Structural and functional studies of the antiviral target proteins HCV p7 and ER alpha-glucosidase II



Mario Hensen
St John's College
University of Oxford

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Michaelmas Term
2019

Word count: 49,443

Abstract

To accomplish the correct folding of N-glycosylated proteins eukaryotes rely on their endoplasmic reticulum quality control pathway. This pathway which consists of glycosyl-hydrolases and -transferases, lectins and chaperones, assists the folding of immature glycoproteins post their translation into the ER lumen. Glycoproteins are processed based on glycan and protein misfold recognition. While correctly matured glycoproteins proceed through the secretory pathway to their final destination, immature glycoproteins are discarded eventually by the ER associated degradation pathway.
Many viruses hijack the glycoprotein quality control pathway after host cell infection in order to produce their own viral glycoproteins, which makes them vulnerable to the interference with the correct performance of the cycle.
Iminosugars which are glucose mimetics inhibit N-linked glycan processing enzymes such as alpha-glucosidase I and II of the glycoprotein quality control pathway, resulting in decreased virion secretion from infected cells and/or virus infectivity.

Presented in this thesis is the protein X-ray crystallography work carried out on ER alpha-glucosidase II in combination with a new class of *N*-alkyloxy iminosugars revealing binding motifs of the latter within the active site of the enzyme. Furthermore, the study demonstrates a new approach of potential alpha-glucosidase II inhibition and the development and use of a fluorescence energy resonance transfer-based high-throughput screen for new drug discovery. The findings enable future studies aimed at improving alpha-glucosidase II inhibition both by developing more specific active site inhibitors as well as via a novel strategy that aims to prevent or disrupt GluII heterodimerisation by targeting the enzyme's α-/β- subunit interface. Long alkyl chain iminosugars do not only inhibit certain host cell enzymes such as ER alpha-glucosidase II but also the hepatitis C virus p7 ion channel, but it is not yet known how this inhibition is achieved. The analysis of how iminosugars interact with their target protein p7 is hindered by the lack of knowledge of the protein channel's oligomeric structure and its location in the ER membrane. Work in this thesis uses electron paramagnetic resonance spectroscopy in combination with spin labelled HCV p7 to analyse the oligomerisation behaviour of the p7 protomer in lipid bilayers. The structural information gained provides the basis for future experiments to better understand drug inhibition of the p7 ion channel and contributes to the still very limited information about viral ion channels in lipid environments.

Acknowledgment

My highest recognition and gratefulness goes to my supervisor Professor Nicole Zitzmann. Not just because she gave me the opportunity to do a DPhil in her truly unique laboratory, no, I am even more grateful for her literally always 'open door', no matter the reason I stepped into her office. Her endless optimism, support and energy – scientific and personal – helped me to build my own ideas and inspired my scientific exploration and learning. All while giving me a huge amount of scientific freedom and associated trust. I deeply enjoyed my time in her laboratory.

What would all this be good for if there wasn't this one colleague who sparks of scientific enthusiasm, who's thirst for knowledge never ends and who is restless until every little detail is solved, no matter what it takes? Dr Juan H. Bolìvar Gonzàles – his love for p7 is contagious. Besides Nicole, Juan supported me from the very first day I set foot into the Zitzmann laboratory and since then never lost his uplifting attitude. He was the one who introduced me to the p7 project with all its details and exhaustiveness and with whom I spent endless days and nights in the lab preparing samples, while cheering each other up in times of setbacks and tiredness, leading to a friendship that has extended beyond the borders of the laboratory…p7!
'A scientist who's equal has yet to be found, with a thirst for great science and cheese' I think this very much describes my co-supervisor Dr Pietro Roversi, to whom I owe all of my knowledge in X-ray crystallography, my interest in scientific discussions and the attitude to never mind a personal challenge. Even after leaving the Zitzmann group, Pietro always provided amazing support and guidance in all parts of the work presented in this study, making it even more complete. I am sure his new colleagues will highly value his endless crystallography knowledge and great personality in all aspects of life.
I would very much like to thank Dr Alessandro Caputo for introducing me to the alpha-glucosidase II project and his will and attitude to impart some of his knowledge to me. Alex's meticulous work, and his tips and tricks for the alpha-glucosidase II project, provided a solid foundation for the research carried out in this study and helped me to reflect and transfer it into other fields of my research.

Without Professor Jason Schnell and Dr Benjamin P. Oestringer, among others, the p7 field would still continue to discuss the oligomeric state of p7 in n-dodecylphosphocholin. Their quest for high quality scientific work sets a mark and should encourage other scientists to follow their example. I would very much like to thank them both for their amazing and always supportive collaboration and the opportunity they gave me to contribute to their research in

the field of viral ion channels. I hope that there will be more collaborations to come in the future.

A huge thank you goes to the people of the CAESR facility (Department of Chemistry). Especially Dr Alice Bowen and Dr William Myers who have always been approachable and of great help with their EPR-based guidance and technical support, committing time and effort to make this study possible.

Parts of the work presented in this thesis were carried out in collaboration with Evotec. The great proficiency of people like Dr Tim James, Dr Elena Wiegmann and Christina Reinisch made this a very successful collaboration, supported by a great sense for science and very joyful meetings – thank you.

There are also a number of people, especially current and past members of the Zitzmann laboratory as well as people from the department, that I owe an honest thank you. People like Dr J.L. Kiappes, Dr Anu Chandran, Juliane Brun, Snezana Vasiljević, Dr Johan Hill, Dr Edward Lowe, Dr Marc-Philipp Pfeil, Dr Anastasia Kantsadi and Benoit Darlot, to just name a few. All those people and many more were part of my stay in the Zitzmann laboratory and contribut ed to this lovely and special experience in many different ways. Whether it was their technical expertise, their company during café breaks, their support during the ups and downs of my DPhil, or just as a climbing partner to relax and get a break from working, I thank them.

Where would I be without the absolute overwhelming support of Susanne and the sacrifices she made in order for me to be able to do my DPhil in Oxford without us living apart? She stayed by my side and joined me on my journey to the UK, accepting that she will be separated from her own, beloved family by many hundreds of kilometres. During the whole time she never complained and always supported me with such strength that I am still astonished where this graceful person takes all this energy from. I love you Susanne and I always will.

Finally, special thanks go to my family who encouraged me throughout my entire life and without whom I would not be where I am now. Just knowing that they always will be there for me made me feel safe and gave me the confidence to carry on step by step – thank you so much!

List of Contents

ABSTRACT ..... **II**

ACKNOWLEDGMENT ..... **III**

LIST OF FIGURES ..... **X**

LIST OF TABLES ..... **XVI**

LIST OF NOMENCLATURE ..... **XVIII**

1. INTRODUCTION **1**

1.1. IMINOSUGARS 2

1.1.1. Iminosugars as therapeutics *4*

1.1.2. Use of iminosugars as antivirals *7*

1.2. VIRAL ION CHANNELS – VIROPORINS 11

1.2.1. Hepatitis C virus *16*

1.2.2. The p7 viroporin *20*

1.3. THE ER GLYCOPROTEIN FOLDING QUALITY CONTROL CYCLE 28

1.3.1. The core of the ERQC – the calnexin/calreticulin cycle *32*

1.3.2. Alpha-glucosidase II *33*

1.4. PROJECT AIMS 40

2. EXPERIMENTAL PRINCIPLES **41**

2.1. CONTINUOUS WAVE ELECTRON PARAMAGNETIC RESONANCE 41

2.2. 4-PULSE DOUBLE ELECTRON-ELECTRON RESONANCE 53

2.3. FÖRSTER ENERGY RESONANCE TRANSFER 59

3. MATERIALS AND METHODS – HCV P7 ..... **61**

3.1. HCV P7-CLONING 61

3.1.1. p7-constructs *61*

3.1.2. Ligation of p7-constructs into vector pMM-LR6 ..... *63*

3.1.3. Agarose gels *65*

3.2. PROTEIN EXPRESSION AND PURIFICATION 66

3.2.1. Transformation of pMM-LR6/p7 in E. coli 66

3.2.2. Expression screening in bacterial cells 66

3.2.3. Sodium dodecyl sulphate – polyacrylamide gel electrophoresis 67

3.2.4. Large-scale expression in E. coli BL21 68

3.2.5. Large-scale expression of ^{15}N-labelled p7 in E. coli BL21 68

3.2.6. Bacterial cell/inclusion body disruption 69

3.2.7. CNBr-cleavage of the p7-fusion protein 70

3.2.8. Size exclusion chromatography 71

3.3. EPR SAMPLE PREPARATION 72

3.3.1. MTSL-labelling of p7-cysteine mutants 72

3.3.2. p7 reconstitution into lipid bilayers 72

3.3.3. Circular dichroism 74

3.3.4. Dynamic light scattering 74

3.3.5. Fluorescent resonance energy transfer 75

3.3.6. Proton flux assay 76

3.3.7. Electron paramagnetic resonance 78

3.3.8. Double electron-electron resonance 80

3.4. SAMPLE PREPARATION OF HCV P7 IN DETERGENT 82

3.4.1. Nuclear magnetic resonance spectroscopy 82

3.4.2. Size exclusion – small angle laser light scattering 83

4. RESULTS AND DISCUSSION – HCV P7 85

4.1. P7 PROTEIN MUTAGENESIS 85

4.2. P7 PROTEIN EXPRESSION AND PURIFICATION 90

4.3. HCV P7 IN N-DODECYLPHOSPHOCHOLINE 94

4.3.1. Nuclear magnetic resonance spectroscopy 95

4.3.2. Size exclusion – Multi-angle light scattering 102

4.4. P7 IN LIPID BILAYERS 109

4.4.1. Sucrose density gradient ultracentrifugation 111

4.4.2. Circular dichroism ... *112*

4.4.3. Continuous wave EPR ... *116*

4.4.3.1. Spin labelling of p7 ... 116

4.4.3.2. Spin label dynamics are influenced by the lipid to p7 molar ratio ... 118

4.4.3.3. Spin label dynamics at different positions within the p7 sequence ... 123

4.4.3.4. CwEPR spectra in the frozen state – spin-spin distance ... 130

4.4.3.5. Conclusions from cwEPR analysis ... 132

4.4.4. Double electron-electron resonance evaluation of p7 in lipid bilayers ... *133*

4.4.4.1. Inter molecule distances and protomer arrangements by DEER ... 136

4.4.4.2. Comparison of DEER shortest distance ... 146

4.4.4.3. 4-pulse DEER on MTSL-spin labelled p7 – Conclusion ... 149

4.4.5. Proton flux assay ... *151*

5. MATERIALS AND METHODS – ALPHA-GLUCOSIDASE II ... 160

5.1. GluII constructs ... 160

5.1.1. Cloning ... *161*

5.1.2. MmGluII subcloning into destination vector ... *161*

5.1.3. Gibson assembly ... *162*

5.1.4. Site-directed mutagenesis ... *163*

5.2. Expression and Purification ... 165

5.2.1. Transfection of mammalian cells ... *165*

5.2.2. Processing of mammalian cell culture supernatant ... *165*

5.2.3. Immobilised metal affinity chromatography ... *165*

5.2.4. Trypsin digestion for protein crystallography ... *166*

5.2.5. Size exclusion chromatography ... *167*

5.2.6. Purification by ion exchange chromatography ... *167*

5.3. Protein concentration determination ... 169

5.4. Gel electrophoresis and Blotting ... 169

5.4.1. SDS-PAGE ... *169*

5.4.2. Western blotting ... *170*

5.4.3. Agarose gels ... *170*

5.5. GluII activity measurements 171
5.5.1. Continuous 4-MUG assay 171
5.6. Stability measurement 172
5.6.1. Differential scanning fluorimetry 172
5.6.2. Time resolved fluorescent resonance energy transfer 173
5.7. X-ray crystallography 175
5.7.1. Crystallisation 175
5.7.1.1. *Mm*GluII$_{tryps}$ crystallography 175
5.7.1.2. Crystal seeding 175
5.7.1.3. GluII crystal compound soaking 176
5.7.1.4. *Mm*GluII$_{LDLRa}$ crystallography 177
5.7.2. Data collection 177
5.7.3. Data processing 179
6. RESULTS AND DISCUSSION ER ALPHA-GLUCOSIDASE II 180
6.1. *Mm*GluII cloning 180
6.2. GluII expression and purification 183
6.3. Activity determination by continuous 4MUG assay 187
6.4. X-ray crystallography of *Mm*GluII α-βLDLRa 188
6.4.1. Crystallography 189
6.4.2. Data processing and refinement 191
6.4.3. Discussion 199
6.5. *Mm*GluII biophysical characterisation for protein crystallography 200
6.5.1. Differential scanning fluorimetry 200
6.6. X-ray crystallography of *Mm*GluII active site inhibitors 202
6.6.1. Crystallisation 205
6.6.2. Data processing and refinement 206
6.7. Fluorescent resonance energy transfer 214
6.7.1. GluII TR-FRET assay development 220
6.7.2. GluII TR-FRET compound screening 231
6.7.3. Assay evaluation 234

6.7.4. GluII TR-FRET assay conclusion 243

6.8. CRYSTALLISATION OF *MM*GLUII WITH α-/β-INTERFACE DISRUPTORS 247

6.8.1. Conclusion of X-ray crystallography carried out on TR-FRET GluII active compounds 253

7. CONCLUSION AND FUTURE DIRECTIONS 255

7.1. HCV P7 255

7.2. ALPHA-GLUCOSIDASE II 257

8. APPENDIX 259

8.1. **APPENDIX A** – HCV P7 259

8.2. **APPENDIX B** – ER ALPHA-GLUCOSIDASE II 274

9. BIBLIOGRAPHY 280

List of Figures

Figure 1: Structures of glucose (A), 1-deoxynojirimycin (DNJ, B), 1-deoxygalactonojirimycin (DGJ, C), *N*-butyl-1-deoxynojirimycin (*N*B-DNJ, D) and castanospermine (CST, E). 3

Figure 2: Illustration of host cell organelles affected by viroporins and their cytopathic effects within the virus life cycle. 15

Figure 3: Illustration of the RNA Hepatitis C virus (HCV) virion. 18

Figure 4: Solution NMR derived structures of the p7 monomer as described in the literature. 23

Figure 5: The p7 oligomeric arrangements as described in the literature. 25

Figure 6: Pre-assembled N-glycan linked to an asparagine residue in the N-glycosylation consensus sequence. 29

Figure 7: Illustration of the calnexin/calreticulin cycle. 31

Figure 8: Example for the GluII hydrolase reaction of a diglucosylated substrate. 35

Figure 9: Illustration of the GluII heterodimer domains. 38

Figure 10: Microwave absorption spectra by an unpaired electron in an applied magnetic field. 44

Figure 11: EPR spin label S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) carrying an unpaired electron at the nitroxide oxygen ... 45

Figure 12: Single electron energy state landscape in an external magnetic field (B_0). 46

Figure 13: Microwave absorption spectra of MTSL in a changing magnetic field, influenced by hyperfine splitting. 46

Figure 14: CwEPR spectra simulation at X-band and 298 K of a single nitroxide spin at varying correlation times (τ_c) 49

Figure 15: Example of EPR spectral convolution. (A) and (B) show two hypothetical different states of an oligomer forming protein (blue dot), with the corresponding EPR spectra. 50

Figure 16: Characteristics of an EPR spectrum: ΔH_0 – linewidth of the centre peak, A_{zz} – apparent hyperfine splitting. 51

Figure 17: DEER methodology 55

Figure 18.: Illustration of the DEER evaluation process. 58

Figure 19: Förster resonance energy transfer (FRET) phenomenon. 60

Figure 20: 3D representation of the putative p7 monomer with highlighted cysteine mutation sites used for cwEPR/DEER. 63

Figure 21: 4-pulse DEER pulse trains at observer (ω_A) and pump pulse frequencies (ω_B). 81

Figure 22: p7-fusion protein (mutant 3b) overexpression screening of *E. coli* BL21 transformed cells using SDS-PAGE. 91

Figure 23: Gel filtration chromatography of p7 mutant 1 after CNBr cleavage of the fusion tag. 91

Figure 24: SDS-PAGE of p7 mutant 3 (I) and 6 (II) under reducing conditions. Lyophilised p7 powder was resuspended in 27 mM CF-6. 92

Figure 25: NMR spectroscopy analysis of p7 genotype 5a (mutant 0a) in DPC micelles 99

Figure 26: NOESY NMR of p7 in DPC micelles. 101

Figure 27: SEC–MALLS of p7 genotype 5a in varying DPC concentrations, as well as empty detergent micelle analysis. 105

Figure 28: p7 reconstitution into lipid bilayers, basic work flow. 111

Figure 29: Sucrose density gradient ultracentrifugation analysis. 112

Figure 30: Circular dichroism of different p7 mutants before and after lipid reconstitution. 113

Figure 31: Standard curve of TEMPO for MTSL labelling efficiency determination. 118

Figure 32: MTSL labelling efficiency of p7 single cysteine mutants on SDS-PAGE 118

Figure 33: CwEPR 1D-field sweep of p7 J4 mutant 1. 121

Figure 34: CwEPR deconvolution series of MTSL-labelled p7 mutant 1 at LPR 250:1. 122

Figure 35: Evaluation of the cwEPR spectra deconvolution (Figure 34). 123

Figure 36: CwEPR 1D-field sweep of different p7 J4 mutants at defined LPRs. 127

Figure 37: Central linewidth (ΔH_0) analysis of measurements carried out on p7 mutants 1, 3b, 5, 6 and 11 in DOPC:POPE at LPR 2000:1 and 250:1 129

Figure 38: CwEPR 1D-field sweep of p7 J4 mutant 3b (A) and 6 (B) at 100 K. 131

Figure 39: Q-band 4-pulse DEER measurements of fully MTSL-labelled p7 mutants 1, 3b, 5 and 6 reconstituted into DOPC:POPE (4:1, w/w) mixtures at LPR 250:1 (with the exception of p7 mutant 1 which is at an LPR of 150:1). 139

Figure 40.: Overlay of DEER distance distribution obtained in Figure 39.. 140

Figure 41: Q-band 4-pulse DEER measurements of fully MTSL-labelled p7 mutants 1, 3b, 5 and 6 reconstituted into DOPC:POPE (4:1, w/w) mixtures at LPR 1000:1 141

Figure 42.: 4-pulsed DEER measurements of MTSL-labelled p7 in DOPC:POPE (4:1, w/w) at LPRs range between 1000 and 2000:1 (A-B) 142

Figure 43.: 4-pulsed DEER comparison of MTSL-labelled p7 in DOPC:POPE (4:1, w/w) at high and low LPRs. 143

Figure 44.: (A) Schematic of four different oligomers, from left to right: tetramer, pentamer, hexamer and heptamer. 145

Figure 45.: MMM[227] distance predictions based on the p7 oligomer model published by Montserret *et al.*[74] and the unusual NMR structure by Ouyang *et al*.[165] PDBID: 2M6X) in comparison to experimental DEER distances obtained in this study. 148

Figure 46.: Schematic of proton flux assay carried out on p7 in unilamellar liposomes. 152

Figure 47.: Proton flux assay using unilamellar liposomes of varying lipid compositions, in the presence and absence of p7 mutant 0 (corresponding LPR is stated in figure legend). 156

Figure 48: Purification of full length *Mm*GluII. (A) IMAC chromatogram of 900 ml of cell culture supernatant purified using a 5 ml HisTrap Excel column. 184

Figure 49: Purification of untagged humanised *Mm*GluII β-subunit (Table 30, *Mm* β_h). 186

Figure 50: 4MUG activity assay of different *Mm*GluII batches. 187

Figure 51: *Mm*GluII α-β_LDLRa crystals found in the Morpheus HT-96 screen (Molecular Dimensions). 191

Figure 52: The κ =180° section of the self-rotation function of the *Mm*GluII α-βLDLRa crystal, calculated from the data processed in spacegroup $P2_1$ 193

Figure 53: Two orthogonal views of two consecutive cells of the $P2_1$. 195

Figure 54: *Mm*GluII α-β_LDLRa α-subunit loop region aa 351-369. 198

Figure 55: Electrostatic potential[313] contoured from -5.0 to 5.0 k_bT/e (red to blue) on the surface of *Mm*GluII α-β_LDLRa in the presence (A) and absence (B) of loop region aa 351-369 (α-subunit). 198

Figure 56: Differential scanning fluorimetry of *Mm*GluII in 20 mM HEPES pH 7.4, 120 mM NaCl at increasing DMSO concentrations (%, v/v). Negative Δ˚C indicates a decrease in protein stability. 201

Figure 57: Orthorhombic *Mm*GluII$_{tryps}$ crystal rods. 205

Figure 58: The κ =180° section of the self-rotation function of a *Mm*GluII$_{tryps}$ crystal calculated from data processed in $P2_1\,2_1\,2$. Figure generated using Molrep (CCP4 program suit). 208

Figure 59: 2-dimensional diagram of deoxynojirimycin (DNJ) in the *Mm*GluII active site. 209

Figure 60: *Mm*GluII active site bound with different N-alkyloxy deoxynojirimycin analouges. 211

Figure 61: Mouse GluII α-/ β-interface polar interactions. 216

Figure 62: GluII α-/ β-interface interference concept 217

Figure 63: TR-FRET high throughput screening concept. 218

Figure 64: Humanisation of the mouse GluII heterodimer interface 219

Figure 65: TR-FRET antibody titration over increasing concentrations of control peptide (His-AQ-FLAG). 221

Figure 66: TR-FRET of three different GluII heterodimer mutants at varrying concentrations at constant mAb mix (1 nM anti-His-Tb and 10 nM anti-FLAG-XL665). 222

Figure 67: TR-FRET titration experiment of humanized full length GluII (GluII α-β_hHF), using 1 nM anti-His-Tb and 10 nM anti-FLAG-XL665 and a 1h incubation time. 225

Figure 68: TR-FRET assay with double titration of anti-His-Tb (0.0-5.0 nM) and anti-FLAG-XL655 antibody (0.0-50 nM) using 3 nM GluII α-β_hHF. 226

Figure 69: TR-FRET of GluII α-β_hHF at varying incubation times. 227

Figure 70: TR-FRET GluII α-β_hHF DMSO tolerance. 228

Figure 71: TR-FRET assay positive control. Shown is the change in activity (decrease in flourescence) over the titration of GluII β_h into GluII α-β_hHF loaded wells. 230

Figure 72: TR-FRET titration experiment of increasing GluII β_h concentration over 3 nM GluII α-β_hHF. 231

Figure 73: Structures of GluII bound ligands found in previous, X-ray crystallography based studies. 234

Figure 74: Illustration of the four most common types of TR-FRET readouts observed in this study. 236

Figure 75: TR-FRET assay hits. Compounds are ordered by decreasing GluII interface splitting efficiency. 239

Figure 76: TR-FRET assay-based structure activity relationship study of EV-JJL001-060-001 derivatives. 242

Figure 77: TR-FRET assay-based structure activity relationship study of IF_ZN2 analogues. 243

Figure 78: *Mm*GluII$_{tryps}$ crystal growing attempts using the Morpheus HT-96 screen (Molecular Dimensions). 249

Figure 79: Different morphologies of *Mm*GluII$_{tryps}$ crystals. 250

Figure A. 1. Mass spectrometry of different p7 single and double cysteine mutants (see Table 2).... 265

Figure A. 2. Size exclusion – multi angle laser light scattering of p7 (GT5a/JFH-1) mutant 0a solubilised in 10 and 80 mM sodium lauryl-sulphate..... 266

Figure A. 3. MTSL labelling test of different p7 single and double cysteine mutants in the presence and absence of spin label. 266

Figure A. 4. Basic deconvolution of two component Nitroxide cwEPR spectrum (2 -components).. 271

Figure A. 5. DEER data processing parameters used in DeerAnalysis2019[237,278] according graphs and distance distributions shown in section 4.4.4 of this study. 272

Figure A. 6. Stoichiometric analysis of FRET measurements carried out by Dr Marc-Philipp Pfeil . 273

Figure B. 1. DNA sequence of *mus musculus* alpha-glucosidaseII α-subunit used in this study. Uniprot: Q8BHN3, *GANAB_mouse.* 274

Figure B. 2. DNA sequence of *mus musculus* alpha-glucosidaseII β-subunit used in this study. Uniprot: O08795, *PRKCSH_mouse.* 275

Figure B. 3. DNA sequence of *homo sapiens* alpha-glucosidaseII α-subunit used in this study. Uniprot: Q14697, *GANAB_human*. 275

Figure B. 4. DNA sequence of *homo sapiens* alpha-glucosidaseII β-subunit used in this study. Uniprot: P14314, *GLU2B_HUMAN.* 276

Figure B. 5. Amino acid sequence alignment of full length mouse and human GluII (UniProt: Q8BHN3/P14314 and Q14697/P14314, respectivly). 278

Figure B. 6. TR-FRET assay plate layout 279

Figure B. 7. Interferometric scattering mass spectrometry (ISCAM)[336] conduced by Dr Weston Stuwe on *Mm*GluII in 20 mM HEPES, 120 mM NaCl, pH 7.4. 279

List of Tables

Table 1: Example viruses and corresponding virus families showing reduced virus infectivity upon iminosugar treatment 10

Table 2: Single and double cysteine p7 mutants analysed in this research, using the amino acid sequence from genotype 1b, isolate J4 as a template (first row, WT = wild type). 62

Table 3: Reaction mix used for vector restriction. The cleavage took place at 37°C, overnight and was stopped by transferring the sample to -20°C 64

Table 4: Reaction mix used for the ligation of digested p7 sequences into the open reading frame of the vector pMM-LR6. 64

Table 5: FRET donor- and acceptor- molar ratios prepared for the detection of p7 oligomeric arrangements within one LPR. 76

Table 6: EPR settings for the BRUKER EMX using the software XENON™. 79

Table 7: Final sample composition used for NMR spectroscopy. 82

Table 8: Dialysis buffer used for p7 sample preparation in detergent. 83

Table 9: Buffer systems used for SEC-MALLS analysis. 84

Table 10: Evaluation of potential p7 cysteine mutant positions to identify suitable positions for the attachment of EPR spin labels. 88

Table 11: HCV p7 mutant purification work flow at its current state. 93

Table 12: NMR and SEC-MALLS evaluation of particle mass contribution corresponding to detergent and protein. 101

Table 13: Detailed list of samples used for SEC-MALLS analysis described in section 4.3.2. 106

Table 14: DICHOWEB[261] secondary structure analysis of p7 solubilised in detergent-lipid and lipid only mixtures. 114

Table 15: Dynamic light scattering (DLS) analysis of different lipid and lipid-protein samples. 155

Table 16: List of lipid compositions used in this study. 157

Table 17: DNA-primers (Gibson assembly ready) and temperatures used for target gene amplification. 161

Table 18: DNA-primers and temperatures used for site directed mutagenesis. 164

Table 19: KLD reaction mix used for vector circularisation. 164

Table 20: Nickel affinity chromatography buffers. 166

Table 21: Reaction mix for *Mm*GluII trypsin digestion 167

Table 22: Buffers used for size exclusion chromatography. 167

Table 23: Buffers used for anion exchange chromatography 168

Table 24: Molar absorption coefficient (ε) assuming that all cysteines are in a disulphate conformation and molecular weight (Mw) of the different *Mm*GluII constructs. 169

Table 25: Reaction mix used for DSF. 172

Table 26: TR-FRET settings 174

Table 27: Crystallisation conditions for *Mm*GluII$_{tryps}$. 176

Table 28: *Mm*GluII$_{tryps}$ crystal soaks carried out in this study 177

Table 29: Data collection parameters used for X-ray radiation of GluII crystal soaks. 178

Table 30: *Mm*GluII DNA constructs designed for the experiments described in this thesis. 181

Table 31: Morpheus HT-96 screen crystallisation conditions for *Mm*GluII α-β_LDLRa. 190

Table 32: Diffraction data processing statistics for *Mm*GluII α-βLDLRa. 192

Table 33: Data refinement statistics of the *Mm*GluII α-β_LDLRa mutant. 194

Table 34: *N*-alyloxy deoxynojirimycin analouges used in this study 204

Table 35: Diffraction data processing statistics for *Mm*GluII$_{tryps}$ soaked with *N*-alkyloxy deoxynojirimycins. 206

Table 36: *N*-alkyloxy DNJ analouges occupancy refinement according to Figure 60. 212

Table 37: IC_{50}- and Ki values of *N*-alkyloxy iminosugars in combination with *Mm*GluII. 213

Table 38: Summary of GluII crystal soaks with TR-FRET active compounds 253

List of nomenclature

4MUG	4-Methylumbelliferyl α-D-glucopyranoside
Å	Angstrom
aa	Amino acid
Ab	Antibody
Asp	Aspartic acid
asu	Asymmetric unit
AWG	Arbitrary waveform generator
BLM	Black lipid membrane
BSA	Bovine serum albumin
BVDV	Bovine viral diarrhoea virus
CA	Cardiolipin
CADD	Computer aided drug design
$CC_{1/2}$	Correlation coefficient $_{1/2}$
CCCP	Carbonyl cyanide m-chlorophenylhydrazone
CD	Circular dichroism
CF	Correction factor
CF-6	Cyclofos-6, 6-cyclohexyl-1-hexylphosphocholine
CMC	Critical micelle concentration
CNX	Calnexin
CRT	Calreticulin
CST	Castanospermine
CV	Column volume
Cy3	Sulfo-cyamine3 N-hydroxysuccinimide ester
Cy5	Sulfo-cyamine5 N-hydroxysuccinimide ester
CFTR	Cystic fibrosis transmembrane conductance regulator
DAA	Direct-acting antiviral agent
DGJ	1-deoxygalactonojirimycin

DHepPC	1,2-diheptanoyl-*sn*-glycero-3-phosphocholine
DHexPC	1,2-dihexanoyl-*sn*-glycero-3-phosphocholine
DLS	Dynamic light scattering
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNJ	1-deoxynojirimycin
DOPC	1,2-dioleoyl-*sn*-glycero-3-phosphocholine
DPC	*n*-dodecylphosphocholine
DPR	Detergent-to-protein ratio
DSF	Differential scanning fluorimetry
DSS	Sodium trimethylsilylpropanesulfonate
EDEM	ER degradation enhancing α-mannosidase-like protein
EM	Electron microscopy
Ems.	Emission
ER	Endoplasmic reticulum
ERAD	ER-associated degradation
ERQC	Endoplasmic reticulum glycoprotein quality control
Exc.	Excitation
FBLD	Fragment based lead discovery
FRAP	Fluorescence recovery after photobleaching
FTIR	Fourier transform infrared
GA	Gibson assembly
Glc	Glucose
GlcNAc	N-acetylglucosamine
GluI	ER alpha-glucosidase I
GT	Genotype
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HTRF	Homogeneous time resolved fluorescence

HVR1	Hypervariable region 1
IAV	Influenza A virus
IC_{50}	Half maximum inhibitory concentration
IFN	Interferon
IEC	Ion exchange chromatography
ISCAM	Interferometric scattering mass spectrometry
k_{cat}	Catalytic constant
K_d	Dissociation coefficient
K_m	Michaelis-Menten constant
LE	Labelling efficiency
LPR	Lipid to protein (molar) ratio
mAb	Monoclinal antibody
Man	Mannose
ManI	α1,2-mannosidase I
MES	2-(N-morpholino) ethanesulfonic acid
Mm	*Mus musculus*
MR	Molecular replacement
MS	Mass spectroscopy
MTSL	S-(1-oxyl-2,2,2,2-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl-methansulfono-thionate
MPL	Multilamellar proteoliposomes
MWCO	Molecular weight cut off
*N*B-DNJ	*N*-butyl-deoxynojirimycin
NOEs	Nuclear Overhauser effects
NMR	Nuclear magnetic resonance
*N*N-DNJ	*N*-Nonyl-deoxynojirimycin
OCC	Occupancy
OST	Oligosaccharyltransferase
p7	“protein of 7 kDa”
PA	Phosphatidic acid
PBS	Phosphate buffered saline

PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PFA	Proton flux assay
PI	Phosphatidylinositol
PMSF	Phenylmethylsulfonyl fluoride
POPE	1-Palmitoyl-2-oleoyl-*sn*- glycero-3-phosphoethanolamine
PRE	Paramagnetic relaxation enhancement
PS	Phosphatidylserine
PVDF	Polyvinylidene difluoride
RBV	Ribavirin
RNA	Ribonucleic acid
RT	Room temperature
TEMPO	(2,2,6,6-tetramethylpiperidin-1-yl)oxidanyl
TFE	Trifuoroethanol
Thr	Threonine
T_m	Phase transition temperature
TR	Tikhonov regularisation
TR-FRET	Time resolved fluorescent resonance energy transfer
SEC-MALLS	Size exclusion chromatography – small angle laser light scattering
Ser	Serine
SDS	Sodium dodecyl sulphate
SM	Sphingomyelin
SNR	Signal-to-noise ratio
SRT	Shot repetition time
STD	Saturation-transfer differences (NMR)
SVR	Sustained virologic response
Tb	Terbium
TEM	Transmission electron microscopy
TFE	2,2,2-Trifluoroethanol

Tm	Melting temperature
TR	Tikhonov regularisation
UCF	Ultracentrifugation
UGGT	UDP-glucose:glycoprotein glucosyltransferase
V_{max}	Maximal velocity

1. Introduction

The absence of organelles for self-replication in viruses compels them to make use of the host cell's reproduction machinery. Viruses do this by hijacking synthesis pathways within the host cell which are of key importance for their own DNA/RNA replication, as well as for the synthesis and maturation of non-structural (e.g. viral proteases and polymerases) and structural viral proteins (e.g. capsid and envelope proteins). To ensure functional envelope assembly some viruses heavily rely on the correct folding of their N-linked envelope glycoproteins by the host's endoplasmic reticulum quality control (ERQC) machinery. The ERQC is found in eukaryotes and consists of several glycoprotein processing enzymes, as well as lectins and chaperones assisting glycoprotein folding[1,2].

Currently available virostatic agents lack the ability to function as broad-spectrum antiviral agents, making their use virus specific and case depended. Iminosugars, which are derivatives of monosaccharide-analogues, could potentially fill this gap and serve as broad-spectrum antivirals. Iminosugars inhibit host cell enzymes needed for virus replication (e.g. ER alpha glucosidase II), and in some cases also, and in parallel, viral proteins (e.g. hepatitis C virus ion channel protein p7)[3-5], resulting in either a

reduction of secretion and/or decreased virus infectivity of secreted virions, contributing to host organism wellbeing/survival. This thesis focuses on both, ER alpha-glucosidase II and HCV p7, providing structural as well as some functional insights aimed at aiding antiviral drug development.

1.1. Iminosugars

Iminosugars are glycomimetics. The first ever isolated iminosugar – nojirimycin, found in *Streptomyces roseochromogenes* and *Streptomyces lavendulae* – was described in 1966 as a D-glucose mimetic with the endocyclic oxygen atom being substituted by nitrogen, a key feature of iminosugars[6]. The nitrogen substitution prevents the ring from being cleaved by carbohydrate metabolising enzymes and lends itself to further modifications (e.g. the addition of alkyl chains). The rising interest in iminosugars and their until then poor stability due to a hydroxy group in the C_1 position, led to the development of new iminosugars of the glucose-, galactose- and mannose families. A more stable iminosugar, 1-deoxynojirimycin (DNJ, Figure 1B), lacking the C_1 hydroxy group, was chemically synthesised. It was later also isolated as a naturally occurring compound from the roots of the mulberry tree and was named moranoline[7]. Naturally occurring iminosugars are not exclusively products of plants as several strains of *Streptomyces* and *Bacillus* are able to produce different varieties of iminosugars[8-10].

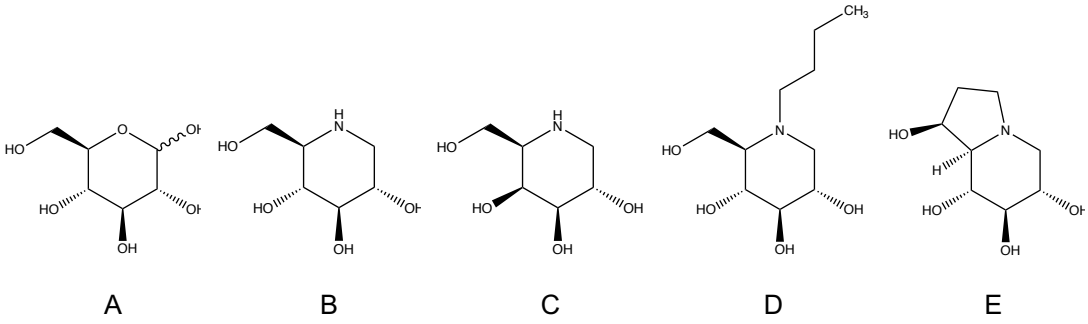


Figure 1: Structures of glucose (A), 1-deoxynojirimycin (DNJ, B), 1-deoxygalactonojirimycin (DGJ, C), *N*-butyl-1-deoxynojirimycin (*N*B-DNJ, D) and castanospermine (CST, E).

Over the years many different DNJ analogues have been described including those in Figure 1. Alkylation of the endocyclic nitrogen can be achieved by chemical synthesis and also occurs naturally[7]. The addition of aliphatic carbon chains increases the bioavailability of iminosugars while at the same time increasing their cell toxicity levels, depending on the chain length attached[11]. The first described naturally occurring *N*-alkyl DNJ (*N*-methyl DNJ) was isolated from the leaves of the mulberry tree, which also co-produces several other iminosugar derivatives. In addition to the mentioned DNJ and *N*-methyl DNJ the plant produces *O*-α-D-glucosides, *O*-ß-D-glucosides and *O*-α-D-galactosides[12,13]. In addition to *N*- and *O*- alkylated iminosugars, castanospermine (CST) is also a naturally occurring bicyclic derivative, (Figure 1E), that was first described in 1980 by Hohenschutz *et al.*[14], found in the seeds of *Castanospermum australe.* Castanospermine causes severe gastrointestinal irritation when taken up orally by livestock and can lead to death.

The potential of iminosugars to act as monosaccharide mimetics is reflected in their ability to serve as broad-spectrum inhibitors for many carbohydrate processing enzymes. Their use affects key metabolic pathways involving glycosidases of the intestinal digestion metabolism, synthesis of glycoconjugates and post-translational glycoprotein processing[15] (see section 1.3). Specifically, DNJ-based iminosugars

affect enzyme activity of endoplasmic reticulum (ER) located glycosidases such as alpha-glucosidase I and II[16,17] as well as intestinal glucosidases[7]. Stereochemical modification of the glucose mimetics' hydroxy groups allowed to create new mannose- and galactose-based iminosugars with shifted inhibition profiles, e.g. mannose analogues, such as kifunensine or swainsonine, inhibit the ER mannosidase I and Golgi α-mannosidase II, respectively[18,19] and galactose analogues, like *N*-butyl-1-deoxygalactonojirimycin (*N*B-DGJ, Figure 1D), the ceramide glycosyltransferases essential for glycolipid synthesis. Andersson *et al.*[20] showed that the ceramide glycosyltransferase inhibition is introduced by the iminosugars *N*-alkyl chain (minimum four carbons) and less by the corresponding headgroup (e.g. galactose- or glucose mimetic). The latter being responsible for its glycosyltransferase specifity, as *N*B-DGJ does not inhibit other enzymes such as ER alpha-glucosidase I and II. The bicyclic glucose mimetic castanospermine has increased specificity for sucrase over intestinal glucosidases[21]

1.1.1. Iminosugars as therapeutics

Before iminosugars found their way into laboratories and clinical studies their therapeutic properties had unconsciously been used for a long time. In traditional ancient Chinese medicine leaves of the mulberry tree were used to cure 'Xiao-ke' (diabetes)[7]. Later, the active compounds in the leaf extract were found to be iminosugars and their mechanism of action became apparent; the inhibition of intestinal glucosidases delays the digestion of complex carbohydrates and thus prevents a sudden rise of glucose in the bloodstream after a meal[3]. Their use for the treatment of type 2 diabetes mellitus continued and especially the development of *N*-

hydroxyethyl DNJ (tradename: Miglitol) for diabetic treatment increased the interest in iminosugar-based drugs when entering the market in 1996[22-24], though these drugs suffer from side-effects such as osmotic diarrhea due to the inhibition of intestinal α-1,4-glucosidases and the resulting accumulation of oligosaccharides in the small intestine[4].

In 2001 another *N*-alkyl DNJ analogue with a slightly longer aliphatic chain, *N*-butyl DNJ (*N*B-DNJ, tradename: Zavesca) was approved for the treatment of Type I Gaucher's disease, a lysosomal storage disease[25]. Due to an inherited inactive or partially inactive glucocerebrosidase in their lysosomes patients suffer a severe increase in glucocerebroside, also known as glucosylceramide, a glycolipid that is usually broken down into ceramide and glucose by the enzyme. Inhibition of acid galactocerebrosidase with *N*B-DNJ prevents such substrate accumulation (substrate reduction therapy) and alleviates disease symptoms. An alternative therapy for the treatment of Type I Gaucher's disease is based on enzyme replacement with recombinant enzyme, which in comparison to *N*B-DNJ treatment is a more expensive strategy[26,27] and not all patients are eligible for this.

Niemann-Pick disease type C is, like Gaucher's disease, a lysosomal storage disorder, characterised by a defect in lipid-trafficking and the resulting accumulation of glycosphingolipids. The administration of *N*B-DNJ causes the inhibition of glycosyltransferases found in the primary steps of the glycosphingolipid biosynthesis pathway and thus reduces the substrate availability for the disease affected β-glucosidase. Substrate reduction therapy does not cure the disease, but it helps patients to live with fewer disease related symptoms[28].

