, dec 2015. ISSN 18791123. doi: 10.1007/s13361-015-1235-6.

199. Nina Morgner and Carol V. Robinson. Mass ign: An assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Analytical Chemistry*, 84(6):2939–2948, mar 2012. ISSN 00032700. doi: 10.1021/ac300056a.

200. Lukasz G. Migas, Aidan P. France, Bruno Bellina, and Perdita E. Barran. ORIGAMI: A software suite for activated ion mobility mass spectrometry (aIM-MS) applied to multimeric protein assemblies. *International Journal of Mass Spectrometry*, 427:20–28, apr 2018. ISSN 13873806. doi: 10.1016/j.ijms.2017.08.014.

201. Timothy M. Allison, Eamonn Reading, Idlir Liko, Andrew J. Baldwin, Arthur Laganowsky, and Carol V. Robinson. Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nature Communications*, 6:1–10, 2015. ISSN 20411723. doi: 10.1038/ncomms9551.

202. Joseph D. Eschweiler, Jessica N. Rabuck-Gibbons, Yuwei Tian, and Brandon T. Ruotolo. CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Analytical Chemistry*, 87(22):11516–11522, oct 2015. ISSN 15206882. doi: 10.1021/acs.analchem.5b03292.

203. R.P. Feynman, R.B. Leighton, and M.L. Sands. *The Feynman Lectures on Physics*. Addison-Wesley world student series. Addison-Wesley Publishing Company, 1963. ISBN 9780805390490. URL `https://books.google.co.uk/books?id=_6XvAAAAMAAJ`.

204. Enrico Fermi, P Pasta, Stanislaw Ulam, and M Tsingou. Studies of the nonlinear problems. Technical report, Los Alamos Scientific Lab., N. Mex., 1955.

205. Berni Julian Alder and Thomas Everett Wainwright. Phase transition for a hard sphere system. *The Journal of chemical physics*, 27(5):1208–1209, 1957.

206. Berni J Alder and Thomas Everett Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959.

207. Aneesur Rahman. Correlations in the motion of atoms in liquid argon. *Physical review*, 136(2A):A405, 1964.

208. J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977. ISSN 00280836. doi: 10.1038/267585a0.

209. Michael Feig, Isseki Yu, Po Hung Wang, Grzegorz Nawrocki, and Yuji Sugita. Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. *Journal of Physical Chemistry B*, 121(34):8009–8025, aug 2017. ISSN 15205207. doi: 10.1021/acs.jpcb.7b03570.

210. Michael Feig, Grzegorz Nawrocki, Isseki Yu, Po Hung Wang, and Yuji Sugita. Challenges and opportunities in connecting simulations with experiments via molecular dynamics of cellular environments. In *Journal of Physics: Conference Series*, volume 1036, page 12010. Institute of Physics Publishing, jun 2018. doi: 10.1088/1742-6596/1036/1/012010.

211. The Nobel Foundation. The nobel prize in chemistry 2013. `https://www.nobelprize.org/prizes/chemistry/2013/summary/`, 2013. Accessed: 2021-03-14.

212. David E Shaw, Martin M Deneroff, Ron O Dror, Jeffrey S Kuskin, Richard H Larson, John K Salmon, Cliff Young, Brannon Batson, Kevin J Bowers, Jack C Chao, et al. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*, 51(7):91–97, 2008.

213. Joshua A Anderson, Chris D Lorenz, and Alex Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of computational physics*, 227(10):5342–5359, 2008.

214. P Van der Ploeg and HJC Berendsen. Molecular dynamics simulation of a bilayer membrane. *The Journal of Chemical Physics*, 76(6):3271–3276, 1982.

215. Thomas B Woolf and Benoit Roux. Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. *Proceedings of the National Academy of Sciences*, 91(24):11631–11635, 1994.

216. Kenneth Goossens and Hans De Winter. Molecular Dynamics Simulations of Membrane Proteins: An Overview. 2018. doi: 10.1021/acs.jcim.8b00639.

217. Rie Nygaard, Jonathan Kim, and Filippo Mancia. Cryo-electron microscopy analysis of small membrane proteins. *Current Opinion in Structural Biology*, 64:26–33, 2020.

218. Carsten Kutzner, Helmut Grubmüller, Bert L De Groot, and Ulrich Zachariae. Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophysical journal*, 101(4):809–817, 2011.

219. Robin A Corey, Phillip J Stansfeld, and Mark SP Sansom. The energetics of protein–lipid interactions as viewed by molecular simulations. *Biochemical Society Transactions*, 48(1):25–37, 2020.

220. David J Huggins, Philip C Biggin, Marc A Dämgen, Jonathan W Essex, Sarah A Harris, Richard H Henchman, Syma Khalid, Antonija Kuzmanic, Charles A Laughton, Julien Michel, et al. Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 9(3):e1393, 2019.

221. Wanling Song, Hsin Yung Yen, Carol V. Robinson, and Mark S.P. Sansom. State-dependent Lipid Interactions with the A2a Receptor Revealed by MD Simulations Using In Vivo-Mimetic Membranes. *Structure*, 27(2):392–403.e3, feb 2019. ISSN 18784186. doi: 10.1016/j.str.2018.10.024.

222. Tanadet Pipatpolkai, Robin A. Corey, Peter Proks, Frances M. Ashcroft, and Phillip J. Stansfeld. Evaluating inositol phospholipid interactions with inward rectifier potassium channels and characterising their role in disease. *Communications Chemistry*, 3(1):1–10, dec 2020. ISSN 23993669. doi: 10.1038/s42004-020-00391-0.

223. Anna L. Duncan, Wanling Song, and Mark S.P. Sansom. Lipid-dependent regulation of ion channels and g protein-coupled receptors: Insights from structures and simulations, jan 2020. ISSN 15454304.

224. Zhixiong Deng, Bing Yuan, and Kai Yang. Cardiolipin Selectively Binds to the Interface of VsSemiSWEET and Regulates Its Dimerization. *The Journal of Physical Chemistry Letters*, 12:1940–1946, feb 2021. ISSN 1948-7185. doi: 10.1021/acs.jpclett.1c00022.

225. Weria Pezeshkian, Melanie König, Siewert J. Marrink, and John H. Ipsen. A Multi-Scale Approach to Membrane Remodeling Processes. *Frontiers in Molecular Biosciences*, 6:59, jul 2019. ISSN 2296889X. doi: 10.3389/fmolb.2019.00059.

226. Weria Pezeshkian, Melanie König, Tsjerk A Wassenaar, and Siewert J Marrink. Backmapping triangulated surfaces to coarse- grained membrane models. *Nature Communications*, (2020). ISSN 2041-1723. doi: 10.1038/s41467-020-16094-y.

227. Weria Pezeshkian and Siewert J. Marrink. Simulating realistic membrane shapes, aug 2021. ISSN 18790410.

228. Laura Domicevica, Heidi Koldsø, and Philip C Biggin. Multiscale molecular dynamics simulations of lipid interactions with p-glycoprotein in a complex membrane. *Journal of Molecular Graphics and Modelling*, 80:147–156, 2018.

229. Marc A. Dämgen and Philip C. Biggin. Computational methods to examine conformational changes and ligand-binding properties: Examples in neurobiology, may 2019. ISSN 18727972.

230. Marc A Dämgen and Philip C Biggin. A refined open state of the glycine receptor obtained via molecular dynamics simulations. *Structure*, 28(1):130–139, 2020.

231. Matteo Aldeghi, Alexander Heifetz, Michael J Bodkin, Stefan Knapp, and Philip C Biggin. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical science*, 7(1):207–218, 2016.