Type 1 Gaucher's disease and Niemann-Pick disease type C are just two examples of many other storage diseases, such as Fabry's disease, Pompe Disease, Tays-Sachs disease and Sandhoff disease, where an iminosugar-based substrate reduction therapy has been suggested as a possible treatment strategy[3].

In 2006 Norez *et al.*[29] showed that a respiratory epithelial cell line carrying a mutation that causes cystic fibrosis (CF, deletion of F508 in the cystic fibrosis transmembrane conductance regulator (CFTR) gene) can acquire a non CF-like phenotype when treated with low dose *N*B-DNJ. The authors hypothesise that *N*B-DNJ prevents interactions between the glycoprotein CFTR and calnexin (CNX), thus allowing the mutated but still functional regulator to leave the ER, while totally misfolded receptors get recognised by chaperones of the ER glycoprotein quality control cycle (ERQC) and degraded by the ER associated degradation (ERAD) pathway[5,30]. For more detailed information about CNX, the ERQC and ERAD see section 1.3. More recently nasal administration of the same iminosugar showed similar effects in cystic fibrosis mice[31].

The proposed application of iminosugars as anti-cancer compounds has been discussed in more detail for the two compounds swainsonine and CST[32-35]. While the former was tested in phase I and II clinical trials it eventually was stopped after showing little therapeutic effects in phase II[36,37].

Iminosugar analogues have also been tested for the treatment of obesity in insulin resistant obese rodent models, showing that *N*-(5′-adamantane-1′-yl-methoxy)-pentyl-1-deoxynojirimycin (AMP-DNJ) normalises elevated tissue glucosylceramide levels, lowers glucose levels and increases oral glucose tolerance, in addition to leading to higher insulin sensitivity in liver and muscle cells[38,39].

1.1.2. Use of iminosugars as antivirals

Over the years many studies showed that a broad-spectrum of carbohydrate processing enzymes is inhibited by iminosugars and their derivatives (see section 1.1.1 and 1.1.2). Due to host cell as well as viral protein inhibition DNJ- and DGJ analogues have potential as broad-spectrum virostatic agents against different virus families.

The development of virostatic agents is often virus specific and depends on the ability of the drug to recognise nucleotide-, carbohydrate- or peptide patterns of a specific virus strain in order to decrease virus infectivity. The enormous adaptive capacity of RNA viruses to their surrounding environment is a result of their viral polymerase lacking the ability to proofread, manifested in the high mutation rate of these viruses[40]. The high mutation frequency of RNA viruses compared to DNA organisms makes direct acting antiviral (DAA) drug design a difficult task.

In order to replicate, viruses are host cell dependent. After infection viruses hijack the cell's biosynthetic pathways and utilise them for their own production of viral RNA/proteins, which later assemble into new, infectious virus particles. Many viral envelope proteins are glycoproteins relying on folding assistance by chaperones in the ER of the infected cells. Iminosugars exploit the virus' dependence on its host by targeting primary ER alpha-glucosidases in the ERQC pathway. Different viruses, such as influenza A virus[41,42], hepatitis B virus[43], human cytomegalovirus[44], Sindbis virus[45], vesicular stomatosis virus[46], dengue virus[47-50] and many others, have reduced infectivity levels after iminosugar treatment. The reduction is usually caused by inhibition of the ER α -glucosidases I and II at the entry of the ERQC, preventing viral glycoproteins from being processed, leading to immature viral envelope glycoproteins. The affected viral glycoproteins are either impaired in forming functional virus particles (e.g. due to misfolding) or get degraded via chaperones of the ERAD pathway

(see section 1.3). However, the antiviral effect of α -glucosidase inhibition is not seen for all viruses, e.g. Rous Sarcoma virus and Semliki forest virus show no effect upon iminosugar treatment[3], and needs to be assessed on a case to case basis.

In some cases inhibition of ER alpha glucosidases leads to the gross misfolding of viral envelope glycoproteins with a subsequent reduction in secretion of viral particles[45,51-54], whereas in others it may only lead to a slight misfolding of envelope glycoproteins, as is the case for HIV. HIV encodes two envelope glycoproteins, gp120 and gp41. Which are expressed in the form of a precursor protein (gp160) in the ER lumen and get proteolytically cleaved after translocation into the cis-Golgi[55]. Non-covalent interactions between the two cleavage products keep gp120 bound to the membrane associated gp41. During host cell infection gp120 on the viral particle binds to the cellular receptor CD4 and after conformational changes exposes gp41 which leads to viral/host cell membrane fusion events. The highly glycosylated gp160 is affected by *N*B-DNJ treatment in a way that does not trigger its degradation nor prevent virion assembly. Changes in the glycosylation pattern and a slight protein misfolding caused by the inactivation of the ER glucosidases prevents gp120 from undergoing conformational changes needed for the exposure of gp41 during cell entry resulting in a decrease of virus infectivity[56]. Clinical trials of *N*B-DNJ for the treatment of HIV were stopped at phase II due to difficulties in maintaining therapeutic concentrations of *N*B-DNJ in serum and the side effect of osmotic diarrhea[5].

Besides host cell inhibition, long alkyl chain iminosugars (e.g. *N*-nonyl-1-desoxynorjiromycin, *N*N-DNJ) also affect some viral proteins directly. *In vitro* ion channel recordings of the hepatitis C virus (HCV) p7 viroporin indicated a loss in channel activity after addition of increasing concentrations of *N*N-DNJ[57,58]. The

mechanism of ion channel inhibition by long alkyl chain DNJ analogues is yet to be elucidated. The effect seems to be *N*-alkyl chain length depended as a shorter *N*-alkyl DNJ derivative (*N*B-DNJ) has no inhibitory effect on ion channel activity. In a study from 2001[59] *in vitro* experiments of bovine viral diarrhoea (BVDV) virus – at the time a pestivirus model for HCV, which could not be grown in cell culture until 2005 – infected Madin-Darby bovine kidney (MDBK) cells showed a correlation between the DNJ *N*-alkyl chain length and BVDV infectivity. Out of the nine DNJ analogues tested *N*N-DNJ showed the strongest inhibitory effects (*N*-alkyl chain length used in that study: C_4 to C_{18}).

Examples of viruses that show reduced virus infectivity upon iminosugar treatment are listed in Table 1. The number of different virus families illustrates the potential of iminosugars to serve as broad-spectrum virostatic agents. Nevertheless, more work is needed to increase their efficacy against additional virus targets, in parallel with the reduction of undesired side effects (e.g. increased gas production and diarrhoea due to unintentional inhibition of intestine glucosidases).

Table 1: Example viruses and corresponding virus families showing reduced virus infectivity upon iminosugar treatment. *Asterik* – experiments carried out on virus pseudotyped particles. *dMM* – manno-1-deoxynojirimycin, *HNJ* – homonojirimycin, *MON-DNJ* – N-(9-methoxynonyl)-1-deoxynojirimycin, *2THO-DNJ* – N-8'-(2''-tetrahydrofuranyl)-octyl-deoxynojirimycin, *UV-5* – (2R,3R,4R,5S)-1-(6-(4-azido-2 nitrophenyl)amino)hexyl)-2 (hydroxymethyl)piperidine-3,4,5 triol.

Example of virus tested	**Virus family**	**Iminosugar**		**Reference**
		Name	**Concentrations tested [µM]**	
Bovine viral diarrhoea virus	*Flaviviridae*	*N*N -DNJ	1-5000	59
		*N*N-DGJ	1-5000	
Dengue virus	*Flaviviridae*	*N*N -DNJ	1-100	47-49
		*N*B-DNJ	1-100	
		DNJ	500	
		CST	500	
Hepatitis B virus	*Hepadnaviridae*	*N*B-DNJ	900-4500	43
Hepatitis C virus	*Flaviviridae*	*N*N -DNJ	5-500	58,60
			1-100*	
		*N*B-DNJ	5-500*	
Human cytomegalovirus	*Herpesviridae*	CST	1250-2500	44
		*N*N -DNJ	2000	
Human immunodeficiency virus	*Retroviridae*	*N*B-DNJ	2000	5
Influenza A virus	*Orthomyxoviridae*	dMM	1000-4000	41,42,61-63
		DNJ	2000	
		*N*B-DNJ	0.1-321.5	
		*N*N -DNJ	0.1-62.5	
		M*O*N-DNJ	82-500	
		2THO-DNJ	1-250	
Sindbis virus	*Togaviridae*	DNJ	1000	45
Vesicular stomatitis virus	*Rhabdoviridae*	DNJ	1000	46
Zika virus	*Flaviviridae*	UV-5	2-100	64

1.2. Viral ion channels – viroporins

Most hydrophilic molecules or substances cannot permeate a membrane without the help of specialised membrane transporters. The transport is energy dependent and two main transporter classes have been established; i) active transporters, which require the use of energy, such as ATP, for the transport of molecules against an electrochemical gradient and ii) passive transporters, which facilitate transport down an electrochemical gradient across a membrane. Ion channels are a type of passive transporters that form a selective channel across the membrane that can open or close to facilitate the passage of ions down their electrochemical gradient.

Viral ion channels are generally referred to as viroporins. This unique class of channels tends to be formed by short peptides of about 60-120 amino acids (aa) which comprise of at least one hydrophobic transmembrane domain. Upon oligomerisation the monomeric subunits form oligomers with a narrow hydrophilic channel at the core. The channel facilitates the selective transport of ions driven by the electrochemical potential established by the membrane[65,66].

Due to the small size of ion channel-forming viroporins, oligomerisation is a necessity in order for them to form functional ion channels. However, their oligomeric order, ion selectivity and gating are very varied. Two of the best characterised viroporins, M2 from influenza A virus and p7 from HCV exemplify this. M2 forms tetramers, is highly selective for protons and opens at low pH[67]. Single channel recordings of M2 embedded in lipid bilayers show clearly separated opening and closing events[68,69]. In contrast, p7 has been reported to form hexamers or heptamers[70-72], shows selectivity for cations[57,70,73-75] and a gating mechanism has not been identified. Depending on the lipid environment channel opening events show ‘burst-like’ or clear opening and closing patterns[57,76]. P7 proton conductance seems to be favoured at low pH[75].

Furthermore, in the case of HCV p7, it has been established that the specific lipid composition of the membrane markedly influences its ion currents[76]. The complexity of viroporins is further increased by the potential coexistence of oligomers of different order[71,72,77], which might be influenced by environmental changes.

The low sequence homology between viroporins and channels of prokaryotic or eukaryotic origin make their discovery quite challenging[78]. The first described viral protein with ion channel activity was the M2 viroporin from the influenza A virus (IAV) in 1992[69]. Over the years many studies followed, making M2 one of the best understood viroporins today and leading to the discovery of other RNA- as well as DNA virus-encoded ion channels[79,80]. For an overview of viruses known to encode a viroporin see Figure 2. To date M2 remains the only viroporin for which high resolution structures of the oligomeric complex exist[67]. A model of the Vpu ion channel from HIV based on nuclear magnetic resonance (NMR) data from the monomer has been described[81,82], as well as for p7 oligomers[71,72,77,83]. A controversial high resolution structure of an HCV p7 hexamer has recently been disproven partly with data presented in this thesis[84]. For more information see section 4.3.

In 2012[85] a classification system for viroporins was proposed based on the number of transmembrane domains of each monomer and the topology of its N- and C-termini. Viroporins with just one transmembrane helix are classified as Class I (e.g. Vpu from HIV-1) and those with two transmembrane helices as Class II viroporins (e.g. 2B from poliovirus). Depending on the orientation of the viroporin's N-terminus the prefix 'A' for ER luminal or 'B' for cytosolic is added to the classification (e.g. Class IIA). A classification for 2+ transmembrane helices has yet to be established, as e.g. viroporins like E5 from the human papillomavirus-16 have three transmembrane helices[86]. In

addition, the system relies on structural relations between viroporins neglecting that even structurally related viroporins often have different functions within virus infected cells, as is the case for example for the two class I viroporins E encoded by the coronavirus and M2 encoded by IAV. The former is a sodium/potassium selective ion channel found in the intermediate compartment of the Golgi and involved in virion formation, while the latter is a proton channel found in the Golgi and cell membrane of virus secreting cells, in the cell-free virus, as well as in endosomes of newly infected cells[81,85] (Figure 2).

As illustrated in Figure 2 and in common with other virally encoded proteins, viroporins have different functions within the infected cell. Their functional variability is not only found between viruses, a viroporin of a particular virus can fulfil several different functions within a cell, ranging from depolarisation of the cell membrane, over de-acidification of certain cell organelles (e.g. Golgi), acidification of others (e.g. of viral particle inside endosomes), to assistance in virion assembly and release[85,87]. Especially the calcium leakage from or the uptake into intracellular stores, like the ER, Golgi or mitochondria, can highly impact cell and cell compartment functionality[88,89]. High calcium concentrations in the cytosol and mitochondria trigger the expression of cytochrome C, and upon release of cytochrome C into the cytoplasm it induces cell apoptosis, as shown for the Sindbis virus 6k viroporin[90,91] (Figure 2B). Viroporins affect a wide range of physiologically relevant cellular processes, such as intracellular trafficking, signalling and reproduction, yet comparably little is known about how they achieve these various functions mechanistically.

The high versatility of viroporins and their involvement in essential viral processes, such as virus particle assembly and release, makes them an interesting drug target.

Over the years several drugs targeting viroporin functionality have been investigated. The first one to be licensed for IAV treatment was amantadine in the 1960s. Although at the time the mechanism of action was unknown and only became apparent several years later, when amantadine was shown to inhibit the influenza A virus M2 viroporin[92-96]. Later still, high resolution structures of the M2 tetramer based on X-ray crystallography, solution NMR and solid state NMR presented conflicting data regarding the inhibitory binding site of amantadine and the structurally related rimantadine[67,97]. After years of controversy, it became apparent that the sensitivity of M2 to its environment affected its interaction with the drugs, and that in the context of a 1,2-dimyristoyl-sn-glycero-3-phos-phoscholine (DMPC) lipid membrane, the transmembrane helices arranged in such a way that allowed the drugs to fit in the pore of the channel, blocking ion conductance[97,98]. An alternative binding site outside the channel was shown to be of lower affinity in the context of the membrane[67,97].

Vpu, encoded by HIV-1, is another drug target within the viroporin family. The channel's cation transfer could be shown to be inhibited by the 5-(N,N-hexamethylene) amiloride (HMA) *in vitro* and *in vivo*, with reduced viral particle formation in cell culture[99-102]. Parts of the structure of Vpu's larger cytosolic domain[103,104], as well as its transmembrane domain[105,106] have been solved. Nevertheless, a binding motif for HMA could not be identified experimentally due to the lack of structural information. Based on a model proposed by Kim *et al.* in 2006[102] of a pentameric Vpu transmembrane domain arrangement, it was suggested that HMA hydrogen-bonds with the pore lining residue S24.

Over the years many *in vitro* and *in vivo* studies helped to further understand viroporins and their function within virus infected cells. The information gained shows once more the importance of these tiny 60-120 amino acid long proteins and their essential role

in the virus' life cycle. Even though viroporins remain poorly understood, they represent an attractive therapeutic target for the development of new drugs.

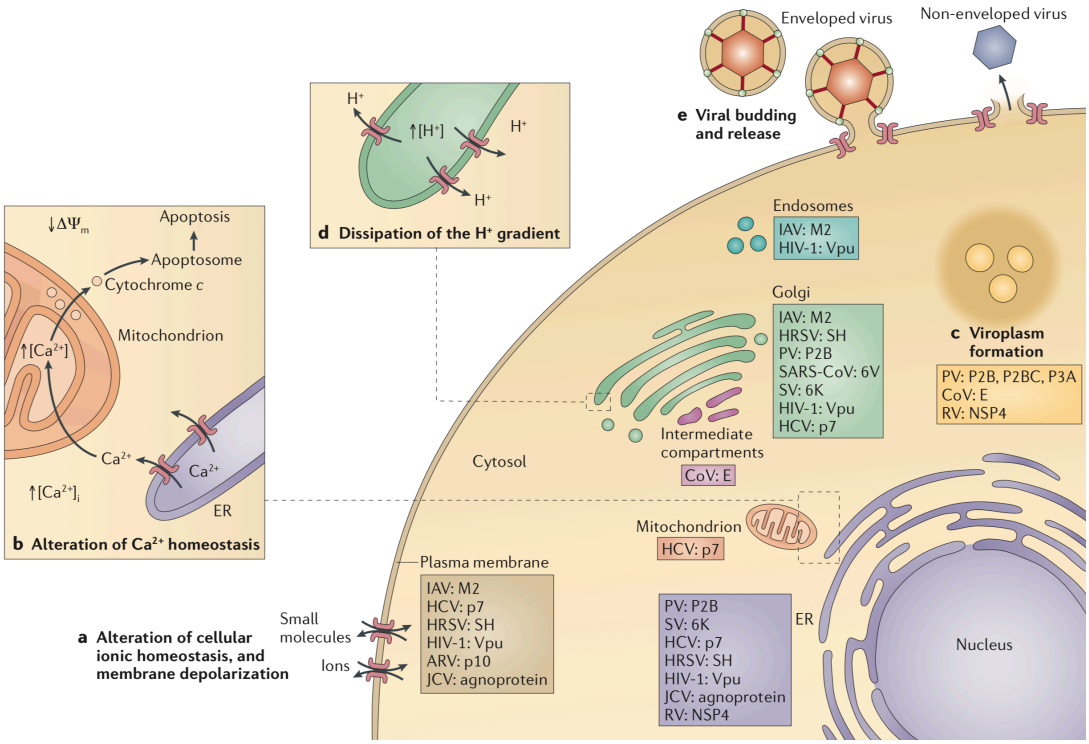


Figure 2: Illustration of host cell organelles affected by viroporins and their cytopathic effects within the virus life cycle. (a) Depolarisation of the plasma membrane's ionic potental due to ion and small molecule permeable viroporins. (b) Release of calcium from the ER lumen increases the cytosolic as well as mitochondrial calcium concentration. The latter activates the production of cytochrome C in the mitochondrion which upon release triggers the cell's apoptosis pathway. (c) Some viroporins are involved in intracellular membrane remodelling leading to the formation of viroplasms (membrane vesicles, viral replication sites). (d) Proton dissipation in the Golgi and trans-Golgi causing de-acidification of vesicular compartments caused by the equilibration of protons with the cytosol. (e) Viroporins play an essential role in the assembly and release of viral particles towards the end of the virus replication cycle. ARV – avian reovirus, CoV – coronavirus, HCV – hepatitis C virus, HIV-1 human immunodeficiency virus 1, HRSV – human respiratory syncytial virus, IAV – influenza A virus, JCV – JC polyomavirus, PV – polyovirus, RV – rotvirus, SARS-CoV – severe acute respiratory syndrome CoV, SV – Sindbis virus. Figure adopted from Nieva *et al.* 2012[85]

1.2.1. Hepatitis C virus

The hepatitis C virus is an enveloped, single positive-stranded RNA virus belonging to the genus *Hepacivirus*, a member of the *Flaviviridae* family[107]. HCV was discovered in 1989[108] and causes acute and chronic hepatitis, liver cirrhosis and in some rare cases liver cell carcinoma[109]. More than 71 million people, 2.35% of the world population, are chronically infected with HCV. The virus is classified into eight genotypes (GT 1-8), GT8 having been recently identified[110], with over 100 subtypes (a, b, c, etc.) known today[111-113], giving evidence to its high mutation rate[114] and genetic variability. GT1 is the most prevalent genotype found worldwide, with sub-genotypes GT1b and GT1a being the most common in Europe and the USA, respectively[115]. Of the 170 million people chronically infected with HCV 1.34 million died in 2015 due to HCV-related chronic liver disease[116].

The HCV genome encodes genes for a 3,000 amino acid precursor (Figure 3, NH_2-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-COOH), which is co- and post-translationally cleaved by cellular and viral proteases into 10 mature viral proteins. The first three proteins of the polypeptide constitute the structural proteins: core (C), of which around 400 form the icosahedral capsid enclosing the viral RNA strand[117], and the two envelope glycoproteins E1 and E2 which are embedded in the host derived lipid envelope of mature HCV virions. E1 and E2 are proposed to stay in a disulphide bond-stabilised complex when exposed at the virus surface. The hypervariable region 1 (HVR1) of E2 is proposed to have a dual function; facilitating virus entry by interactions with the host cell SRBI and CD81 receptors and protecting E2 from being recognised by the humoral immune response due to HVR1 shielding upon CD81 binding. The high variability of the amino acids in the HVR1 region shows high

chemical, conformational and physical conservation of E2 across all HCV genotypes[118-121].

The seven non-structural proteins are required for viral propagation. NS2 is essential for virus assembly, interacting with the E1E2 glycoprotein complex. NS2 also interacts with p7, NS3 and NS5A, helping to cluster viral proteins for virion development[77,122-124]. NS3 has an N-terminal serine protease activity while its C-terminal region contains an RNA-helicase activity[125-127]. NS3 binds NS4A, a small hydrophobic protein, which serves as cofactor for the NS3 protease. The N-terminal part of NS4A facilitates membrane association of the NS3NS4A complex[128]. NS4B is poorly characterised, but one of its functions is to rearrange intracellular membranes for scaffolding of the HCV replication complex[129]. The second last protein encoded by the gene of HCV is NS5A, a highly phosphorylated 56 to 58 kDa protein, depending on its phosphorylation state[130]. Its phosphorylation state can be correlated to HCV replication efficacy[131-134]. NS5B is the viral RNA-dependent RNA-polymerase which is essential for the synthesis of negative-strand RNA from the genomic positive-strand of the virus, as well as for the creation of more positive-stranded viral progeny[135-137].

For more information regarding the viral p7 protein, located between the structural- and non-structural section of the HCV polypeptide, see section 1.2 and 1.2.2.

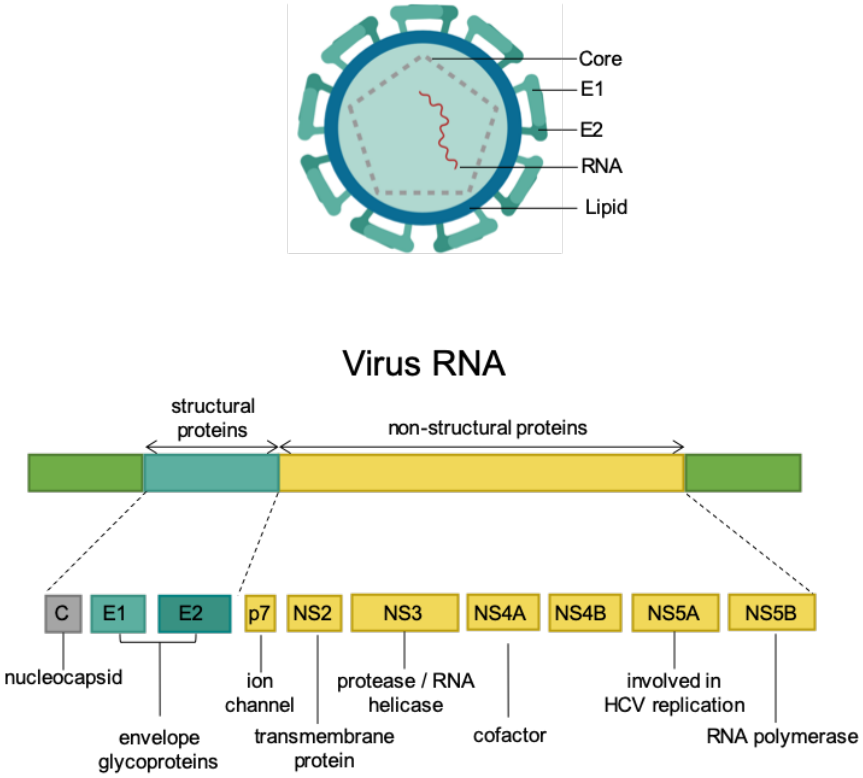


Figure 3: Illustration of the RNA Hepatitis C virus (HCV) virion (top). *Middle* – virus encapsulated genome organisation. *Bottom* – monomeric viral proteins after post-translational cleavage by viral and host cell proteases.

Until 2005 laboratories worldwide had difficulties to grow HCV in cell culture. Only one HCV strain could be grown successfully without adaptive mutations of the virus for the expression of virus particles (GT2a isolate JFH-1) and subsequent tests in functional assays; and even that strain could only be grown in a specific subset of Huh7 cells[138] . In 2005 Lindenbach *et al.*[139] used the replication machinery of JFH-1 (the so-called subgenomic replicon) to create chimeras carrying structural proteins also of other genotypes, for the first time achieving good expression levels also of those virions, showing the importance of that particular viral strain harvested from a Japanese patient suffering from acute fulminant hepatitis C, for *in vitro* RNA replication and thus virus particle production.

Since 2011 guidelines and therapy strategies for patients with chronic HCV infection have changed drastically. Before and around 2011 therapies were based on a combination of pegylated interferon-α (IFN- α) and ribavirin (RBV) which could incorporate specific antiviral drugs targeting e.g. the viral serine protease NS3[140,141]. Over the years interferon-based treatments showed low rates of success paired with serious side effects, caused by the oral application of pegylated interferon-α[142].

Newly developed direct-acting antiviral agents (DAAs) interfere with viral replication by targeting the virus' non-structural proteins. DAAs have a very good sustained virologic response (SVR, > 90%) outcompeting older IFN/RBV-based treatments. Overall DAAs have fewer side-effects than interferon based treatments[143]. Between 2014 and 2016 five DAAs were approved by regulatory agencies in the US and Europe[144], listing e.g. combinations of Sofosbuvir/Ledipasvir (an NS5B polymerase and NS5A inhibitor, respectively), Simiprevir/Daclatasvir (an NS3/4A and NS5A inhibitor, respectively), a more complex regimen comprising of Paritaprevir/Ombitasvir/Dasabuvir (NS3/4A serine protease/NS5A/NS5B inhibition) and Grazoprevir/Elbasvir (an NS3/4A and NS5A inhibitor, respectively). The newly described DAAs cover all HCV GTs, and some are pan-genotypic[143,145].

The inhibition of non-structural HCV proteins through DAAs has the disadvantage that due to the very high mutation rate of the virus, and hence the great genetic variability, HCV can and does develop escape mutants against DAA-based treatment strategies[110-114], as shown for the main DAA targets like NS3, NS5A and NS5B[146,147].

1.2.2. The p7 viroporin

In the viral genome, the HCV p7 protein is found encoded between the structural and non-structural proteins (Figure 3). The p7 monomer consists of 63 amino acids. Upon host infection it is mainly located in the membrane of the ER, and the nuclear envelope of infected cells[148]. The protein has two transmembrane spanning helices connected by a short basic loop region, while the C- and N- termini are pointing towards the ER lumen[149]. In addition an L-shape conformation of the viroporin has been proposed where the C-terminal transmembrane helix is located ER luminal orientated parallel to the lipid bilayer normal[150,151]. This is especially found in the context of incomplete E2-p7 cleavage, occurring naturally in HCV infected cells[152,153], as shown by immunofluorescent microscopy[151].

In vitro, p7 forms cation-selective ion channels in planar lipid bilayers[57,70,73,74]. It is classified as a viroporin class IIA, other members of which include e.g. picornavirus 2B or 6k of alphaviruses.

The protein p7 is essential for the release of infectious virus *in vivo*[154] and *in vitro*[58,155,156]. It is not involved in HCV RNA synthesis, but plays an important role in the later steps of viral particle assembly. In infected cells, p7 has been reported to function as a proton channel dissipating the low pH of acidic compartments such as those of the secretory pathway[75]. P7 performs other essential functions independent of its ion channel activity which involve interaction with other viral or host proteins[148,157,158,159]. It is very likely that p7 interacts with the structural proteins core and E2 as well as the viral protease NS2 and NS5A (Figure 3), although exactly how they interact is yet to be described[160-162]. In the context of the L-shape conformation of p7 and its association to E2, as a result of non-processed E2-p7 precursor cleavage[151-

[153], the complex was hypothesised to be incorporated into secreted HCV virions and thus potentially being involved in virus host cell entry[150]. Lee *et al.*[163] analysed the potential of p7 to be involved in membrane-to-membrane adhesion, in addition to its known ion channel activity. The authors use confocal microscopy to visualise adhesion of unilamellar liposomes in the absence and presence of fluorescently labelled p7. They show an involvement of the viroporin in liposome fusion depending on lipid composition chosen, but fail to explain how p7 mediates fusion. Lee *et al.*[163] refer to the p7 L-shape conformation and an observed aggregation of p7 molecules in their liposome adhesion assay as a potential starting point for membrane-to-membrane adhesion/fusion.

By using confocal microscopy and different fluorescently labelled core-, core-E1 and core-E1-E2 protein constructs in the presence and absence of p7 co-transfection of Huh7.5 cells Boson *et al.*[157] showed that p7 is involved in localising core at the ER. However, the nature of this interaction is yet to be determined.

In the literature, several high-resolution structures of the p7 monomer derived from NMR experiments in different membrane mimetic environments have been described (Figure 4). All these structures predict a 'hairpin-like' conformation where the monomer contains two transmembrane sections separated by a basic loop (residues ~33-37). Penin and co-workers'[77] structure of p7 in 50% trifluoroethanol (TFE, Figure 4A) shows that the first segment of the hairpin consists of two α-helices divided by a kink (residues 14 to 19). Molecular modelling and dynamics simulations of this monomer in POPC bilayers place the first α-helix crossing the water-bilayer interface, and the second α-helix crossing the bilayer to the other side of the membrane (N-terminal transmembrane helix). The second part of the hairpin also contains two α-

helices, with a very subtle kink (~ residue 47) and the last Ct segment (~ residues 57-63) is located at the interface of the membrane in an unstructured fashion. For p7 solubilised in the detergent 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHexPC), Cook *et al.*[164] describe, very similarly, that each of the two sections of the hairpin also contains two helices. However, in this case, the kinks dividing the helices are more pronounced, and the first N-terminal α-helix is positioned perpendicular to what would be the bilayer normal, and presumably would be located at the bilayer interface (see Figure 4B). A third NMR structure of a FLAG-tagged p7 monomer dissolved in methanol reveals a continuous N-terminal transmembrane helix parallel to what would be the bilayer normal, and a continuous C-terminal transmembrane helix. A small section at the end of the C-terminus, which is mostly unstructured in the other structures, forms a short α-helix that is perpendicular to what would be the bilayer normal. Also, the basic loop separating the two transmembrane segments itself contains a short α-helix[165] (Figure 4C). The latest solution NMR structure published by Oestringer *et al.*[166] of p7 solubilised in the detergent 6-cyclohexyl-1-hexylphosphocholine (cyclofos-6, CF-6), is most similar to that of Cook *et al.*[164], except that it presents a small helical segment towards the C-terminal end of the protein (Figure 4D).

The differences in all the monomeric structures could potentially be the result of the different environments used in each study. The more similar structures are those in detergents, which are considered better mimetics of the lipid membrane. The possibility that p7 might be able to change its topology to an L-shape conformation in which the C-terminal transmembrane region lies perpendicular to the bilayer normal has also been proposed based on antibody detection of the N- and C-terminus in cells overexpressing p7[150].

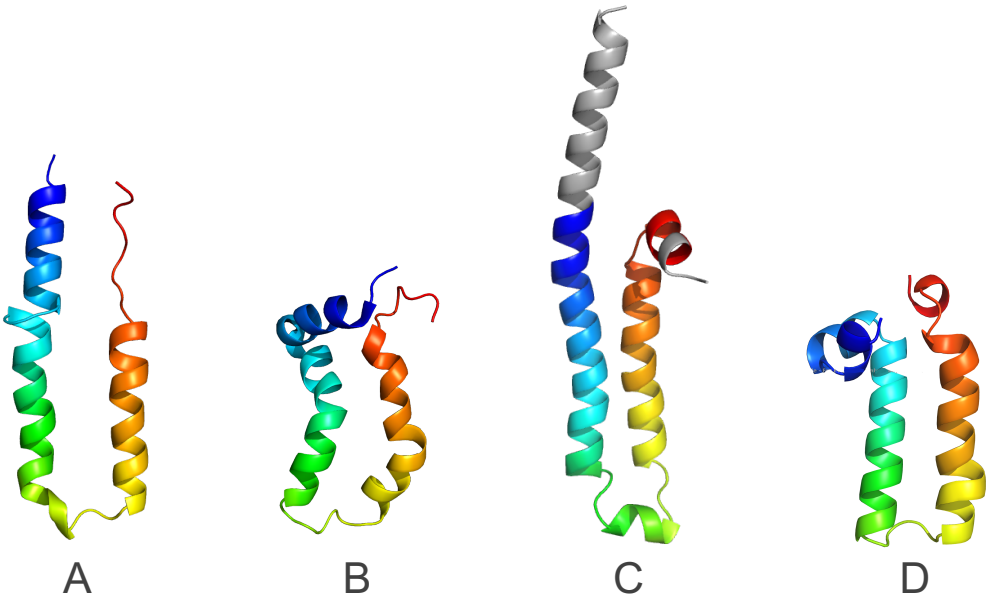


Figure 4: Solution NMR derived structures of the p7 monomer as described in the literature. Structures are shown in cartoon representation with the N-terminus on the left. (A) Montserret *et al.* 2010[77] determined the structure in 50% (v/v) trifluoroethanol (TFE) and modelled it into a POPC bilayer, (B) Cook *et al*. 2013[164] used 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHexPC), (C) the Foster *et al.* 2014[165] structure is in methanol (*grey domains* – additional amino acids introduced by the authors) and (D) Oestringer's *et al.* 2019[166] in cyclo-fos-6 (CF-6).

For the p7 channel/oligomer structural variations are also observed (Figure 5). The N-terminal transmembrane helix is thought to form the pore of the cation selective ion channel. In addition, electrophysiological measurements in black lipid membranes (BLMs) show the inhibition of the p7 ion channel activity by copper. The metal ion is proposed to interact with the pore-lining residue H17 of each of the protomers forming a copper chelating arrangement[167].

Transmission electron microscopy (TEM) studies suggest that p7 protomers assemble into hexamers[70] (Figure 5B-C) or heptamers[71] (Figure 5A) in lipid bilayers, and an EM 16 Å resolution structure in 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHepPC) shows a hexameric arrangement[72] (Figure 5B). Both studies suggest a hairpin-like arrangement of the protomers within the oligomer. Chou and co-workers published an unusual structure of the p7 channel derived from their NMR results in fos-choline-12

(DPC) detergent micelles[168] (Figure 5D). The authors suggested a hexameric p7 structure where a protomer (i) not only interacts with the neighbouring protomer i+1, but also with the protomer i+2 and i+3 in an entangled manner. Additionally, the secondary structural elements and topology in the protomers forming the channel are very different from those described by the other NMR structures of isolated protomers. The structure was disproven by Oestringer *et al.* 2018[84] showing that the authors of the 2013[168] paper misinterpreted intramolecular nuclear Overhauser effects (NOEs) as intermolecular NOEs in their NMR datasets. *In silico* studies of the p7 oligomers further support a hairpin-like conformation[83,169]. For more information regarding this topic see section 4.3 of this dissertation.

Summarising all these published data it becomes obvious that there is a lack of crucial structural information about the p7 mono- and oligomer in lipid bilayers. Methods to analyse the structure of membrane embedded proteins in their physiological environment are rare and need to overcome several obstacles that make the analysis difficult.

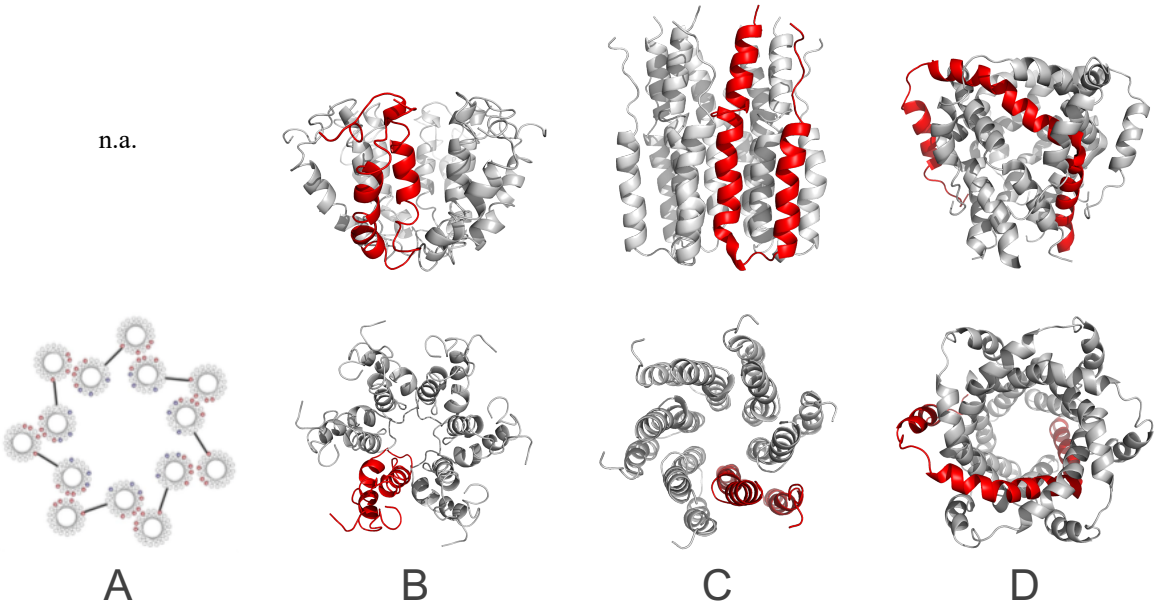


Figure 5: The p7 oligomeric arrangements as described in the literature. Structures are shown in the cartoon representation with a single monomer highlighted in red. (A-C) p7 monomers modelled into hexa- and heptameric arrangements: (A) Diagram taken from Clarke *et al.* 2006[71] based on TEM data of p7 in lipid membranes, (B) model structure fitted into the 16 Å EM volume of p7 in DHepPC from Luik *et al*. 2009[72] and (C) an oligomeric model built by Chandler *et al*. [83], based on the NMR monomer from p7 in 50% TFE from Montserret *et al.* 2010[77]. The NMR derived structure by OuYang *et al.* 2013[168] (D) was shown to be wrong in 2018/2019 by Oestringer *et al*.[84,166]

In 2011 Whitfield *et al.*[76] compared the secondary structure of p7 using circular dichroism (CD) as well as its ion channel activity using single channel recordings in different mixtures of phosphatidylcholine (PC) and phosphatidylethanolamine (PE), the main constituents of the ER membrane. It was observed that in PC-rich membranes the frequency of ion channel openings and closings increased strongly and the CD measurements indicated a loss of α-helicity in the secondary structure of p7 (see Figure 3 of the corresponding study). The authors describe the spectrum in PC-rich environment as an "NB-like" spectrum, due to its similarity to that of one of the transmembrane domains of the ion channel NB from influenza B virus, which has the same unusual shape[170]. A similar behaviour of p7 – changing from a more α-helical- to a more β-sheet-like conformation – reconstituted into lipid bilayers consisting of DMPC or 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) are described by

Fourier transform infrared (FTIR) spectroscopy[171,172], however, in this study the change in secondary structure is attributed to different sample processing, rather than to the specific lipid components.

As mentioned in section 1.2.2 viroporins are important for the virus life cycle and thus are potential targets for therapeutic drug development. The response of p7 to several drugs has always been quite different across genotypes. Rimantadine and *N*N-DNJ show a good degree of inhibition against several p7 isolates[58,173]. In parallel rimantadine was shown to be an effective drug against the M2 ion channel of influenza A virus. While rimantadine is most likely to act by blocking the inner pore of M2, a lower affinity rimantadine binding site outside the channel has also been described[67,97]. Using molecular modelling of p7 it has been proposed that rimantadine could inhibit channel activity through an allosteric site outside the pore[165], but structural data is missing to support this. The same lack of information is encountered for the mechanism of interaction between p7 and *N*-alkyl DNJ analogues, as previously described in section 1.1.2 of this dissertation. Three scenarios of how long alkyl chain iminosugars may inhibit the p7 ion channel could be imagined: i) active pore blockage by iminosugars binding to residues forming the pore, ii) an allosteric inhibition locking the channel in a closed conformation as initially described for rimantadine and M2[67], iii) or the separation of the oligomer by prevention of p7 protomer-protomer interactions by binding of iminosugars to the protomers.

Early studies show inhibitors block p7 ion channel activity *in vitro*[57,70,73] and later studies reveal a correlation between p7 channel activity and the inhibition of production of infectious viral particles *in vivo*[58,75,173-175]. Attempts to understand the mechanism of interaction between p7 and its inhibitors rimantadine, *N*N-DNJ and

amantadine however failed. One reason for this could be the absence of p7 oligomers under the experimental conditions used, which might be needed in order for the drugs to bind.

1.3. The ER glycoprotein folding quality control cycle

Proteins and enzymes need to reach their native secondary structure, usually described by their most energetically favourable state, in order to function properly. They have to fold in a highly specific manner. If folding fails further processing/translocation or transport within the cell is prevented and the protein eventually degraded. To minimize protein degradation cells have developed mechanisms assisting the maturation of proteins[1].

About half of the proteins found in eukaryotes have been estimated to be N-linked glycoproteins, having an oligosaccharide (glycan) attached to one or more of their surface located asparagines[176]. N-linked glycosylation can occur co-translationally or immediately post-translation of nascent polypeptides into the ER lumen. Immediately after an N-glycosylation consensus sequence within a nascent protein – Asp-X-Ser/Thr-Y, where X can be any amino acid except proline – leaves the Sec61 channel of the ER membrane located translocon complex, it is recognized by the tightly associated oligosaccharyltransferase (OST). OST then transfers a pre-assembled precursor N-glycan ($Glc_3Man_9GlcNAc_2$-R, where R represents the protein backbone, Figure 6) onto the asparagine side chain of the consensus sequence[177]. Mellquist *et al.*[178] showed that the consensus sequence glycosylation efficacy is the highest when the Y position is occupied by either arginine, serine or cysteine.

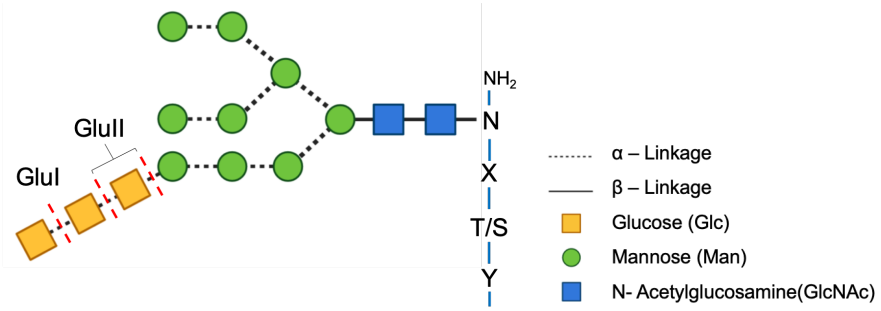


Figure 6: Pre-assembled N-glycan linked to an asparagine residue in the N-glycosylation consensus sequence. N-glycans are transferred by OST onto nascent glycoproteins co-translationally in the ER lumen. ER alpha-glucosidase cleavage sites are indicated by red lines.

The ERQC involves ER located enzymes in combination with chaperones and lectins for N-linked glycoprotein maturation assistance on the basis of glycan- and protein misfold recognition. A schematic of the N-linked glycoprotein processing can be found in Figure 7. ER alpha-glucosidase I (GluI) is the first enzyme of the ER quality control (ERQC) machinery of N-linked glycoproteins and hydrolyses the terminal α(1,2)-linked glucose of N-glycans attached to newly made proteins, leaving $Glc_2Man_9GlcNAc_2$-R[179] (Figure 6). Malectin, the most recently discovered member of the ERQC was identified due to sequence homology analysis of carbohydrate binding enzymes[180]. The membrane associated lectin binds specifically diglucosylated N-linked species and is proposed to be one of the first ERQC glycoprotein misfolding sensors[5,181].

After GluI processing, the diglucosylated glycoprotein species becomes a substrate either for malectin or ER alpha glucosidase II (GluII, for more information see section 1.3.2), the latter removes the middle α(1,3)-linked glucose of the original N-glycan precursor, resulting in $Glc_1Man_9GlcNAc_2$-R (Figure 6). The product – a monoglucosylated glycan carrying glycoprotein – serves as a substrate for either a second hydrolase step by GluII or for recognition by calnexin (CNX)/calreticulin

(CRT, a membrane bound homologue of CNX)[182,183]. The chaperones CNX and CRT mediate interactions with other co-chaperones[119,129]. Subsequent removal of the glycan's innermost α(1,3)-linked glucose by GluII (leaving $Man_9GlcNAc_2$-R) prevents further binding to CNX/CRT. After the second GluII cleavage, there are generally speaking two potential fates for the glycoprotein;

i) The processed glycoprotein is correctly folded (native stage) and gets translocated towards the Golgi apparatus (Figure 7). Along the secretory pathway, towards its site of action, additional glycan processing can occur leading to glycoproteins with e.g. 'complex glycans'. The glycans are most often modified by the addition/substitution of different monosaccharide species e.g. galactose, fructose, mannose, N-acetylglucosamine or sialic acid[184]. Alternatively, proteins may carry oligomannose or hybrid type N-glycan structures (with one or more antennae carrying mannose only, while the others are processed to complex type glycans).

ii) The glycoprotein has still not reached its native stage and is recognised by UDP-glucose:glycoprotein glucosyltransferase (UGGT) which transfers a glucose back onto the glycan, resulting in $Glc_1Man_9GlcNAc_2$-R. The glycan modification enables the slightly misfolded glycoprotein to become a substrate for CNX/CRT again and re-enter the calnexin cycle, eventually reaching its native stage with further processing as described in (i)[185,186]. If a glycoprotein is terminally misfolded, slow-acting mannosidases (e.g. α1,2-mannosidase I, ManI)[187] eventually remove a mannose from the N-linked glycan ($Glc_1Man_8GlcNAc_2$-R) thereby tagging it for recognition by enzymes involved in the ER associated degradation (ERAD) pathway, such as ER degradation enhancing α-mannosidase-like proteins 1-3 (EDEM1-3), which facilitate misfolded glycoprotein retrotranslocation into the cell's cytoplasm for ERAD[5,30].

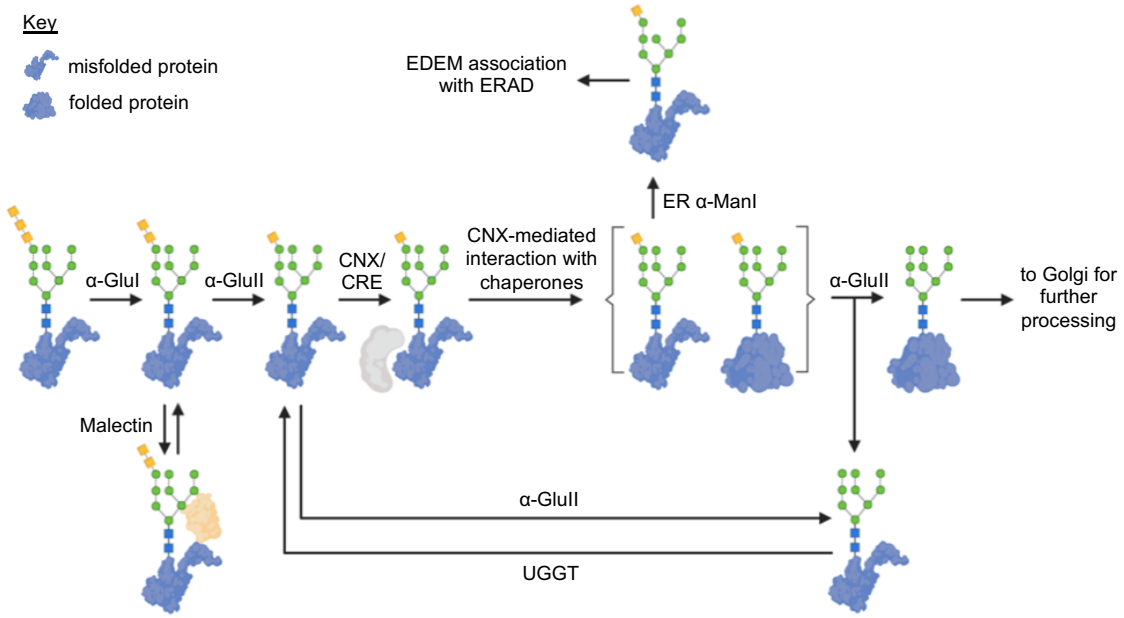


Figure 7: Illustration of the calnexin/calreticulin cycle.

The high number of N-linked glycoproteins estimated in eukaryotes and their associated processing in the ERQC implies a vulnerability of the system for defects or mutations within proteins and enzymes acting in the ERQC. Congenital disorders of glycosylation tend to be early fatal or manifest in rare diseases leading to a non-functional or fragile organism. A clinical proband, a newborn child with a defect in the GluI gene for example survived embryonic development but passed away within 74 days of life with a high degree of neurological defects[188]. Defects in the GluII gene on the other hand led to autosomal dominant polycystic liver and kidney disease[2,189,190].

The usually advantageous ERQC can, on the other hand, also lead to severe disease when e.g. misfold-introducing mutations occur in otherwise still functional secreted glycoproteins, as such proteins get detected by ERQC associated lectins/chaperones and are sent for degradation, preventing the glycoprotein from reaching its site of action within the organism[191,192]. In such cases, partial inhibition of the ERQC can

reverse such symptoms in mature organisms (as e.g. proposed for cystic fibrosis, see section 1.1.1).

1.3.1. The core of the ERQC – the calnexin/calreticulin cycle

The calnexin/calreticulin cycle was first described in 1994 by Hammond et al.[193]. Since then several studies contributed to the understanding of the cycle and the involvement of other partners including soluble and membrane associated lectins, chaperones and other enzymes (see section 1.3 and 1.3.2).

Hammond *et al*. described CNX as a membrane associated ER chaperone found in the lumen of the organelle. CNX and CRT recognize monoglucosylated ($Glc_1Man_9GlcNAc_2$-R) glycans of unfolded or partially unfolded N-linked glycoproteins. Glycan binding is mediated by a lectin-like region of CNX/CRT[194,195]. The chaperones are proposed to prevent protein aggregation and facilitate the recruitment of other co-chaperones assisting correct folding of substrate glycoproteins. Among those co-chaperones are protein disulphide isomerases and peptidyl-prolyl cis/trans isomerases which interact with the P-domain of CNX/CRT[196,197].

When a substrate glycoprotein is not binding to CNX/CRT the last glucose of its monoglucosylated N-glycan can be hydrolysed by GluII, as described previously in section 1.3. After that, correctly folded glycoproteins proceed to their final destination, while immature/slightly misfolded glycoproteins are recognised by UGGT. This enzyme plays a key role in the CNX/CRT cycle as it is the sole glycoprotein misfold sensor that adds a single glucose unit onto an N-glycan which allows the glycoprotein to re-enter the CNX/CRT assisted folding cycle. UGGT's catalytic domain GT24 carries the glycan recognition site as well as a binding pocket for uridine-diphosphate-glucose (UDP-Glc), the activated monosaccharide needed for its glucosyltransferase

activity. The mechanism of substrate protein misfold recognition has yet to be elucidated. It has been proposed that a certain degree of flexibility and a hydrophobic patch recognizing domain is needed in order for the protein to sense misfolded glycoproteins. The full-length crystal structure of *Chaetomium thermophilum* UGGT was published recently[198].

1.3.2. Alpha-glucosidase II

ER alpha-glucosidase II (GluII) is an essential part of the ERQC as previously described (section 1.3.1). The enzyme is acting on N-linked glycoproteins controlling entry and exit from the CNX/CRT cycle. GluII is a ~170 kDa heterodimer consisting of an α-subunit carrying the catalytic site and a tightly associated β-subunit with catalysis supporting properties and an ER retrieval signal[199,200].

The ~110 kDa GluII α-subunit is encoded by the GANAB gene. Due to the catalytic properties and sequence similarities of the α-subunit to other carbohydrate processing enzymes GluII is assigned to the GH31-family. The α-subunit consists of four domains; the N-terminal domain carrying an N-glycosylation site at αN97[201], the GH31 domain containing the catalytic site of the enzyme, the proximal C-terminal domain, followed by the distal C-terminal domain (Figure 9). The majority of the α-subunit structure of *Mus musculus* (*Mm*) GluII was determined by Caputo *et al*. in 2016[201]. The structure is based on trypsin treated full length *Mm*GluII lacking amino acid residues 186-243 and 351-369. The highly conserved active site of members of the GH31 family consists of a WXDMNE amino acid motif with the aspartate being the nucleophile for the catalysed reaction of glycosidases. GH31 enzymes show an archetypal Koshland double displacement retaining hydrolase activity (Figure 8) where a glycosidic bond is attacked by a nucleophile leading to a covalent linkage

between nucleophile and substrate via an oxocarbenium-like transition state. A general acid/base catalyst, often in form of an aspartate, provides the proton for the first half of the reaction and serves as a base upon water addition and hydrolysis. Due to the nucleophile-introduced transition state and the base-facilitated location of the water the monosaccharide retains the α-anomeric conformation upon release[202,203]. The α(1,3) glycosidic hydrolase activity of *Mm*GluII was recently determined by X-ray crystallography identifying αD564 and αD640 as the enzyme's nucleophile and acid/base catalyst, respectively. Within the same study the authors show that the α- over β-anomer selection of GluII is mediated by the stereochemical hindrance of β-anomer substrates by residues αR617 and αM565. The truncated version of *Mm*GluII shows reduced activity in a 4-methylumbelliferyl α-D-glucopyranoside (4MUG) cleavage assay of about 30% compared to wild type, indicating some direct or allosteric involvement of the removed β-subunit domains[201]. *In vivo* GluII catalyses the sequential cleavage of glucose moieties from $Glc_2Man_9GlcNAc_2$-R and $Glc_1Man_9GlcNAc_2$-R N-linked glycoproteins[204]. Between the cleavage of the middle and the innermost glucose, of an originally triglucosylated N-glycan (Figure 6), a conformational reorientation of the substrate in the active site has to occur for GluII to catalyse the glucose hydrolysis from the monoglucosylated species[205]. Additional *in vivo* studies indicate that catalysis of diglucosylated species processes rapidly while cleavage of the monoglucosylated species proceeds more slowly[206,207]. This difference is proposed to form a catalytic delay which allows CNX/CRT to interact with $Glc_1Man_9GlcNAc_2$-R N-linked glycoproteins to assist their correct folding[208], illustrated in Figure 7.

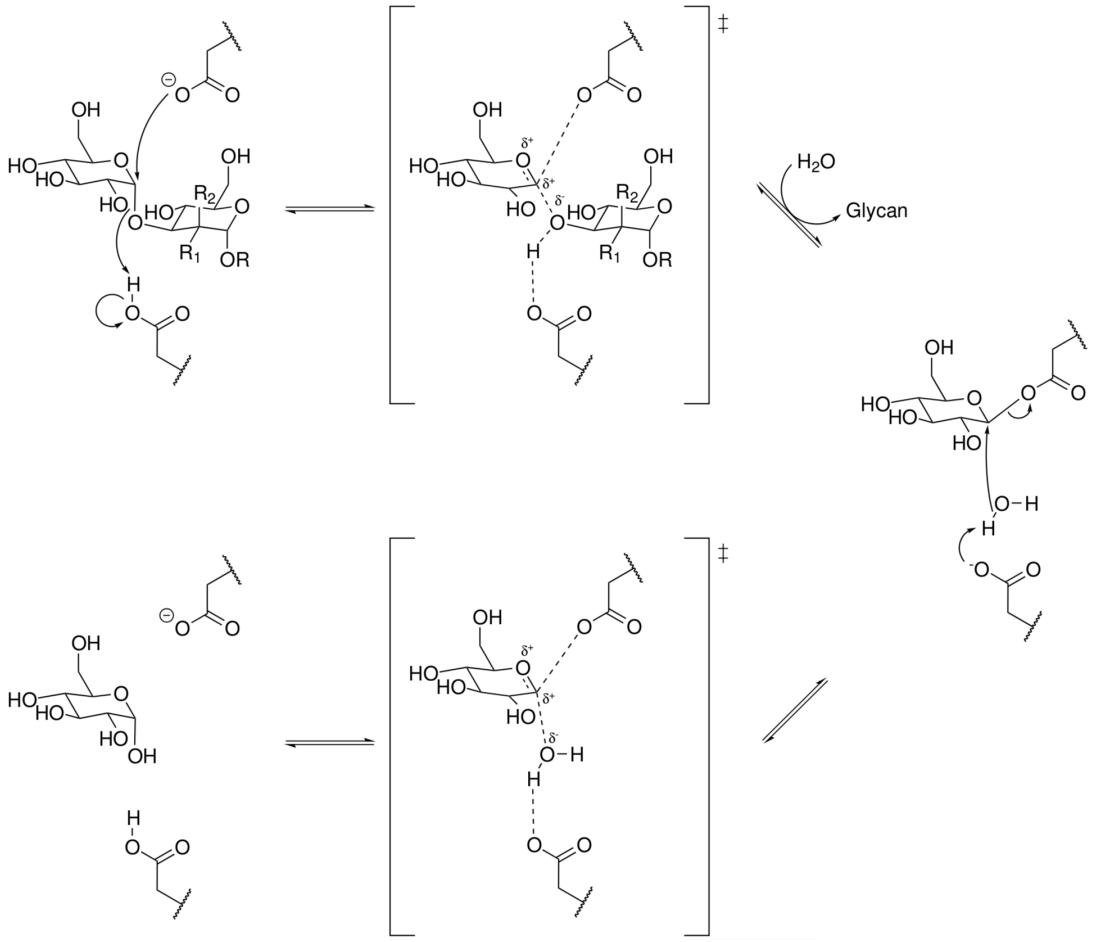


Figure 8: Example for the GluII hydrolase reaction of a diglucosylated substrate. The schematic shows the nucleophile and acid/bace catalyst involved, as well as the retainment of the α-anomeric conformation of the released monosaccharide. *Brackets* – generalised transition state, *R* – $Man_{9(8)}GlcNAc_2$-protein with R_1 = OH and R_2 = H in the case of $Man_9GlcNAc_2$-protein and R_1 = H and R_2 = OH in the case of $Man_8GlcNAc_2$-protein. Figure courtesy of Dr A. Caputo.

The ~60 kDa GluII β-subunit is encoded by the protein kinase C substrate 80K-H (PRKCSH) gene. The gene name stems from the initial discovery of the encoded protein and its sequence similarities to enzymes of the protein kinase C family as well as its property to migrate at about 80 kDa on SDS-PAGE under reducing conditions. This behaviour is likely to be caused by negatively charged residue patches along the amino acid sequence[200]. About 75% of the β-subunit's elongated structure[201] are still unknown. The N-terminal domain of the GluII β-subunit was solved together with the majority of the α-subunit in 2016 by Caputo *et al.*[201]. It shows two tandem repeats of

low-density lipoprotein receptor type-A (LDLRa) domains (Figure 9) around central Ca^{2+} ions. Due to the use of a serine protease during sample preparation other parts of the β-subunit were proteolysed and are not present in the published structure. The LDLRa domains are tightly associated to the proximal and distal C-terminal domains of the α-subunit through non-covalent binding facilitated by up to 16 salt bridges and 13 hydrogen bonds. For additional information about the α-/β-interface see section 6.7. The β-subunit domains following the LDLRa sequence are proposed to show structural features of two tandem EF-hands followed by an unstructured acidic stretch. Structural properties of the C-terminal mannose-6-phosphate receptor homology (MRH) domain were determined in 2013 and 2015 using NMR and x-ray crystallography, respectively[209,210]. The β-subunit's MRH domain shows a fattened β-barrel structure and has been proposed to be involved in glycoprotein substrate recognition, due to its ability of mannose binding and its potentially close proximity to the GluII active site[209,210]. Small angle X-ray scattering (SAXS) analysis of the GluII heterodimer supports this assumption. However, if the same N-glycan is proposed to bind to the MRH domain while having a glucose residue removed by the active site of the same MRH domain containing GluII enzyme, the authors have to postulate that the β-subunit undergoes large conformational changes in order to be able to bind to terminal mannoses of the same glycan[201]. In their 2015 study Olsen *et al*. describe the MRH mannose binding pocket to be formed by four conserved residues: Q384, R414, E433 and Y439 (numbering according to *Schizosaccharomyces pombe*) with a short amino acid loop region, in close proximity to the binding pocket, changing orientation upon mannose binding. The conserved tryptophan in the loop is proposed to interact with structural features of either the α- or β-subunit of GluII, due to it being found in

intermolecular interactions between neighbouring copies of the protein in the crystal packing[210].

The β-subunit carries a C-terminal HDEL motif keeping the GluII heterodimer associated with the ER membrane bound HDEL-receptor and thus preventing it from entering the secretory pathway via the Golgi[211-214].

The previously described re-orientation of the monoglucosylated glycan within the GluII active site was associated with stabilising/retaining effects of the β-subunit[204,215]. Furthermore, a recent study shows the involvement of the β-subunit in the ER stress response and its function as a regulator for the activation of inositol-requiring enzyme 1α (IRE1α) in the unfolded protein response (UPR)[199].

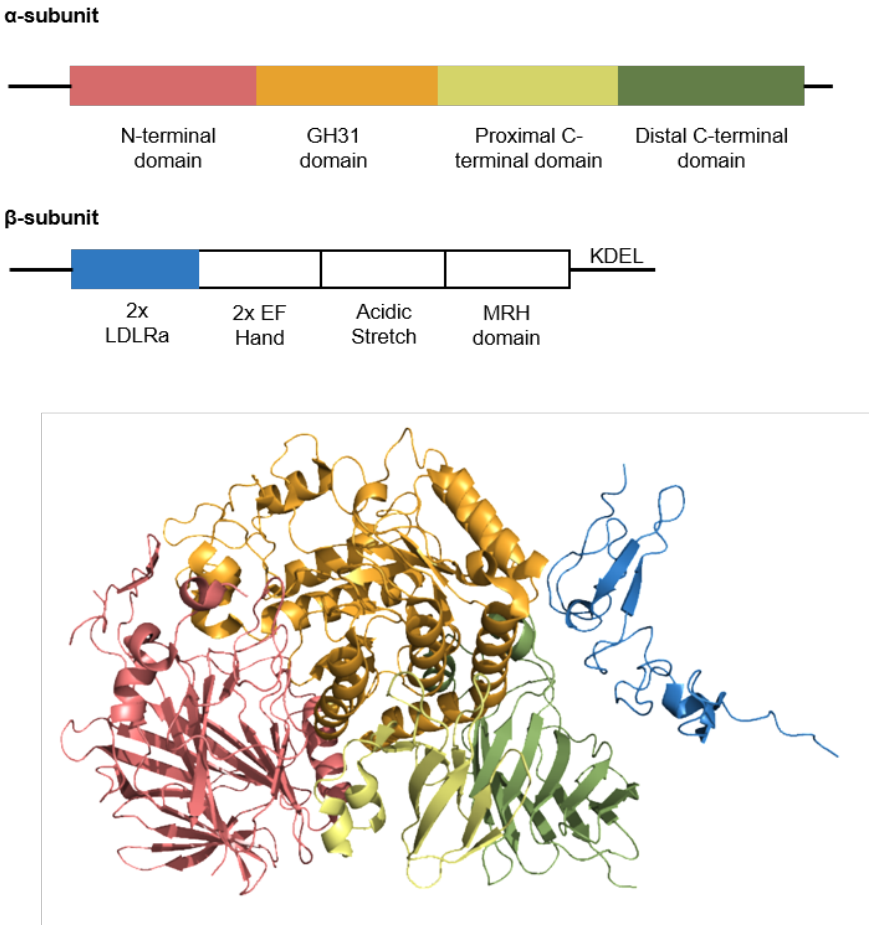


Figure 9: Illustration of the GluII heterodimer domains. *Top* – domains are colour coded matching the corresponding GluII structure domains in the cartoon repesentation at the botom (Structre PDBID: 5IEG).

In summary, the three key structural features of the GluII heterodimer, which are essential for the protein's correct *in vivo* function in the ERQC are the di- and monoglucosylated glycan processing region of the GluII α-subunit, the tightly associated α-/ β-interface formed by C-terminal domains of the α-subunit and the two LDLRa domains of the β-subunit (alongside potentially further interaction surfaces with the as yet unsolved parts of the β-subunit), as well as the ER retrieval signal encoded by the β-subunit's C-terminus, keeping the whole GluII heterodimer in the ER lumen[216]. Though the *Saccharomyces cerevisiae* GluII β-subunit does not contain an ER retrieval signal. This finding implies another mechanism of ER retrieval for this

GluII enzyme, like potentially the interaction of GluII with other ER membrane associated proteins or complexes[200,217].

Studies have shown that mutation within the GANAB (GluII, α-subunit) or PRKCSH (GluII, β-subunit) gene lead to autosomal dominant polycystic liver and kidney disease[2,189]. Patients with genetic diseases based on the deletion of the entire enzyme or either of its subunits are not reported, indicating that GluII is crucial during early organism development.

The literature highlights the importance of GluII as a 'gate keeper' for the ERQC of N-linked glycoproteins. The enzyme controls entry and exit of glycoproteins into and from the cycle. Recent studies help to better understand structural as well as enzymatic features of this delicate quality control system, but currently available drugs are not specific enough for GluII to avoid off-target effects often caused by the inhibition of glucosidases in other compartments of the eukaryotic body. The development of highly specific GluII inhibitors based on active site- as well as allosteric inhibition would be of interest for the treatment of many diseases, including their potential to serve as a broad-spectrum antiviral agents.

1.4. Project Aims

Long alkyl chain iminosugars inhibit the activity of the HCV viroporin p7. Due to the absence of relevant structural information of the p7 oligomer in artificial membrane-mimicking environments, the interactions between iminosugars and the p7 protein have yet to be determined at the molecular level. In this study nuclear magnetic resonance (NMR) spectroscopy as well as small-angle laser light scattering (SEC-MALLS) are used to analyse the p7 oligomeric state in the detergent n-dodecylphosphocholin (DPC). In addition electron paramagnetic resonance (EPR) and double electron-electron resonance (DEER) in combination with circular dichroism (CD) are the techniques of choice to gain more structural information about the viroporin p7, and to establish a basis for the analysis of p7-iminosugar interaction in lipid bilayers.

For ER alpha-glucosidase II, two aspects in particular were investigated: i) The scope of improving iminosugars as active site inhibitors of the enzyme, by either improving their efficacy and/or lowering their toxicity, in particular by introducing oxygen in various places along the alkyl side chain and investigating the effect this has on binding to the enzyme. For this, X-ray crystallography was used to gain structural information about the binding motif of a new class of *N*-alkyloxy DNJs within the active site of GluII. ii) A novel approach of GluII inhibition by disrupting its heterodimerisation is being explored. For this, a high-throughput time-resolved Förster resonance energy transfer (TR-FRET) assay was developed in order to identify and describe potential interface active compounds *in vitro*, as a first step for the detection of novel GluII inhibitors.

2. Experimental principles

2.1. Continuous wave electron paramagnetic resonance

Electron paramagnetic resonance (EPR), or electron spin resonance (EPR), is a technique sensitive to the motion of molecules with paramagnetic centres such as unpaired electrons or metal ions. EPR detects their absorption of energy from microwave irradiation in a strong magnetic field.

Over the years several methods were established in the field of protein structure analysis. The most powerful and commonly used ones are NMR and protein X-ray crystallography. But even these powerful methods reach their limits when used for the analysis of high molecular weight proteins or protein folding in real-time. They are also not particularly useful for the analysis of membrane embedded proteins. To date only a few examples for NMR structure analysis of membrane proteins can be found in the literature. X-ray crystallography approaches to membrane protein structure elucidation are based largely on the formation of a liquid cubic phase adopted by lipids of the monoolein family. These have led to the discovery of over 700 unique structures in the last two decades[218], however the number of membrane embedded proteins

solved by X-ray crystallography is still low compared to the number of soluble proteins. In general, information for this important protein type is sparse due to the insolubility and instability of their transmembrane domains, which make structural studies extremely challenging[219,220].

Since the publication of Hubbell and co-workers in 1990[221] site-directed spin labelling (SDSL) in combination with EPR has become an effective tool for analysing proteins. EPR measurements are not limited by the high molecular weight of proteins. The technique can also be used to analyse proteins in membranes and even in multilamellar liposomes, which allows analysis of biomolecules under native-like conditions. With an EPR time scale dynamic sensitivity of 0.1 up to 300 ns this method also allows the analysis of protein folding in real-time.

A common strategy in SDSL for the attachment of EPR spin labels to the protein backbone is site-directed mutagenesis. This allows attaching spin labels at defined sites within the protein amino acid sequence, e.g. in the active site or at positions with high motion of the protein backbone. In general, single amino acids in the original sequence are replaced by cysteine, followed by modification of the sulfhydryl group with a spin label reagent[33].

The spin labels often carry an unpaired electron in form of a nitroxide group or are composed of metal ions in chelating complexes. An unpaired electron has a magnetic moment with a defined energy level (Figure 12, $B_0 = 0$). If a strong external magnetic field is applied to the electron the magnetic field causes splitting of the energy level as the electron aligns parallel (Figure 12, ms = -1/2) or antiparallel (Figure 12, ms = +1/2) to the magnetic field. By increasing the magnetic field (Figure 12, $B_0>0$) the splitting

between the two energy levels gets more pronounced. This effect was discovered in 1882 by Pieter Zeeman and was named after him ‘Zeeman effect’.

The Zeeman effect is used in EPR to split the energy level of an unpaired electron while applying a strong magnetic field (≈ 0.35 Tesla). It can be described with the spin Hamiltonian ($\widehat{H}_z$):

$$\widehat{H}_z = \mu_B * \vec{B} * g * \hat{S} \quad (1)$$

where μ_B is the Bohr magneton, $\vec{B}$ the magnetic field tensor, g the electron’s g-tensor, and $\hat{S}$ the spin operator. By applying the eigenvalues the energy level of the spin can be described by:

$$E = \pm\frac{1}{2} g \mu_B B \quad (2)$$

When the correct microwave frequency (v) is applied equation 2 can be used to describe the energy transition (ΔE) between the two spin populations, by:

$$\Delta E = hv = g\mu_B B \quad (3)$$

where h corresponds to the Planck’s constant. Equation 3 is true, when the absorbed microwaves allow the spin from the lower energy level to reorientate to the higher spin state in the Boltzmann distribution (Figure 04). This absorbance of microwaves can be detected and is generally shown as the 1st derivative of the absorbance as a function of the strength of the magnetic field (Figure 10).

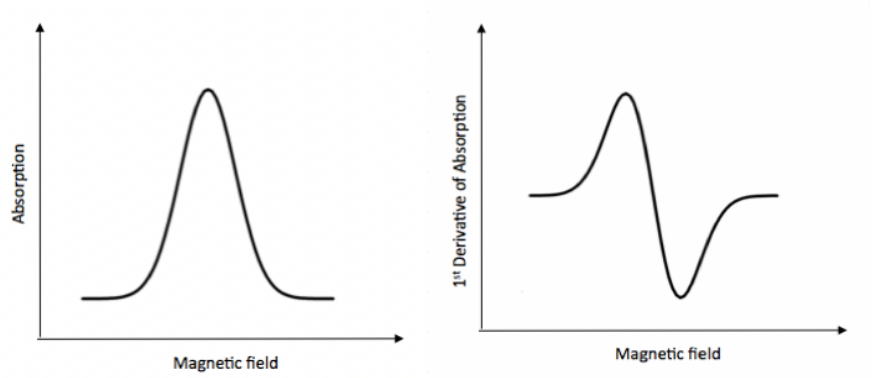


Figure 10: Microwave absorption spectra by an unpaired electron in an applied magnetic field. *Left* – Microwave absorption over change in a magnetic field. *Right* – 1st derivative of the microwave absorption shown over the change in a magnetic field, conventional representation of an EPR spectrum.

In this study S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl) methyl methanesulfonothioate (MTSL, Figure 11) was used as a spin label probe for EPR-based techniques. This nitroxide spin label is highly sulfhydryl specific and can be attached to cysteine side chains of cysteine containing proteins by formation of a disulphide bond (Figure 11). MTSL with a molecular weight of 264.3 g/mol is one of the most commonly used spin labels in SDSL studies. It has a small molecular volume and when attached to a cysteine the size is similar to that of a tryptophan side chain. The distance from the NO in the nitroxide group to the C_β in the amino acid side chain varies between 5-8 Å depending on the flexibility of the spin label side chain[222]. Several experiments like CD, activity measurements and thermal unfolding studies indicate that MTSL has no influence on protein folding[223].

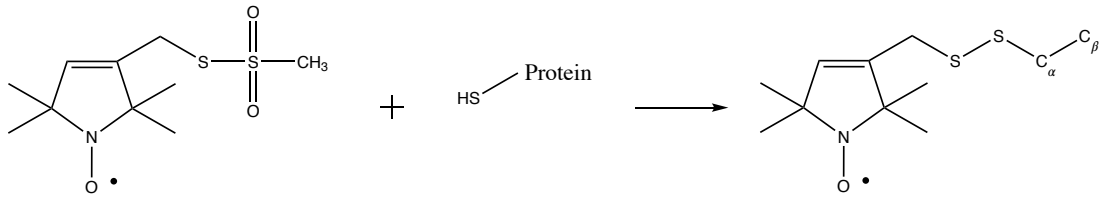


Figure 11: EPR spin label S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) carrying an unpaired electron at the nitroxide oxygen (little black dot). MTSL can be attached to cysteine side chains by formation of disulphide bonds.

MTSL carries a stable unpaired electron at its nitroxide oxygen (N-O•). The unpaired electron is influenced by the nuclear magnetic moment of the nitrogen in the nitroxide's N-O group[222]. Nitrogen has a nuclear spin (I) of 1 which causes a splitting of the unpaired electron energy levels at the high and low Zeeman states, the so-called hyperfine splitting of the energy states[224] (Figure 12). Under constant microwave power and a variable magnetic field equation 3 becomes true, resulting in three absorption peaks in the EPR spectrum (Figure 13). Hyperfine splitting comes into action when nuclei with a spin > 0 (e.g. nitrogen, fluorine, hydrogen) are in close proximity to unpaired electrons and can be described by:

$$\hat{H} = \mu_B * \vec{B} * g * \hat{S} + \hat{I} * A * \hat{S} \qquad (4)$$

where A represents the stand of the hyperfine tensor and $\hat{I}$ the number of spin operator ($\hat{I}$ is calculated: 2I+1).

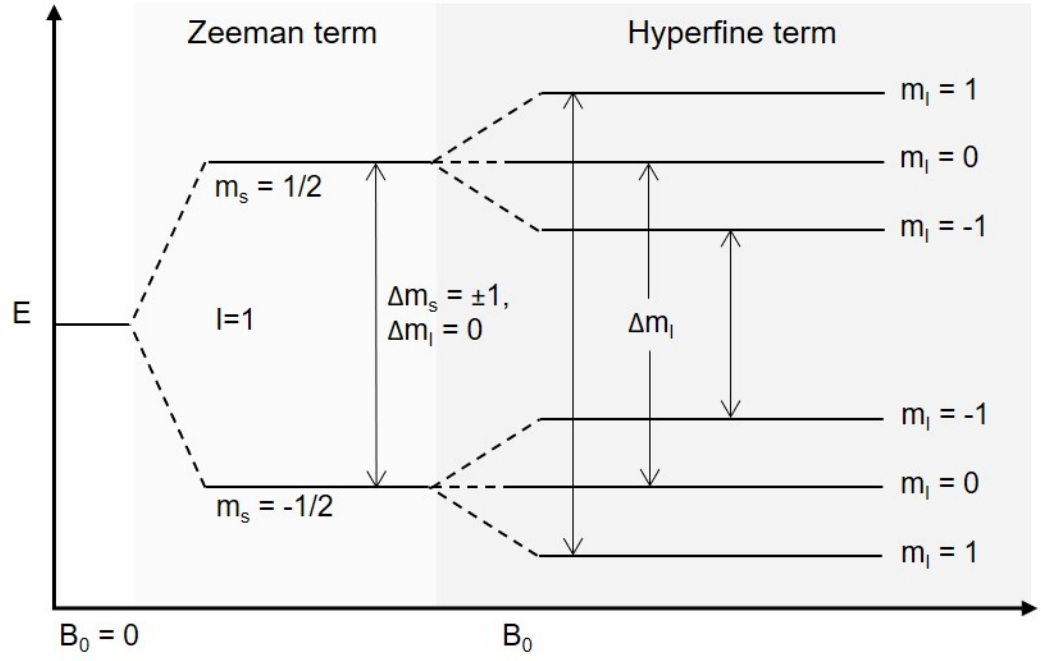


Figure 12: Single electron energy state landscape in an external magnetic field (B_0). *Left* – Low (antiparallel, ms = +1/2) and high energy state (parallel, ms = -1/2) of an unpaired electron in relation to a changing magnetic field. The separation into two different energy levels is called Zeemann effect. *Right* – Representation of the Zeemann effect described on the left, with additional hyperfine splitting. In the case of the nitroxide group the additional splitting is caused by the nucleus spin of nitrogen (I = 1) in close proximity to the radical carrying oxygen.

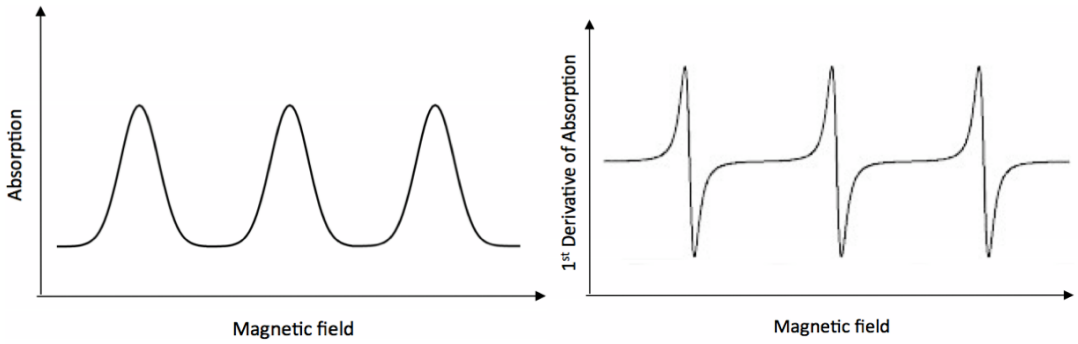


Figure 13: Microwave absorption spectra of MTSL in a changing magnetic field, influenced by hyperfine splitting. *Left* – absorption. *Right* – 1st derivative of absorption. Conventional representation of EPR spectra.

Continuous wave (cw) EPR applies constant microwave frequencies over an increasing magnetic field, as described above, and combined SDSL can give information about the spin label side chain local dynamics, the solvent accessibility,

the polarity of its immediate environment and intra-/intermolecular distances between two or more spin labels[222]. The polarity of the nitroxide local environment is represented by an increase in hyperfine splitting. A more polar environment causes the unpaired electron to populate the π_Z orbital of the nitrogen to a larger extent resulting in a stronger hyperfine splitting.

Equation 4 can be used to describe the electron's dynamics within a given spin system. It is defined by the g (g-tensors), describing the interactions of the electron magnetic moment with B_0, and the electron's hyperfine interactions A (A-tensors). Due to the intrinsic directionality of the electron accumulating oxygen *p-orbital* both magnetic as well as hyperfine interactions are influenced with a degree of axial anisotropy[225]. The anisotropy can by described by:

$$g = \begin{Bmatrix} g_{xx} & g_{xy} & g_{xz} \\ g_{yx} & g_{yy} & g_{yz} \\ g_{zx} & g_{zy} & g_{zz} \end{Bmatrix} \quad A = \begin{Bmatrix} A_{xx} & A_{xy} & A_{xz} \\ A_{yx} & A_{yy} & A_{yz} \\ A_{zx} & A_{zy} & A_{zz} \end{Bmatrix}$$

whereby the x-axis is in alignment with the N-O bond vector[226].

Due to the general anisotropic nature of nitroxide spin labels their Hamiltonian describes a partial average of the anisotropies, reporting on local dynamics of the spin label. The recorded spectrum thus will be influenced by isotropic as well as anisotropic motions, both contributing to the spin's anisotropy. In the fast-motion regime (isotropic motion) nitroxide EPR spectra will appear sharp with three distinct line modulations (see hyperfine splitting Figure 13 and Figure 14A), whereas slow-motion (anisotropic motion, compared to the EPR timescale) will result in a broadening of the line shape[226].

For nitroxide labelled proteins the spin label local dynamics are mainly influenced by three components: i) the relative motion of the nitroxide compared to the attached protein, ii) the overall protein secondary structure movements, which might be very different between proteins due to their secondary, tertiary and quaternary arrangement and iii) the general dynamics of the protein environment, the latter being negligible for proteins >15 kDa and membrane associated proteins, as their overall correlation times are too slow and thus can be regarded as 'static' on the EPR timescale (0.01 to 300 ns)[227]. Figure 14 shows spectra simulations of a nitroxide spin at different correlation times. The spectra show the line shape broadening due to changes in the spin label's dynamics, occurring when changing from an isotropic (A, fast correlation times) to an anisotropic motion regime (D, slow correlation times).