232. Robin A. Corey, Owen N. Vickery, Mark S.P. Sansom, and Phillip J. Stansfeld. Insights into Membrane Protein-Lipid Interactions from Free Energy Calculations. *Journal of Chemical Theory and Computation*, 15(10):5727–5736, oct 2019. ISSN 15499626. doi: 10.1021/acs.jctc.9b00548.

233. C Arnarez, SJ Marrink, and X Periole. Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. *Scientific reports*, 3(1):1–9, 2013.

234. Clement Arnarez, Jean-Pierre Mazat, Juan Elezgaray, Siewert-J Marrink, and Xavier Periole. Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc 1. *Journal of the American Chemical Society*, 135 (8):3112–3120, 2013.

235. Qinrui Wang, Robin A Corey, George Hedger, Prafulla Aryal, Mariana Grieben, Chady Nasrallah, Agnese Baronina, Ashley CW Pike, Jiye Shi, Elisabeth P Carpenter, et al. Lipid interactions of a ciliary membrane trp channel: Simulation and structural studies of polycystin-2. *Structure*, 28(2):169–184, 2020.

236. Jan Domański, George Hedger, Robert B. Best, Phillip J. Stansfeld, and Mark S.P. Sansom. Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association. *Journal of Physical Chemistry B*, 121(15):3364–3375, 2017. ISSN 15205207. doi: 10.1021/acs.jpcb.6b08445.

237. Matteo Aldeghi, Joseph P Bluck, and Philip C Biggin. Absolute alchemical free energy calculations for ligand binding: A beginner's guide. In *Computational Drug Discovery and Design*, pages 199–232. Springer, 2018.

238. Chen Fan, Nattakan Sukomon, Emelie Flood, Jan Rheinberger, Toby W. Allen, and Crina M. Nimigean. Ball-and-chain inactivation in a calcium-gated potassium channel. *Nature*, 580(September 2019):23–25, 2020. ISSN 0028-0836. doi: 10.1038/s41586-020-2116-0.

239. Laura M Wingler, Meredith A Skiba, Conor McMahon, Dean P Staus, Alissa LW Kleinhenz, Carl-Mikael Suomivuori, Naomi R Latorraca, Ron O Dror, Robert J Lefkowitz, and Andrew C Kruse. Angiotensin and biased analogs induce structurally distinct active conformations within a gpcr. *Science*, 367(6480):888–892, 2020.

240. Duncan Laverty, Philip Thomas, Martin Field, Ole J Andersen, Matthew G Gold, Philip C Biggin, Marc Gielen, and Trevor G Smart. Crystal structures of a gaba a-receptor chimera reveal new endogenous neurosteroid-binding sites. *Nature structural & molecular biology*, 24(11):977, 2017.

241. Ruyu Jia, Chloe Martens, Mrinal Shekhar, Shashank Pant, Grant A Pellowe, Andy M Lau, Heather E Findlay, Nicola J Harris, Emad Tajkhorshid, Paula J Booth, et al. Hydrogen-deuterium exchange mass spectrometry captures distinct dynamics upon substrate and inhibitor binding to a transporter. *Nature communications*, 11(1):1–10, 2020.

242. Idlir Liko, Matteo T Degiacomi, Sejeong Lee, Thomas D Newport, Joseph Gault, Eamonn Reading, Jonathan TS Hopper, Nicholas G Housden, Paul White, Matthew Colledge, et al. Lipid binding attenuates channel closure of the outer membrane protein ompf. *Proceedings of the National Academy of Sciences*, 115(26):6691–6696, 2018.

243. Erwin Schroedinger. An undulatory theory of the mechanics of atoms and molecules. *Physical review*, 28(6):1049, 1926.

244. Max Born and Robert Oppenheimer. Zur quantentheorie der molekeln. *Annalen der physik*, 389(20):457–484, 1927.

245. GROMACS development team. Molecular dynamics simulations. `https://manual.gromacs.org/current/reference-manual/introduction.html`, 2021. Accessed: 2021-03-14.

246. Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

247. Kenno Vanommeslaeghe, Elizabeth Hatcher, Chayan Acharya, Sibsankar Kundu, Shijun Zhong, Jihyun Shim, Eva Darian, Olgun Guvench, P Lopes, Igor Vorobyov, et al. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–690, 2010.

248. Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

249. Paul Robustelli, Stefano Piana, and David E Shaw. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*, 115(21):E4758–E4766, 2018.

250. VS Sandeep Inakollu, Daan P Geerke, Christopher N Rowley, and Haibo Yu. Polarisable force fields: what do they add in biomolecular simulations? *Current opinion in structural biology*, 61:182–190, 2020.

251. Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

252. Alexey Onufriev. Implicit solvent models in molecular dynamics simulations: A brief overview. *Annual Reports in Computational Chemistry*, 4:125–137, 2008.

253. Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations. *Journal of Physical Chemistry B*, 111(June): 7812–7824, 2007. ISSN 1520-6106. doi: 10.1021/jp071097f.

254. Siewert J. Marrink, Alex H. De Vries, and Alan E. Mark. Coarse Grained Model for Semiquantitative Lipid Simulations. *Journal of Physical Chemistry B*, 108(2):750–760, jan 2004. ISSN 15206106. doi: 10.1021/jp036508g.

255. Xavier Periole, Marco Cavalli, Siewert Jan Marrink, and Marco A. Ceruso. Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition. *Journal of Chemical Theory and Computation*, 5(9):2531–2543, 2009. ISSN 15499618. doi: 10.1021/ct9002114.

256. Clément Arnarez, Jaakko J. Uusitalo, Marcelo F. Masman, Helgi I. Ingólfsson, Djurre H. De Jong, Manuel N. Melo, Xavier Periole, Alex H. De Vries, and Siewert J. Marrink. Dry martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *Journal of Chemical Theory and Computation*, 11(1):260–275, 2015. ISSN 15499626. doi: 10.1021/ct500477k.

257. Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert Jan Marrink. The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008. ISSN 15499618. doi: 10.1021/ct700324x.

258. Djurre H. De Jong, Gurpreet Singh, W. F.Drew Bennett, Clement Arnarez, Tsjerk A. Wassenaar, Lars V. Schäfer, Xavier Periole, D. Peter Tieleman, and Siewert J. Marrink. Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation*, 9(1):687–697, jan 2013. ISSN 15499618. doi: 10.1021/ct300646g.

259. Matti Javanainen, Hector Martinez-Seara, and Ilpo Vattulainen. Excessive aggregation of membrane proteins in the martini model. *PLoS One*, 12(11): e0187936, 2017.

260. Riccardo Alessandri, Paulo CT Souza, Sebastian Thallmair, Manuel N Melo, Alex H De Vries, and Siewert J Marrink. Pitfalls of the martini model. *Journal of chemical theory and computation*, 15(10):5448–5460, 2019.

261. Paulo CT Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart MH Bruininks, Tsjerk A Wassenaar, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods*, 18(4):382–388, 2021.

262. Riccardo Alessandri, Fabian Grünewald, and Siewert J. Marrink. The Martini Model in Materials Science. *Advanced Materials*, page 2008635, may 2021. ISSN 0935-9648. doi: 10.1002/adma.202008635.

263. Xavier Periole and Siewert-Jan Marrink. The Martini Coarse-Grained Force Field. pages 533–565. 2013. doi: 10.1007/978-1-62703-017-5_20.

264. Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

265. Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52 (12):7182–7190, 1981.

266. Shuichi Nosé and ML Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5):1055–1076, 1983.

267. Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

268. Paul P Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

269. Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

270. R. W. Hockney, S. P. Goel, and J. W. Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, feb 1974. ISSN 10902716. doi: 10.1016/0021-9991(74)90010-2.