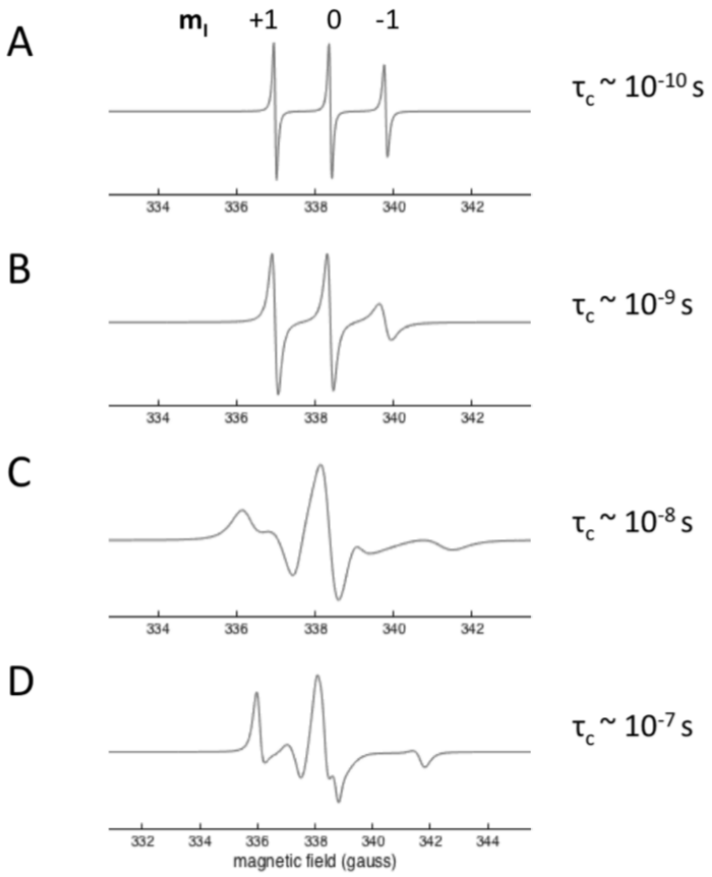


Figure 14: CwEPR spectra simulation at X-band and 298 K of a single nitroxide spin at varying correlation times (τ_c); 10^{-10} s, 10^{-9} s, 10^{-8} s and 10^{-7} s, A-D, respectively. The nitroxide's mostly isotropic motion regime (static) in A) shows a line shape broadening towards slower correlation times in D). The spin label's radical is described by the g-value tensors: $g_{xx} = 2.0085$; $g_{yy} = 2.0063$, $g_{zz} = 2.002$ and the axial anisotropy set to: $a_{xx} = 5.8$; $a_{yy} = 5.8$ and $a_{zz} = 30.8$ Gauss. Figure courtesy of Dr. M. P. Pfeil.

The mixture of two or more populations of spin labels with different motions will result in a two or more component spectrum which is a convolution of the spectra arising from the different populations of spin labels if the exchange rate between such populations is slower than about 100 ns. To illustrate this, an example is shown in Figure 15. (A) and (B) show two different states of an oligomer forming protein. (A) shows the oligomeric state of a putative p7 hexamer with the corresponding EPR spectrum. The motion of the attached spin label (red dot) is restricted due to protein-protein or protein-lipid interactions in the oligomer. (B) shows the monomeric state with fast-motion of the spin label. Both states are in equilibrium with an exchange rate

slower than 100 ns. The analysis of this mixture would result in a two-component EPR spectrum plotted in (C). Both spectra – the fast- and slow-motion – will be added to each other as the exchange rate between the two populations is very slow (>100 ns) with respect to the EPR time scale.

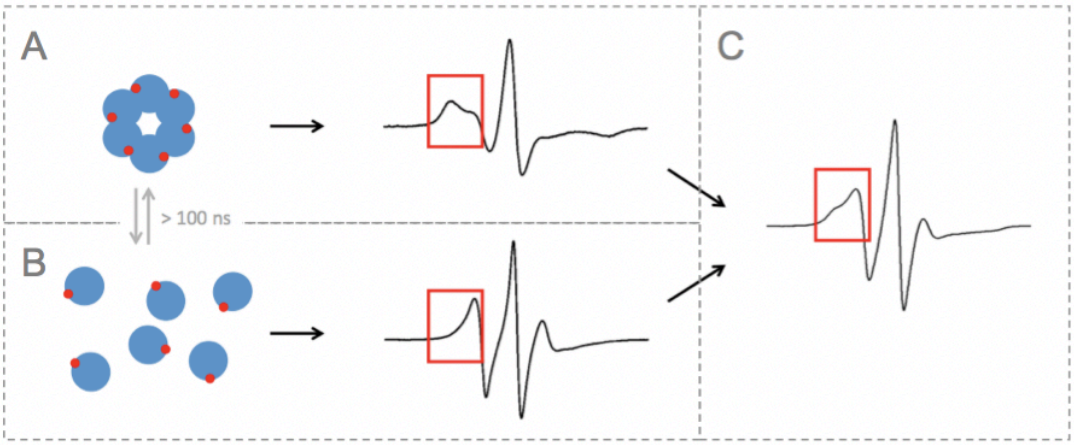


Figure 15: Example of EPR spectral convolution. (A) and (B) show two hypothetical different states of an oligomer forming protein (blue dot), with the corresponding EPR spectra. The red dot represents the attached spin label. If we were to assume that both protein states are in equilibrium and the exchange rate between these two states is slower than 100 ns, then (C) shows the spectrum expected for such a sample, which would result from the convolution of (A) and (B). The red square highlights the region where spectra (A) and (B) differ the most. In the same highlighted region in (C) two peaks arise from the convolution of (A) and (B).

The environment of the spin label therefore influences its dynamics and thus the resulting EPR spectrum. If the mobility of the spin label side chain is restricted e.g. by the backbone of the attached protein, or other molecules in close proximity, the central line width (ΔH_0, Figure 16) and the apparent hyperfine splitting at low and high field (A_{zz}, Figure 16) are increased. "For the MTSL side chain in the interior of a protein, the reorientational motion of the nitroxide is completely restricted and the EPR spectral shape resembles that of a so-called: powder spectrum."[222] An example for a powder spectrum is shown in Figure 16.

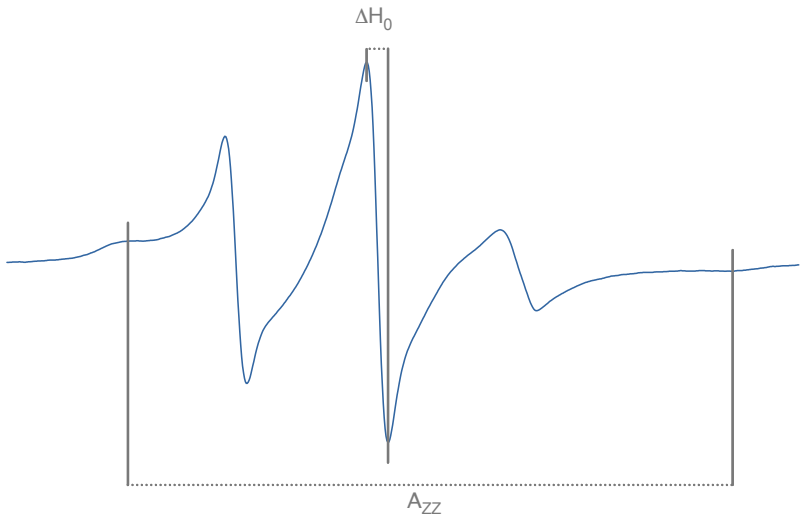


Figure 16: Characteristics of an EPR spectrum: ΔH_0 – line width of the centre field peak, A_{zz} – apparent hyperfine splitting.

Continuous wave EPR can also give information about the intra- or intermolecular distances of coupled spins by identifying spin-spin interaction. Spin-spin interaction can be caused by two different kinds of interaction; exchange- and magnetic dipol-dipol interaction between unpaired electrons of spin labels.

"The exchange interaction leads to an exchange of the spin states of two interacting radicals and thus induces flip-flops between opposite orientated spins."[228].

Dipol-dipol interaction can be seen as the interaction between an electron and the spin (I) of a nucleus (see above). The magnetic moment of an electron can influence the magnetic moment of another electron in its vicinity.

Heisenberg spin exchange causes a broadening of the EPR spectrum when the two spins are in close proximity to each other (< 1.5 nm) and their orbitals start to overlap. In this extreme case the electrons are in a "near covalent status". The Heisenberg exchange coupling for organic radicals can be described in an isotropic way by:

$$\widehat{H}_{AZZ} = J_{1,2} S_1 S_2 \quad (5)$$

where J is the exchange coupling tensor between the two spin populations (S) 1 and 2[229,230]. The line broadening caused by Heisenberg exchange coupling can be used to visually obtain structural information about the spin system. If a very strong exchange (short distance) is observed the spectrum broadens to the extent that the triple peak spectrum caused by the hyperfine splitting (described in Figure 12) is almost cancelled out, due to strong distortions of the central field peak.

2.2. 4-pulse double electron-electron resonance

Unlike continuous wave EPR (see section 2.1), where a static continuous wave radio frequency and a variable magnetic field are used to measure the spin label's microwave absorbance, pulsed EPR uses a static magnetic field in combination with timed microwave pulses on the nanosecond time scale. The technique introduced by Milov *et al.* in 1981[231,232] was modified in subsequent years to become the most widely used 4-pulse double electron-electron resonance (DEER) spectroscopy technique from the beginning of the 2000s. The technique is alternatively known as pulsed electron double resonance (PELDOR).

DEER is based on a Hahn echo[233] introduced by a simple microwave pulse sequence (Figure 17A) which allows to resolve dipolar-coupling between two paramagnetic centres which are in close proximity to each other; i.e. up to 10 nm for deuterated soluble proteins and up to 6 nm for membrane proteins[234,235]. By analysing the recorded DEER time trace (Figure 18A) distance distributions of paramagnetic centres can be resolved (Figure 18B-C).

The Hahn echo is generated by two successive pulses at a specific magnetic field (Figure 17A). In the corresponding rotary frame representation (Figure 17B), a $\pi/2$ or 90˚ pulse flips the bulk of the magnetisation into the x/y-plane while the external magnetic field (B_0) is aligned with the z-axis. The inter-pulse delay (τ) between the two pulses causes a dephasing of the individual spins. The subsequent π or 180˚ pulse flips the spins by 180˚, causing a rephasing of the individual spins, which then relaxes back into the original alignment with the external magnetic field (z-axis). The whole process of flipping, dephasing, flipping, rephasing and final relaxation creates a detectable spin echo[236].

For 4-pulse DEER the two pulse frequencies shown in Figure 17C are applied at given frequencies that depend on the paramagnetic probe used: observer – (ω_A) and pump (ω_B) frequency, respectively. Figure 17D shows a 1D-field sweep of a nitroxide spin probe at Q-band (~34 GHz, 50 K) and the corresponding pump- and observer frequencies chosen. The π-pulse at pump frequency is used to excite a sub-population of the spins in the sample and has to occur at a certain frequency offset (~ 80-100 Mhz, $\Delta\omega$, Figure 17C) to avoid overlap of the two excitation windows. In the event that the two separately excited spin populations are coupled, the pulse at ω_B will influence the local field of the observer spins, resulting in a change in spin phase during the Hahn echo sequence. Variation in the pump-pulse offset time (t, Figure 17C) allows the detection of the coupling caused by the two spin populations stimulated[231] (Figure 18).

Earlier DEER approaches relied on a three-pulse sequence which lacked the second refocusing π-pulse at ω_A and showed a pump-pulse (ω_B) application timed between the π/2- and π-pulse at ω_A. The short time interval between the π/2-pulse on ω_A and the π-pulse on ω_B caused a temporal overlap of the two, which distorted early time points in the DEER time trace and thus made the analysis of short distances – which are resolved at short t-times during DEER measurements – unreliable. The distortion is referred to as the 'dead-time' problem and was eventually overcome by the development of the 4-pulsed DEER technique described above[237].

In addition, in 4-pulsed DEER a second refocusing π-pulse is implemented after the first Hahn echo at ω_A (Figure 17C). The additional pulse allows the pump-pulse to be applied at t = 0 without overlap with other pulses of the ω_A pulse train. This improves measurements at short distances while slightly reducing signal-to-noise levels.

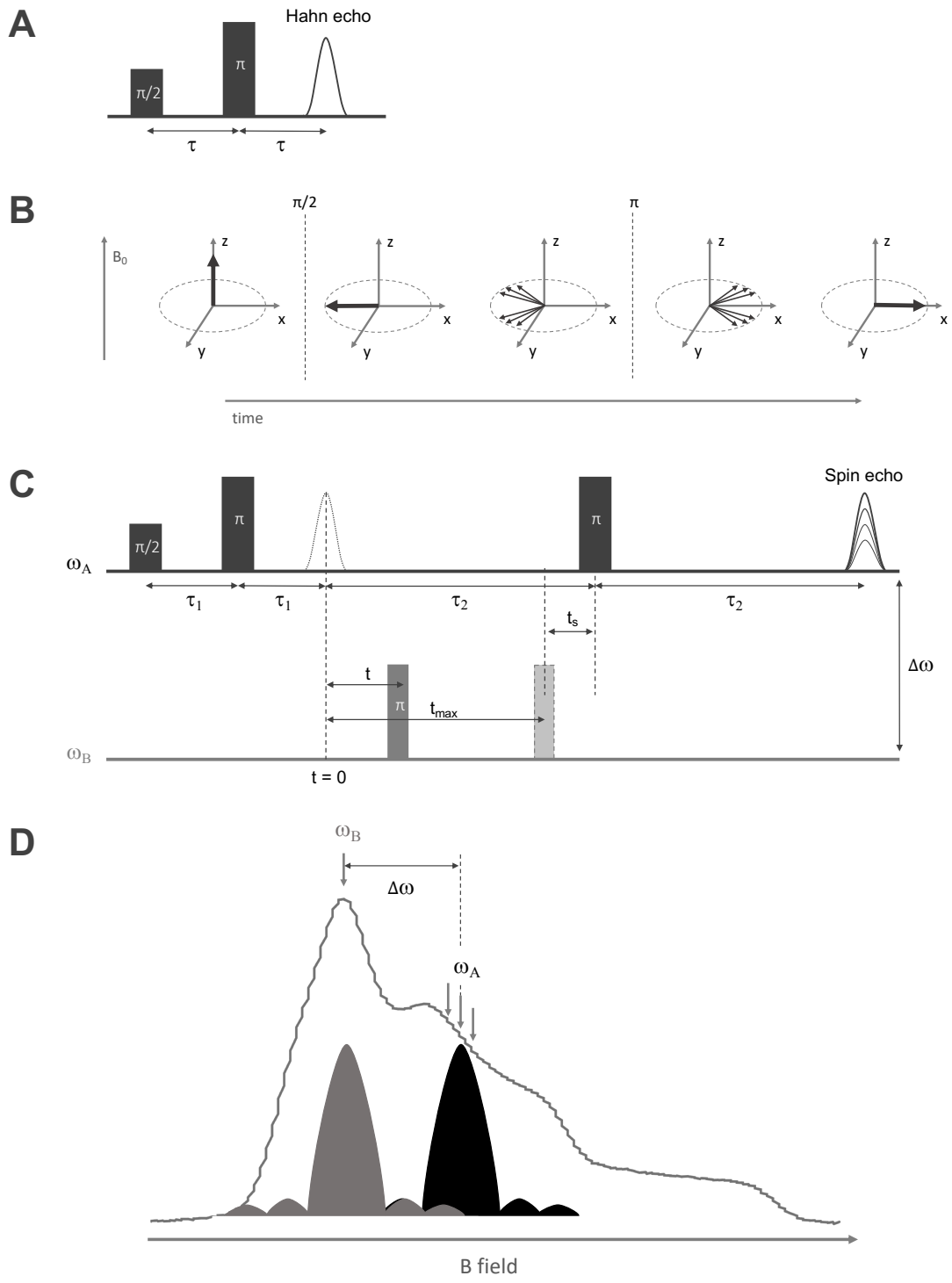


Figure 17: DEER methodology. (A) Illustration of a Hahn-echo pulse sequence. (B) Corresponding electron magnetisation in a rotary frame representation of the pulse sequence shown in A). (C) 4-pulse DEER pulse trains at observer (ω_A) and pump pulse frequencies (ω_B). (D) Nitroxide absorption spectrum at Q-band of MTSL-labelled protein at 30 K. Black and grey shaded areas indicate ω_A and ω_B pulse efficiencies, respectively. Small bumps indicate off target resonances.

Under the assumption that each A spin in a nanoobject is coupling with more than one B spin the recorded DEER time trace takes the form of:

$$D(t) = F(t)B(t) \tag{6}$$

whereby, D(t) is the dipolar evolution function and represents the second echo intensity as a function of time, B(t) the background contribution which consists of spin interactions between the coupled spin pair and spins in the bulk environment and F(t) describes the form factor product of all possible spin pair contributions[238]. F(t) is defined as:

$$F(t) = 1 - \lambda\left[1 - \int_0^1 \cos\left(\frac{C_i}{r_i^3}(1 - 3\,cos^2\,\theta_i)\,t\right) d\cos\theta\right] \tag{7}$$

with λ being the pump pulse efficiency, C_i the concentration of an i-spin population, r_i the radius between coupled i-spins and θ_i the corresponding angle between the spin-spin operator and the external magnetic field. The background function takes the form of:

$$B(t) = \exp\left(-C_B t^{\frac{D}{3}}\right) \tag{8}$$

C_B corresponds to the overall spin concentration and D to the background dimension in the sample of interest. For soluble nanoobjects following a homogenous distribution the theoretical background dimension is ~3 and for nanoobjects embedded in lipid membranes ~2, as the spins within the sample follow a less homogenous distribution[239]. Due to B(t) being highly depended on the excitation profile (pump

pulse efficiency) and EPR line shape (under experimental conditions) often the real background distribution has to be adjusted accordingly in the range of D = 2-3.

A general DEER evaluation process of a recorded dipolar evolution function, also referred to as DEER time trace, is shown in Figure 18. In order to isolate the form factor from the dipolar evolution function the background correlation is analysed (A) and subtracted from D(t). The resulting form factor (B) shows the oscillation frequency of the coupled spin-pair with the corresponding modulation depth (Δ). In the past, distance distributions from DEER data have been extracted using Fourier transformation and the Pake pattern, whereby the two 'spikes' of the spectrum represent the dipolar coupling distance between the two spins[240]. The method is still used in the software package DEERAnalysis to approximate distance distributions[241], but was generally replaced by a Tikhonov regularisation (TR) approach which delivers more accurate distances[241,242]. Within the software the Tikhonov approach is used to find the best parameter estimation for a given dataset with multiple solutions, while tolerating a small degree of bias.

Furthermore, a strong modulation depth Δ (see Figure 18B and equation 4) improves background substation. Δ increases with the number of coupled spins (n) within a given system and is highly dependent on the pump pulse efficiency (λ).

$$\Delta = 1 - (1 - \lambda)^{n-1} \tag{9}$$

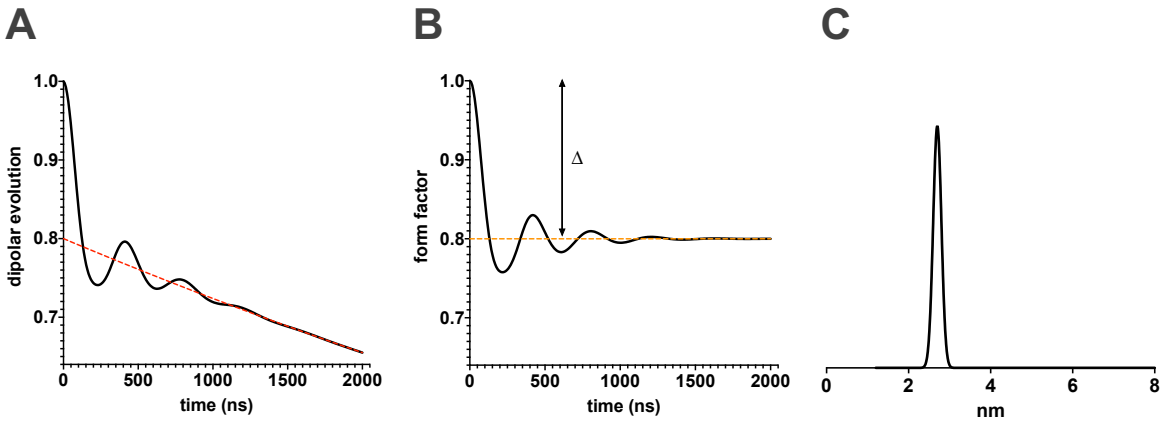


Figure 18.: Illustration of the DEER evaluation process. (A) Experimental dipolar evolution function (D(t)) with background correction (B(t)) shown in red. (B) Background corrected from factor (F(t)) of A). Orange line indicates the modulation depth (Δ) which is calculated by 1-Δ. (C) Distance distribution showing a single distance at 2.7 nm. The graph is generated by applying a Tikhonov regularisation[242] to the modulation frequencies shown in B). Graphs are simulated using the software tool SimPel (https://pypi.org/project/SimPel/)

2.3. Fluorescent energy resonance transfer

Time resolved – fluorescent resonance energy transfer (TR – FRET), a biophysical technique capable of measuring distances between fluorophores with overlapping absorption and emission spectra, was chosen for the development of a high-throughput ligand screen[243] for ER GluII (section 6.7). In FRET the energy transfer from a fluorophore A (donor) to a fluorophore B (acceptor) can occur without the need of direct contact between molecules, as long as the acceptor is within the Förster distance of the donor[244]. This makes FRET an ideal technique to study protein interactions and binding-induced structural changes. As a result of the donor emission being absorbed by and transferred to the acceptor, the measurable intensity and lifetime of the donor emission is decreasing with increasing FRET efficiency (Figure 19). The term "time resolved" stems from the use of FRET in modern devices. Here the detection of the acceptor emission is delayed by several hundred milliseconds after the excitation pulse for the donor, avoiding false-positive/negative interpretation of the measurements due to overlapping emission/absorption spectra of the fluorophores. Furthermore, the technique allows to minimize background noise by reducing all non-specific, short-lived fluorescent emission introduced by sample componence/dust introduced. The coexistence of long-lived fluorescent compounds and/or proteins is something that has to be considered during FRET evaluation as those can have a high impact on the achieved SNR[243]. Furthermore, fluorescent labels and here especially the use of fluorescent proteins, are highly influenced by environmental conditions. The influence of salt, pH, buffer system or the addition of other substances to the sample composition are recognised parameters affecting the label's fluorescent performance. In systems were fluorescent proteins are used as donors as well as acceptors, fluorophore sample

compositions can have different effects on the performance of the fluorophores chosen[245].

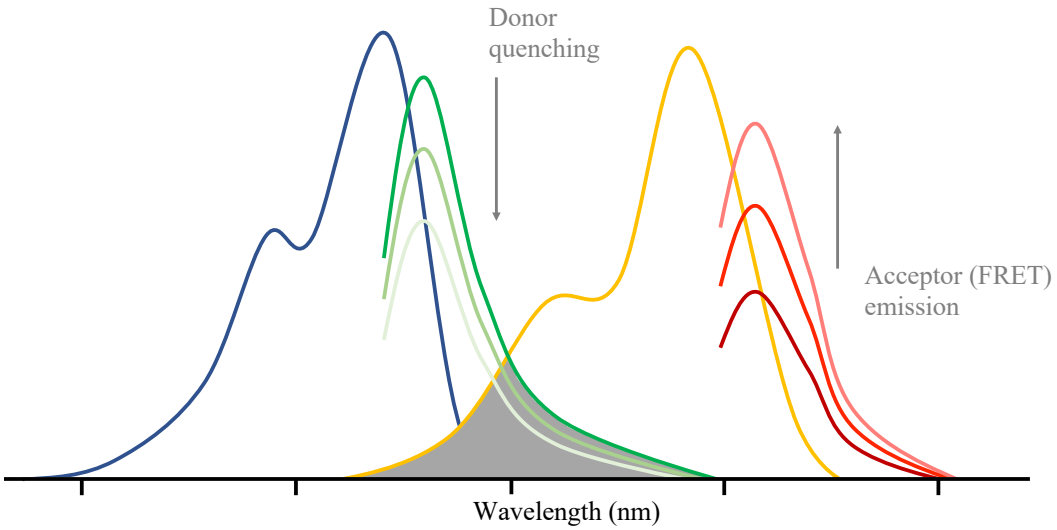


Figure 19: Fluorescent resonance energy transfer (FRET) phenomenon. The exitation of a donor fluorophore (blue spectrum) triggers fluorescence emission (green spectrum) which is, due to a spectral overlap (grey area), absorbed by the acceptor fluorophore (yellow spectrum). The exited accaptor then emits light in form of fluorescence (FRET).

3. Materials and Methods – HCV p7

3.1. HCV p7-cloning

3.1.1. p7-constructs

DNA sequences containing the required mutations (see Table 2) were ordered from GeneArt®, Thermo Fisher Scientific, as linear strings of double stranded DNA. All sequences were provided with restriction sites for HinDIII (N-terminus) and BamHI (C-terminus) for ligation purposes. At each side of the sequences nine extra pairs of bases were included to allow efficient digestion with the restriction enzymes.

Table 2 shows the modified single and double cysteine mutant p7 amino acid sequences prepared in this study. Within the table each mutation is allocated with a number. Henceforth the mutations are referred to as the numbers.

For ease of understanding five p7 cysteine mutants, which are extensively used in the cwEPR and DEER section of this thesis, are visualised in a 3D representation in Figure 20.

Table 2: Single and double cysteine p7 mutants analysed in this research, using the amino acid sequence from genotype 1b, isolate J4 as a template (first row, WT = wild type). Highlighted in yellow is the wild type cysteine (C27) mutated to a serine. Highlighted in green are the sites mutated to cysteine for attachment of the spin label. The linking glycines at the Nt and Ct ends are highlighted in grey.

No.	Position of Mutation	Amino Acid Sequence
	Genotype 1b (J4)	
WT		ALENLVVLNAASVAGAHGILSFLVFFCAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
0	C27S	ALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
1	C-1_G0_C27S	CGALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
2	L2C_C27S	ACENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
3a	S20C_C27S	ALENLVVLNAASVAGAHGICSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
3b	S21C_C27S	ALENLVVLNAASVAGAHGILCFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
3c	S22C_C27S	ALENLVVLNAASVAGAHGILSCLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
3d	S23C_C27S	ALENLVVLNAASVAGAHGILSFCVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
4	C27S_A28C	ALENLVVLNAASVAGAHGILSFLVFFSCAWYIKGRLAPGAAYAFYGVWPLLLLLLALPPRAYA
5	C27S_L36C	ALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRCAPGAAYAFYGVWPLLLLLLALPPRAYA
6	C27S_L51C	ALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPCLLLLLALPPRAYA
7	C27S_P59C	ALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLLALCPRAYA
8	C27S_G64_C65	ALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLALPPRAYAGC
9	C-1_G0_C27S_G64_C65	CGALENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLALPPRAYAGC
10	L2C_C27S_P59C	ACENLVVLNAASVAGAHGILSFLVFFSAAWYIKGRLAPGAAYAFYGVWPLLLLLALCPRAYA
11	S21C_C27S_L51C	ALENLVVLNAASVAGAHGILCFLVFFSAAWYIKGRLAPGAAYAFYGVWPCLLLLLALPPRAYA
	Genotype 5a	
0a		GAKNVIVLNAASAAGNHGFFWGLLVVTLAWHVKGRLVPGATYLSLGVWPLLLVRLLRPHRALA
3b	W21C	GAKNVIVLNAASAAGNHGFFCGLLVVTLAWHVKGRLVPGATYLSLGVWPLLLVRLLRPHRALA
6	L51C	GAKNVIVLNAASAAGNHGFFWGLLVVTLAWHVKGRLVPGATYLSLGVWPLCLVRLLRPHRALA
11	W21C_L51C	GAKNVIVLNAASAAGNHGFFCGLLVVTLAWHVKGRLVPGATYLSLGVWPLCLVRLLRPHRALA
13	A2C_L51C	GCKNVIVLNAASAAGNHGFFWGLLVVTLAWHVKGRLVPGATYLSLGVWPLCLVRLLRPHRALA
	Genotype 1a	
H77	M47V	ALENLVILNAASLAGTHGLVSFLVFFCFAWYLKGRWVPGAVYAFYGVWPLLLLLALPQRAYA

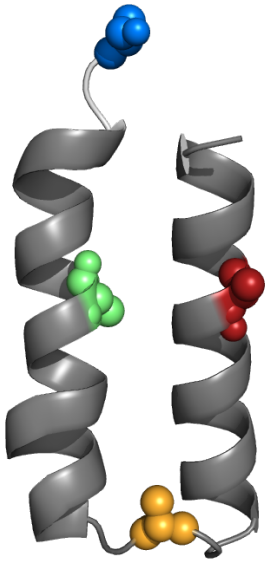


Figure 20: 3D representation of the putative p7 monomer with highlighted cysteine mutation sites used for cwEPR/DEER. The p7 mutants 1 (blue), 3b (green), 5 (orange) and 6 (red) are shown in spheres. The double cysteine mutant 11 is represented by the combination of mutant 3b and 6. *Dark grey* – cartoon representation of the hairpin-like conformation of a p7 monomer. *Light grey* – N-terminal extension, see Table 2 mutant 1.

3.1.2. Ligation of p7-constructs into vector pMM-LR6

The BamHI- and HinDIII-linearised vector pMM-LR6 and p7-constructs released by the same restriction enzymes from section 3.1 were used for ligation (for reaction mix see Table 3).

The pMM-LR6 vector contains the TrpΔLE sequence of the *E. coli* tryptophan operon and an His^{9}- fusion protein N-terminal of the open reading frame[246-248]. A C-terminal methionine added to the C-terminus of the fusion protein serves as cleavage site for the release of the target protein. For more information about cyanogen bromide - cleavage see section 3.2.7.

Table 3: Reaction mix used for vector restriction. The cleavage took place at 37°C, overnight and was stopped by transferring the sample to -20°C

Component [-]	Volume [µl]
DNA (1 µg)	X
Cut-Smart buffer	5
Enzyme BamHI (20k Units)	3
Enzyme HinDIII (20k Units)	3
dH_2O	top up to 50

The ligation was performed as described in the New England BioLabs™ "Ligation Protocol with T4 DNA Ligase (M0202)". Previous experiments had shown that a 1:2 (vector:insert) molar ratio results in a good ligation efficiency, thus the protocol was slightly modified as follows:

The ligation reaction outlined in Table 4 was set up in a microcentrifuge tube on ice. The sample was mixed by pipetting up and down gently and incubated at 16°C overnight. The reaction was stopped by incubating at 65°C for 10 min.

Table 4: Reaction mix used for the ligation of digested p7 sequences into the open reading frame of the vector pMM-LR6.

Component	Amount
10x T4 DNA ligase buffer	2 µl
Vector DNA	0.02 pmol
Insert DNA	0.04 pmol
T4 DNA ligase	1 µl
Nuclease free water	to 20 µl

3.1.3. Agarose gels

Agarose gels for the quantitative analysis of DNA constructs were made of 0.8-1.2% w/v agarose in 1x TAE depending on DNA-construct size to be separated. After dissolving the agarose using heat and the addition of 1:10000 GelRed nucleic acid gel stain (Biotium) gels were poured into moulds. Once cured, samples mixed with 6x Purple loading dye (NEB) were loaded and the gels run at 150 V for 50-120 min in 1x TAE buffer. After samples were run, DNA bands were visualised by UV excitation.

3.2. Protein expression and purification

3.2.1. Transformation of pMM-LR6/p7 in *E. coli*

The following sections 3.2.1-3.2.8 are based on the p7 expression- and purification protocol published by Opella and co-workers[249] including process-related modifications.

After the ligation took place the modified vectors were transformed in *E. coli* cells, using a standard heat shock transformation.

50 µl of a competent *E. coli* BL21 (DE3) pLysS glycerol stock were thawed on ice. 3 µl of the ligation mix (section 3.1.2) were added and the reaction solution was mixed gently by sucking it up and down with a pipette. Before the heat shock took place the reaction mix was stored on ice for an additional 60 min. Afterwards the cells were heated up for 45 sec at 42°C. This was followed by incubation for 2 min on ice, before 250 µl of SOC-medium were added and the cells incubated for 30 min at 37°C shaking. 50 µl of each culture were plated on LB-agar plates (10 g/l trypton, 51 g/l yeast extract, 10 g/l NaCl, 20 g/l glucose and 15 g/l agar) including 1:1000 kanamycin (25 mg/ml stock) and chloramphenicol (50 mg/ml stock) and stored at 37°C upside down until cell growth was observed.

3.2.2. Expression screening in bacterial cells

Single colonies were picked (see section 3.2.1) and transferred into 50 ml tubes containing 7 ml LB-medium, including 1:1000 kanamycin (25 mg/ml stock) and chloramphenicol (50 mg/ml stock). The cultures were incubated for 16 h at 37°C and 250 rpm. Afterwards 70 µl of the LB-culture were used to inoculate 7 ml of TB-medium (13.33 g/l tryptone, 26.67 g/l yeast extract, 0.8% glycerol (v/v), 23.1 g/l

KH_2PO_4, 125.4 g/l K_2HPO_4) including 1:1000 kanamycin and chloramphenicol. In addition, cryostocks of the single cell cultures were produced by mixing 500 µl LB-culture with 500 µl glycerol and stored at -80°C. The left-over culture was pelleted at 4000 rpm for 30 min. The supernatant was discarded and the pellets were stored at -20°C for further analysis. The TB-cultures were incubated as described before. As soon as their OD_{600} reached 1.2 the expression of the target protein was induced by the addition of 3.5 µl isopropyl-β-D-thiogalactopyranosid (IPTG, 1M stock).

For SDS-PAGE 20 µl of cell culture were resuspended in gel loading buffer at a ratio of 1:3. For more information, about the gel loading buffer and the general protocol of SDS-PAGE see Section 3.2.3.

For DNA sequencing, the pellet of the 7 ml LB-cultures (see above) was isolated using the "QIAprep® Spin Miniprep Kit (250)" from QUIAGEN™.

3.2.3. Sodium dodecyl sulphate – polyacrylamide gel electrophoresis

For the discontinuous sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE) samples were mixed with minimal 0.25-fold (v/v) gel loading buffer (4.92 M urea, 230 µl 4x LDS loading buffer, 26.72 mM SDS, 77 mM DTT). The reaction mix was heated up for 5 min at 95°C and a volume corresponding to 2.5 ng pure p7 protein transferred into the pocket of a precast "4-12% NuPAGE Bis-Tris gel" from Life-Technologies™. "Sharp Pre Stained Protein Standard" - Novex™ was used as a protein marker. Electrophoresis was performed at 200 volts and 120 mA for 30 min in MES buffer (50 mM MES, 50 mM TRIS base, 0.1% SDS, 1 mM EDTA, pH 7.3). Afterwards the gels were washed with Milli-Q water and stained with "InstantBlue" from expdeon™.

3.2.4. Large-scale expression in *E. coli* BL21

10µl of the *E. coli*-p7 cryostocks prepared in section 3.2.2 were used to inoculate 100 ml LB-medium containing the antibiotics chloramphenicol and kanamycin (concentrations same as above). Cell cultures were incubated overnight at 37°C and 180 rpm.

5 ml of the overnight culture were then used to inoculate 500 ml TB-medium containing the antibiotics at the concentrations above. Cells were grown at 37°C and 150 rpm until their OD_{600} reached 1.2. At this cell concentration the target protein expression was induced by the addition of 250 µl 1M IPTG. Subsequently, the cells were incubated at 25°C, 150 rpm for 18 h.

For cell harvest the culture was transferred into 1 l centrifuge buckets and spun at 7500 rpm for 25 min. The pellets were collected in 20 g portions and stored at -20 °C until cell disruption.

3.2.5. Large-scale expression of ^{15}N-labelled p7 in *E. coli* BL21

For the expression of ^{15}N-labelled p7 LB-cultures were prepared as described in section 3.2.4. For a uniform ^{15}N-labelling of p7 5 ml of the LB-culture were used to inoculate 500 ml isotope-enriched M9 minimal medium including 1g/L ^{15}N-H_4Cl (ammonium chloride, Sigma-Aldrich) and 1:1000 kanamycin (25 mg/ml stock).

Expression of the target protein as well as cell harvesting were conducted as described above.

3.2.6. Bacterial cell/inclusion body disruption

As previously described in section 3.1.2 p7 was expressed as a TrpΔLE-/His^9-tag fusion protein. The TrpΔLE-tag forces *E. coli* to collect/store the expressed p7 proteins in inclusion bodies, which makes cell disruption and isolation of p7 much more efficient. Furthermore, it protects the cells from a potential toxic effect of the target protein.

For cell disruption approximately 80g cell pellet (see Section 3.2.4) were resuspended in 240 ml resuspension buffer I (1 mM EDTA, 50 mM Tris pH 8, 1 mM NaN_3, 15% glycerol) using a 140 ml homogenizer. Afterwards the suspension was sonicated on ice twice for 5 min using pulses of 0.7s on and 0.3s off (total lapsed time 7 min) with an applied 65% intensity, in a Status US 70 sonicator (Philip Harris Scientific). To ensure optimal cell disruption around 70 ml of sample were sonicated at a time using a titanium probe with 1.3 cm diameter. To collect undissolved, heavy cell compounds from the crude extract which included inclusion bodies, the solution was centrifuged for approximately 25 min at 17,000 rpm and 4°C. The supernatant was discarded and the pellet resuspended in 240 ml resuspension buffer II (1 mM EDTA, 50 mM Tris pH 8, 1 mM NaN_3, 1% (w/v) deoxycholic acid, 1% IGEPAL Ca-630) using a homogenizer. The suspension was sonicated again as described before and spun for 25 min and 17,000 rpm at 4°C. Afterwards the supernatant was discarded and the total pellet resuspended in 60 ml resuspension buffer III (6 M guanidine HCl, 0.5 M NaCl, 20 mM Tris pH 8, 5 mM imidazole), homogenized, incubated and stirred at 4°C, overnight. For the full disruption of the inclusion bodies the solution was sonicated as described before, and spun at 12,000 rpm, approximately 25 min, at 4°C. The disrupted inclusion bodies in the supernatant were checked for purity by SDS-PAGE (methanol-chloroform precipitation of a 100 µl aliquot from the supernatant was taken

to perform the SDS-PAGE). The disrupted inclusion bodies in the supernatant were transferred into dialysing tubes with a MWCO of 3.5 kDa. Approximately 100 ml of the sample were dialysed against 5 l of Milli-Q water and the water was exchanged twenty times, over 2 or more days. Afterwards the sample was spun down to collect the precipitant and the supernatant was discarded. The whole sample was frozen using liquid nitrogen and subsequently lyophilised. To check purity of the sample SDS-PAGE was performed (see section 3.2.3). The lyophilised p7-fusion protein was sealed under argon and stored at -20°C until further purification.

3.2.7. CNBr-cleavage of the p7-fusion protein

200 mg of lyophilised fusion protein (see section 3.2.6) were solubilised in 5 ml 90% formic acid and 2 g of CNBr were added before vortexing for 2 min. The reaction mixture was put under nitrogen flow, protected from light and stirred for 2 h. The cleavage was stopped by adding 20 ml 1 M NaOH and an additional 15 ml of Milli-Q water. Immediately the sample was transferred into dialysing tubes with a MWCO of 1 kDa and dialysed against 5 l of Milli-Q water. The water was changed after 30 min for the first time and subsequently three times every hour on that day. The dialysis water was then changed up to 15 times over one or more days. After the dialysis the whole sample was lyophilised.

3.2.8. Size exclusion chromatography

After cleavage of the fusion protein TrpΔLE/His^9-p7 the full length p7 was purified using size exclusion chromatography (SEC) FPLC.

For sample preparation approximately 200 mg of lyophilised cleavage product (see section 3.2.7) were resuspended in 16 ml 10% SDS and treated 15 min in a sonication bath until the sample was clear. The sample was diluted with 16 ml of FPLC running buffer (20 mM sodium phosphate pH 8.2, 4 mM SDS, 1 mM EDTA, 1 mM NaN_3, sterile filtered) to give an absorbance at 280 nm of ≈ 4-6. Depending on the p7 mutation 1-50 mM reducing agent was added and the sample mixed gently. The pH was adjusted to 8.2 when necessary. A "HiPrep 26/60 Sephacryl S-200 HR" - column from GE Healthcare™ was equilibrated with 1.5 column volume (CV) running buffer at a flow rate of 2.5 ml/min. 5 ml sample were applied to the column at a time and the FPLC was run for 1.3 CV at a flow rate of 0.75 ml/min. Peaks at 280 nm were analysed using SDS-PAGE by mixing 18.65 µl of each fraction with 6.25 µl of gel loading buffer (without reducing agent) following the protocol in section 3.2.3. After SEC the fractions containing pure p7 were pooled and 100 ml aliquots were dialysed against 5 l Milli-Q water using 1 kDa MWCO dialysing tubes. The water was changed every three h up to 20 times over several days. Precipitant was collected by centrifugation at 4000 rpm for 20 min and the pellets lyophilised.

3.3. EPR sample preparation

3.3.1. MTSL-labelling of p7-cysteine mutants

For EPR/DEER p7 single and double cysteine mutants were labelled using the spin label MTSL.

The required amount of lyophilised p7 powder (section 3.2.8) was solubilised in 10 mM sodium phosphate pH 8.0, 11 mM SDS, 1.5 M urea and 15-fold excess of MTSL compared to protein. The sample was sealed under argon and incubated rotating, 3 h at room temperature (RT). Subsequently all samples were dialysed in 10 kDa MWCO dialysing tubes against 5 l of 10 mM sodium phosphate pH 7.2, 11 mM SDS for 9 h, dark at 20˚C and twice against 10 mM sodium phosphate pH 7.2 and 4 mM SDS, 9 h each. Sample was collected, spun and the supernatant stored at RT until further processing.