271. Phillip J Stansfeld, Joseph E Goose, Martin Caffrey, Elisabeth P Carpenter, Joanne L Parker, Simon Newstead, and Mark SP Sansom. Memprotmd: automated insertion of membrane protein structures into explicit lipid membranes. *Structure*, 23(7):1350–1361, 2015.

272. Tsjerk A. Wassenaar, Helgi I. Ingólfsson, Rainer A. Böckmann, D. Peter Tieleman, and Siewert J. Marrink. Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. *Journal of Chemical Theory and Computation*, 11(5):2144–2155, 2015. ISSN 15499626. doi: 10.1021/acs.jctc.5b00209.

273. Brooke E. Husic and Vijay S. Pande. Markov State Models: From an Art to a Science, feb 2018. ISSN 15205126.

274. R. B. Best, G. Hummer, and W. A. Eaton. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*, 110(44):17874–17879, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1311599110.

275. Vladimir N. Maiorov and Gordon M. Crippen. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*, 235(2):625–634, 1994. ISSN 00222836. doi: 10.1006/jmbi.1994.1017.

276. J. V. Iribarne and B. A. Thomson. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, 64(6):2287–2294, aug 1976. ISSN 00219606. doi: 10.1063/1.432536.

277. Lars Konermann, Elias Ahadi, Antony D. Rodriguez, and Siavash Vahidi. Unraveling the mechanism of electrospray ionization. *Analytical Chemistry*, 85 (1):2–9, jan 2013. ISSN 00032700. doi: 10.1021/ac302789c.

278. Elnaz Aliyari and Lars Konermann. Formation of gaseous proteins via the ion evaporation model (iem) in electrospray mass spectrometry. *Analytical Chemistry*, 92(15):10807–10814, 2020.

279. J. Fernandez De La Mora. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta*, 406(1):93–104, feb 2000. ISSN 00032670. doi: 10.1016/S0003-2670(99) 00601-7.

280. Haidy Metwally and Lars Konermann. Crown Ether Effects on the Location of Charge Carriers in Electrospray Droplets: Implications for the Mechanism of Protein Charging and Supercharging. *Analytical Chemistry*, 90(6):4126–4134, 2018. ISSN 15206882. doi: 10.1021/acs.analchem.8b00099.

281. David C. Gale and Richard D. Smith. Small volume and low flow-rate electrospray lonization mass spectrometry of aqueous samples. *Rapid Communications in Mass Spectrometry*, 7(11):1017–1021, nov 1993. ISSN 0951-4198. doi: 10.1002/rcm.1290071111.

282. Mark R Emmett and Richard M Caprioli. Micro-Electrospray Mass Spectrometry: Ultra-High-Sensitivity Analysis of Peptides and Proteins. Technical report, 1994.

283. Matthias S. Wilm and Matthias Mann. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes*, 136(2-3):167–180, sep 1994. ISSN 01681176. doi: 10.1016/0168-1176(94)04024-9.

284. Frank Sobott and Carol V. Robinson. Characterising electrosprayed biomolecules using tandem-MS - The noncovalent GroEL chaperonin assembly. *International Journal of Mass Spectrometry*, 236(1-3):25–32, aug 2004. ISSN 13873806. doi: 10.1016/j.ijms.2004.05.010.

285. Adam R. McKay, Brandon T. Ruotolo, Leopold L. Ilag, and Carol V. Robinson. Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *Journal of the American Chemical Society*, 128(35):11433–11442, sep 2006. ISSN 00027863. doi: 10.1021/ja061468q.

286. W Paul and H Steinwedel. A new mass spectrometer without a magnetic field. *Zeitschrift fuer Naturforschung (West Germany) Divided into Z. Nautrforsch., A, and Z. Naturforsch., B: Anorg. Chem., Org. Chem., Biochem., Biophys.*,, 8, 1953.

287. Steven A Hofstadler, Ray Bakhtiar, and Richard D Smith. Electrospray ionization mass spectroscopy: Part i. instrumentation and spectral interpretation. *Journal of Chemical Education*, 73(4):A82, 1996.

288. Frank Sobott, Justin L.P. Benesch, Elizabeth Vierling, and Carol V. Robinson. Subunit exchange of multimeric protein complexes: Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *Journal of Biological Chemistry*, 277(41):38921–38929, oct 2002. ISSN 00219258. doi: 10.1074/jbc.M206060200.

289. Donald J. Douglas, Aaron J. Frank, and Dunmin Mao. Linear ion traps in mass spectrometry, jan 2005. ISSN 02777037.

290. Robert J Cotter. Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Analytical chemistry*, 64(21):1027A–1039A, 1992.

291. Alexander Makarov. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chemistry*, 72(6):1156–1162, 2000. ISSN 00032700. doi: 10.1021/ac991131p.

292. KH Kingdon. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Physical Review*, 21(4):408, 1923.

293. Alexander Makarov, Eduard Denisov, Alexander Kholomeev, Wilko Balschun, Oliver Lange, Kerstin Strupat, and Stevan Horning. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical Chemistry*, 78(7):2113–2120, apr 2006. ISSN 00032700. doi: 10.1021/ac0518811.

294. Roman A. Zubarev and Alexander Makarov. Orbitrap mass spectrometry. *Analytical Chemistry*, 85(11):5288–5296, jun 2013. ISSN 00032700. doi: 10.1021/ac4001223.

295. Avia Rosenhouse-Dantsker, Dolly Mehta, and Irena Levitan. Regulation of ion channels by membrane lipids. *Comprehensive Physiology*, 2(1):31–68, 2012. ISSN 20404603. doi: 10.1002/cphy.c110001.

296. Andrey Dyachenko, Guanbo Wang, Mike Belov, Alexander Makarov, Rob N. De Jong, Ewald T.J. Van Den Bremer, Paul W.H.I. Parren, and Albert J.R. Heck. Tandem Native Mass-Spectrometry on Antibody-Drug Conjugates and Submillion da Antibody-Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Analytical Chemistry*, 87(12):6095–6102, jun 2015. ISSN 15206882. doi: 10.1021/acs.analchem.5b00788.

297. Kyle L. Fort, Michiel Van De Waterbeemd, Dmitriy Boll, Maria Reinhardt-Szyba, Mikhail E. Belov, Eita Sasaki, Reinhard Zschoche, Donald Hilvert, Alexander A. Makarov, and Albert J.R. Heck. Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. *Analyst*, 143(1):100–105, jan 2018. ISSN 13645528. doi: 10.1039/c7an01629h.

298. Yehia M Ibrahim, Sandilya VB Garimella, Spencer A Prost, Roza Wojcik, Randolph V Norheim, Erin S Baker, Ivan Rusyn, and Richard D Smith. Development of an ion mobility spectrometry-orbitrap mass spectrometer platform. *Analytical chemistry*, 88(24):12152–12160, 2016.

299. Joel D Keelor, Stephen Zambrzycki, Anyin Li, Brian H Clowers, and Facundo M Fernandez. Atmospheric pressure drift tube ion mobility–orbitrap mass spectrometry: initial performance characterization. *Analytical chemistry*, 89(21):11301–11309, 2017.

300. Michael L Poltash, Jacob W McCabe, Mehdi Shirzadeh, Arthur Laganowsky, Brian H Clowers, and David H Russell. Fourier transform-ion mobility-orbitrap mass spectrometer: a next-generation instrument for native mass spectrometry. *Analytical chemistry*, 90(17):10472–10478, 2018.

301. Brandon T Ruotolo, Justin L P Benesch, Alan M Sandercock, Suk-Joon Hyung, and Carol V Robinson. Ion mobility–mass spectrometry analysis of large protein complexes. *Nature Protocols*, 3(7):1139–1152, 2008. ISSN 1754-2189. doi: 10.1038/nprot.2008.78.