3.3.2. p7 reconstitution into lipid bilayers

Spin labelled p7 (see section 3.3.1) was mixed with solubilised lipid for reconstitution into bilayers. Lipids from chloroform stocks were mixed at molar ratios intended and dried in glass vials as a thin film under a stream of nitrogen and stored under vacuum for 15 h to fully remove remaining solvent. The thin film of lipids was resuspended in 27 mM CF-6, 10 mM sodium phosphate pH 7.2 and incubated for 30 min, shaking at 37˚C. The solubilised lipids were split into 1.5 ml aliquots and sonicated in a bath sonicator for up to 9 min at 28-35˚C until the solution became clear. Before mixing the solubilised lipids with the protein sample, the latter was spiked to a final concentration of 27 mM CF-6. For reconstitution steps all components were pre-heated to 37°C. Then the required amount of solubilised lipid and p7 protein in SDS/CF-6 were mixed and incubated for 30 min at 37°C to achieve the desired

lipid:protein molar ratios (LPR). The ratio is based on a hypothetical oligomeric assembly of the protomers and thus corresponds to 1/6 of the actual p7 concentration (monomer) within a given sample. Reconstitution was achieved by removing the detergent using dialysis. For this, the samples were transferred into 3.5 kDa MWCO snake skin dialysing tubes and dialysed against 5 l of 10 mM sodium phosphate pH 7.2 (36 h, dark, RT) replacing the buffer after approximately 16 h and 25 h.

Reconstituted proteoliposomes with an LPR < 500:1 were collected by ultracentrifugation (SW41Ti rotor; 36000 rpm, 2 h, 37°C), the pellet resuspended in 1.2 ml 10 mM sodium phosphate pH 7.2 containing 20% (v/v) ethylene glycol and 70% deuterium oxide (EPR-buffer) and incubated for 10 min at 37°C. The proteoliposomes were pelleted again in a bench top centrifuge at 20,000 xg and resuspended in EPR-buffer (including incubation at 37°C) three more times. A minimum volume of 10 µl for DEER and precisely 10 µl for a cwEPR-based spin labelling efficiency test, was loaded into Q-Band suprasil EPR quartz tubes (1.6x1.1x100 mm, round bottom, Goss) and spun a few sec at 1000 rpm to collect the sample at the bottom. The samples were kept at RT until measurement.

For proteoliposomes with an LPR > 500:1 the dialysed sample (see first paragraph of this section) was transferred into a 10 kDa MWCO spin concentrator and reduced to a final volume of 100 µl. The buffer was exchanged by adding 4x the sample volume of EPR-buffer and the whole sample concentrated to a final volume of 25-50 µl. EPR quartz capillaries were loaded with sample as described above.

3.3.3. Circular dichroism

Circular dichroism spectra were recorded in a JASCO J-815 CD-spectropolarimeter. The device was fitted with a computer-controlled Peltier temperature control unit, allowing consistent measurements at 20°C.

The spectra were recorded in a range of 260 down to 185 nm wavelengths with a 1 s data integration time, a 20 nm/min scanning speed and a 0.5 nm data pitch. In case of a low protein concentration the spectrum was scanned up to five times to reduce the signal-to-noise ratio. Samples with a protein concentration of approximately 15-500 µM monomeric p7 were measured. Therefore, different quartz cuvettes with an inner diameter of 0.1, 1.0 or 10 mm had to be used. JASCO Spectra Manager Version 2.0 software was used and the results plotted in mdeg of ellipticity [Θ] against the wavelength in nm. For sample preparation see section 3.3.2.

3.3.4. Dynamic light scattering

To be able to determine the size of proteoliposomes prepared in section 3.3.2 dynamic light scattering (DLS) was carried out.

A Viscotek Model 802 DLS equipped with a 660nm laser and an Avalanche Photo Diode (APD) was used to measure scattered light at 90° at a rate of one photon per µsec. For the measurement a minimum of 20 µl sample were transferred into a micro quartz cuvette and the cuvette introduced into the 20˚C pre-heated sample holder. After the sample temperature had stabilised ten 10 sec measurements were taken while applying the automated data filter implemented in the device's software. In cases where the amount of light scattering exceeded the detector limit serial dilutions of the sample were prepared using corresponding buffer.

3.3.5. Fluorescent resonance energy transfer

For the analysis of the formation of putative p7 ion channels in a lipid environment Förster or fluorescent resonance energy transfer (FRET) in combination with site directed mutagenesis of the target protein was conducted in collaboration with Dr Marc-Philip Pfeil (former member of Prof Anthony Watts' group, Biochemistry Department, University of Oxford, UK). Ligation, expression and purification of the target protein specifically modified for FRET analysis were carried out in the Zitzmann laboratory.

Lyophilised p7 mutant C27S_K33R was solubilised in 2.35 M urea, 11 mM SDS, 20 mM sodium phosphate pH 8.0, aiming for a final protein concentration of 800 µM. 5-10-fold excess of the corresponding dye (sulfo-Cyamine3 NHS ester (Cy3, donor) or sulfo-Cyamine5 NHS ester (Cy5, acceptor) solubilised in water) were added and the sample incubated for 3h at RT. Subsequently the samples were dialysed as described in section 3.2.8 and the precipitant lyophilised.

Labelling efficiency was determined by solubilising a small flake of dye-labelled p7 powder in 2,2,2-trifluoroethanol (TFE). A standard curve of unreacted dye was prepared in TFE and the correction factor (CF) evaluated for both dyes at A280 to A_{dye} max. The labelling efficiency (LE) was calculated according to the following equation:

$$LE = \frac{\varepsilon_{dye}}{\varepsilon_{protein}} = \frac{A_{dye}\ x\ \varepsilon_{dye}}{A_{protein}\ x\ \varepsilon_{protein} - CF\ x\ \varepsilon_{dye}} \qquad (10)$$

For FRET sample preparation dye-labelled p7 powder was taken up in 27 mM CF-6 in water, at 570 µM for Cy5-labelled and 350 µM Cy3-labelled p7, respectively.

200 pmol dye-labelled p7 reconstituted into lipid bilayers at varying LPRs with a final volume of 1 ml were used for FRET experiments. Sample reconstitution was performed as described in section 3.3.2. For each LPR nine different samples with varying donor concentrations (Cy3-labelled p7) ranging from 20 to 180 pmols were prepared as stated in Table 5. In addition, lipid-only, donor-only and acceptor-only samples were prepared for FRET corrections.

Table 5: FRET donor- and acceptor- molar ratios prepared for the detection of p7 oligomeric arrangements within one LPR. Each lane corresponds to one sample of the LPR-series which were needed for FRET evaluation

Donor [pmol]	Acceptor [pmol]
180	20
160	40
140	60
120	80
100	100
80	120
60	140
40	160
20	180
-	200
200	-
-	-

3.3.6. Proton flux assay

Proteoliposomes were prepared as described in section 3.3.2. After the 1st dialysis step the dialysis tubes, including the sample, were transferred into 5 l of internal liposome buffer (IL-buffer; 60 mM KH_2PO_4, 60 mM K_2HPO_4, 150 mM NaCl and 20 mM KCl, pH 7.0) and dialysed for additional 18 h at RT. The buffer was exchanged three times over the duration of the dialysis.

The samples were harvested and extruded 21 times using a glass syringe extruder setup equipped with two 0.2 µm polycarbonate membranes separated by a support membrane. In case of a high proteoliposome density an additional extrusion step with a 0.4 µm polycarbonate membrane was performed before the final extrusion at 0.2 µm. The external IL-buffer was exchanged by running the samples through a 5 ml HiTrap buffer exchange column (GE Healthcare) equilibrated with external liposome buffer (EL-buffer; 190 mM Na_2SO_4, 0.1 mM KH_2PO_4 and 0.1 mM K_2HPO_4, pH 7.0). Liposome containing fractions were collected and diluted to the required lipid concentrations using EL-buffer. Final sample volumes ranged from 1.5 to 3 ml, containing 0.27 – 6.1 mg/ml lipids. Until measurement samples were stored at 4˚C.

The pH change over time was monitored by a pH microelectrode (Mettler Toledo, InLab) while rapidly mixing the sample using a microstir bar. A change of the electrochemical potential across the liposome membrane was achieved by decreasing the pH of the EL-buffer from pH 7.0 to 5.5-6.0, adding 1 M HCl to the sample. When the pH had stabilised (after about 3 min) a final concentration of 25 nM of the potassium ionophore valinomycin was added and the pH recorded for additional 3 min. The experiment was terminated by addition of 0.5 µM (final concentration) of the ionophore carbonyl cyanide m-chlorophenylhydrazone (CCCP) and the pH recording was stopped after a total of 9 min.

3.3.7. Electron paramagnetic resonance

Continuous wave electron paramagnetic resonance (cw-EPR) measurements were conducted on a Bruker "EMX" spectrometer using X-band (~9.3 GHz). Instrument settings are summarised in Table 6. A nitrogen-cryostat in combination with a variable temperature control unit (Oxford Instruments) was connected to the resonator for temperature control.

For labelling efficiency measurements the sample volume was reduced to 10 µl (see sections 3.3.1 and 3.3.2). Detergent- as well as lipid-based samples were stored in the dark at RT, until the EPR measurement were taken.

Table 6: EPR settings for the BRUKER EMX using the software XENON™.

1D - Field Sweep		
Magnetic field		
Centre field	3295.95	[G]
Sweep width	150	[G]
Sweep time	45-180	[s]
Sample G-factor	2.000000	[-]
Signal channel		
Receiver gain	45-60	[dB]
Modulation amplitude	0.5-1	[G]
Number of scans	3-10	[-]
Offset	0	[%]
Microwave		
Attenuation	13	[dB]
Power	10	[mW]
Digital filter		
Mode	auto	
Number of points	10-20	
Options		
Magnetic field		
Number of points	1500	
Field settings	wait LED off	
Setting delay	0.0	[s]
Sweep direction	up	
Field flyback	on	
Signal channel		
Modulation frequency	100.00	[kHz]
Modulation phase	0.0	
Conv. Time	30.00	[ms]
Time constant	2.56	[ms]
Dual detection	off	
pts / mod. Amp.	10-20	

3.3.8. Double electron-electron resonance

All DEER measurements were performed at 50K on a Bruker ELEXYS E580 spectrometer. The device was equipped with a SuperQ-FT bridge (Bruker) and a 2 mm split-ring resonator (EN-5107D2, Bruker). Pulses were generated at Q-band (~33.4 GHz) using an arbitrary waveform generator (AWG). The standard four-pulsed DEER (see Figure 21) observer sequence (see equation 1) and pump sequence (see equation 2) were used.

$$\pi/2 \overset{\tau 1}{\rightarrow} \pi \overset{\tau 1}{\rightarrow} echo \overset{\tau 2}{\rightarrow} \pi \overset{\tau 2}{\rightarrow} ref.echo \tag{11}$$

$$echo \overset{t}{\rightarrow} \pi \tag{12}$$

To obtain the absorption spectrum of the nitroxide spin label a Hahn echo sequence (see equation 13) was used to record the echo intensity as a function of the magnetic field.

$$\pi/2 \overset{\tau}{\rightarrow} \pi \overset{\tau}{\rightarrow} echo \tag{13}$$

Longitudinal relaxation (T_2) was determined using a two pulsed echo experiment with an incremented inter-pulsed delay (τ) (see equation below).

$$\pi/2 \overset{t_0+\tau}{\longrightarrow} \pi \overset{t_0+\tau}{\longrightarrow} echo \tag{14}$$

An inversion recovery experiment was used to determine transversal relaxation (T_1) under similar settings found for T_2. To specify the recycle delay for the DEER experiment T_1 was set to 1.3times T_1. The pump pulse was always defined as 12 ns,

whereas the observer pulse varied in length depending on the sample analysed (typically values were between 12-18 ns).

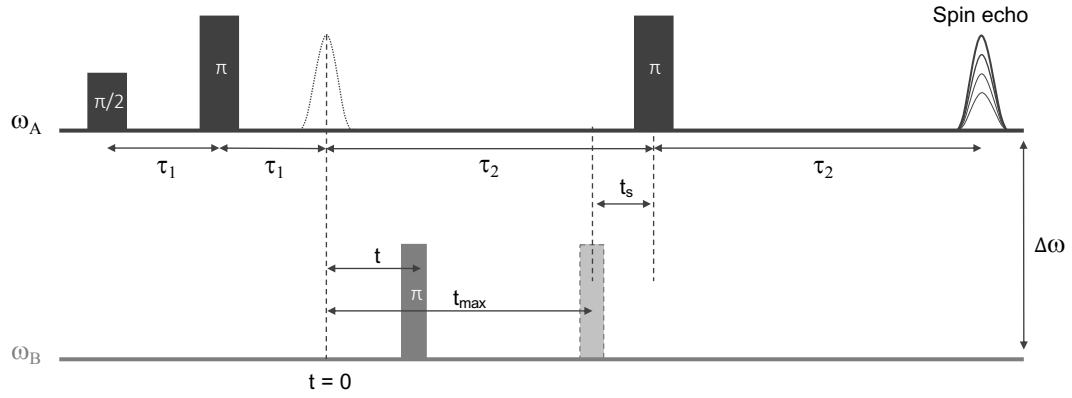


Figure 21: 4-pulse DEER pulse trains at observer (ω_A) and pump pulse frequencies (ω_B).

3.4. Sample preparation of HCV p7 in detergent

For size exclusion – small angle laser light scattering (SEC-MALLS) and NMR spectroscopy p7 was required to be solubilised in n-dodecylphosphocholine (DPC, Avanti Polar Lipids Inc.). Sample preparation was based on protocols published in OuYang *et al.* 2013[168] and Gen *et al.* 2015[250] including process-related modifications.

3.4.1. Nuclear magnetic resonance spectroscopy

For NMR spectroscopy ^{15}N-labelled p7 5a mutation 0a (see section 3.1.1, Table 2 and section 3.2.5) was solubilised using 200 mM DPC, 6M guanidine HCl in water, aiming for a protein concentration of 400 µM. Samples were incubated for 30 min at RT and transferred into 2 kDa MWCO dialysis cups with a maximum sample volume of 130 µl, each. The samples were dialysed against 2 l of 25 mM MES pH 6.5 for 6 h with a total of three buffer changes. Samples were transferred into new 12 kDa MWCO dialysis tubes (Pur-A-Lyzer, SIGMA-ALDRICH™) and dialysed against 20 ml of 50 mM DPC, 25 mM MES pH 6.5 for 24 h. Samples were harvested, protein concentration determined and the sample diluted to the final concentrations stated in Table 7.

Table 7: Final sample composition used for NMR spectroscopy.

Component	**Amount**
^{15}N-labelled p7	5 / 200 µM
DPC	50 / 200 mM
MES pH 6.5	25 mM
DSS	1 mM
D_2O	5% (v/v)
Final volume	550 µl

^{1}H-^{15}N-TROSY NMR measurements on ^{15}N-labelled p7 in detergent were carried out by Prof Jason Schnell (Biochemistry Department – University of Oxford, UK) on a 750 MHz spectrometer equipped with a CryoProbe and a Bruker - Avance III HD console.

3.4.2. Size exclusion – small angle laser light scattering

Required amounts of lyophilised p7 5a mutation 0a (see section 3.1.1, Table 2 and section 3.2.5) were solubilised in 200 mM DPC, 6 M guanidine HCl in water or 20 mM sodium phosphate, 80 mM SDS, 1.25 M urea and incubated 1 h, shaking at RT. Samples were spun, the supernatant collected, and the absorbance measured at 280 nm. Subsequently up to two dialysis steps were carried out following the protocol in section 3.4.1 using the dialysis buffer stated in Table 8.

Table 8: Dialysis buffer used for p7 sample preparation in detergent. Samples were analysed using SEC-MALLS.

1st dialysis step	**2nd dialysis step**
25 mM MES pH 6.5	50 mM DPC, 25 mM MES pH 6.5
	3 mM DPC, 25 mM MES pH 6.5
80 mM SDS, 20 mM Na Phos. pH 7.2	-
10 mM SDS, 20 mM Na Phos. pH 7.2	-

SEC-MALLS was carried out on a “Dawn HELEOS-II” (Wyatt Technologies) 8-angle light scattering detector and “Optilab T-rEX” (Wyatt Technologies) refractive index detector linked to a Shimadzu HPLC system comprising an LC-20AD pump and a SPD20A UV/Vis detector. Prior to sample injection the two “Superdex™-200 5/150 GL”-columns (GE Healthcare) connected in series were equilibrated with the corresponding buffer (see Table 9).

Table 9: Buffer systems used for SEC-MALLS analysis.

Buffer composition
50 mM DPC, 25 mM MES pH 6.5
3 mM DPC, 25 mM MES pH 6.5
80 mM SDS, 20 mM sodium phosphate pH 7.2
10 mM SDS, 20 mM sodium phosphate pH 7.2

Molar masses of the samples measured were analysed using the protein conjugate analysis tool implemented in ASTRA v.6.1.1.17 (Wyatt). For detergent-only samples the standard analysis tool was applied, based on reflective index and light scattering. Analysis was carried out by using the following d*n*/d*c* values; DPC 0.1398 ml g-1, SDS 0.1100 ml g-1 and protein 0.185 ml g-1.

4. Results and Discussion – HCV p7

4.1. p7 protein mutagenesis

HCV presents very high mutation rates and great genetic variability[114]. At present, HCV is classified into 8 genotypes (GT1-8), with GT8 having been only recently identified[110], and with over a 100 subtypes described (e.g. a, b, etc)[111-113]. This variability is reflected in the p7 protein variants encoded by different HCV genotypes and makes it a complicated task to decide which one(s) to use for structural studies. Sequence alignments of p7 from the different genotypes reveal very few strictly conserved residues across the whole amino acid sequence. Just nine out of 63 amino acid side chains are conserved with another 12 being highly similar but not found across all genotypes aligned. Physicochemical properties of the amino acid sequences reveal a more conserved pattern with similar hydrophobic and hydrophilic patches along the sequences[77,149]. In an extensive study Scull *et al.*[251] introduced point mutations in the p7 sequence to analyse their effect on virus production in cell culture using the HCV chimeric J6/JFH-1 (GT2a) system. First, an alanine residue was

inserted at every third position across the amino acid sequence of p7. The authors found that, out of the 19 insertions, none altered RNA replication of the virus, but two (6A, 60A) inhibited production of infectious virus particles below detectable levels and three (27A, 39A, 51A) significantly reduced production of infectious virus particles. The authors focused their attention next on the second half of the N-terminal TMH as well as sections of the loop region where they substituted residues 1 - 43 with a bulky tryptophan one at a time. They showed no HCV RNA copy reduction for any of the positions tested but described a significant drop in infectious virus particle release for mutants A1W, L2W, E3W, K4W, V6W, H9W, A10W, A11W, S12W, A13W F26W, A28W and V36W. The extensive study performed by Scull *et al*. highlights the tolerability of p7 function to substitutions with a bulky tryptophan residue, and helps identify many sites that could well accommodate a spin label probe like MTSL, similar in size to tryptophan. The authors concluded that the N-terminal region of p7 is more sensitive to substitutions by bulky residues such as tryptophan compared to alanine. Several other studies have analysed the effects of point mutations on p7 in different ways[174,252,175,253,165].

The sequence analysis by Montserret *et al.*[77] and Carrere-Kremer *et al.*[149], and the mutagenesis studies performed by Scull *et al.*[251] in combination with the observations made by members of the Zitzmann laboratory listed in Table 10 reveal that p7 tolerates a high degree of sequence variability, as long as the physicochemical properties are preserved.

Previous p7 studies carried out in the Zitzmann laboratory were based on GT1b isolate J4 and GT1a isolate H77, as were several publications during the early days of p7 research. The use of those particular HCV strains originated from the difficulties/inability of laboratories worldwide to grow HCV in cell culture (for further information see section 1.2.1). I continued for the main part of the structural and functional studies with GT1b isolate J4 so that previous findings made in the group could contribute to our understanding of the viroporin p7. This choice was not based on structurally relevant features of the different p7 strains and was driven by historic reasons (and to enable comparison to previous datasets) and experience of the Zitzmann laboratory and a comparably small number of labs who were involved in early p7 work.

Our efforts to re-evaluate the NMR studies of p7 solubilised in DPC (section 4.3), that led to the publication of an 'unusual' p7 structure published in 2013 and 2018 by Chou and co-workers[168,254], meant that I had to also express and purify mutants of p7 genotype 5a (isolate EUH1480, which contains five artificially introduced mutations to increase expression levels) to enable the direct comparison of our results. These were added to the list of p7 constructs expressed and purified in this dissertation.

For a complete list of all p7 wildtype proteins and related mutants prepared in this study, as well as a model showing the positions of the mutations within the context of the p7 monomer see Figure 20 and Table 11.

Table 10: Evaluation of potential p7 cysteine mutant positions to identify suitable positions for the attachment of EPR spin labels. P7 amino acid sequence (left column) based on GT1b isform J4. Note the naturally occuring cysteine 27 of isolate J4 was substituted with a serine (C27S) in order to avoid double labelling with the EPR spin label. If not indicated differently all obersvations are based on sequence alignments published by Carrere-Kremer *et al.*[149], Foster *et al*.[174] and Montserret *et al.*[77]. Table adapted from Dr Juan H. Bolivar Gonzales (former member of the Zitzmann laboratory).

Mutation	Comment
L2C	GT5a/EUH1480 is the only one that shows a C at position 2. Other GTs show L, V, M. Residue is not highly conserved, but always hydrophobic.
N4C	No C found at this position, but bulky polar amino acids (K, R).
A16C	Several isolates in GT2 and 6 contain C at this position. GT6 even has a bulky K. Residue forms part of the triple mutant N15A:C16A:H17A that has found to show reduced virion release and infectivity[160].
H17C	No C reported at this position. StGelais *et al.*[175] showed importance of H17 for p7 ion channel function in their liposome based dye release assay. Furthermore H17 is part of the triple mutant described above[160].
L20C	GT6 carries a C at this position. Numerous W, F for other GTs, variable position. Position is involved in the interaction with amantadine as shown by several studies[174,252].
S21C	GT2, 4 and 6 contain C at this position. A very variable position with bulky amino acids like W, H, Y and F. StGlais *et al*. 2009[175] report that S21A has no effect on the liposome dye release assay using FLAG-tagged p7.
F22C	Several isolates of GT6 carry C at this position. A very variable position with bulky (Y, F) and polar (T, S) amino acids. Liposome dye release assays using a F22:F25A:F26A triple mutant caused a 60% increase in dye release compared to wild type p7[175].
C27S	C at position 27 has to be substituted in order to avoid unwanted labelling of this position by EPR spin label. Amino acids naturally occurring in other GTs: V,T and I. In StGelais *et al.*[175] C27A has no influence on liposome based dye release assay compared to wild type p7. NMR studies of a p7 C27S mutant are reported in Cook *et al.*[249] and Osteringer *et al.*[166].
A28C	Several GT6 isolates have C. Variable with bulky amino acids (W, F).
Y31C	GT2a/BEBE1 shows C at position 31. Otherwise only H is found across several GTs.
K33C	No C reported at position 33. Amino acid located in hydrophobic section of the protein. *In vitro* assays of a R33A:R35A p7 double mutant showed a decrease in virion release and infectivity compared to wild type[160].
R35C	Same as K33C.
L36C	No C reported at this position. Probably because it is a hydrophobic region of the protein. It contains bulky amino acids like W (GT1), F (GT4) and other hydrophobics like V, I and A.
A37C	Same as position 36, but some strains do contain a T (within GT1). Mostly smaller hydrophobic amino acids, although GT3 carries a F at this position.

A40C	GT6a/EUHK2 and GT6b/Th580 have a C at this position. In addition residues like L, V, T and I are found naturally across other GTs. Position seems to be very variable.
S43C	C found in GT2a/JFH-1, as well as A, T, G, L and M.
F44C	GT5 (two isolates) and 6 (one isolate) have C at position 44. Variable with bulky (T, F) or hydrophobic amino acids. GT6 carries a W in one isoform.
Y45C	GT4 has many isolates with C. GT7 also carries a C at this position . Bulky amino acid (W in GT1), Y, polar and hydrophobic amino acids are found at this position. Very variable position.
L47C	GT2c/BEBE1 shows C mutation, otherwise hydrophobic residues like M and V are found.
L50C	Only one GT4 isolate has a C. Not very variable position, but a bulky F found in many ioslates of GT2 and 7. L50-55A had no effect on the liposome dye release assay performed by StGlais *et al*. 2009[175], but made the protein used in this study insensitive to amantadine.
L51C	GT2 has several isoforms with a C at this position. Variation of bulky, hydrophobic or polar residues are found. For L50-55A see L50C.
A56C	GT6 has six isolates with C. Mostly hydrophobic small residues, but GT7 carries an S and two isolates of GT6 carry a T at position 56.
P59C	No C residue reported at position 59. Very variable position, with larger amino acids like H, R, E, Y. Overall, more polar residues are found here.
Addition of C-1G0 at the Nt	Walter *et al*. 2016[253] use an N-terminal HAHA-tag attached with a GGGSG-linker. Furthermore in Foster *et al.* 2013[165] the authors claim that mutation C0 resulted in NMR chemical shifts unaffected by the reaction state of the cysteine introduced. Therefore the addition of a small linker (a single G) at the N-terminus was chosen to not disturb the p7 amino acid sequence.
Addition of G64C65 at the Ct	The C-terminal amino acid sequence of p7 (AYA) is highly conserved throughout all GTs[77,149,251]. To minimally perturb this region an approach was chosen similar to the one described for the N-terminal region (see above).

4.2. p7 protein expression and purification

Within this study the 22 different p7 mutants described in Table 11 were successfully ligated into expression vectors and transformed into bacterial cell systems for recombinant protein expression.

Some of the techniques described in this thesis (i.e. EPR-based techniques and sucrose density ultracentrifugation) require high quantities of protein (mg). Therefore, transformed *E. coli* BL21 colonies were screened for p7 overexpression and only those with expression levels above average, as determined by SDS-PAGE, were carried forward (Figure 22). Previous purification approaches based on synthetic protein synthesis and subsequent HPLC purification[255] were not suitable for the purification of recombinant p7 at a large scale. Large scale expression and purification was carried out according to a protocol described by Opella *et al.* 2011[249] with minor adaptations to the protocol due to the introduction of single- and double cysteine mutants. The presence of free cysteines required the use of reducing agent (TCEP or DTT) throughout the whole purification process. It was kept present until the protein precipitated in the final dialysis step to prevent the formation of non-native disulphide bonds. This was essential for subsequent spin labelling (see section 3.3.1).

One litre of bacterial cell culture resulted in about 1.8 g of lyophilised inclusion body extract containing the p7-fusion protein. After further processing (i.e. CNBr cleavage and SEC purification, Figure 23) about 45 mg of pure p7 protein could be obtained. The purity and mass of the p7 mutants were determined by SDS-page and MS (Figure 24 and appendix 8.1, Figure A. 1) showing very little impurities. Table 11 shows the expression/purification process performed for this thesis.

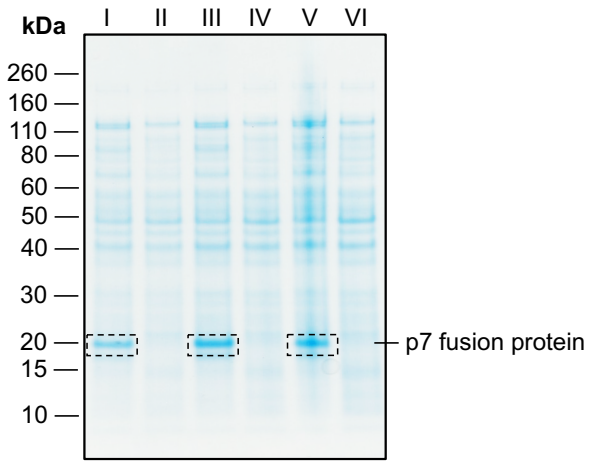


Figure 22: p7-fusion protein (mutant 3b) overexpression screening of *E. coli* BL21 transformed cells using SDS-PAGE. Molecular weight markers indicated on the left. *Lanes I-VI* – lysates of single colonies from expression trials. *Dashed boxes* – Protein bands showing overexpression of p7-fusion protein (~ 20 kDa).

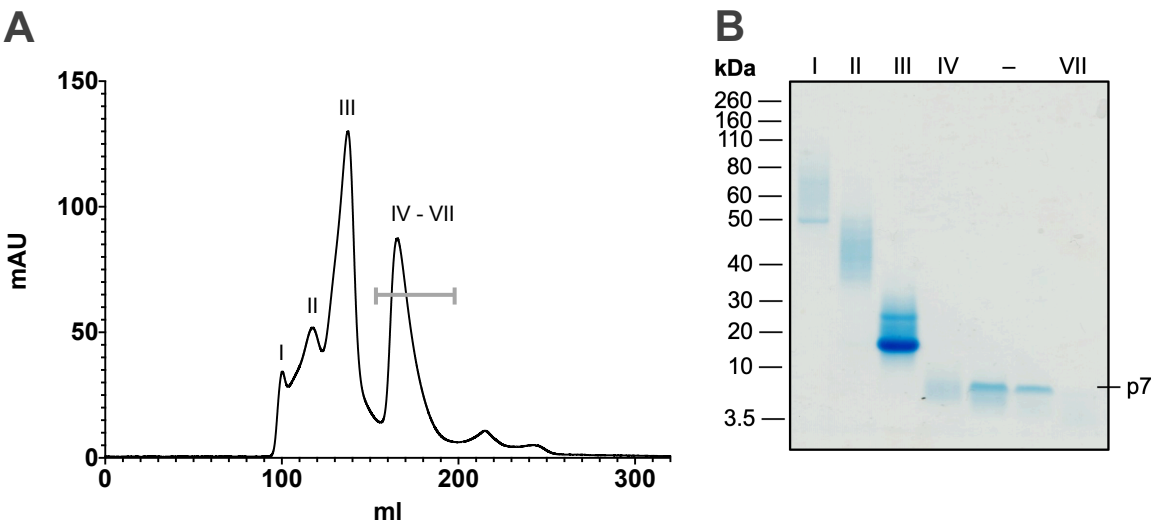


Figure 23: Gel filtration chromatography of p7 mutant 1 after CNBr cleavage of the fusion tag. (A) Sephacryl 200 26/60 HQ gel filtration. Grey line – fractions pooled for subsequent detergent removal. (B) SDS-PAGE under reducing conditions of fractions highlighted in (A).

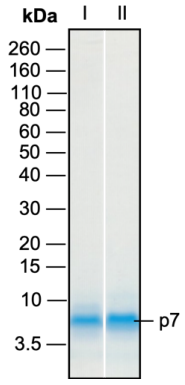


Figure 24: SDS-PAGE of p7 mutant 3 (I) and 6 (II) under reducing conditions. Lyophilised p7 powder was resuspended in 27 mM CF-6.

Table 11: HCV p7 mutant purification work flow at its current state. *Asterisk* – mutations listed are in addition to modifications of p7 genotype 5a described in OuYang *et al*. 2013[168]. *Double asterisk* – protein expressed in M9 minimal medium with the addition of $^{15}NH_4Cl$ for isotope labelling of the protein's nitroxide groups. *Triple asterisk* – p7 mutant not used in this study.

No.	Position of Mutation	Cloning	Bacterial expression	Fusion protein cleavage	SEC purification
	Genotype 1b (J4)				
WT		Done	Done		
0	C27S	Done	Done	Done	Done
1	C-1_G0_C27S	Done	Done	Done	Done
2	L2C_C27S	Done	Done	Done	
3a	L20C_C27S	Done			
3b	S21C_C27S	Done	Done	Done	Done
3c	F22C_C27S	Done			
3d	L23C_C27S	Done			
4	C27S_A28C	Done			
5	C27S_L36C	Done	Done	Done	Done
6	C27S_L51C	Done	Done	Done	Done
7	C27S_P59C	Done			
8	C27S_G64_C65	Done	Done	Done	Done
9	C-1_G0_C27S_G64_C65	Done			
10	L2C_C27S_P59C	Done	Done	Done	Done
11	S21C_C27S_L51C	Done	Done	Done	Done
12	C27S_K33R	Done	Done	Done	Done
	Genotype 5a*				
0a	-	Done	Done**	Done	Done
3e	W21C	Done	Done	Done	Done
6a	L51C	Done	Done	Done	Done
11a	W21C_L51C	Done	Done	Done	Done
13	A2C_L51C	Done	Done		
	Genotype 1a***				
H77	M47V	Done	Done		

4.3. HCV p7 in n-dodecylphosphocholine

During HCV infection p7 is exclusively found embedded in cellular membranes[148,150]. The high hydrophobicity of the protein makes the use of solvents or detergents necessary in order to study p7 outside of its natural environment. The use of lipid bilayers often complicates or even prevents the analysis by standard techniques like CD, NMR or X-ray crystallography. Many different solvents and detergents have been described for structural analysis of monomeric as well as oligomeric p7[77,164,165,167,168]. As mentioned in the introduction (section 1.2.2) the only high resolution structure of a p7 oligomer was published by OuYang *et al*. in 2013[168]. In this study and three subsequent publications[250,256,257] the authors solubilised the protein in varying concentrations of DPC and analysed it using NMR. By evaluating the recorded nuclear Overhauser effects (NOEs) of neighbouring protons the authors claim to have found intermolecular relations between p7 protomers within an oligomeric arrangement which were used to model the unusual oligomeric p7 structure (PDBID: 2M6X). The structure caused controversy in the field as it is in conflict with the results of several other studies. The following features drew particular attention:

(i) His17, which is critical for p7 ion conductance[167,175], is located at a position where its side chain is not involved in channel pore formation. Instead, it is pointing towards the membrane bilayer. (ii) The positioning of the p7 termini are in contradiction to immuno electron microscopy based N- and C-terminus detection by p7-specific F_{ab}-fragments[72] and (iii) the 'unusual' oligomeric structure would expose patches of charged residues (H17/H59/R60, p7 GT5a/EUH1480) into the hydrophobic core region of the surrounding membrane[258,259].

I was part of the joined effort between the Zitzmann lab (Dr Ben Oestringer and Dr Juan H. Bolivar Gonzalez) and Prof Jason Schnell (Biochemistry Department,

University of Oxford) to disprove the published structure. I expressed and purified the same p7 construct as used by OuYang *et al*. [168] and prepared a number of samples for the NMR and the SEC-MALLS experiments carried out on p7 solubilized in DPC and SDS under varying conditions. Subsequent analyses performed by Oestringer and Schnell confirmed our hypothesis that the NOEs reported in the OuYang paper[168] might have been misinterpreted. Instead of looking at intermolecular NOEs the authors were actually looking at intramolecular NOEs, which led to wrong conclusions. Our work was published in Nature[84] as a brief communication arising from OuYang *et al*. 2018[168].

4.3.1. Nuclear magnetic resonance spectroscopy

In order to be able to quantify the NMR data shown in OuYang *et al*.[168] and Dev *et al*.[250] a sample of ^{15}N-labelled p7 mutant 0a was prepared following the protocol described in the OuYang study. Recorded ^{1}H-^{15}N-correlation NMR spectra overlaid with data shown in both studies confirm the successful reproduction of their sample conditions[168,250] (Figure 25A).

Two subsequently conducted measurements were carried out on the very same sample: (i) *Conventional, heteronuclear single quantum correlation (HSQC) NMR* – The sample allowed the recording of high quality HSQC data at 37˚C (Figure 25B, left). Assuming an overall particle size of 60-80 kDa (hexameric p7 in combination with associated detergent shell) the achievement of high quality data using this technique was very unexpected[260]. To rule out changes in the protein's secondary structure due to the increased temperature (initial ^{1}H–^{15}N-TROSY NMR experiments were carried out at 30˚C) measurements were repeated between 30-37˚C without any changes observed (Figure 25B, right).

(ii) *Particle size estimation by rotational correlation time analysis based on NMR ^{15}N R_1- and R^2-relaxation rates* – The correlation time of p7 solubilised in 200 mM DPC was calculated to be 10.1 ns at 37˚C. This corresponds to an effective particle mass of about 39 kDa, which would be in good agreement with a p7 hexameric arrangement (~ 42 kDa) when totally neglecting the detergent shell associated to the protein. As shown further down by SEC-MALLS analysis (Section 4.3.2), about 78% of the particle mass contribution corresponds to the detergent (Table 12). The observation makes a hexameric arrangement of the protein under the conditions described in OuYang *et al.*[168]/Dev *et al.*[250] unlikely and therefore contradicts their findings. Alternatively, a similar rotational correlation time could be achieved when having a p7 monomer associated with a detergent shell of about 30 kDa (calculations based on SEC-MALLS analysis of p7 in 50 mM DPC, see section 4.3.2).

The structure of the p7 oligomer by OuYang *et al.*[168] is based on the interpretation of putative intermolecular NOEs resulting from a 3D ^{15}N-edited NOE spectroscopy with transverse relaxation-optimized spectroscopy (NOESY-TROSY) experiment, of a 1:1 mixture of ^{15}N-^{2}H- and ^{13}C-labelled p7. The technique in general is capable of informing on intermolecular NOEs between exchangeable protons of ^{15}N-^{2}H-labelled protein with non-exchangeable aliphatic protons of the ^{13}C-labelled protein species. However, control experiments of unmixed, pure ^{15}N-^{2}H-labelled protein are needed in order to eliminate intramolecular NOEs arising from incomplete deuteration. This is because unintended proton sources, such as starting culture, air, residual protons in D_2O and deuterated glucose used for recombinant cell expression, cannot be entirely ruled out. OuYang *et al.*[168] did not perform such control experiment in their study. In support of the possibility of a wrong assignment of NOEs, Oestringer *et al.*[84] show a parallel comparison between the authors'[168] mixed-labelled spectrum and their

unmixed ^{15}N-^{2}H-labelled spectrum which reveal that strong intramolecular NOE peaks in the unmixed sample coincide with seven of the twenty-two putative intermolecular NOEs in the mixed sample (Figure 26).

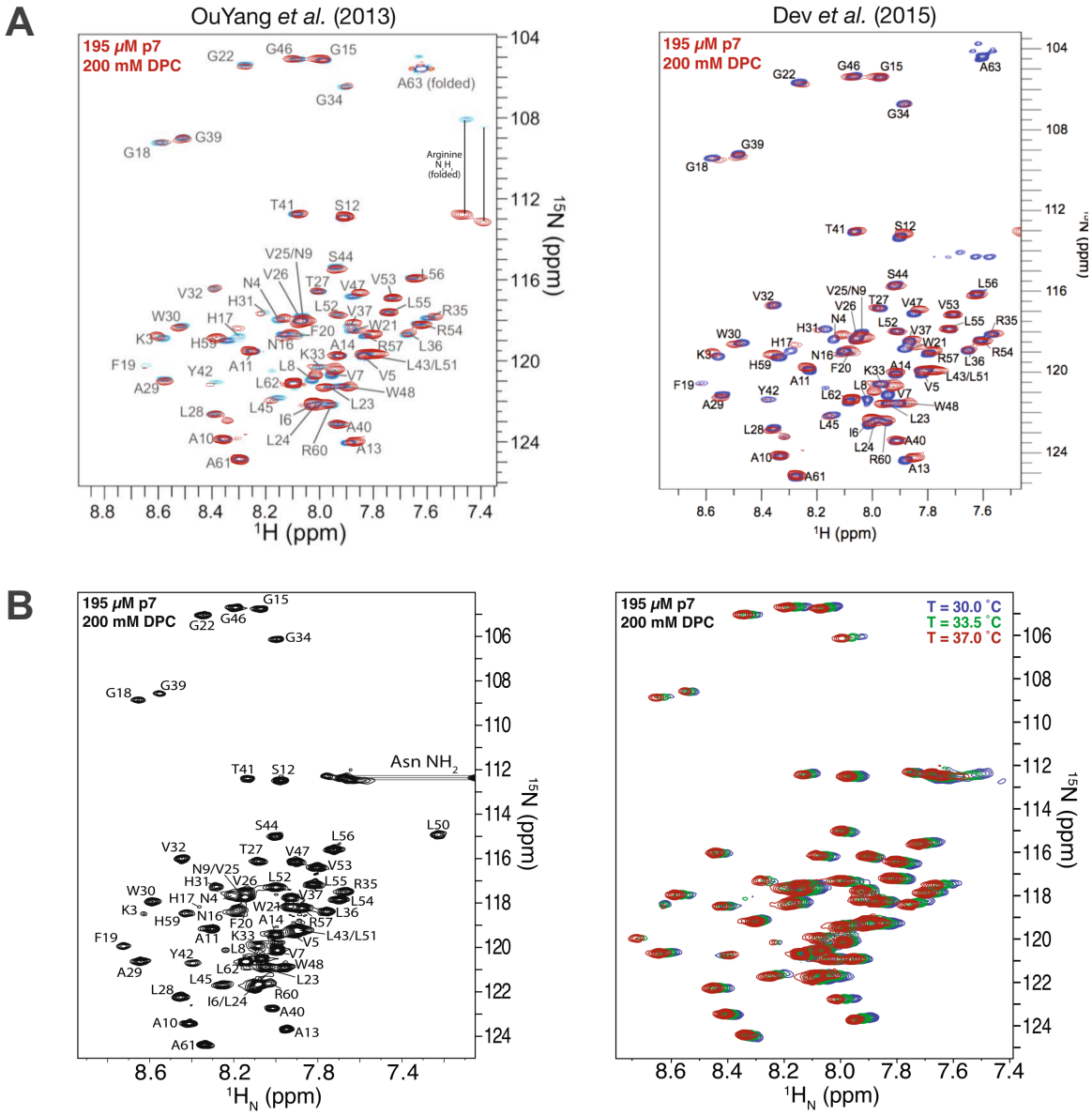
The mixed-labelled NOEs show a slight (about 0.01 ppm) up-field shift, which is typical for mostly deuterated proteins, caused by e.g. methyl groups hosting deuterium and protons simultaneously (R-$C^2H_2{}^1H$). Deuterium introduced shifts can be misinterpreted for long range NOEs due to them being as large as 0.04 ppm (e.g. alanine methyls)[261]. About a third of the p7 backbone amines show NOEs with signs for incomplete deuteration (Figure 26A). Further evaluation revealed the coexistence of those NOEs with amino acids having side chain located protons in close proximity to the protein's backbone amines. In general alanine is prone to cause such cross-peaks as it has three protons close to the C_α-chain, which are all subject to insufficient deuteration. And indeed, out of the eight alanine-based NOEs observed by OuYang *et al.*[168] six show cross-peaks potentially caused by residual protonation.