302. Edward A. Mason and Homer W. Schamp. Mobility of gaseous lons in weak electric fields. *Annals of Physics*, 4(3):233–270, jul 1958. ISSN 1096035X. doi: 10.1016/0003-4916(58)90049-6.

303. Steven D. Pringle, Kevin Giles, Jason L. Wildgoose, Jonathan P. Williams, Susan E. Slade, Konstantinos Thalassinos, Robert H. Bateman, Michael T. Bowers, and James H. Scrivens. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry*, 261(1): 1–12, mar 2007. ISSN 13873806. doi: 10.1016/j.ijms.2006.07.021.

304. Kevin Giles, Steven D Pringle, Kenneth R Worthington, David Little, Jason L Wildgoose, and Robert H Bateman. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry*, 18(20):2401–2414, 2004.

305. James N Dodds, Jody C May, and John A McLean. Correlating resolving power, resolution, and collision cross section: unifying cross-platform assessment of separation efficiency in ion mobility spectrometry. *Analytical chemistry*, 89(22):12176–12184, 2017.

306. Alyssa Q Stiving, Benjamin J Jones, Jakub Ujma, Kevin Giles, and Vicki H Wysocki. Collision cross sections of charge-reduced proteins and protein complexes: A database for collision cross section calibration. *Analytical chemistry*, 92(6):4475–4483, 2020.

307. Timothy M Allison, Perdita Barran, Sarah Cianférani, Matteo T Degiacomi, Valérie Gabelica, Rita Grandori, Erik G Marklund, Thomas Menneteau, Lukasz G Migas, Argyris Politis, et al. Computational strategies and challenges for using native ion mobility mass spectrometry in biophysics and structural biology. *Analytical Chemistry*, 92(16):10872–10880, 2020.

308. Zoe Hall, Argyris Politis, Matthew F. Bush, Lorna J. Smith, and Carol V. Robinson. Charge-state dependent compaction and dissociation of protein complexes: Insights from ion mobility and molecular dynamics. *Journal of the American Chemical Society*, 134(7):3429–3438, feb 2012. ISSN 00027863. doi: 10.1021/ja2096859.

309. Justin L P Benesch. Collisional Activation of Protein Complexes: Picking Up the Pieces. *JAM*, 20:341–348, 2009. doi: 10.1016/j.jasms.2008.11.014.

310. Surajith N. Wanasundara and Mark Thachuk. Theoretical Investigations of the Dissociation of Charged Protein Complexes in the Gas Phase. *Journal of the American Society for Mass Spectrometry*, 18(12):2242–2253, dec 2007. ISSN 10440305. doi: 10.1016/j.jasms.2007.09.022.

311. Justin L.P. Benesch, J. Andrew Aquilina, Brandon T. Ruotolo, Frank Sobott, and Carol V. Robinson. Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. *Chemistry and Biology*, 13(6): 597–605, jun 2006. ISSN 10745521. doi: 10.1016/j.chembiol.2006.04.006.

312. Sarah K. Fegan and Mark Thachuk. Controlling dissociation channels of gas-phase protein complexes using charge manipulation. *Journal of the American Society for Mass Spectrometry*, 25(5):722–728, feb 2014. ISSN 18791123. doi: 10.1007/s13361-014-0831-1.

313. Mark Thachuk, Sarah K. Fegan, and Nigare Raheem. Description and control of dissociation channels in gas-phase protein complexes. *Journal of Chemical Physics*, 145(6):065101, aug 2016. ISSN 00219606. doi: 10.1063/1.4960615.

314. R. G. Cooks, D. T. Terwilliger, T. Ast, J. H. Beynon, and T. Keough. Surface Modified Mass Spectrometry, mar 1975. ISSN 15205126.

315. Md A. Mabud, Michael J. Dekrey, and R. Graham Cooks. Surface-induced dissociation of molecular ions. *International Journal of Mass Spectrometry and Ion Processes*, 67(3):285–294, nov 1985. ISSN 01681176. doi: 10.1016/0168-1176(85)83024-X.

316. Dalton T. Snyder, Erin M. Panczyk, Arpad Somogyi, Desmond A. Kaplan, and Vicki Wysocki. Simple and minimally invasive SID devices for native mass spectrometry. *Analytical Chemistry*, 92(16):11195–11203, aug 2020. ISSN 15206882. doi: 10.1021/acs.analchem.0c01657.

317. Verena Grill, Jianwei Shen, Chris Evans, and R. Graham Cooks. Collisions of ions with surfaces at chemically relevant energies: Instrumentation and phenomena, aug 2001. ISSN 00346748.

318. Christopher M. Jones, Richard L. Beardsley, Asiri S. Galhena, Shai Dagan, Guilong Cheng, and Vicki H. Wysocki. Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions. *Journal of the American Chemical Society*, 128(47):15044–15045, nov 2006. ISSN 00027863. doi: 10.1021/ja064586m. URL `http://pubs.acs.org.`

319. Vicki H. Wysocki, Karen E. Joyce, Christopher M. Jones, and Richard L. Beardsley. Surface-Induced Dissociation of Small Molecules, Peptides, and Non-Covalent Protein Complexes. *Journal of the American Society for Mass Spectrometry*, 19(2):190–208, feb 2008. ISSN 10440305. doi: 10.1016/j.jasms.2007.11.005.

320. Sophie R. Harvey, Justin T. Seffernick, Royston S. Quintyn, Yang Song, Yue Ju, Jing Yan, Aniruddha N. Sahasrabuddhe, Andrew Norris, Mowei Zhou, Edward J. Behrman, Steffen Lindert, and Vicki H. Wysocki. Relative interfacial cleavage energetics of protein complexes revealed by surface collisions. *Proceedings of the National Academy of Sciences of the United States of America*, 116(17):8143–8148, apr 2019. ISSN 10916490. doi: 10.1073/pnas.1817632116.

321. Emmanuel D. Levy, Elisabetta Boeri Erba, Carol V. Robinson, and Sarah A. Teichmann. Assembly reflects evolution of protein complexes. *Nature*, 453 (7199):1262–1265, jun 2008. ISSN 0028-0836. doi: 10.1038/nature06942.

322. Gabriel Villar, Alex W. Wilber, Alex J. Williamson, Parvinder Thiara, Jonathan P.K. Doye, Ard A. Louis, Mara N. Jochum, Anna C.F. Lewis, and Emmanuel D. Levy. Self-assembly and evolution of homomeric protein complexes. *Physical Review Letters*, 102(11), mar 2009. ISSN 00319007. doi: 10.1103/PhysRevLett.102.118106.

323. Yang Song, Micah T. Nelp, Vahe Bandarian, and Vicki H. Wysocki. Refining the structural model of a heterohexameric protein complex: Surface induced dissociation and ion mobility provide key connectivity and topology information. *ACS Central Science*, 1(9):477–487, dec 2015. ISSN 23747951. doi: 10.1021/acscentsci.5b00251.

324. Justin T Seffernick, Sophie R Harvey, Vicki H Wysocki, and Steffen Lindert. Predicting Protein Complex Structure from Surface-Induced Dissociation Mass Spectrometry Data. 2019. doi: 10.1021/acscentsci.8b00912.

325. Arthur Laganowsky, Eamonn Reading, Jonathan T S Hopper, and Carol V Robinson. Mass spectrometry of intact membrane protein complexes. *Nature Protocols*, 8(4):639–651, 2013. ISSN 1754-2189. doi: 10.1038/nprot.2013.024. URL `http://www.nature.com/doifinder/10.1038/nprot.2013.024`.