Section 1.2.1 describes the currently known drugs found to be active against the viroporin p7. One of them – rimantadine – is used by OuYang *et al.*[168] to validate drug interactions with their suggested p7 oligomeric arrangement.

Figure 26B shows the p7 amino acids and their corresponding intermolecular NOEs to rimantadine as outlined by OuYang *et al.*[168] (Figure 26B, right columns). The intermolecular NOEs specifically associated to drug binding are represented by cross-peaks in the mixed-labelled NOESY and only ^{15}N-^{2}H-lablled NOESY as well (Figure 26B, left and middle columns, respectively), bearing in mind that those samples do not contain any drug. Backbone amide chemical-shift differences analysed based on the recorded spectra by OuYang *et al.*[168] in the presence and absence of rimantadine (5 mM) show unspecific drug binding all along the p7 sequence. Figure 25D visualises

the chemical-shift calculated in this study while indicating single amino acids (pick bars in Figure 25D) which have been interpreted by OuYang *et al.*[168] to form the rimantadine binding pocket.

This again supports the evidence that the data published by OuYang *et al.*, Dev *et al.* and Chen *et al.*[168,250,254] are being misinterpreted by the authors.



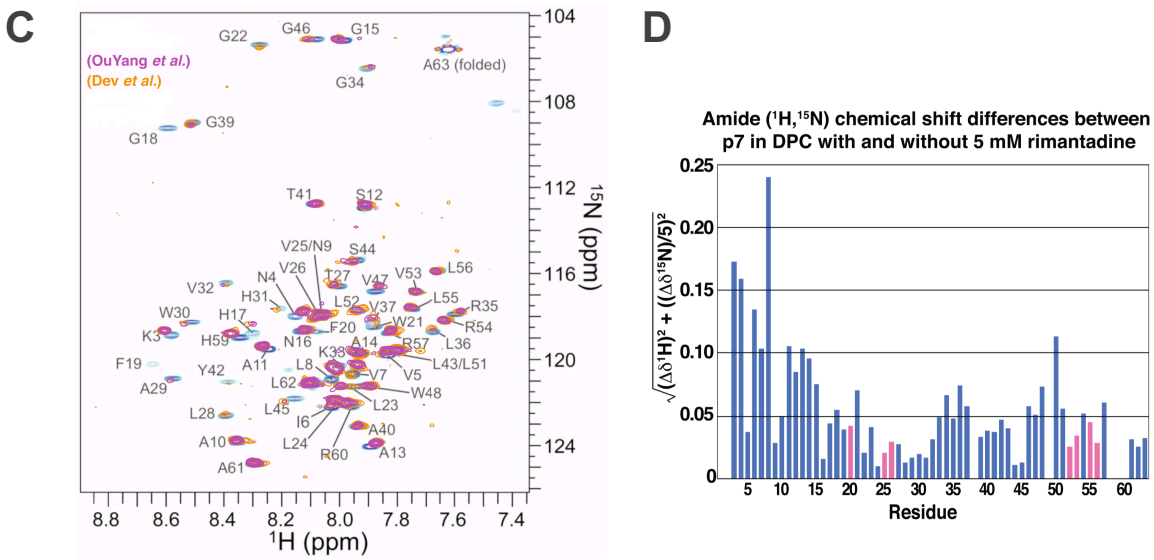
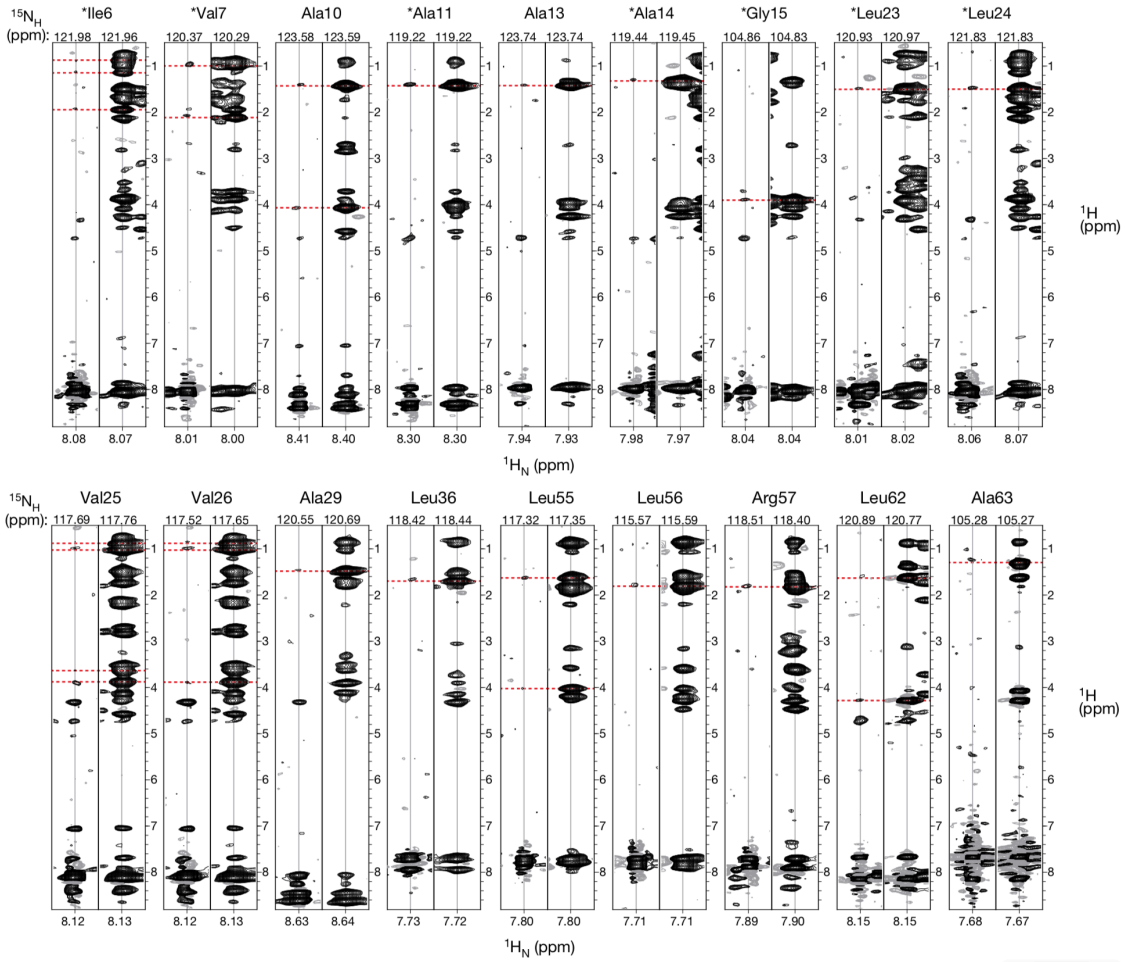


Figure 25: NMR spectroscopy analysis of p7 genotype 5a (mutant 0a) in DPC micelles. (A) 2D ^{1}H-^{15}N-TROSY NMR manual comparison between p7 prepared as described above and previously published spectra[168,250]. Left, NMR spectrum overlay of p7 in DPC recorded in this study (red peaks, 30 °C, ^{1}H frequency of 600 MHz) and data previously published as indicated above spectra (blue peaks). Right, same red p7 spectrum (red cross peaks) as shown on the left, overlaid with another published spectrum as indicated above spectra (blue peaks). (B) ^{1}H–^{15}N HSQC of p7 in DPC at varying temperatures. Left, high quality ^{1}H–^{15}N HSQC data on the sample described in (A). Measurements carried out at 37˚C. Right, temperature sensitivity analysis of p7 in 200 mM DPC at 30, 33.5 and 37˚C (blue, green and red, respectively). (C) NMR spectra overlay of p7 in DPC. *Magenta* – spectrum by OuYang *et al*. 2013[168], *Orange* – spectrum by Dev et al. 2015[250], *Blue* – SEC-MALLS peak elution conditions, after injection sample was prepared according to OuYang *et* al.[168] containing 195 µM p7 mutant 0a in 200 mM DPC. At time of elution calculated sample concentrations were as follows: 5 µM p7, 50 mM DPC. For band-selective excitation shot-transient transverse relaxation-optimized spectroscopy (BEST-TROSY), with a total recording time of about 20 h in order to obtain sufficient signal-to-noise ratio (SNR), the following sample was prepared: 5 µM p7 mutant 0a, 25 mM MES at pH 6.5, 50 mM DPC, 1 mM DSS and 5% D2O. (D) Backbone amide chemical-shift differences of p7 in DPC analysed based on the recorded NOESY spectra by OuYang *et al.*[168] in the presence and absence of rimantadine (5 mM). *Blue columns* – Chemical-shifts calculated in this study. *Pink columns* – Single amino acids which have been interpreted by OuYang *et al.*[168] to form the rimantadine binding pocket. Figures adopted from Oestringer *et al.* 2018[84].

A


$^{15}N_H$ (ppm):
*Ile6 121.98 121.96
*Val7 120.37 120.29
Ala10 123.58 123.59
*Ala11 119.22 119.22
Ala13 123.74 123.74
*Ala14 119.44 119.45
*Gly15 104.86 104.83
*Leu23 120.93 120.97
*Leu24 121.83 121.83
8.08 8.07 8.01 8.00 8.41 8.40 8.30 8.30 7.94 7.93 7.98 7.97 8.04 8.04 8.01 8.02 8.06 8.07
^1H_N (ppm)
^1H (ppm)
$^{15}N_H$ (ppm):
Val25 117.69 117.76
Val26 117.52 117.65
Ala29 120.55 120.69
Leu36 118.42 118.44
Leu55 117.32 117.35
Leu56 115.57 115.59
Arg57 118.51 118.40
Leu62 120.89 120.77
Ala63 105.28 105.27
8.12 8.13 8.12 8.13 8.63 8.64 7.73 7.72 7.80 7.80 7.71 7.71 7.89 7.90 8.15 8.15 7.68 7.67
^1H_N (ppm)
^1H (ppm)

B

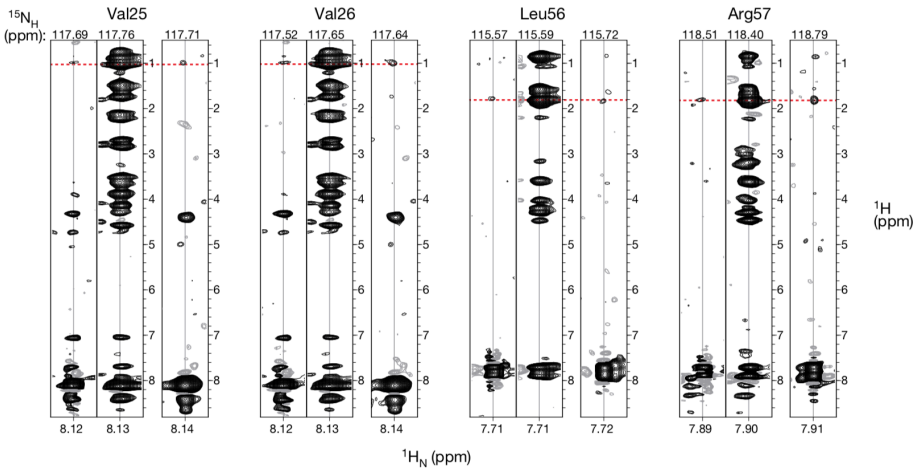


Figure 26: NOESY NMR of p7 in DPC micelles. (A) 3D ^{15}N-edited NOESY spectra of OuYang *et al.*[168] aligned along the indirect ^{1}H-dimension. *Left strip* – 1:1 mix of ^{15}N–^{2}H- and ^{13}C-labelled p7, of amino acid indicated at the top. *Right strip* – ^{15}N–^{1}H-labelled sample. Positive and negative contours are in black and grey, respectively. *Red lines* – chemical-shift correlations between measurements. Based on chemical shifts of intramolecular NOEs shown in the fully protonated sample (right strip). *Asterisks* – Amino acids from which OuYang *et al.*[168] identified putative intermolecular NOEs. (B) NMR analysis of amino acids involved in rimantadine binding as indicated by OuYang at al.. *Right strip* – putative intermolecular NOEs of drug and indicated amino acid (top of figure) from a ^{15}N–^{2}H-labelled sample containing 5 mM rimantadine. *Left and middle strips* – corresponding amino acid NOESYs shown in A). Figures adopted from Oestringer *et al*. 2018[84].

Table 12: NMR and SEC-MALLS evaluation of particle mass contribution corresponding to detergent and protein. *Asterix* – Hypothetical arrangement of p7, due to limitation of technique to inform on detergent and protein contributions. Measurements done on p7 mutant 0a suspended in 200 mM DPC. *Cross* – calculated value, based on protein-detergent distributions found with SEC-MALLS of p7 in 50 mM DPC, see section 4.3.2.

Methodology	Effective protein mass (kDa)	Effective detergent mass (kDa)
^{15}N R_1 and R_2 relaxation rates (NMR), hexameric p7*	39.3	-
^{15}N R_1 and R_2 relaxation rates (NMR), monomeric p7*	7 ± 2.7†	30.6 ± 2.7†
SEC-MALLS (50 mM DPC)	6.4 ± 1.1	23.5 ± 0.5
SEC-MALLS (3 mM DPC)	8.4 ± 0.7	42.4 ± 3.7

4.3.2. Size exclusion – Multi-angle light scattering

To further support the findings in section 4.3.1 – i.e. that p7 genotype 5a adopts a monomeric arrangement in 50 or 200 mM DPC – size exclusion chromatography combined with SEC-MALLS was carried out on samples similar to the ones prepared in Dev *et al.*[250]. The technique decomposes detected masses into protein and detergent distributions utilising the high molar absorbance coefficient of p7 (GT5a: 17,990 M^{-1} cm^{-1}) and the relatively good scattering properties of the detergent-protein complex[262]. First rounds of SEC-MALLS had shown that by using the NMR detergent concentrations described in OuYang *et al.* 2013[168] (200 mM DPC) in the running buffer, the measurements were inapplicable due to the excessive light scattering of the numerous detergent micelles (data not shown). This resulted in a change in SEC-MALLS strategy by lowering the running buffer detergent concentrations down to 50 mM, thereby matching NMR detergent concentrations described in Dev *et al*. 2015[250]. In theory this change should not have any implications on the p7 oligomeric state as the two studies[168,250] show identical NMR p7 spectra under the two detergent concentrations mentioned above.

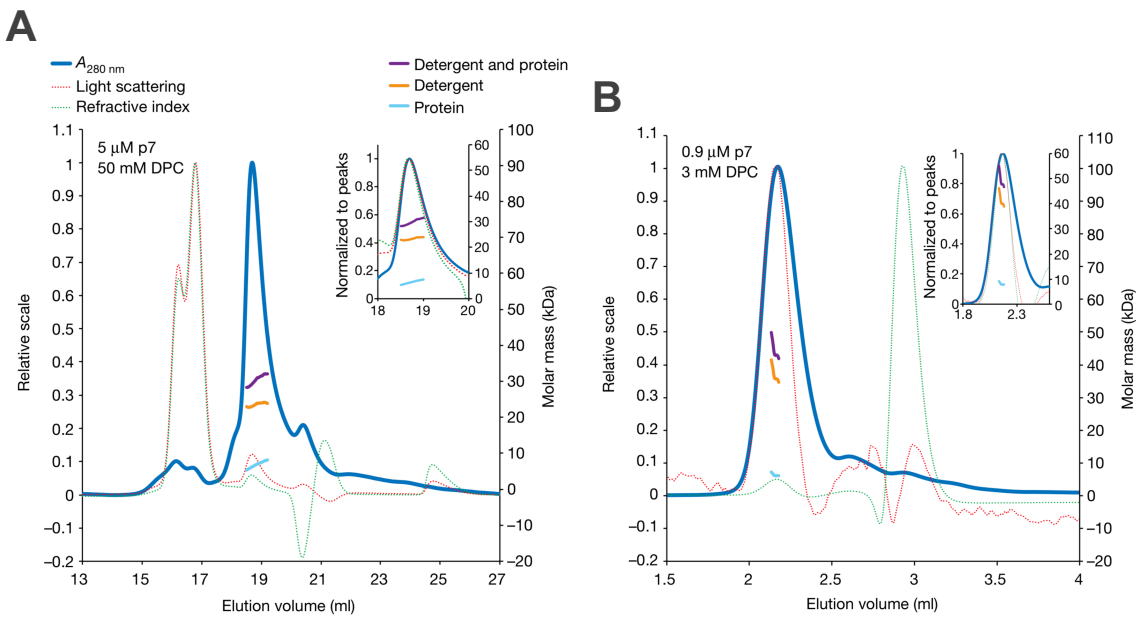
Figure 27A shows p7 genotype 5a (Table 11, mutant 0a) refolded in 200 mM DPC following the protocol of OuYang *et al.* 2013[168] with the sample subsequently being dialysed into buffer containing 50 mM DPC compensating detector saturation by detergent micelle light scattering. The sample's final detergent-to-protein (DPR) molar ratio was 250:1. Evaluation of the SEC-MALLS elution profile shows minor protein aggregation at about 16 ml followed by the main protein-detergent complex at about 19 ml. For the latter the calculated mass values deconvolute into 6.4 ± 1.1 kDa for protein (p7 protein mass: ~7 kDa) and 23.5 ± 0.5 kDa for the associated detergent shell. The values are in a good agreement with p7 adopting a monomeric conformation

under SEC-MALLS elution conditions. Another experimental sample composition described by the authors[168], which was used for the analysis of putative p7 hexamers in 3 mM DPC, were electron microscopy (EM) studies shown in Figure 27B. Here p7 was refolded as previously described with excess detergent being removed using SEC on a Superdex 200 10/300 column equilibrated in 3 mM DPC running buffer. The run showed a small peak of protein aggregation similar to the one in A. The later eluting, monodispersed peak was collected and subjected to SEC-MALLS analysis. Besides matching detergent concentrations it has to be noted that the protein concentration in our sample was slightly higher than in the originally described EM experiment, resulting in a DPR of 200:1 pre-injection. The evaluation of the SEC-MALLS run (Figure 27B) shows masses corresponding to 8.4 ± 0.7 kDa of protein and 42 ± 3.7 kDa of associated detergent, which again is in good agreement with p7 being monomeric in the sample prepared. The last sample condition tested was the one described in Dev *et at.* 2015[250] indicating p7 and detergent masses of 7.8 ± 1.1 kDa and 24.7 ± 0.6 kDa, respectively (C and Table 13).

Table 13A, B and C describe the protein and detergent concentrations as used for SEC-MALLS-based mass analysis at peak elution summit. Due to dilution effects during SEC, DPRs at sample injection time differ significantly to DPRs at peak elution, with approximately a 15-fold dilution taking place during the SEC run (see Table 13). This could bring about a change in p7 oligomeric state. To further address this question NMR samples were prepared matching the peak elution conditions found in Figure 27A (5 µM p7, 25 mM MES at pH 6.5, 50 mM DPC). The comparison with published NMR spectra – used to model the p7 oligomer described by OuYang *et al.* 2013[168] and Dev *et al*. 2015[250] – shows a good overlay of all three spectra (Figure 25C), supporting our hypothesis that p7 solubilised in DPC at DPRs 200-10,000:1 is monomeric.

Furthermore, Table 13 illustrates another phenomena which is caused by the detergent concentration used for SEC-MALLS analysis. When comparing SEC-MALLS measurements at detergent concentrations close to the DPC's critical micelle concentration (CMC; ~ 1.5 mM, Figure 27B) and concentrations several-fold above the corresponding CMC (Figure 27C) the mass of protein associated detergent increases with decreasing detergent concentrations. This phenomena was also observed with p7 solubilised in 10 or 80 mM SDS (CMC SDS ~ 8 mM, Figure 27 D-E) and SEC-MALLS experiments carried out on detergent-only samples (calculated protein and/or detergent masses listed in Table 13). The detergent micelle sizes evaluated at lower detergent concentrations are markedly and consistently higher and values found are in close agreement with masses reported in the literature: DPC ~ 19.0 kDa and SDS ~ 17.3 kDa[263]. The reason for this micelle behaviour is not yet described in the literature, but it is potentially caused by the change in hydrophobicity of the micelle surrounding buffer solution.

P7 appears unambiguously monomeric in the SDS DPRs tested showing a similar trend seen for p7 in DPC.



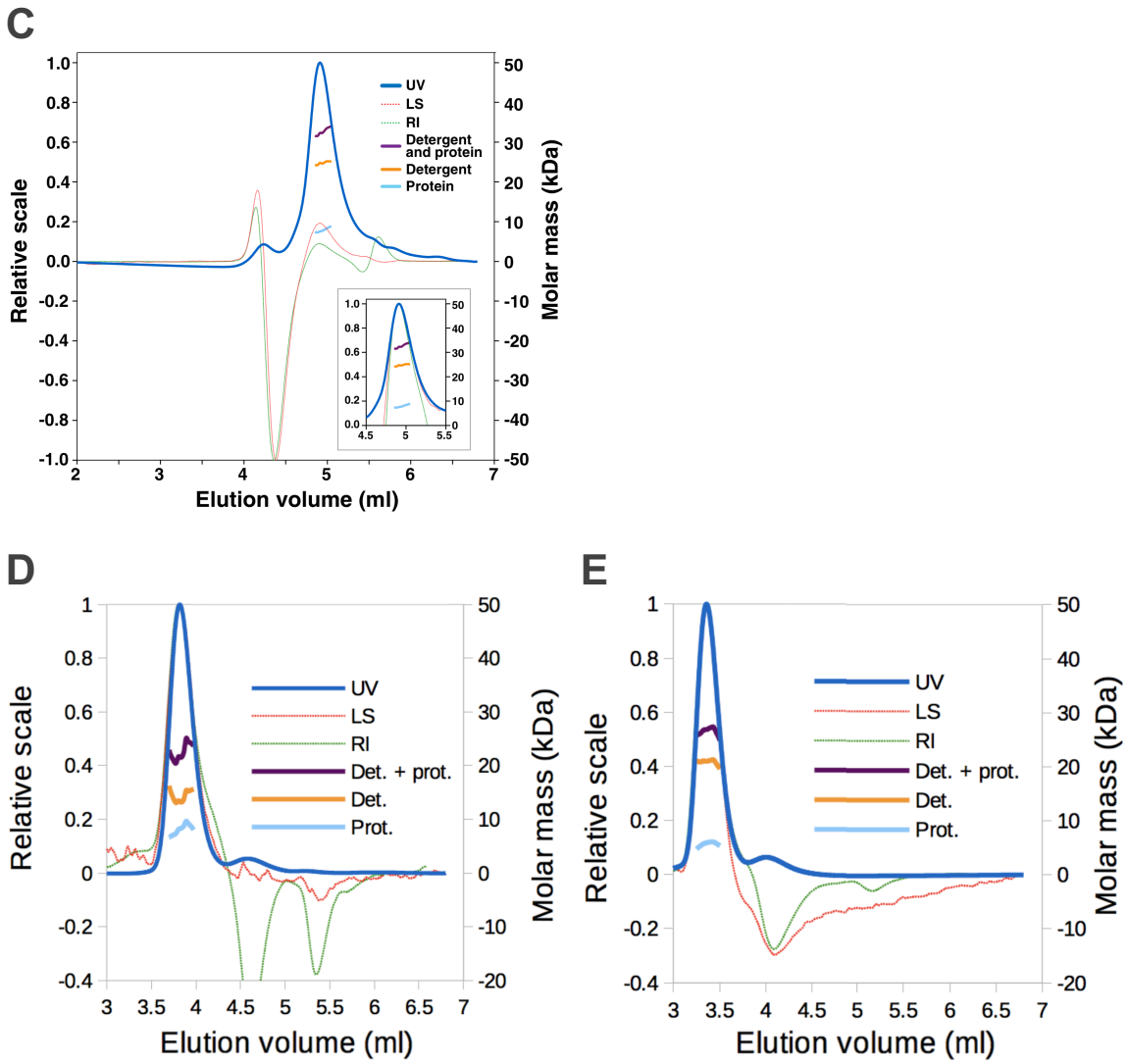


Figure 27: SEC–MALLS of p7 genotype 5a in varying DPC concentrations, as well as empty detergent micelle analysis. (A) SEC–MALLS of a sample prepared following the protocol in OuYang *et al.* 2013[168] with subsequent dialysis against 50 mM DPC to match detergent concentrations described in Dev *et al.* 2015[250]. SEC–MALLS was performed using a Superdex 200 10/300 column. Reflective index, light scattering and absorbance at 280 nm, as well as protein molar masses, associated detergent shell and detergent-protein complexes mass are represented by the left and right axis, respectively. The calculated p7 (molar, monomer) and DPC concentration at elution peak summit are indicated in the upper left corner. Further information about samples injected and calculated protein and detergent masses are summarized in Table 13. (B) Sample as prepared for electron microscopy studies described by OuYang *et al.* 2013[168] (3 mM DPC, 300nM protein), but with higher protein concentration. The protein was refolded in 200 mM DPC. Excess detergent was removed by applying the sample to a Superdex 200 10/300 column equilibrated in 3 mM DPC. The collected peak at 280 nm was analysed by SEC–MALLS on a Superdex 200 5/150 column with the system being equilibrated with running buffer containing 3 mM DPC. (C) Sample prepared as described in Dev *et al.* 2015[250]. SEC-MALLS analysis took place on a Superdex 200 5/150 column (D) p7 refolded in 80 mM SDS (10-fold corresponding CMC) 10mM sodium phosphate pH 7.2, chromatography was carried out on a Superdex 200 5/150 column. (E) Same as D) but with only 10 mM SDS in the sample as well as in the chromatography system. (A-E) Further information is provided in Table 13. Figures adopted from Oestringer *et al.* 2018[84]

Table 13: Detailed list of samples used for SEC-MALLS analysis described in section 4.3.2. *Asterix* – corresponding sample detergent concentration (mM). *†* – Superdex 200 10/300 column, *††* – Superdex 200 5/15 column, *†††* –two Superdex 200 5/15 columns in series.

	Pre-injection			Elution (peak summit)				
Corresponding figure	**Detergent conc. running buffer**	**p7 conc. (µM, monomeric)**	**Detergent-to-protein ratio**	**Sample injection Vol. (µl)**	**p7 conc. (µM, monomeric)**	**Detergent-to-protein ratio**	**Protein mass (kDa)**	**P7 associated detergent (kDa)**
27A	50 mM DPC †	200	250	70	5	10,000	6.4 ± 1.1	23.5 ± 0.5
27B	3 mM DPC ††	15	200	15	0.9	3,333	8.4 ± 0.7	42.4 ± 3.7
27C	50 mM DPC †††	140	357	50	5	10,000	7.8 ± 1.1	24.4 ± 0.6
27D	80 mM SDS †††	464	172	30	36	2,222	7.3 ± 1.8	10.2 ± 0.7
27E	10 mM SDS †††	58	172	30	5	2,000	5.9 ± 0.8	24.1 ± 2.3
A. 2.	80 mM SDS †††	58	1,379	30	6	13,333	8.3 ± 1.7	13.2 ± 3.1
A. 2.	10 mM SDS †††	120	83	30	9	1,111	5.9 ± 1.2	21.0 ± 1.5
Oestringer *et al.* 2018	3 mM DPC ††	-	8*	5	-	-	-	20.2 ± 3.2
Oestringer *et al.* 2018	50 mM DPC †	-	55*	100	-	-	-	16.8 ± 1.8
Oestringer *et al.* 2018	10 mM SDS †††	-	16*	30	-	-	-	17.2 ± 3.4
Oestringer *et al.* 2018	80 mM SDS †††	-	80*	30	-	-	-	6.9 ± 3.1

In summary, SEC-MALLS analyses of p7 solubilised in DPC described in this dissertation and in Oestringer *et al*. 2018[84] are supporting the assumption that p7 solubilised in DPC is monomeric at a wide range of detergent-to-protein ratios (DPC 200:1 - 10,000:1 tested). The evaluation of SEC-MALLS experiments for samples prepared following the NMR protocols published in OuYang *et al.* 2013[168] and Dev *et al*. 2015[250] show protein and detergent masses of 6.4 ± 1.1 / 23.5 ± 0.5 kDa and 7.8 ± 1.1 / 24.4 ± 0.6 kDa, respectively. This is in close relation to the molecular mass of monomeric p7 (~ 7 kDa). To rule out misinterpretation of the data, due to dilution effects during SEC, sample composition at elution peak summit (Figure 25A) were reproduced and analysed by NMR. The overlay of this measurement with NMR measurements of the putative p7 oligomer published by OuYang *et al.*[168] and Dev *et* al.[250] show good coverage of the 1H-15N-peaks for all three spectra (Figure 25C).

In response to the work described by Oestringer *et al.* 2018[84] Chou and co-workers published a study emphasising the importance of the DPR for the formation of p7 oligomers in DPC[257]. Furthermore, the authors claim that the 40-fold discrepancy in the DPRs between their 2013 study[168] and the study described by Oestringer *et al.* [84] is the cause for the different conclusions presented.

To my concern, the authors do compare the SEC-MALLS elution-peak concentrations from Oestringer *et al.* [84] with their own SEC injection concentrations. Thus, the central statement of Chen *et al.* [257] – the 40-fold difference between the two studies – cannot be correct, as both samples are prepared following the same protocol and are agreeing in their DPR before injection.

With increasing DPRs (DPR > 127:1) Chen *et al.* [257] describe a brake-down of the p7 oligomer into monomers. At DPR 10,000:1, at which both Oestringer *et al.* [84] and Chen

et al. [257] agree p7 is monomeric, the NMR spectra of p7 (shown by Oestringer *at al.* [84]) is the same as at DPR concentrations used by the Chou and co-workers in their structural studies of oligomeric p7 published in 2013 and 2015[168,250].

The incorrect comparison of their SEC injection concentrations with the SEC-MALLS elution-peak concentrations of Oestringer *et al.* [84], in addition to contradicting statements made by the authors and the use of inadequate techniques is an indisputable and fundamental error that causes confusion for the p7 field that has to be addressed.

4.4. p7 in lipid bilayers

As p7 is primarily found in the host's ER membrane during HCV infection a lipid composition close to the one naturally found in the ER membrane of mammalian cells would reflect the most physiologically relevant environment for structural studies of p7. According to the literature[264], the four main lipids found in the ER membrane of eukaryotes are phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI) and phosphatidylserine (PS). With PC and PE being about 55% and 30% of the ER's total phospholipid content, respectively, followed by about 10% PI and 3% PS.

Artificial DOPC:POPE lipid bilayers support p7 channel ion currents in planar lipid bilayers. In particular mixtures at a 4:1 molar ratio showed increased ion channel openings resulting in a burst-like pattern with openings lasting up to 0.5 s[76].

At present, only three publications have reported experimental structural data of p7 in a lipid environment[76,171,254]. The use of lipids for the structural analysis of membrane proteins is still challenging as most of the commonly used techniques (e.g. NMR, SEC or X-ray crystallography) cannot analyse samples provided in lipids or the disadvantages, such as slow correlation times, loss in sensitivity or high light or X-ray scattering, outweigh the gains. Therefore, detergents or solvents are often used to replace lipids for the analysis of membrane proteins, providing some but not all lipid-mimicking properties. Chou and co-workers[254] used NMR and solvent paramagnetic relaxation enhancement (PRE) analysis to validate the putative p7 oligomer in bicelles (DMPC/DHPC) mimicking a lipid bilayer. Unfortunately the authors omit to report whether the bicelles stay intact upon protein addition. To support the validity of their results the authors compare their NMR results with an NMR spectrum of p7 solubilised

in DPC similar to the one shown in OuYang *et al*. 2013[168], which we have shown not to be representative of oligomeric p7[84] (see section 4.3).

It has been shown that the environment in which p7 is analysed influences its secondary structure and thus might affect properties such as its tendency to oligomerise or its ability to conduct ions. Single channel recordings carried out by Whitfield *et al*.[76] suggest a correlation between p7 activity and the lipid composition chosen for reconstitution. This underpins the need to analyse p7 under conditions which are close to its natural environment – the ER membrane. Only when the conditions provided are in favour for the correct formation of the p7 secondary structure one can start to understand the HCV p7 viroporin and its function within virus infected cells. Structural information is required in order to carry out experiments aimed at understanding p7 inhibition by drugs such as amantadine or long alkyl chain iminosugars. As the structural information of monomeric as well as oligomeric p7 in lipid environments is still very limited the experiments in this dissertation will focus on the structural analysis and the formation of p7 ion channels in lipid bilayers composed of DOPC/POPE, the main bulk lipids making up the ER membrane.

Due to the design of the sample preparation protocol and the way p7 was reconstituted into lipid membranes all experiments conducted in section 4.4 are based on the sample preparation process outlined in Figure 28. Samples for the individual techniques were taken at different time points of the p7 sample preparation protocol as illustrated by small arrows. This allowed monitoring of the p7 reconstitution progress throughout each step of the protocol.

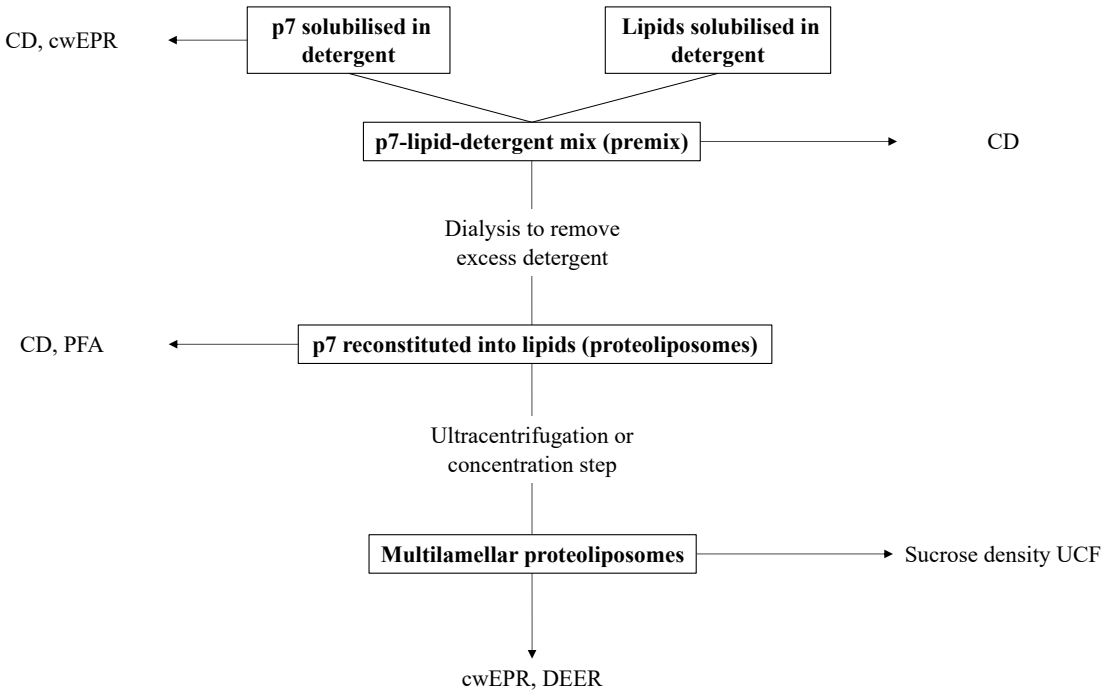


Figure 28: p7 reconstitution into lipid bilayers, basic work flow. *Arrows* – indicate removal of sample aliquots for indicated analysis during the preparation process. *CD* – circular dichroism, *cwEPR* – continuous wave electron paramagnetic resonance, *PFA* – proton flux assay, *UCF* – ultracentrifugation , *DEER* – double electron-electron resonance spectroscopy

4.4.1. Sucrose density gradient ultracentrifugation

A number of p7 mutants (mutants 0, 1, 3b, 5 and 6), were reconstituted into multilamellar liposomes consisting of DOPC:POPE (4:1, w/w) at lipid-to-hexameric protein molar ratios (LPRs) of 150:1, 250:1 and 2000:1. The resulting sharp bands in the sucrose density gradient ultracentrifugation (UCF) confirm the formation of proteoliposomes (PL) of a homogeneous density (Figure 29). As expected for a decrease in LPR the PLs formed travel further within in the gradient, e.g. PLs at LPR 150:1 (Figure 29, A) have a higher protein content and thus a higher density compared to PLs at LPR 250:1 or 2000:1 (Figure 29, B and C). The results of the sucrose density gradient UCF show a successful reconstitution of p7 at different LPRs into lipid bilayers of DOPC:POPE, but do not inform about the precise LPRs achieved. In order

to analyse LPRs achieved the sample contained a small fraction of rhodamine-PE (rPE, absorption at 564 nm) and its absorbance was compared to that of the protein (280 nm). LPRs achieved at the end of the sample preparation showed an error margin of up to ± 13% compared to the anticipated LPR.

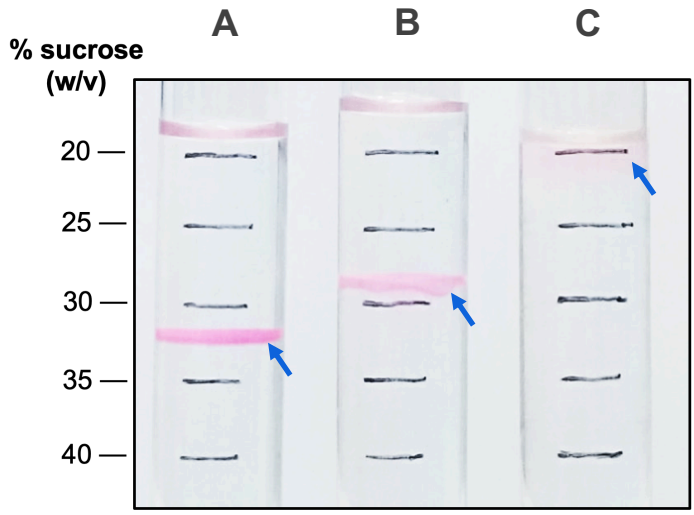


Figure 29: Sucrose density gradient ultracentrifugation analysis. Samples show p7 reconstituted into multilamellar proteoliposomes (MPL) consisting of DOPC:POPE (4:1, w/w). (A) LPR 150:1, (B) LPR 250:1 and (C) LPR 2000:1. For visualisation and absorbance measurements rhodamine-PE (rPE) was added (0.6:1 rPE to p7 monomer). Sucrose concentrations range from 20-40% (w/v). *Arrow* – MPL layer.

4.4.2. Circular dichroism

Several high-resolution NMR structures derived from monomeric p7 in different membrane mimetic environments are discussed in the literature[77,164-166]. All authors come to similar conclusions, where the p7 monomer adopts two main transmembrane helices connected by a short loop region with an overall very high content of α-helicity. In this thesis CD was used to monitor the reconstitution process of p7 into DOPC:POPE lipid bilayers. Figure 30A shows p7 labelled with MTSL (except mutant 0) solubilised in 27 mM CF-6 prior to detergent remove by dialysis for reconstitution

into lipid bilayers. The spectra of the different p7 mutants overlay (Figure 30) nicely and clearly indicate α-helicity with characteristic dips at about 208 and 222 nm. Evaluation of the spectra utilising the DICHOWEB web server[265] resulted in 68 to 89% α-helix, 1 to 11% β-sheet 4 to 10% turns and 4 to 11% unordered with NRMSDs < 0.01 (Table 14). The spectra overlay indicates that the mutations introduced have little influence on the CD spectra line shape and thus on p7 secondary structure.