326. David S. Goodsell and Arthur J. Olson. Structural symmetry and protein function, nov 2000. ISSN 10568700.

327. David S Goodsell. Inside a living cell. *Trends in biochemical sciences*, 16: 203–206, 1991.

328. Mayssam H. Ali and Barbara Imperiali. Protein oligomerization: How and why, sep 2005. ISSN 09680896.

329. Yuan Hu Xuan, Yi Bing Hu, Li Qing Chen, Davide Sosso, Daniel C. Ducat, Bi Huei Hou, and Wolf B. Frommer. Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39):E3685–E3694, sep 2013. ISSN 00278424. doi: 10.1073/pnas.1311244110.

330. Salvador Sierra, Rudy Toneatti, and Javier González-Maeso. Class a gpcr oligomerization: reasons of controversy. In *GPCRs*, pages 121–140. Elsevier, 2020.

331. Katarzyna Mlyniec, Dominika Siodłak, Urszula Doboszewska, and Gabriel Nowak. Gpcr oligomerization as a target for antidepressants: Focus on gpr39. *Pharmacology & Therapeutics*, page 107842, 2021.

332. Sue-Hwa Lin and Guido Guidotti. Purification of membrane proteins. *Methods in enzymology*, 463:619–629, 2009.

333. S. Singer, D. Jonathan Losos, K.A. Mason, P. George B. Johnson, and P.H. Raven. *Carbohydrates: Energy storage and structural molecules*. McGraw-Hill Education, 10 edition, 2016.

334. Li Qing Chen, Bi Huei Hou, Sylvie Lalonde, Hitomi Takanaga, Mara L. Hartung, Xiao Qing Qu, Woei Jiun Guo, Jung Gun Kim, William Underwood, Bhavna Chaudhuri, Diane Chermak, Ginny Antony, Frank F. White, Shauna C. Somerville, Mary Beth Mudgett, and Wolf B. Frommer. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*, 468 (7323):527–532, nov 2010. ISSN 00280836. doi: 10.1038/nature09606.

335. Liang Feng and Wolf B. Frommer. Structure and function of SemiSWEET and SWEET sugar transporters. *Trends in Biochemical Sciences*, 40(8):480–486, 2015. ISSN 13624326. doi: 10.1016/j.tibs.2015.05.005.

336. Li Qing Chen, I. Winnie Lin, Xiao Qing Qu, Davide Sosso, Heather E. McFarlane, Alejandra Londoño, A. Lacey Samuels, and Wolf B. Frommer. A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the arabidopsis embryo. *Plant Cell*, 27(3):607–619, 2015. ISSN 1532298X. doi: 10.1105/tpc.114.134585.

337. Zhiguang Jia, Mahdieh Yazdani, Guohui Zhang, Jianmin Cui, and Jianhan Chen. Hydrophobic gating in BK channels. *Nature Communications*, 9(1), dec 2018. ISSN 20411723. doi: 10.1038/s41467-018-05970-3.

338. Yan Xu, Yuyong Tao, Lily S. Cheung, Chao Fan, Li Qing Chen, Sophia Xu, Kay Perry, Wolf B. Frommer, and Liang Feng. Structures of bacterial homologues of SWEET transporters in two distinct conformations. *Nature*, 515(7527):448–452, nov 2014. ISSN 14764687. doi: 10.1038/nature13670.

339. Yongchan Lee, Tomohiro Nishizawa, Keitaro Yamashita, Ryuichiro Ishitani, and Osamu Nureki. Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. *Nature Communications*, 6(1):1–8, jan 2015. ISSN 20411723. doi: 10.1038/ncomms7112.

340. Simon Newstead and Francis Barr. Molecular basis for KDEL-mediated retrieval of escaped ER-resident proteins – SWEET talking the COPs. *Journal of Cell Science*, 133(19), oct 2020. ISSN 14779137. doi: 10.1242/jcs.250100.

341. IngMarie Nilsson, Annika SaÈaÈf, Paul Whitley, Guro Gafvelin, Cecilia Waller, and Gunnar von Heijne. Proline-induced disruption of a transmembrane α-helix in its natural environment. *Journal of molecular biology*, 284(4):1165–1175, 1998.

342. Joseph Donlan. *The influence of lipid binding on the oligomeric state of membrane protein complexes*. PhD thesis, University of Oxford, 2020.

343. Michael Landreh, Erik G. Marklund, Povilas Uzdavinys, Matteo T. Degiacomi, Mathieu Coincon, Joseph Gault, Kallol Gupta, Idlir Liko, Justin L. P. Benesch, David Drew, and Carol V. Robinson. Integrating mass spectrometry with MD simulations reveals the role of lipids in Na+/H+ antiporters. *Nature Communications*, 8:13993, 2017. ISSN 2041-1723. doi: 10.1038/ncomms13993.

344. Royston S Quintyn, Jing Yan, and Vicki H Wysocki. Surface-induced dissociation of homotetramers with d2 symmetry yields their assembly pathways and characterizes the effect of ligand binding. *Chemistry & biology*, 22(5):583–592, 2015.

345. Ronald A Fisher. Xv.—the correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.

346. Ronald Aylmer Fisher. Statistical methods for research workers. In *Breakthroughs in statistics*, pages 66–70. Springer, 1992.

347. Inc. Wikimedia Foundation. Analysis of variance, 2021. URL `https://en.wikipedia.org/wiki/Analysis_of_variance`.

348. Martin Dahlberg and Arnold Maliniak. Mechanical properties of coarse-grained bilayers formed by cardiolipin and zwitterionic lipids. *Journal of chemical theory and computation*, 6(5):1638–1649, 2010.

349. Naomi R. Latorraca, Nathan M. Fastman, A. J. Venkatakrishnan, Wolf B. Frommer, Ron O. Dror, and Liang Feng. Mechanism of Substrate Translocation in an Alternating Access Transporter. *Cell*, 169(1):96–107.e12, 2017. ISSN 10974172. doi: 10.1016/j.cell.2017.03.010.

350. Erik G Marklund, Matteo T Degiacomi, Carol V Robinson, Andrew J Baldwin, and Justin LP Benesch. Collision cross sections for structural proteomics. *Structure*, 23(4):791–799, 2015.

351. Moitrayee Bhattacharyya, Chanda R Bhat, and Saraswathi Vishveshwara. An automated approach to network features of protein structure ensembles. *Protein Science*, 22(10):1399–1416, 2013.

352. Joanne L Parker, Robin A Corey, Phillip J Stansfeld, and Simon Newstead. Structural basis for substrate specificity and regulation of nucleotide sugar transporters in the lipid bilayer. *Nature communications*, 10(1):1–10, 2019.

353. Yuyong Tao, Lily S Cheung, Shuo Li, Joon-Seob Eom, Li-Qing Chen, Yan Xu, Kay Perry, Wolf B Frommer, and Liang Feng. Structure of a eukaryotic sweet transporter in a homotrimeric complex. *Nature*, 527(7577):259–263, 2015.

354. Mark James Abraham, Teemu Murtola, Roland Schulz, Szilard Pall, Jeremy C. Smith, Berk Hess, and Erik Lindah. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015. ISSN 23527110. doi: 10.1016/j.softx.2015.06.001.

355. Ilario G Tironi, René Sperb, Paul E Smith, and Wilfred F van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *The Journal of chemical physics*, 102(13):5451–5459, 1995.

356. Szilárd Páll and Berk Hess. A flexible algorithm for calculating pair interactions on simd architectures. *Computer Physics Communications*, 184(12):2641–2650, 2013.

357. Berk Hess. P-lincs: A parallel linear constraint solver for molecular simulation. *Journal of chemical theory and computation*, 4(1):116–122, 2008.

358. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

359. Richard J Gowers, Max Linke, Jonathan Barnoud, Tyler John Edward Reddy, Manuel N Melo, Sean L Seyler, Jan Domanski, David L Dotson, Sébastien Buchoux, Ian M Kenney, et al. Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2019.

360. Wanling Song, Robin A Corey, T Bertie Ansell, C Keith Cassidy, Michael R Horrell, Anna L Duncan, Phillip J Stansfeld, and Mark SP Sansom. Pylipid: A python package for analysis of protein–lipid interactions from molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 2022.

361. Phillip J Stansfeld and Mark SP Sansom. Molecular simulation approaches to membrane proteins. *Structure*, 19(11):1562–1572, 2011.

362. Florian Kiefer, Konstantin Arnold, Michael Künzli, Lorenza Bordoli, and Torsten Schwede. The swiss-model repository and associated resources. *Nucleic acids research*, 37(suppl_1):D387–D392, 2009.

363. Navendu Vasavada. One-way anova (analysis of variance) with post-hoc tukey hsd (honestly significant difference) test calculator for comparing multiple treatments, 2016. URL `https://astatsa.com/OneWay_Anova_with_TukeyHSD/`.

364. Gustavo F. Contreras, Karen Castillo, Nicolás Enrique, Willy Carrasquel-Ursulaez, Juan Pablo Castillo, Verónica Milesi, Alan Neely, Osvaldo Alvarez, Gonzalo Ferreira, Carlos Gonzalez, and Ramon Latorre. A BK (Slo1) channel journey from molecule to physiology. *Channels*, 7(6), 2013. ISSN 19336969. doi: 10.4161/chan.26242.

365. Lawrence Salkoff, Alice Butler, Gonzalo Ferreira, Celia Santi, and Aguan Wei. High-conductance potassium channels of the slo family. *Nature Reviews Neuroscience*, 7(12):921–931, 2006.

366. Bo H Bentzen, Søren-Peter Olesen, Lars CB Rønn, and Morten Grunnet. Bk channel activators and their therapeutic perspectives. *Frontiers in physiology*, 5:389, 2014.

367. Joseph E Brayden and Mark T Nelson. Regulation of arterial tone by activation of calcium-dependent potassium channels. *Science*, 256(5056):532–535, 1992.

368. Robert Brenner, Guillermo J Peréz, Adrian D Bonev, Delrae M Eckman, Jon C Kosek, Steven W Wiler, Andrew J Patterson, Mark T Nelson, and Richard W Aldrich. Vasoregulation by the $\beta 1$ subunit of the calcium-activated potassium channel. *Nature*, 407(6806):870–876, 2000.

369. Wei Du, Jocelyn F Bautista, Huanghe Yang, Ana Diez-Sampedro, Sun-Ah You, Lejin Wang, Prakash Kotagal, Hans O Lüders, Jingyi Shi, Jianmin Cui, et al. Calcium-sensitive potassium channelopathy in human epilepsy and paroxysmal movement disorder. *Nature genetics*, 37(7):733–738, 2005.

370. Junqiu Yang, Gayathri Krishnamoorthy, Akansha Saxena, Guohui Zhang, Jingyi Shi, Huanghe Yang, Kelli Delaloye, David Sept, and Jianmin Cui. An epilepsy/dyskinesia-associated mutation enhances bk channel activation by potentiating ca2+ sensing. *Neuron*, 66(6):871–883, 2010.

371. Huanghe Yang, Guohui Zhang, and Jianmin Cui. BK channels: Multiple sensors, one activation gate. *Frontiers in Physiology*, 6(FEB):1–16, 2015. ISSN 1664042X. doi: 10.3389/fphys.2015.00029.

372. Guillermo J Pérez, Adrian D Bonev, and Mark T Nelson. Micromolar ca2+ from sparks activates ca2+-sensitive k+ channels in rat cerebral artery smooth muscle. *American Journal of Physiology-Cell Physiology*, 281(6):C1769–C1775, 2001.

373. Xiao Tao, Richard K Hite, and Roderick MacKinnon. Cryo-em structure of the open high-conductance ca 2+-activated k+ channel. *Nature*, 541(7635):46–51, 2017.

374. Douglas R. Smith, Lynn A. Doucette-Stamm, Craig Deloughery, Hongmei Lee, Joann Dubois, Tyler Aldredge, Romina Bashirzadeh, Derron Blakely, Robin Cook, Katie Gilbert, Dawn Harrison, Lieu Hoang, Pamela Keagle, Wendy Lumm, Bryan Pothier, Dayong Qiu, Rob Spadafora, Rita Vicaire, Ying Wang, Jamey Wierzbowski, Rene Gibson, Nilofer Jiwani, Anthony Caruso, David Bush, Hershel Safer, Donivan Patwell, Shashi Prabhakar, Steve Mcdougall, George Shimer, Anil Goyal, Shmuel Pietrokovski, George M. Church, Charles J. Daniels, Jen I. Mao, Phil Rice, Jörk Nölling, and John N. Reeve. Complete genome sequence of Methanobacterium thermoautotrophicum ΔH: Functional analysis and comparative genomics. *Journal of Bacteriology*, 179(22):7135–7155, 1997. ISSN 00219193. doi: 10.1128/jb.179.22.7135-7155.1997.

375. Y Jiang. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature*, 417:515–522, 2002. doi: 10.1038/417515a.

376. Wojciech Kopec, David A. Köpfer, Owen N. Vickery, Anna S. Bondarenko, Thomas L.C. Jansen, Bert L. de Groot, and Ulrich Zachariae. Direct knock-on of desolvated ions governs strict ion selectivity in K+ channels. *Nature Chemistry*, 10(8):813–820, 2018. ISSN 17554349. doi: 10.1038/s41557-018-0105-9.

377. Sheng Ye, Yang Li, Liping Chen, and Youxing Jiang. Crystal structures of a ligand-free mthk gating ring: insights into the ligand gating mechanism of k+ channels. *Cell*, 126(6):1161–1173, 2006.

378. Brittany Zadek and Crina M. Nimigean. Calcium-dependent Gating of MthK, a Prokaryotic Potassium Channel. *The Journal of General Physiology*, 127 (6):673–685, 2006. ISSN 0022-1295. doi: 10.1085/jgp.200609534. URL `http://www.jgp.org/lookup/doi/10.1085/jgp.200609534`.

379. Lyubov V. Parfenova, Brittany M. Crane, and Brad S. Rothberg. Modulation of MthK potassium channel activity at the intracellular entrance to the pore. *Journal of Biological Chemistry*, 281(30):21131–21138, 2006. ISSN 00219258. doi: 10.1074/jbc.M603109200.

380. Delfina C. Dominguez. Calcium signalling in bacteria. *Molecular Microbiology*, 54(2):291–297, 2004. ISSN 0950382X. doi: 10.1111/j.1365-2958.2004.04276.x.

381. Delfina C. Domínguez, Manita Guragain, and Marianna Patrauchan. Calcium binding proteins and calcium signaling in prokaryotes. *Cell Calcium*, 57(3): 151–165, 2015. ISSN 15321991. doi: 10.1016/j.ceca.2014.12.006.

382. Ernesto Carafoli and Joachim Krebs. Why calcium? How calcium became the best communicator. *Journal of Biological Chemistry*, 291(40):20849–20857, 2016. ISSN 1083351X. doi: 10.1074/jbc.R116.735894.

383. Frank J. Smith, Victor P T Pau, Gino Cingolani, and Brad S. Rothberg. Structural basis of allosteric interactions among Ca2+-binding sites in a K+channel RCK domain. *Nature Communications*, 4:1–10, 2013. ISSN 20411723. doi: 10.1038/ncomms3621. URL `http://dx.doi.org/10.1038/ncomms3621`.