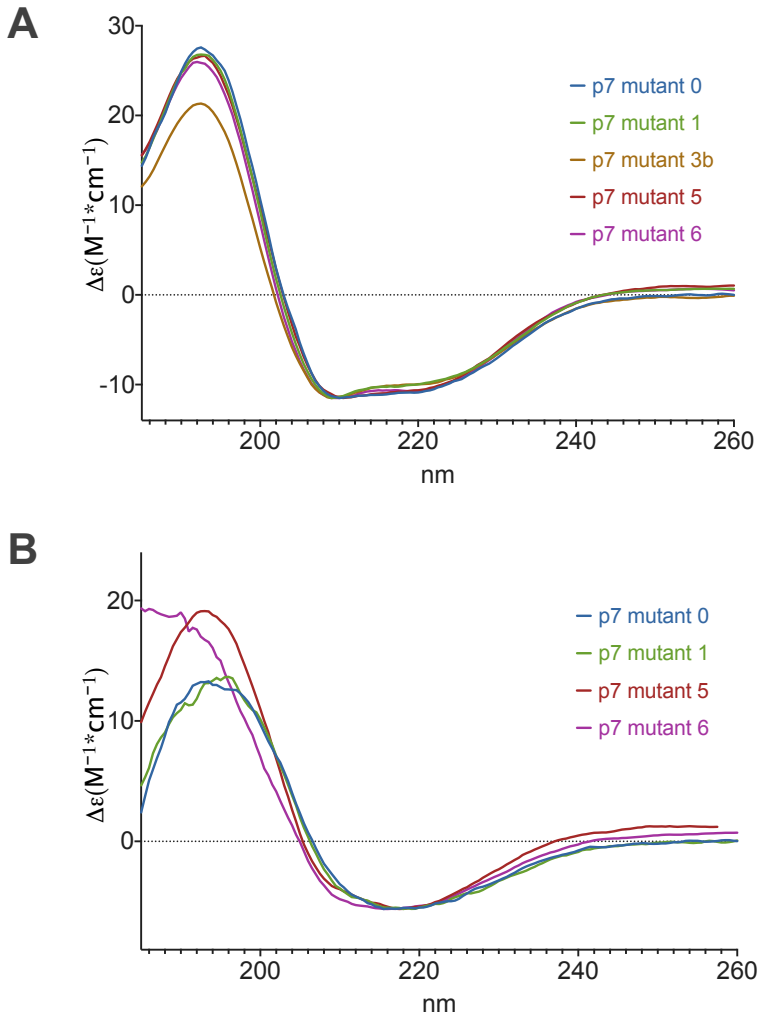


Figure 30: Circular dichroism of different p7 mutants before and after lipid reconstitution. (A) Spectra were measured using a 1 mm quartz cuvette. Protein solubilised in 4 mM SDS, 27 mM CF6 was mixed with 250times molar (hypothetical hexameric arrangement) excess of a DOPC:POPE (4:1, w/w) lipid mixture solubilised in 27 mM CF6. (B) Spectra from (A) after detergent removal using dialysis. Further information is provided in Table 14.

Table 14: DICHOWEB[265] secondary structure analysis of p7 solubilised in detergent-lipid and lipid only mixtures. Spectra analyses were undertaken with reference dataset SMP180[266] and the CDSSTR[267-269] algorithm. In general NRMSD[270] with values > 0.2 indicate a poor fit and should not be considered.

p7 mutant	Environment	α-helicity (%)	Strand (%)	Turns (%)	Unordered (%)	NRMSD	Corresponding Figure
0	detergent	89	1	4	4	0.006	Figure 30A
1	detergent	78	11	5	7	0.001	Figure 30A
3b	detergent	68	3	10	21	0.010	Figure 30A
5	detergent	82	2	5	11	0.004	Figure 30A
6	detergent	77	11	6	7	0.000	Figure 30A
0	lipid	54	21	8	18	0.002	Figure 30B
1	lipid	8	64	5	22	0.011	Figure 30B
5	lipid	67	12	6	14	0.002	Figure 30B
6	lipid	51	23	9	17	0.002	Figure 30B

After p7 reconstitution into lipid bilayers composed of DOPC:POPE (4:1, w/w) differential scattering and the absorbance flattening of the proteoliposomes limited the analysis of some of the samples. The use of a 0.1 mm path length cuvette wherever possible helped mitigate differential scattering effects.

The shape of the recorded spectra in Figure 30B show a marked change in p7 secondary structure compared to p7 solubilised in CF-6 (Figure 30A). The spectrum in membranes resembles more a β-sheet line shape but is missing classical antiparallel β-sheet characteristics normally found at 205 nm (dip) and 188 nm (peak)[267]. Evaluation by DICHROWEB resulted in an average of about 8 to 67 % α-helix, 12 to 64% stand, 5 to 9% turns and 14 to 22% unordered with NRMSDs < 0.011.

The reconstitution at different LPRs (i.e. 1000:1 and 250:1) seems not to have any influence on the p7 secondary structure. Both spectra show the same β-sheet-rich line shape in CD (data not shown), albeit scattering being clearly more pronounced for the higher LPRs, hindering secondary content quantification.

The recorded CD resembled two other cases in the literature where CD spectra in PC-rich environments were recorded[76,170]. In Whitfield *et al*.[76] the authors show that p7 embedded into DOPC:POPE gives rise to the same β-sheet-like spectrum while showing channel activity. A similar CD line shape was described by Fischer *et al*.[170] for the transmembrane domain of the Influenza B ion channel NB in POPC proteoliposomes.

Both – Fischer *et al*.[170] and Whitfield *et al*.[76] – omit to inform on CD spectra evaluation by software tools such as DICHROWEB. If this is due to structural features being underrepresented in the CD database at the time of their studies, or alternatively, the formation of microaggregates by the peptide which distorted the CD spectrum, is unknown.

In another study, using infrared spectroscopy (IR spectroscopy), an alternative, more β-sheet-like conformation of p7 in lipid membranes was reported by Gan *et al*.[171]. The authors could show using a liposome-based dye release assay, that p7 in a more β-sheet-like conformation is able to release dye, while lacking the ability to conduct protons. Gan *et al*. therefore suggest an alternative conformation of p7 with a more β-sheet-like secondary structure. To date it is not known whether this suggested alternative conformation is biological relevant or if it arises from the *in vitro* handling of the protein. Unfortunately, CD measurements of the very same IR spectra are not reported, making a direct comparison of the two techniques impossible.

The ability of p7 to adopt different conformations within its surrounding lipid environment had been hypothesized earlier, but structural measurements are still lacking[150].

The CD experiments show a strong change in p7 secondary structure, by going from a detergent- to a more lipid-based environment and confirm the need for additional structural studies in lipid bilayers. Especially in the context of the hypothesised alternative conformation suggested by Gan *et al.*[171] additional structural as well as functional information is needed to support the findings.

4.4.3. Continuous wave EPR

Continuous wave EPR reports on protein local dynamics via a probe attached with SDSL. In this study a common EPR spin probe, MTSL, was attached via disulphide bonds to p7 cysteine mutant side chains. For more information see section 2.1.

4.4.3.1. Spin labelling of p7

Before the p7 mutants could be reconstituted into lipid membranes of DOPC:POPE the protein had to be solubilised in order to carry out MTSL-labelling of the protein's sulfhydryl groups. The purified p7 lyophilisate was solubilised with a solution containing MTSL, urea and detergent (see section 3.3.1) to unfold the protein and allow immediate labelling. This is to prevent the formation of non-native disulphide bonds between p7 molecules, which otherwise tend to form rapidly. For a SDS-PAGE gel of different p7 single and double cysteine mutants solubilised in the presence and absence MTSL see appendix Figure A. 3. Refolding of the protein as well as removal of urea and unbound spin label was accomplished by extensive dialysis. Evaluation of the labelling efficiency was carried out using cwEPR by taking a defined volume of sample (10 µl, measured with a high precision Hamilton syringe) and recording the 1D field sweep spectrum on X-band, at 22ºC. Of note, all cwEPR experiments on lipids

were carried out at 35˚C, if not stated differently in this thesis. The second integral of the recorded spectrum, representing the number of spins within the sample, and its known protein concentration, in combination with a (2,2,6,6-tetramethylpiperidin-1-yl) oxidanyl (TEMPO) calibration curve were used to determine labelling efficiency (see Figure 32). Commonly observed spin labelling efficiencies for single cysteine mutants ranged between 92-99%. As mentioned above, when the single cysteine mutants are refolded in the absence of MTSL, these tend to form disulphide bonded dimers visible by SDS-PAGE under non-reducing conditions, but as expected, the labelled protein showed almost no dimers or higher order p7 oligomers (Figure 32). The double cysteine mutant p7 S21C_C27S_L51C (mutant 11 in Table 11) showed decreased labelling efficiencies at about 50% compared to single cysteine mutants. A possible explanation for this could be the formation of intramolecular disulphide bonds between the introduced cysteines at aa21 and aa51, “clipping” together the transmembrane segments. The absence of dimers or higher order oligomers on SDS - PAGE under non-reducing conditions does support this notion (see appendix Figure A. 3), and as both residues in the corresponding singly labelled mutants showed near 100% labelling efficiency, it is safe to assume that there is no singly labelled protein present. Importantly, this would imply that the EPR signal from this mutant would arise mostly from doubly labelled protein.

The high degree of p7 spin labelling was crucial for cwEPR as well as DEER experiments in lipid membranes. Low spin labelling efficiency would result in negative effects on the signal strength recorded and therefore contribute to a weak signal-to-noise ratio (SNR).

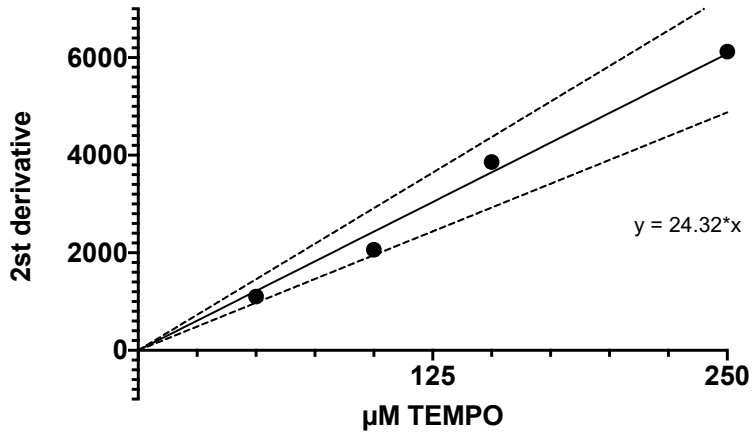


Figure 31: Standard curve of TEMPO for MTSL labelling efficiency determination. Shown is the relation of the 2nd integral of the recorded spectra over the TEMPO concentration. *Solid line* – linear regression, *dotted lines* – 99% confidence bands.

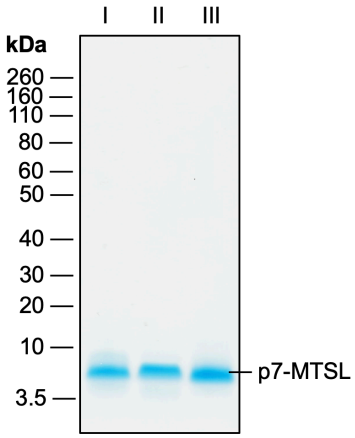


Figure 32: MTSL labelling efficiency of p7 single cysteine mutants on SDS-PAGE (non-reducing conditions). Lanes I, II and III show p7 J4 mutants 3b, 5 and 6, respectively.

4.4.3.2. Spin label dynamics are influenced by the lipid to p7 molar ratio

The cwEPR spectrum reports on the probe's local dynamics which can be influenced by environmental changes. For p7 in an oligomeric arrangement the spin label's local dynamics could change compared to its dynamics in a monomeric state. For example, packing of the monomers in the oligomeric arrangement could restrict the molecular

motion of the spin label resulting in a broader EPR spectrum. For an example see Figure 14.

As p7 has a low tendency to oligomerise in detergent[77], in order to interrogate the p7 oligomerisation behaviour the MTSL-labelled protein was incorporated into lipid membranes at different LPRs while monitoring the spin label's dynamic motions. If not stated differently all EPR spectra were recorded at 35˚C in order to stay above the phase transition temperature (T_m) of the lipids used (DOPC ~ -17˚C, POPE ~ 25˚C). Figure 33 shows the spectra of p7 spin labelled at the N-terminus (mutant 1) reconstituted in bilayers at increasing LPRs. The overall spectrum line shape shows at least two components: one slow-motion component (grey arrows) and one fast-motion component (blue arrows, see Figure 15 for a description of a two-component spectrum). The fast-motion component decreases from high to low LPRs, while conversely the slow-motion component increases. The reduction of the dynamics is more pronounced at LPRs higher than 500:1. Below this ratio the dynamics are changing little and differences between the ratios are less marked.

The presence of at least two components in all of the spectra could be explained by one, or a combination, of three different scenarios which could influence the spin label local dynamics: i) not all of the p7 molecules are in the monomeric state and some are forming oligomeric structures. In this case the sample would contain two different populations of p7 with different motion of the spin label side chain, due to the restriction of movement imposed by protein-protein interactions in the population of p7 molecules in an oligomeric state.

ii) the p7 sample contains only monomeric p7, but the monomers exist in two different conformations. One of the confirmations is compact and shows restricted spin label

motion. The other one is relaxed and thus shows a faster motion of the spin label side chain.

iii) p7 is in a single monomeric conformation, but the spin label side chain is in equilibrium between two local environments/motions and thus can flip between these two states of fast and slow-motion, resulting in a spectrum showing two different kinds of spin label dynamics.

Those spin label properties could contribute to the complexity of the spectra. The change in spin label motion cannot exclusively be explained by the formation of higher ordered p7 oligomers but as the protein is embedded in the lipid bilayer the only change is the LPR and thus the hypothetical distance between embedded proteins in relation to each other.

To gain more detailed information about the dynamic changes in the multi-component spectra of Figure 33, the spectra were deconvoluted using a two-component attempt by fitting a fast- (Figure 34, red trace) and a slow-motion component (Figure 34, yellow trace). The computational deconvolution utilises a generic global search[271] followed by a Levenberg-Marquardt local[272] fit. The results of the deconvolution are shown in Figure 34 and Figure 35, respectively.

Figure 34 shows an example of spectral deconvolution using spectral simulation (EasySpin[273], for more information see appendix Figure A. 4). The simulations yielded overall good RMSD values ranging between 0.009-0.016.

Figure 35 plots the change in slow-motion component as a function of LPRs: the analysis reveals that the change in dynamics fits well to an exponential decay. The change becomes less pronounced as the LPRs increase.

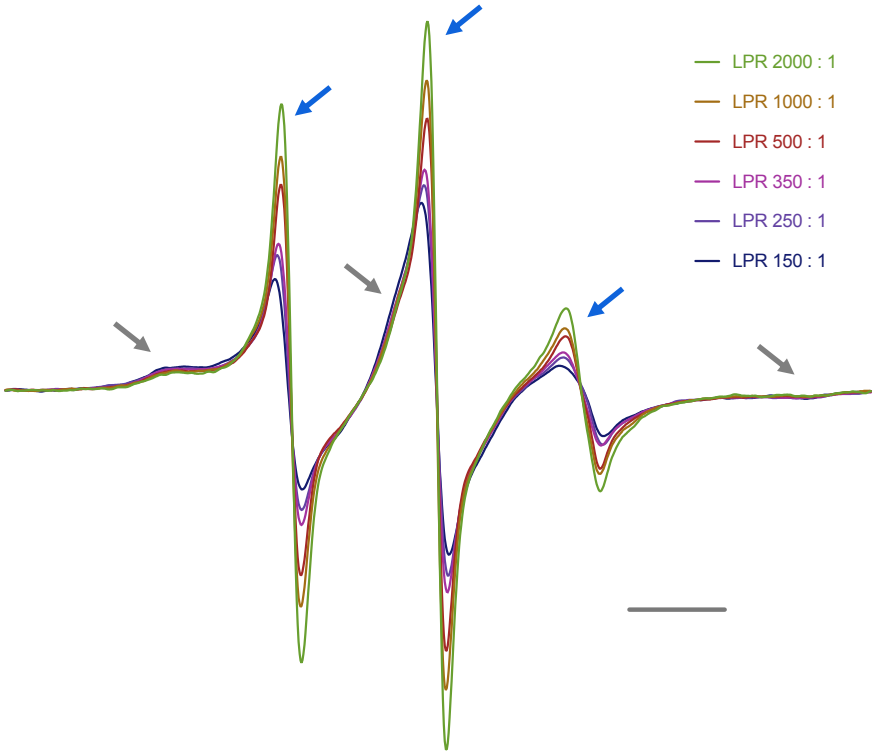


Figure 33: CwEPR 1D-field sweep of p7 J4 mutant 1. Shown are different protein to lipid molar ratios (LPRs) of MTSL-labelled p7 reconstituted into DOPC:POPE. Spectra are normalised to their second integral. Measurements were carried out at 35˚C. *Black bar* – indicates 100 Gauss. *Blue arrows* – spin label fast-motion component. *Grey arrows* – spin label slow-motion component.

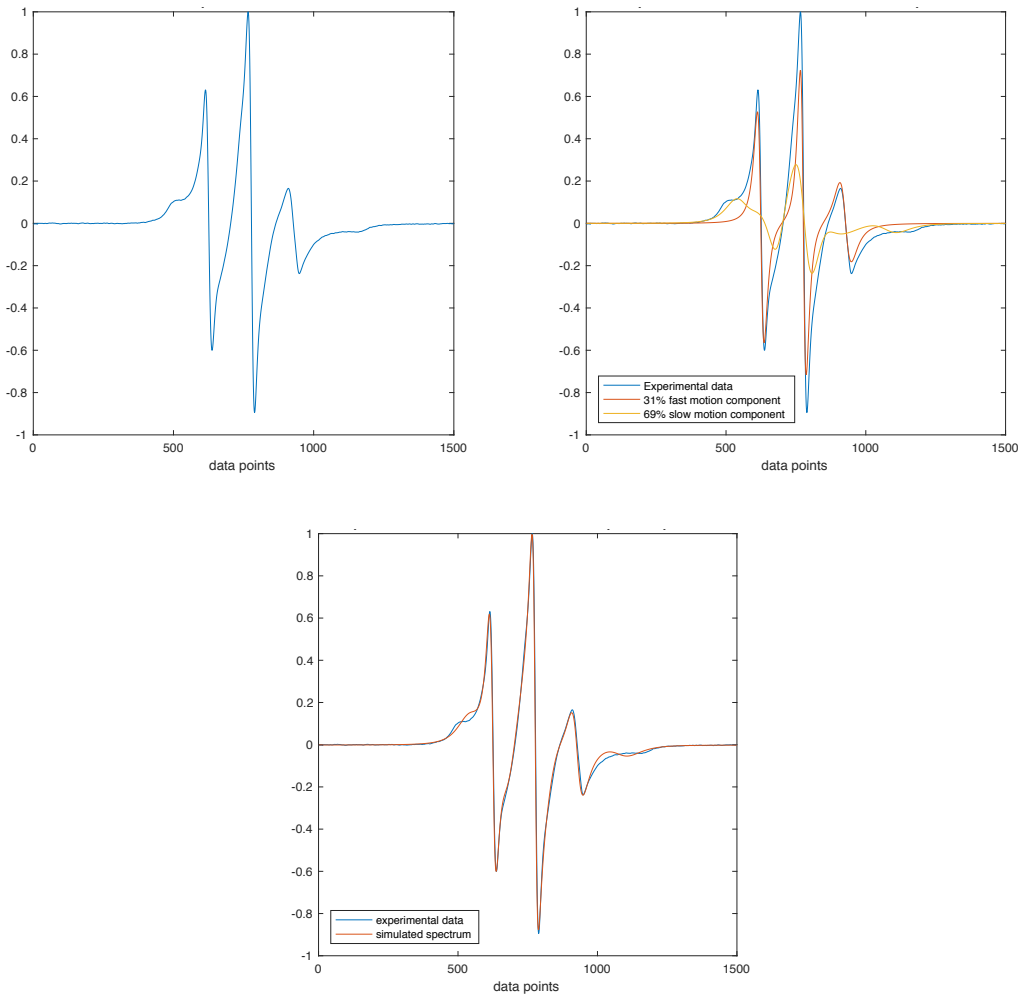


Figure 34: CwEPR deconvolution series of MTSL-labelled p7 mutant 1 at LPR 250:1. *Top left panel* – Multi-component spectrum of experimental data. *Top right panel* – Deconvolution of the experimental data (blue) into a fast- (red) and a slow-motion component (yellow). *Bottom middle panel* – Experimental data overlaid with the sum of the fast-/slow-motion components.

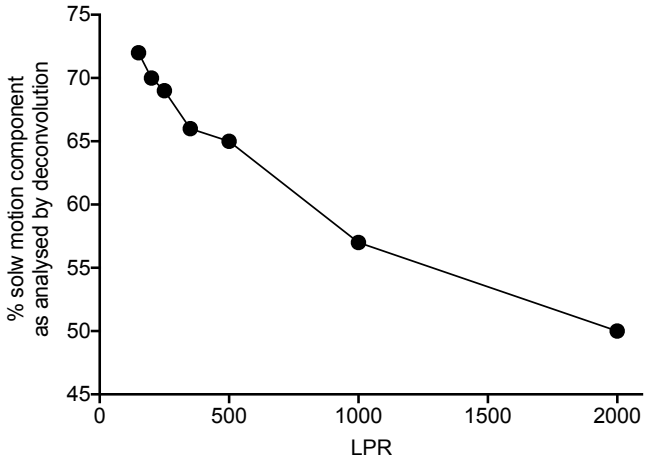


Figure 35: Evaluation of the cwEPR spectra deconvolution (Figure 34). Plotted are the percentages of slow-motion component over the different lipid to protein ratios.

4.4.3.3. Spin label dynamics at different positions within the p7 amino acid sequence

By attaching the spin labels at different positions within the p7 amino acid sequence it is possible to monitor the spin label's local environment to relate it to features of the p7 structure in membranes. The spin label can face a variety of local environments such as buffer solution, lipid membrane, the protein backbone and intra or inter molecular interactions with spins on other protein copies. Extensive NMR studies[77,164-166,168] have investigated structural and dynamic features of monomeric p7 in detergents and solvents, but the information is very limited in lipid membranes, in particular for the oligomeric form of the protein. The p7 single- and double-cysteine mutants described in section 3.1 were designed in such a way that three main regions of the protein were targeted: the C- and N-termini, the two transmembrane helices (TMHs) and the short loop region connecting the TMHs.

The spectra in Figure 36 represent five MTSL-labelled p7 cysteine mutants reconstituted into DOPC:POPE at LPR 250:1 (A) and 2000:1 (B). As described for

Figure 33 a lower LPR correlates with a decrease in fast-motion component and thus a putative higher order oligomeric arrangement of the p7 protomers.

Figure 36A shows p7 mutants 1, 3b, 5, 6 and 11 reconstituted into DOPC:POPE at LPR 250:1. The spectra follow the trend described for Figure 33 by showing a decreased content of fast-motion component when compared to higher LPRs (Figure 36B).

Under the assumption that the measurements are carried out with the majority of p7 in an oligomeric arrangement, p7 mutants 1, 3b and 11 show the same characteristic changes in spin label local dynamics as seen in B) with even stronger reduction in the fast-motion component, which could be caused by the formation of p7 ion channels.

The spin label attachment site of p7 mutant 1 is at the N-terminus (G-1_C-2_MTSL) which is expected to be highly dynamic, according to NMR data in detergent[164,166]. Furthermore, the termini were found to face the same side of the membrane surface, as analysed by epitope-labelled p7 in cells[148], which would expose the label to a more buffer-based environment, leading to fewer dynamic restraints caused by lipid or protein molecules in close proximity (Figure 36A, blue spectrum).

Compared to Figure 33 additional information is provided by the two single cysteine mutants 3b and 6. P7 mutant 3b (Figure 36A, green spectrum) is labelled at position 21, placing the nitroxide within the pore facing N-terminal transmembrane helix. The resulting spin label local dynamics are much more restricted compared to p7 mutant 1 (Figure 36A, blue spectrum) as reflected in an even further decreased fast-motion and an increased slow-motion component. The less dynamic spin probe at the position tested could be embedded in between lipids and α-helical protein backbone, or between monomers in an oligomeric arrangement. Taking this into account the

restricted motion of p7 mutant 3b is in very good agreement with the hypothesised secondary structure features of the protein in a lipid environment.

P7 mutant 6 (Figure 36A, red spectrum), which is the equivalent of mutant 3b but with the cysteine introduced into the C-terminal transmembrane domain of the protein, shows almost the same graph shape as mutant 3b, indicating a strong shift from fast- to slow-motion component potentially introduced by neighbouring lipid or protein molecules.

The double cysteine mutant 11 (Figure 36, orange spectra) carries two single cysteine mutations at position 21 and 51, respectively. Each cysteine is introduced into one of the two transmembrane domains and therefore should provide similar characteristics as found for p7 mutant 3b and 6 (Figure 36, green and red spectra, respectively). The cwEPR spectrum of p7 mutant 11 shows a low content of fast-motion component, which is in good agreement with the predicted p7 secondary structure, as previously described for mutants 3b and 6. Compared to those mutants at low LPR the spin label local dynamics overlay almost perfectly (Figure 36A), especially in the case of mutants 6 and 11.

The p7 mutant 5 covers the third secondary structural feature targeted – the loop region connecting the two transmembrane helices of the protein. Interestingly, the cwEPR spectrum of mutant 5, with the spin label incorporated at position 36, shows an intermediate signal between p7 mutants 1 and 3b (or mutant 6) in terms of spin label local dynamics. Several NMR studies[77,164-166] indicate residues ~33-37 in the monomer form a loop connecting the two transmembrane segments, while molecular dynamics simulations of p7 in bilayers suggest the loop region is at the water-protein interface[77,83,169]. The spin label's dynamics fit well with these notions. Located in a short loop, the spin label is expected to experience faster dynamics than in the context

of an alpha helix, but being at the lipid-water interface, its mobility could be more restricted through interactions with the lipid headgroups in comparison to the more water exposed, unstructured N-terminus.

Figure 36B shows p7 mutants 1, 3b and 11 at an LPR of 2000:1. Overall the spectra show a similar line shape as previously shown in A) but with an increase in fast-motion component and a parallel decrease in slow-motion component. This is in good agreement with the findings shown in Figure 33 and the hypothesized majority of the protein being monomeric, therefore the spin label's dynamics are less influenced by the formation of putative p7 oligomers (Figure 36A). For all mutants the overall dynamics are faster at an LPR of 2000:1, but the differences in relation to the different positions probed – termini, TMHs and loop region – remained the same, suggesting that the overall structural features are retained between a monomer on its own and a monomer involved in an oligomer.

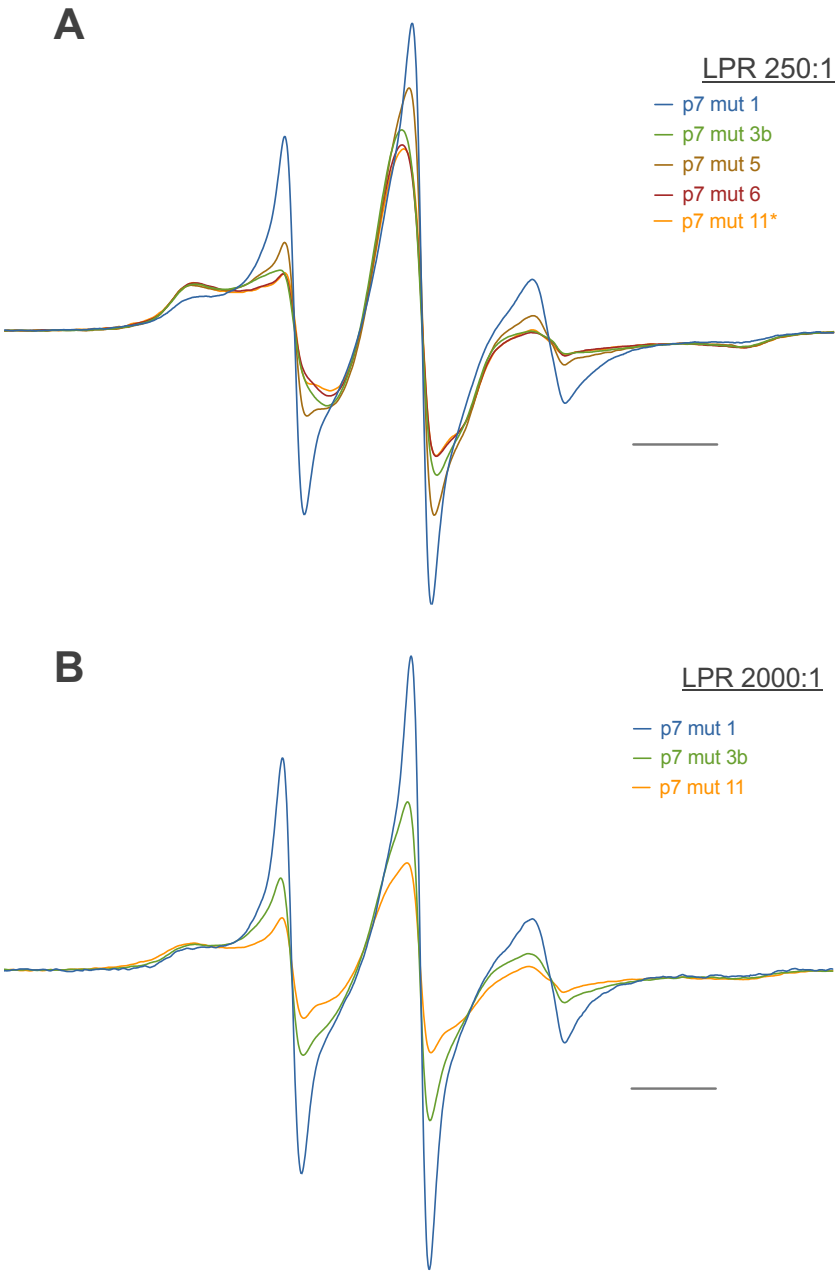


Figure 36: CwEPR 1D-field sweep of different p7 J4 mutants at defined LPRs. Shown are two different protein to lipid molar ratios of p7 reconstituted into DOPC:POPE (A) 2000:1 and (B) 250:1. Spectra are normalised to their second intrgral. Measurements were carried out at 35˚C. *Asterix* – p7 mutant 11 was reconstituted at LPR 150:1 *Grey bar* – indicates 100 Gauss.

The central linewidth (ΔH_0) of the cwEPR spectrum serves as a quantitative measurement of the spin label's segmental acyl chain motion. Figure 37 shows the measured central linewidth for each of the four p7 mutants measured at LPR 2000:1 and 250:1 (Figure 33A and B, respectively). Plotted are the central linewidths over the corresponding p7 cysteine mutation site. The results of the measurement are in agreement with the visual inspection and deconvolution carried out earlier and could represent a hairpin-like conformation of p7 in lipid bilayers. Although the LPRs of A) and B) differ by almost an order of magnitude, the acyl chain motion shows the same characteristics just less pronounced when going from high to low LPRs, indicating the unique acyl chain motion profile for each of the labelling sites chosen. This supports the hypothesis that the change observed in Figure 33 and Figure 36 is caused by a change in p7 oligomeric state (environment) and is not an artefact of the spin label acyl chain motion.

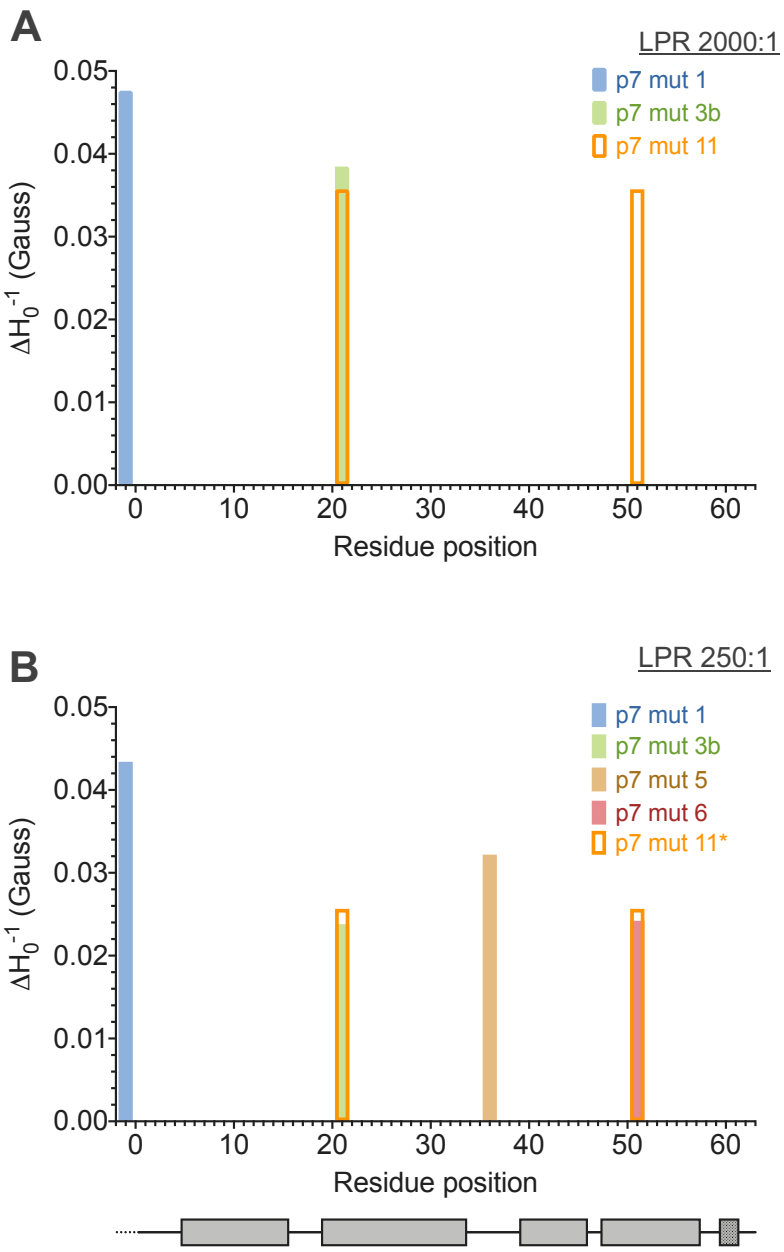


Figure 37: Central linewidth (ΔH_0) analysis of measurements carried out on p7 mutants 1, 3b, 5, 6 and 11 in DOPC:POPE at LPR 2000:1 and 250:1 (A and B, respectively) at 35˚C. Plot of inverse ΔH_0 over p7 J4 amino acid residue position with indicated secondary structure prediction below. Structure prediction is based on Oestringer *et al*. 2019[166] of monomeric p7 in CF-6. The values calculated for p7 mutant 11 are plotted as equal for both of its mutation sites, because it is not known how much each site contributes to ΔH_0. *Black line* – coiled, *grey square* – α-helix, *grey square with pattern* – partial α-helix, *dotted line* – undefined due to amino acid extensions introduced in this study.

4.4.3.4. CwEPR spectra in the frozen state – spin-spin distance

Frozen cwEPR data were recorded on p7 mutants 3b and 6 at constant LPR with changing spin dilutions to analyse spectra broadening due to Heisenberg spin exchange (see section 2.1). To suppress sample mobility contributions to the spectra broadening experiments were carried out at cryogenic temperatures (100 K). The method requires measuring of the same sample under conditions with and without spin-spin interaction in order to distinguish spectra broadening from MTSL background contributions. To address this, background samples were prepared at 14% MTSL spin dilution, aiming for just one p7 monomer being labelled in a hypothetical hexameric arrangement of the protein. Figure 38A shows the spectral line broadening of fully labelled p7 mutant 3b compared to a spin diluted sample. The differences between the spectra are not very pronounced but clearly visible, which leads to the conclusion that some of the cwEPR spectra broadening is caused by spin-spin interactions due to radicals being as close as 1.8 nm to each other. This is in good agreement with computational models that predict the N-terminal TMH to constitute the channel pore and thus place p7 mutant 3b into close proximity to labelled sites of the same kind. In comparison Figure 38B shows cryo-cwEPR data of p7 mutant 6, which carries its spin label attachment site in the second, C-terminal TMH. The differences in spectra broadening seem to be similar to the ones observed for mutant 3b, even though the broadening is more pronounced at the dip of the mid field absorbance peak. Why this difference in broadening occurs is yet to be determined. It could be that difficulties setting the baseline of the 14 % spin labelled sample slightly misaligned the spectrum compared to that of the 100 % labelled sample. The sample is spin diluted to an extent that the SNR during sample recording was still acceptable but considerably lower than the SNR of samples with a

higher percentage of spin labelled protein. With low SNR small errors in baseline correction during data processing become more significant. Another possibility which would explain the differences in spin-spin interactions observed is the orientation of the labelling sites with respect to each other.

Overall the observation of weak spin-spin interactions for both labelling sites is in agreement with the finding that p7 mutant 11 does not form oligomers due to intermolecular disulphate bond formation, as analysed by non-reducing SDS-PAGE (see appendix Figure A. 3). This indicates that the residues of mutants 3b and 6 are in close proximity to each other and favour the intramolecular formation of disulphide bonds ('clipping') between the introduced cysteines in mutant 11.

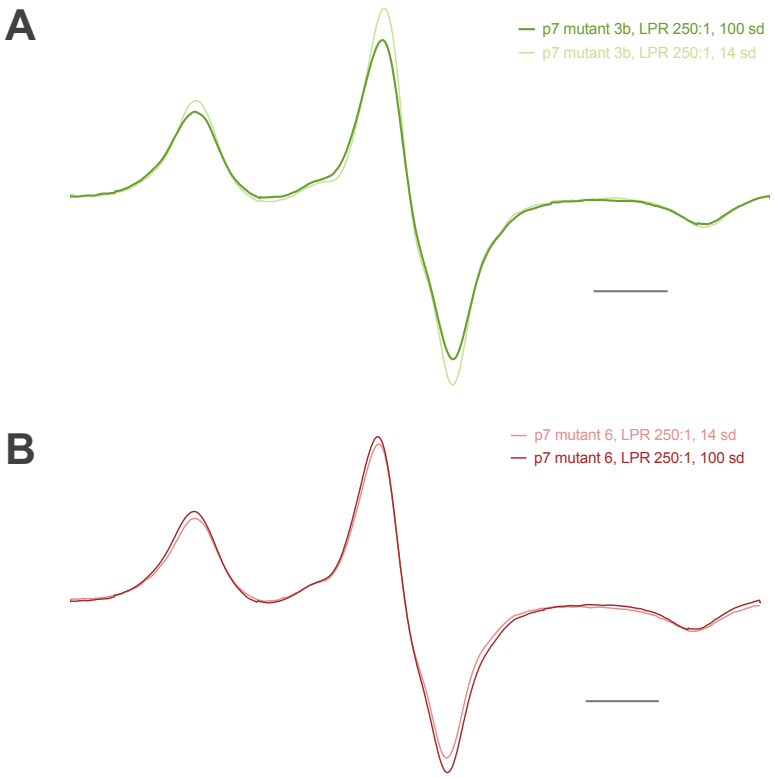


Figure 38: CwEPR 1D-field sweep of p7 J4 mutant 3b (A) and 6 (B) at 100 K. Measurements carried out for the detection of Heisenberg spin-spin interactions. Protein reconstituted in DOPC:POPE (w/w) at LPR 250:1. Spectra are normalised to their second integrals. *sd* – MTSL spin dilution. *Grey bar* – indicates 100 Gauss.

4.4.3.5. Conclusions from cwEPR analysis

Overall it was possible to MTSL-label, record and analyse cwEPR spectra of five different p7 single- and double-cysteine mutants reconstituted into DOPC:POPE at varying lipid to protein molar ratios. The analysis showed that a decrease in LPR can be correlated to an increase in spin label local dynamics restraints. The change in spin label local dynamics is hypothesised to be introduced by the formation of p7 oligomeric arrangements within the lipid bilayer. The hypothesis is supported by FRET experiments carried out by Dr Marc-Philip Pfeil (former member of Professor Anthony Watts' group, Department of Biochemistry – Oxford University) which indicate a hexameric arrangement of fluorophore labelled p7 reconstituted into DOPC:POPE bilayers at LPR 150:1 (for more information see appendix Figure A. 6, data unpublished).

Analysis of the central linewidth of different p7 mutants reconstituted at LPR 250:1 showed that the spin label local dynamics are influenced by its position in the p7 sequence. NMR-based structure and dynamics analysis in detergent of different monomeric p7 genotypes in solvent or detergent[77,164-166] seem to be in good agreement with spin label motion dynamics found for four different labelling sites – N-terminus, transmembrane helices and loop region – of p7 J4 reconstituted into lipid bilayers. Furthermore, a solvent paramagnetic relaxation enhancement (PRE) approach on p7 genotype 5a in DMPC/DHPC bicelles shows a profile that would fit nicely to a hairpin-like conformation of the protein[254] even though the authors' interpretation of their data differs completely from the one presented here. The reconstitution of p7 into bicelles is the closest environmental mimetic to a lipid bilayer published at the time of writing, although the authors did not present evidence that the bicellar structure is retained after incorporation of p7.

4.4.4. Double electron-electron resonance evaluation of p7 in lipid bilayers

As previously described, the main aim of this study is to analyse p7 oligomeric arrangements in lipids. The reconstitution of membrane proteins into lipid bilayers is a difficult task and requires much more time and effort than working with soluble proteins. Furthermore, the prepared samples are not that amenable to structural analysis, as most of the standard techniques like protein X-ray crystallography, NMR, CD, etc. can either not, or only to a very limited extent, be used for the analysis of membrane embedded proteins.

In combination with SDSL, DEER can provide distance measurements between paramagnetic centres within a protein or other nanoobjects. The technique is not limited by the size of the protein to be analysed and can be applied to proteins reconstituted in bilayers. Therefore, it is a powerful tool to shed more light on the p7 ion channel arrangement in DOPC:POPE mixtures. However, the application of DEER to membrane proteins in lipid membranes remains one of the greatest challenges in the EPR field. Studies of the ABC transporter MsbA in bilayers have highlighted that the two-dimensional (2D) disposition of proteins embedded in a liposome strongly limits the distance measurement between coupled spins to maximal 4-5 nm, compared to a distance limit of up to 7-8 nm for proteins in a homogenous three-dimensional (3D) spatial distribution in solution[274]. This is because the background factor from proteins distributed in a 2D space dominates the change in the second echo intensity in the DEER experiment. Additionally, the higher local concentration of protein reduces the second echo intensity, further limiting DEER measurements[231,234,275]. Finally, proteins containing more than two spin pairs, or single labelled proteins in a higher (2+i)

oligomeric arrangement complicate the DEER signal. These multi-spin systems cause signal contributions from difference and sum combinations of dipolar frequencies. Common tools such as the DEERAnalysis package[241] are specifically designed for single spin pair evaluation and thus are error driven by multi-spin systems[276]. Using those software packages limits distance measurements to the shortest distance obtained[277]. All these limitations need to be addressed with careful sample design: protein local concentration and spatial distribution as well as multi-spin effects can be fine-tuned by varying the LPRs and by spin dilution (mixing a specific proportion of labelled and unlabelled protein in an oligomeric system). An alternative approach is to vary the pump pulse efficiency to limit DEER[238,278].