384. Yihao Jiang, Sandipan Chowdhury, and Baron Chanda. Activation of MthK is exquisitely regulated by temperature. pages 1–24, 2020.

385. Ruo Xu Gu and Bert L. de Groot. Lipid-protein interactions modulate the conformational equilibrium of a potassium channel. *Nature Communications*, 11(1):1–10, 2020. ISSN 20411723. doi: 10.1038/s41467-020-15741-8.

386. Bruno Miroux and John E Walker. Over-production of proteins in escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology*, 260(3):289–298, 1996.

387. Inc. Agilent Technologies. Xl1-blue competent cells, 2021. URL `https://www.agilent.com/en/product/mutagenesis-cloning/competent-cells-competent-cell-supplies/competent-cells-for-routine-cloning/xl1-blue-competent-cells-233099`.

388. Marina Schrecker, Dorith Wunnicke, and Inga Hänelt. How RCK domains regulate gating of K+ channels. *Biological Chemistry*, 0(0):1303–1322, 2019. ISSN 1431-6730. doi: 10.1515/hsz-2019-0153.

389. Sriram Vemuri, Cheng-Der Yu, Vuthichai Wangsatorntanakun, and Niek Roosdorp. Large-scale production of liposomes by a microfluidizer. *Drug development and industrial pharmacy*, 16(15):2243–2256, 1990.

390. E Hochuli, W Bannwarth, H Döbeli, R Gentz, and D Stüber. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio/technology*, 6(11):1321–1325, 1988.

391. Thomas P Hopp, Kathryn S Prickett, Virginia L Price, Randell T Libby, Carl J March, Douglas Pat Cerretti, David L Urdal, and Paul J Conlon. A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/technology*, 6(10):1204–1210, 1988.

392. Xiao Tao and Roderick MacKinnon. Molecular structures of the human slo1 k+ channel in complex with $\beta 4$. *Elife*, 8:e51409, 2019.

393. Mark S.P. Sansom, Siewert J. Marrink, Valentina Corradi, Helgi I. Ingólfsson, Paulo C.T. Souza, and D. Peter Tieleman. Computational Modeling of Realistic Cell Membranes. *Chemical Reviews*, 2019. ISSN 0009-2665. doi: 10.1021/acs.chemrev.8b00460.

394. Tahrin Mahmood and Ping-Chang Yang. Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences*, 4(9):429, 2012.

395. Helena Hernandez and Carol V Robinson. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature protocols*, 2(3):715–726, 2007.

396. William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

397. Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

398. Bokeh Development Team. Bokeh: Python library for interactive visualization, 2020. URL `https://bokeh.org/`.

399. Plotly Technologies Inc. Collaborative data science, 2015. URL `https://plot.ly`.

400. J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Appendix

A PDF of the publication of NaViA:

Quetschlich, D., Esser, T. K., Newport, T. D., Fiorentino, F., Shutin, D., Chen, S., Davis, R., Lovera, S., Liko, I., Stansfeld, P. J. & Robinson, C. V. (2021). NaViA: a program for the visual analysis of complex mass spectra. *Bioinformatics*, 37(24), 4876-4878.

Bioinformatics, 37(24), 2021, 4876–4878
doi: 10.1093/bioinformatics/btab436
Advance Access Publication Date: 19 June 2021
Applications Note

Structural bioinformatics

NaViA: a program for the visual analysis of complex mass spectra

Daniel Quetschlich [1,2], Tim K. Esser[1], Thomas D. Newport[2,†], Francesco Fiorentino [1], Denis Shutin[1], Siyun Chen[1], Rachel Davis[3], Silvia Lovera[4], Idlir Liko[5], Phillip J. Stansfeld [2,6,*] and Carol V. Robinson[1,*]

[1]Department of Chemistry, University of Oxford, Oxford OX1 3TF, UK, [2]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK, [3]UCB Pharma, Slough, Berkshire SL1 3WE, UK, [4]UCB Pharma, Chemin du Foriest, 1420 Braine-l'Alleud, Belgium, [5]OMass Therapeutics, Oxford OX4 4GE, UK and [6]School of Life Sciences & Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

*To whom correspondence should be addressed.
Present address: Oxford Nanopore Technologies, Oxford OX4 4DQ, UK.
Associate Editor: Pier Luigi Martelli

Received on February 23, 2021; revised on April 28, 2021; editorial decision on May 23, 2021; advance access date 19 June 2021

Abstract

Motivation: Native mass spectrometry is now a well-established method for the investigation of protein complexes, specifically their subunit stoichiometry and ligand binding properties. Recent advances allowing the analysis of complex mixtures lead to an increasing diversity and complexity in the spectra obtained. These spectra can be time-consuming to tackle through manual assignment and challenging for automated approaches.
Results: Native Mass Spectrometry Visual Analyser is a web-based tool to augment the manual process of peak assignment. In addition to matching masses to the stoichiometry of its component subunits, it allows raw data processing, assignment and annotation and permits mass spectra to be shared with their respective interpretation.
Availability and implementation: NaViA is open-source and can be accessed online under https://navia.ms. The source code and documentation can be accessed at https://github.com/d-que/navia, under the BSD 2-Clause licence.
Contact: phillip.stansfeld@warwick.ac.uk or carol.robinson@chem.ox.ac.uk
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Native mass spectrometry (nMS) is an established technique that allows a label-free analysis of intact proteins and their complexes in a mass spectrometer (Marty *et al.*, 2015). It has proven to be a powerful tool in investigating the stoichiometry and interactions of a wide variety of proteins with unprecedented resolution (Bolla *et al.*, 2019). In nMS, protein complexes are transferred into the mass spectrometer by means of electrospray ionization (Dole *et al.*, 1968). A series of ion selection steps and application of energy through accelerating voltages and collisions with inert gas molecules allow the transmission of folded proteins and interaction partners into a mass analyser (Gupta *et al.*, 2018). Soluble proteins and protein complexes with molecular weights up to the mega-Dalton mass range are amenable to study and examples include ribosomes (McKay *et al.*, 2006) and intact viruses (Snijder *et al.*, 2008). Also, more recently membrane proteins have been studied following their release from their solubilizing agent (e.g. detergent micelles, bicelles, nanodiscs or vesicles) through application of energy prior to mass analysis (Laganowsky *et al.*, 2013). Mass spectra are then analysed manually or by using automated approaches such as UniDec (Bern *et al.*, 2018; Marty *et al.*, 2015). Interpretation of these spectra requires a priori knowledge of the molecules included in the complex from gel electrophoresis, western blot or 'omics techniques (Gault *et al.*, 2016).

Recent progress in sample preparation paved the way for the investigation of increasingly complex environments such as vesicles obtained from cell membranes. Following this procedure, samples are injected directly into the mass spectrometer and analysed through nMS without previous purification or treatment with detergents (SoLVe-MS) (Chorev *et al.*, 2018), to preserve the native conditions for membrane proteins. These experiments yield highly complicated spectra containing a plethora of charge state series corresponding to the different protein complexes present in the vesicle. Analysing these spectra through high-level automated approaches is challenging given the extent of overlapping peak series. Thus, performing a manual analysis of these spectra is necessary yet incredibly time-consuming.

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/bioinformatics/article/37/24/4876/6305832 by guest on 27 February 2022

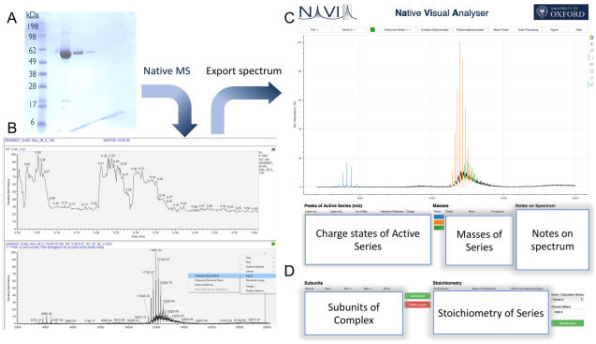

Fig 1. Example of the integration of NaViA into native mass spectrometry workflow. (**A**) In this example purified GroEL purchased from Sigma Aldrich was used as a sample. (**B**) The native mass spectrometry spectrum can be exported as a CSV file. (**C**) Import CSV file into NaViA and analyse by visually selecting peaks for each molecular species. (**D**) Assess stoichiometry for identified species though 'Complex stoichiometry' in NaViA

Motivated by these necessities, we created Native Mass Spectrometry Visual Analyser (NaViA), with the aim of occupying the middle ground between manual and automated analysis. Therefore, NaViA is complementary to existing software (e.g. UniDec). NaViA is a web service that allows a semi-automated analysis of native mass spectra through an easy-to-use interface (Fig. 1). Charge states corresponding to a particular protein are selected visually and masses are assigned automatically following known algorithms. The focus on augmenting a manual assignment is also instructive for scientists entering the field of nMS, since it provides key information on where peaks are to be expected for a certain complex and highlights peak series for this complex. In addition, sessions may be saved comprising both raw data and also processed data with annotations. This permits the nMS community to share both, spectra and the analysis, in a standardized format online as part of a collaboration or publication.

2 Materials and methods

2.1. Implementation

The web service is implemented in Python using the Bokeh visualization package. It is a standalone HTML file without any server in the background. It is implemented in Python 3.6 and JavaScript using the Bokeh framework (version 1.4.0). According to Wellcome and the UK research and innovation's Common Principles on Data Policy on data, software and materials management and sharing, all data supporting this study will be openly available from the software repository.

2.2. Documentation and tutorials

The documentation of NaViA is available on its GitHub repository as well as an introductory video (https://www.youtube.com/channel/UCRvzQxegz0WNkjvvaRJ8PpQ/).

2.3. Input

NaViA currently loads files in a CSV format of the two columns *m/z* and intensity. These can be exported by commonly used software such as MassLynx for Waters Instruments or Xcalibur for ThermoFisher Scientific instruments such as OrbiTraps (Makarov, 2000). Upon loading, the intensities are normalized.

3 Results

3.1 Processing of mass spectra

NaViA enables smoothing and background subtraction of native mass spectra using the same algorithm as standard software, e.g. UniDec. Smoothing is performed through a Gaussian filter and can be performed multiple times. In addition, different modes of background subtraction are facilitated and a filter for 'minimum intensity' can be applied. The background subtraction modes include subtracting a fixed value, a linear interpolation between the intensities of the smallest and largest *m/z* values and a curved subtraction as described previously (Morgner and Robinson, 2012).

3.2 Peak selection and mass calculation

Peaks are selected through a range on the *m/z*-axis using the mouse cursor. Once multiple charge states are selected, the corresponding mass is calculated by the MacSED algorithm (McKay *et al.*, 2006). This algorithm systematically creates a matrix of possible charge states assuming there are no missing charge states in the series. Based on this series a matrix of masses is created where each *m/z* value of a peak is deconvoluted into a mass value. The variance between the masses is then calculated for each series of charges. The series of masses with the minimal variance is then selected to be the true series of charge states. Finally, the mean of the measured mass is set as the true mass and the standard deviation set to be the uncertainty value.

3.3 Tools for augmenting peak assignment

For the analysis of complex native mass spectra two tools have been implemented: a 'Mass Finder', which allows the user to highlight *m/z* regions for a specific mass of a complex, and a 'Peak Prediction' tool, which highlights adjacent peaks for a known series. The 'Mass Finder' makes it possible to see whether a complex with known mass is present. This is especially important for SoLVe spectra where many molecular species are present. For these spectra, 'Peak prediction' augments the differentiation of overlapping peak series.

3.4 Comparing complex mass to subunit masses

The 'Complex Stoichiometry' function makes it possible to see whether a complex with known mass is present. This is especially important for cases where many molecular species are present.

3.5 Sessions in NaViA

NaViA allows sessions to be saved, including raw data, data processing and peak assignment, in a *navia* file in JSON format. This feature allows users to share spectra including assignments as well as collaborating or deposition of mass spectra for enabling raw data sharing. In addition, JSON formats are easily readable for various programming languages and software packages, e.g. Python or Origin. Therefore, *navia* files are well suited for consecutive data analysis. The format is described in the documentation on the GitHub page.

3.6 Compatibility with UniDec

NaViA was developed to be complementary to the current software packages available for native MS data analysis. As part of this, an option to import NaViA sessions has been incorporated into the current UniDec version (4.4.1) (Experimental → Import from NaViA).

3.7 Exporting images

NaViA provides the possibility to export images in 4k quality sufficient for A4 pages in publications as well as vector graphics in the SVG format.

Acknowledgements

We would like to thank the Robinson, Stansfeld and Benesch groups for their support and feedback. In addition, we would like to thank Bryan Van de Ven and Eugene Pakhomov from the Bokeh forum for their extensive support on technical issues. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Downloaded from https://academic.oup.com/bioinformatics/article/37/24/4876/6305832 by guest on 27 February 2022

Funding

Research in the C.V.R. laboratory is supported by a Medical Research Council program grant [MR/N020413/1], a European Research Council Advanced Grant ENABLE [695511] and a Wellcome Senior Investigator Award [104633/Z/14/Z]. Research in P.J.S.'s lab is funded by Wellcome [208361/Z/17/Z], the MRC [MR/S009213/1] and BBSRC [BB/P01948X/1, BB/R002517/1 and BB/S003339/1]. D.Q. and F.F. hold a SABS CDT studentship supported by the EPSRC and MRC [EP/L016044/1].

Conflict of Interest: none declared.

References

Bern,M. *et al.* (2018) Parsimonious charge deconvolution for native mass spectrometry. *J. Proteome Res.*, **17**, 1216–1226.

Bolla,J.R. *et al.* (2019) Membrane protein–lipid interactions probed using mass spectrometry. *Annu. Rev. Biochem.*, **88**, 85–111.

Chorev,D.S. *et al.* (2018) Protein assemblies ejected directly from native membranes yield complexes for mass spectrometry. Science (New York, N.Y.), **362**, 829–834. 10.1126/science.aau0976 30442809

Dole,M. *et al.* (1968) Molecular beams of macroions. *J. Chem. Phys.*, **49**, 2240–2249.

Gault,J. *et al.* (2016) High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods*, **13**, 333–336.

Gupta,K. *et al.* (2018) Identifying key membrane protein lipid interactions using mass spectrometry. *Nat. Protoc.*, **13**, 1106–1120.

Laganowsky,A. *et al.* (2013) Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.*, **8**, 639–651.

Makarov,A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Analytical Chemistry, **72**, 1156–1162. 10.1021/ac991131p 10740853

Marty,M.T. *et al.* (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.*, **87**, 4370–4376.

McKay,A.R. *et al.* (2006) Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *J. Am. Chem. Soc.*, **128**, 11433–11442.

Morgner,N. and Robinson,C.V. (2012) Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.*, **84**, 2939–2948.

Snijder,J. *et al.* (2008) Studying 18 mega Dalton virus assemblies with native mass spectrometry. *Bone*, **23**, 1–7.

Downloaded from https://academic.oup.com/bioinformatics/article/37/24/4876/6305832 by guest on 27 February 2022