Other important considerations in terms of experimental setup and additives for sample preparation are listed here:

i) *Frozen state measurements* – All DEER measurements in this study were acquired at 50 K under the use of liquid helium. Samples contained up to 20 % ethylene glycol as cryoprotectant to minimise the formation of crystals while freezing. The temperature chosen is a compromise between signal intensity (echo strength) and the length of measuring time (pulse sequence repetition time). Cryogenic temperatures affect DEER measurements in three different ways: 1) Restriction of spin label dynamic motions and thus a narrower distance distribution can be achieved. 2) T_1- and T_2- relaxation times increase, which allow longer measuring times that permit the resolution of weak dipolar coupling between longer distance spin labels. 3) The energy difference in spin state in the Boltzmann

distribution increases and therefore enhances the recorded echo intensity[279].

The extension in spin relaxation time implies, conversely, that the number of measurements within a given time frame decreases due to the longer time it takes to perform a measurement (see shot repetition time, SRT, in 'iv')[279].

ii) *Suppression of nuclei induced magnetisation relaxation* – Due to unwanted spin-spin interactions between electrons and close nuclei, samples were enriched with up to 70% deuterated water (D_2O, nuclear spin I = 1, magnetic moment 3.5 times less than that of protons). D_2O has a lower geomagnetic ratio compared to protons and thus causes a reduced magnetisation relaxation of the paramagnetic centres. This allows longer T_2-times to be recorded, its effect is listed under (2) in (i) above.

iii) *Avoidance of nuclear modulation artefacts* – Under the most sensitive DEER conditions (i.e. strong signal with little SNR) for a given sample the slight overlap of observer- and pump pulse causes the pump pulse to excite unintentional electron-nuclear transitions of a given A spin which causes artefacts in the DEER signal. The most prominent signals arise from near nuclei or deuterium and result in peaks between 1.75 and 2.0 nm in DEER distance distribution. Due to the advantages to signal improvement described in (ii) the presence of deuterium could not be avoided. In order to suppress the nuclear modulation an inter pulse delay with 8-steps and $\Delta\tau_1$ = 16 ns and subsequent averaging of the recorded DEER traces were implemented in the experiment[234,280].

iv) *Shot repetition time* – SRT is a time delay between subsequent DEER pulse trains. For a given time frame longer SRTs allow fewer measurements to

be averaged and therefore decrease the SNR. Shorter SRTs will provide better SNRs, but they need to be long enough to allow the bulk magnetisation to return to equilibrium in between pulse trains. SRTs in this study were calculated individually and based on SRT = 1.5 x T_1.

v) *Suppression of echoes introduced by coherent transfer pathways* – The use of arbitrary waveform generators (AWGs) allows the generation of very precise microwave pulses with defined shapes while pump- and observer pulse are in coherence with each other. Due to the slight pulse overlay described in iii) and the resulting artefact, which moves along the observer frequencies with shifting pump-pulse times ('t' in Figure 17C), a generated artefact at the position of the refocused Hahn echo cannot be excluded. Thus a phase cycle for pump- and observer pulses has to be used in order to remove unwanted artefacts. The complexity of the DEER pulse trains require a minimum of 16 phase cycle steps in order to remove unwanted artefacts sufficiently[281].

4.4.4.1. Inter molecule distances and protomer arrangements by DEER

Figure 39 to Figure 43 show DEER experiments on fully MTSL-labelled p7 samples. Figure 39 shows DEER measurements of the four p7 single cysteine mutants 1, 3b, 5 and 6 reconstituted into DOPC:POPE (4:1, w/w) bilayers. All samples are at an LPR between 150 and 250:1, for which FRET and cwEPR analysis indicated that p7 is largely in an oligomeric arrangement. The evaluation was carried out using the software package DeerAnalysis2019[241,282]. The reliability range (background colours: green, yellow, orange, red in Figure 40 and Figure 42) given for each of the distance

distributions obtained is mainly influenced by the dipolar evolution time, which tends to be shorter for samples in lipid bilayers.

Subtraction of the background from the dipolar evolution function was successful, with dimension values (D, see section 2.2, equation 8) ranging from 2.0 to 2.27 and RMSs for the background fit ranging from 0.0011 to 0.0023. In all cases, the resulting form factors start with a sharp decay, indicative of a very short distance and modulation depths ranging from 0.063 to 0.182 (Figure 39A/C/E/G and Figure 41A/C/E). After the first sharp decay, oscillations of small amplitude follow, as expected for multi-spin systems[276]. Tikhonov regularisation could successfully be used to fit F(t) resulting in good distance distributions for all p7 mutants (Figure 39B/D/F/H and Figure 41B/D/F). As indicated above, the presence of multi-spin effects means that only the shortest distance can be extracted reliably[276]. All mutants present a very short inter-spin distance of ~1.7 nm. For p7 mutant 1 the peak corresponding to this 1.7 nm distance has a defined shoulder at + 0.25 nm (Figure 39B) and is broader compared to those of the other mutants in Figure 39. For an overlay of all mutants and ease of comparison see Figure 40. The width could be explained by the positioning of the attached spin probe at the N-terminus of the p7 sequence and its high degree of dynamic motion, as previously described by cwEPR (see section 4.4.3.3).

When having the spin label attached within the N-terminal TMH of the protein (p7 mutant 3b, Figure 39C and D) the corresponding distance distribution shows similarities to p7 mutant 1 for shorter distances (< 3 nm) but differs significantly for distances above this value. The last distance at 4.1 nm could be an artefact introduced by an unresolved oscillation towards the end of the evolution function (Figure 39C). Limitations in the spin relaxation time prevented further evaluation.

The p7 mutants 5 and 6 (Figure 39E/F and G/H, respectively) show a very dominating short distance at about 1.7 nm and up to five low intensity peaks at distances > 2 nm. The low intensities for the longer distances get reflected in the corresponding evolution function (Figure 40E and G, respectively) by an almost flat line shape at > 0.25 µs.

The shortest distances found for mutants 3b and 6 using DEER are in agreement with cryo-cwEPR data of the very same mutants described in section 4.4.3.4. A study conducted by Banham *et al*. 2008[283] describes the sensitivity limits for both techniques by evaluating the inter-spin distances obtained from several double nitroxide-labelled peptides in the range from 1.0-4.5 nm. The authors show that the upper detection limit of cwEPR spectra broadening and the lower limit of DEER distance evaluation by Tikhonov regularisation is around 1.6-1.9 nm, respectively. This puts the shortest distances obtained by DEER in this study perfectly within the higher/lower limit of the techniques used. Banham *et al*. come to the conclusion that distance evaluation can still be reliably obtained by using both techniques in combination.

For ease of visualisation the distance distributions of p7 mutant 1, 3b, 5 and 6 at LPR 150-250:1 are overlaid in Figure 40.

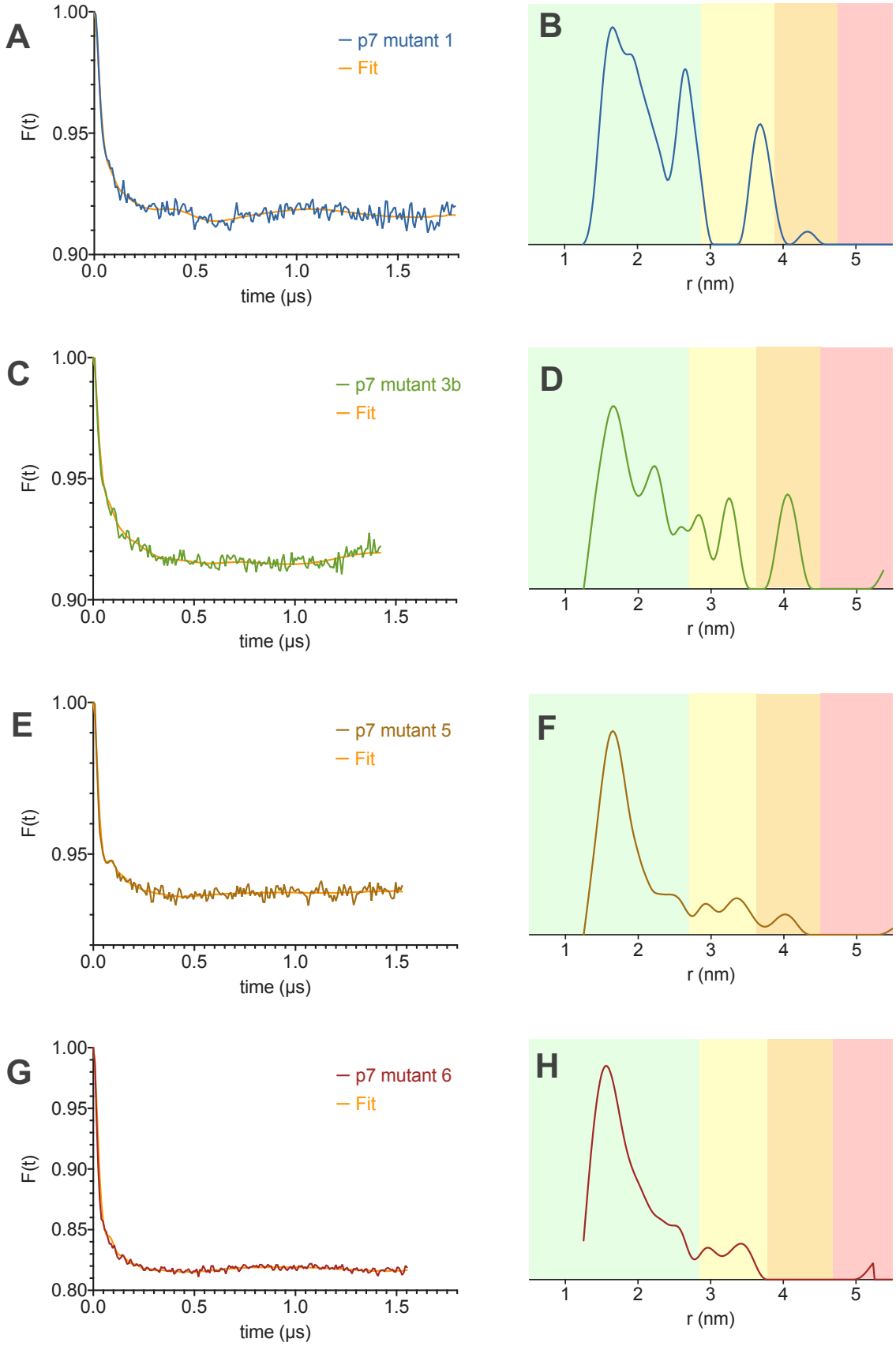


Figure 39: Q-band 4-pulse DEER measurements of fully MTSL-labelled p7 mutants 1, 3b, 5 and 6 reconstituted into DOPC:POPE (4:1, w/w) mixtures at LPR 250:1 (with the exception of p7 mutant 1

which is at an LPR of 150:1). (A/C/E/G) DEER form factor (F(t)) after background correction with corresponding trace fit (orange) by Tikhonov regularisation. (B/D/F/H) Obtained distance distribution calculated on the basis of A/C/E/G, respectively. Background colours indicate peak reliability[282]; *green* – distance distribution is reliable, *yellow* – width and mean distance are reliable, *orange* – only mean distance is reliable and *red* – peaks are unreliable.

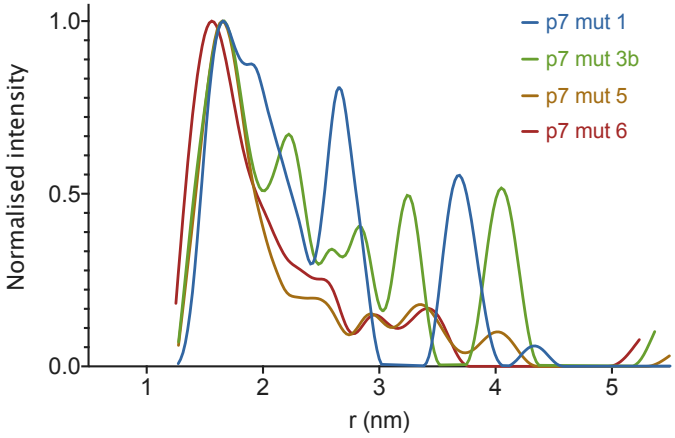


Figure 40.: Overlay of DEER distance distribution obtained in Figure 39. All MTSL-labelled p7 mutants were reconstituted in DOPC:POPE (4:1, w/w) at LPR ranges between 150 and 250:1.

DEER measurements were also carried out on fully labelled p7 mutants 1, 3b and 6, reconstituted at an LPR of 1000:1 or 2000:1 (Figure 41), where there is a lower degree of oligomerisation, according to the FRET and cwEPR data. The presence of monomeric protein is not expected to contribute to the form factor (F(t)), and oligomers of lower order are expected to present less pronounced multi-spin effects[276]. The presence of DEER signal in these samples indicates that oligomers are still present at these ratios. Multi-spin effects are still apparent with non-marked oscillations after the first sharp decay in the background subtracted DEER time trace. Except for p7 mutant 1, the Tikhonov regularisation results in a spin-spin shortest distance of ca. 1.7 nm, similar to that of the 150:1/250:1 LPR samples (Figure 39). The p7 mutant 1 shortest distance was truncated and could not be verified by TR. This might indicate the presence of shorter distances, but it is not clear why this would be the case at the

1000:1/2000:1 LPRs and not in the 150:1/250:1 sample. One possibility is that in the context of lower order oligomers, the dynamics of this mutant are more pronounced, giving a broader distance distribution that reaches shorter distances below the DEER detection limit.

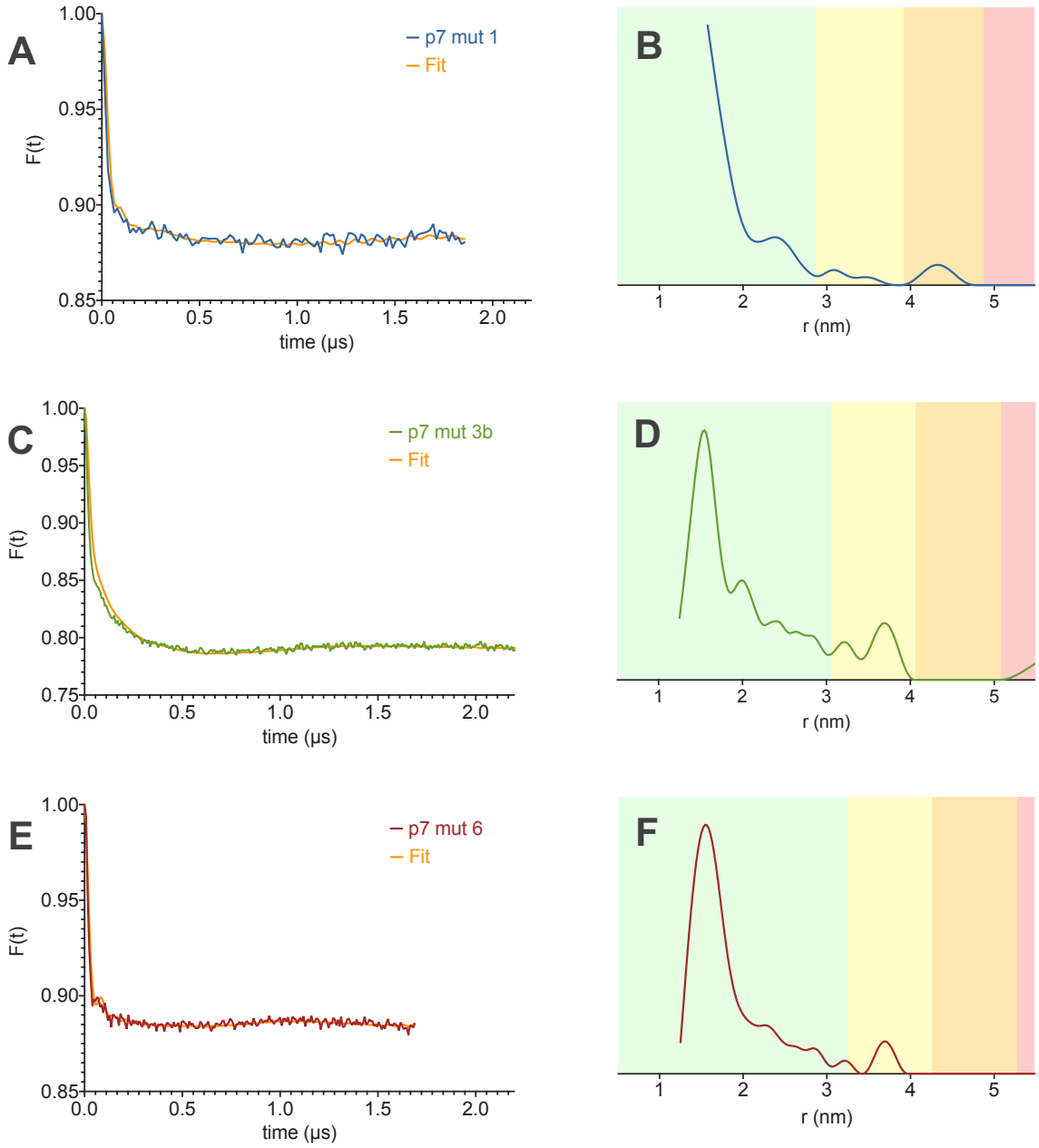


Figure 41: Q-band 4-pulse DEER measurements of fully MTSL-labelled p7 mutants 1, 3b, 5 and 6 reconstituted into DOPC:POPE (4:1, w/w) mixtures at LPR 1000:1. (Exceptions: Mutant 1 at LPR 2000:1, mutant 6 33% spin diluted). (A/C/E) DEER form factor (F(t)) after background correction with corresponding trace fit (orange) by Tikhonov regularisation. (B/D/F) Obtained distance distribution

calculated on the basis of A/C/E, respectively. Background colours indicate peak reliability ; *green* – distance distribution is reliable, *yellow* – width and mean distance are reliable, *orange* – only mean distance is reliable and *red* – peaks are unreliable[282].

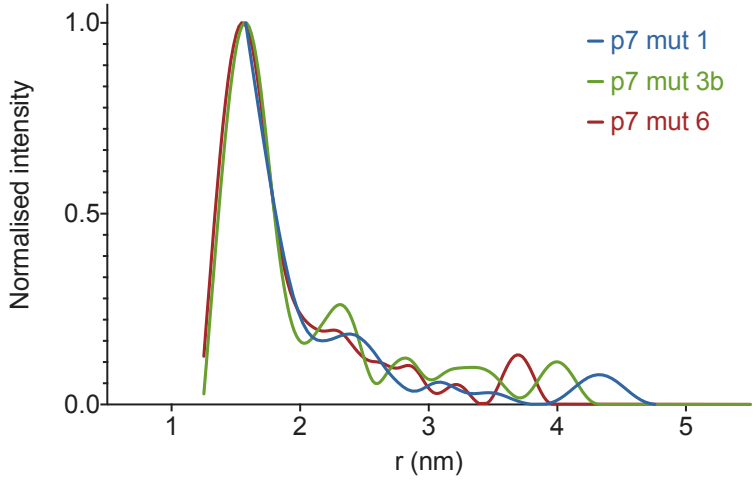


Figure 42.: 4-pulsed DEER measurements of MTSL-labelled p7 in DOPC:POPE (4:1, w/w) at LPRs range between 1000 and 2000:1 (A-B)

The distances from measurements at LPRs 150:1/250:1 and 1000:1/2000:1 are compared individually for p7 mutants 1, 3b and 6 in Figure 43. The peak corresponding to the shortest distance has slightly shifted to a shorter distance for the LPR 1000:1/2000:1 samples. Multi-spin effects change by varying the LPRs, but analysis does not improve at the higher LPRs; in contrast with the reproducible shortest distance, the longer distances vary markedly.

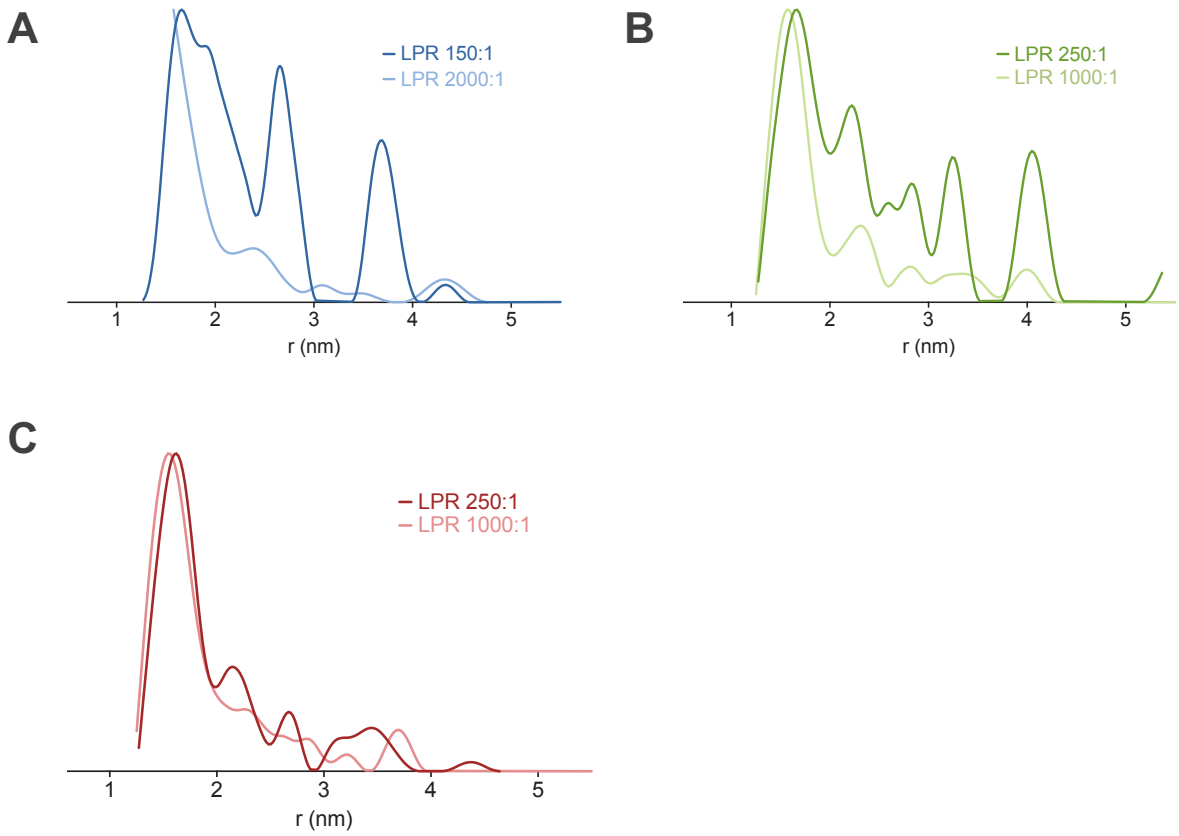


Figure 43.: 4-pulsed DEER comparison of MTSL-labelled p7 in DOPC:POPE (4:1, w/w) at high and low LPRs. The figure shows the distance distribution of p7 mutant 1 (A), mutant 3b (B) and mutant 6 (C) reconstituted at the LPR indicated. Individual graphs are internally normalised to their shortest distance intensity.

Figure 44A and B illustrate the hypothetical distance distributions and corresponding peak intensities for four p7 oligomers under ideal conditions. Up to a heptameric arrangement (Figure 44B, red graph) only three different distances should be detectable. The extensive study from Giannoulis *et al.*[276] simulated how multi-spin effects could distort distance distributions and this is illustrated in Figures 28C and D. The distortions are comparable to those observed in our samples.

In order to suppress the influence of multi-spin effects under the conditions tested experiments were conducted on samples of MTSL-p7 mixed with unlabelled p7 (p7 mutant 0, see Table 2). Based on a hypothetical hexameric arrangement of p7 in lipid bilayers a biradical system in which two out of six p7 monomers are MTSL-labelled (33% spin dilution) was attempted (for more information see section 2.2). Due to the uncertainty of the p7 oligomeric arrangement and the Gaussian distribution of MTSL-p7 within the formed oligomers, only ~ 32% (mean 2 ± 0.5) of the sample's hypothetical hexamers exhibit a biradical-like composition, leaving ~ 37% for higher spin arrangements. To further reduce the influence of the latter, spin dilutions of 14% were examined as well. These resulted in very weak echo intensities and a decrease in SNR which turned out to be unusable for further processing (data not shown).

Figure 44E and F compare DEER experiments of p7 mutant 6 reconstituted at LPR 250:1 with no- and 33% spin dilution. Despite spin-dilution, after the first sharp decay, the time trace shows nearly an identical form factor as observed for the not spin diluted sample, with oscillations remaining cancelled out from multi-spin effects. Additionally, the modulation depth decreased with decreasing number of spins interacting (Figure 44E). This made subtraction of B(t) from dipolar evolution function D(t) more difficult and similarly spin diluted samples prepared for mutants 1, 3b, 5 and 11 at varying LPRs resulted in insufficient fits of TR (not shown). Therefore, Figure 39 to Figure 43 only show samples with a good fit of the background function (D between dimension 2-2.5) and TR RMSs of < 0.004.

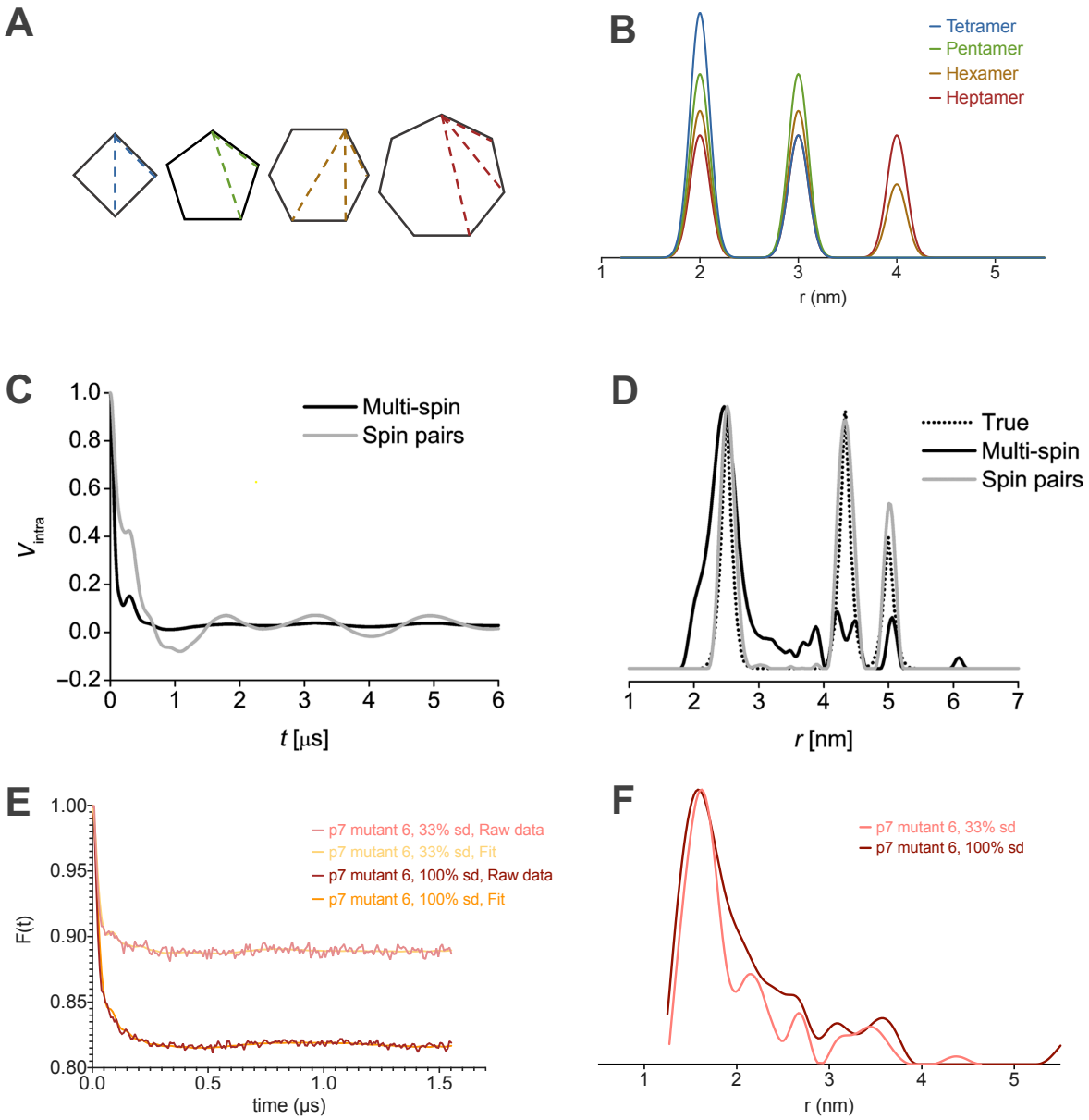


Figure 44.: (A) Schematic of four different oligomers, from left to right: tetramer, pentamer, hexamer and heptamer. Dotted lines indicate distances between hypothetical spin label positions. (B) Simulated distance distribution and corresponding peak intensities expected of oligomers shown in A). (C-D) Example of dipolar evolution function distortion and corresponding distance distribution caused by multi-spin effects. Figure adopted from Giannoulis *et al*. 2013[276] (E) DEER time trace of p7 mutant 6 at two different spin label dilutions (33 and 100%). Shown is the change in modulation depth due to different numbers of spins interacting. (F) DEER distance distribution of E). Evaluation of both graphs took place with regression parameter set to 3.98E+1.

Studies based on gadolinium (Gd^{3+}) containing spin labels such as the chelating complex Gd^{3+}-DOTA (1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) have been proposed as an alternative to nitroxide based systems[284-286]. The use of metal-ions as coupled spins, replacing nitroxides, could help to totally eliminate multi-spin artefacts. Gd^{3+}-DOTA-maleimide was tested for labelling purposes of p7 but resulted in degradation of the maleimide during activation of the chelating complex with $Gd^{3+}Cl_3$. Due to time limitations further investigations had to be postponed.

For the p7 samples presented here dipolar evolution times were limited to about 1.7 µs on average, which limits reliable distance evaluations down to about 3 nm. In order to obtain longer dipolar evolution functions of p7 reconstituted into lipid bilayers a so-called "stitch-approach" first described by Lovett *et al.* 2012[287] could be used. The method makes use of the eliminated dead-time problem in 4-pulse DEER (described in section 2.2) and combines it with the high sensitivity at longer dipolar evolution times of 3-pluse DEER by mathematically 'stitching' both spectra together. However, due to the existence of strong multi-spin effects of MTSL-labelled p7 in an oligomeric arrangement the use of a 'stitch-approach' was abandoned as longer distances are error-prone due to differences in their evolution time frequencies.

4.4.4.2. Comparison of DEER shortest distance

Due to the shortest DEER distances being the only reliable distances an attempt was made to compare the distances obtained in this study with distances calculated from the p7 oligomer model predicted by Montserret *et al.*[77] and the previously described NMR structure of Ouyang *et al*.[168] (PDBID: 2M6X, section 4.3).

MMM[231,282], a spin label rotamer prediction software package for MATLAB, was used to determine the shortest inter- or intramonomer distance possible within the structures

provided. Due to the N-terminal labelling position of p7 mutant 1 and the short linker introduced (G0_C-1) distance comparison were only carried out on mutants 3b, 5 and 6. Within MMM the structures of Montserret *et al.*[77] and Ouyang *et al.*[168] were MTSL-labelled *in sito* at corresponding positions and their spin label rotamers calculated considering structural and physical restraints of the surrounding protein backbone. Figure 45 shows the overlay of the resulting distance distributions with DEER distances obtained in this study.

The evaluation shows high similarities between all two structures for the labelling position of mutant 5 (Figure 45C and D). For mutant 3b (Figure 45A and B), the predicted shortest distance seems to be even slightly shorter than the one obtained in this study. And the experimentally determined distance of mutant 6 (Figure 45E and F) is about 1 nm shorter than the MMM predictions.

In general model and structure used in this evaluation show very similar distance calculations, even though their predicted mono- as well as oligomeric arrangement is very different. For more information regarding this topic see section 1.2.2. The appearance of multiple peaks in the predictions is caused by two or more favourable rotamer states of the attached spin label and the fact that MMM is not being able to take spin label local restraints introduced by surrounding lipids into account.

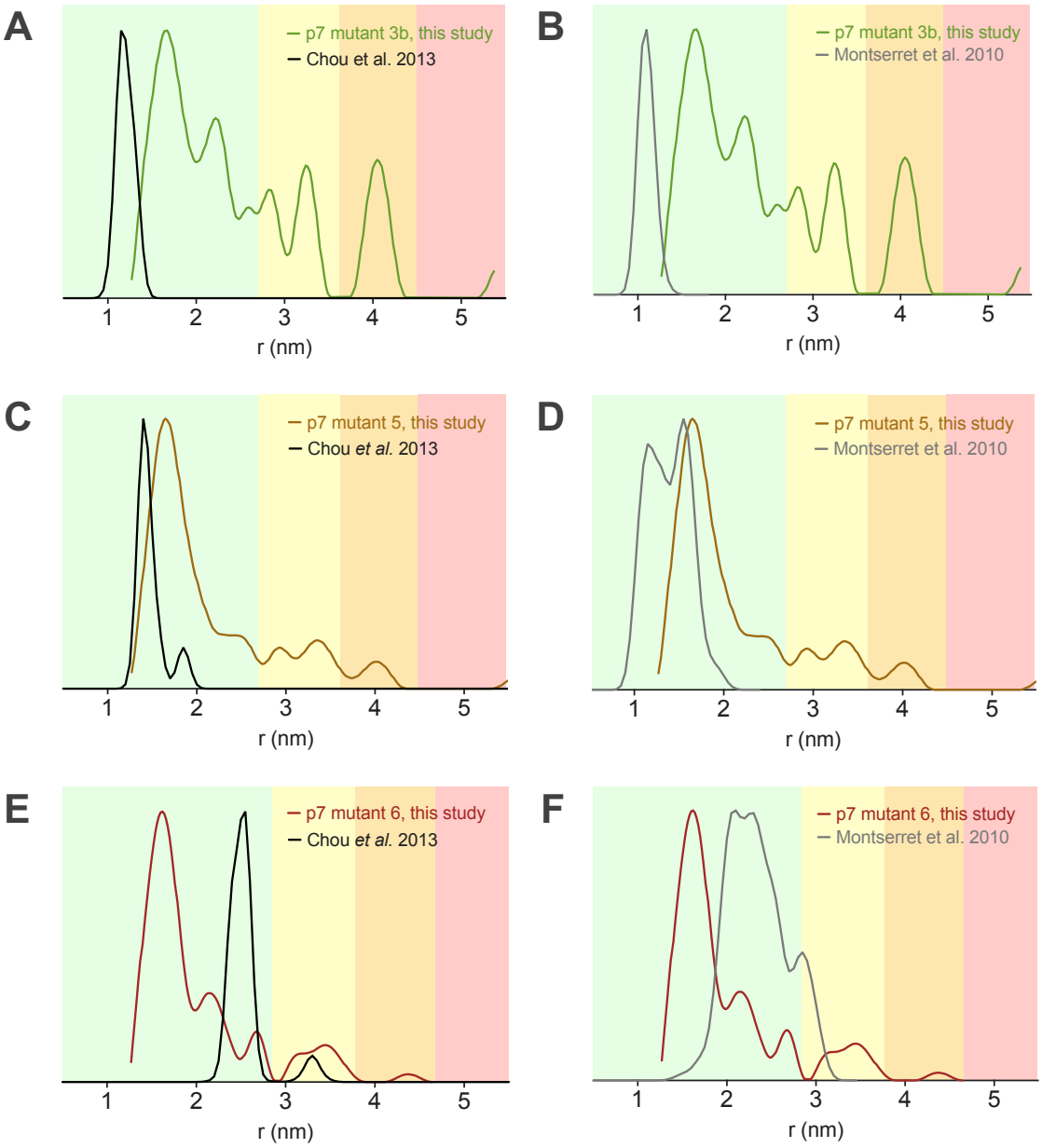


Figure 45.: MMM[231] distance predictions based on the p7 oligomer model published by Montserret *et al.*[77] and the unusual NMR structure by Ouyang *et al.*[168] (PDBID: 2M6X) in comparison to experimental DEER distances obtained in this study. Spin labelling and rotamer calculations were carried out *in situ* using MMM. Shown is only the comparison between the shortest inter-/ intramonomer distance posible due to reliability restraints for longer distances in this study.

4.4.4.3. 4-pulse DEER on MTSL-spin labelled p7 – Conclusion

4-pulsed DEER in combination with side-directed spin labelling has proven to be a powerful tool for distance determination. The technique is capable of analysing distances between biradical spin systems while being little influenced by the presence of lipid bilayers. DEER measurements on MTSL-labelled p7 single cysteine mutants show good results for short distances while being potentially error-prone for longer distances due to multi-spin effects. The measurements acquired on p7 helped to further understand structural relations between p7 monomers in an oligomeric arrangement and would allow to design a model in lipid environment on distances obtained within this study. The study emphasises the novelty of using DEER for structure determinations of membrane proteins. Pulsed EPR is a well-established technique and finds use in a huge variety of studies, going from the analysis of biradical model systems to analysing natural occurring free radicals in photo-active protein complexes. The potential of DEER in context with spin labelled membrane proteins has become more and more popular in recent years and has helped to understand the oligomeric state of several membrane embedded proteins such as the protein-translocating element of the Tat system (TatA), the potassium channel of *streptomyces* A (KcsA), or the bacterial mechanosensitive channel of small conductance (MscS)[275,277,288,289]. The increased interest in membrane proteins stems from the fact that more than 60% of the currently available drugs target membrane embedded proteins such as ion channels or G-protein coupled receptors[290,291]. As described earlier on, oligomer determination always has to deal with the drawback of the presence of multi spin effects and available software tools such as DeerAnalysis are still not capable of disentangling DEER evolution functions arising from multi-spin systems. This shows

the novelty of the studies conducted in this thesis and emphasises the need of development of new accessible software tools and the better understanding of multi-spin interactions for the analysis of multi-spin systems in context of protein oligomer formations in lipid bilayers.

In summary, the DEER experiments conducted on p7 at low and high LPRs show that due to the length of the recorded dipolar evolution function with an average of about 1.7 µs and the circumstances that the existence of multi-spin effects cannot be neglected under the experimental setup used, only the shortest distance at about 1.7 nm can be relied on safely.

Nevertheless, the established p7-DEER system, at its current stage, could still help to understand the mechanism of interactions of N-alkyl iminosugars and other drugs with their target protein p7 in lipid bilayers. This could be accomplished by analysing the samples' modulation depth (Δ) – instead of relying on the distance distribution – of sample duplicates in the presence and absence of drug. The modulation depth of a sample is defined by the pump pulse efficiency used and the number of spins being in a spin-coupled state (see section 2.2, equation 9) and thus being undistorted by multi-spin effects. By comparing the samples' modulation depth, in the presence and absence of drug, a decrease in modulation depth would report on a dissociation of protomers and thus distinguish the hypothesised p7 inhibition by monomer separation from ion channel pore blockage. Investigations to that extent are ongoing and may be finalised within a year.

4.4.5. Proton flux assay

Studies of HCV p7 in artificial bilayers indicate poor cation selectivity of the formed ion channel[57,70,73,74]. In addition Wozniak *et al.*[75] showed in their study on isolated intracellular lipid vesicles derived from HCV infected or p7 only transfected cells that p7 is involved in proton conductance and deacidification of certain cell compartments. Until then p7 ion channel activity was mainly measured by electrophysiological measurements ('single' channel recordings) in artificial membranes such as BLMs[57,70,71,73,76] or by a liposome-based dye release assay first introduced for p7 in 2007 by StGelais *et al.*[292]. BLM measurements are based on measuring current flow between two buffer filled chambers separated by a small, artificial membrane consisting usually of phosphatidylcholine (PC), phosphatidylserine (PS) or Phosphatidylethanolamine (PE) with varying fatty acid chain length and saturation. The target protein is supplied to the system solubilised in solvent or detergent. The technique is expensive, preparation time consuming and requires a high level of expertise, which makes it unsuitable for p7 compound library screening *in vitro*. The liposome-based dye release assay is cheaper and simpler, relying on the release of a fluorescent dye (6-carboxyfluorescein) from dye loaded liposomes after the addition of target protein to the system, making its use in a high-throughput compound library screen more feasible. However, it is unclear how a dye of that size (376 Da) is supposed to specifically use a correctly assembled p7 ion channel, i.e. control experiments ruling out unspecific dye release due to potential random aggregation of the p7 protein destabilising the membrane would be essential.

The proton flux assay (PFA) conducted in this thesis is a measure of p7 channel activity in lipid vesicles. The assay is based on work carried out by Schroeder,

Nguitragool and Moffat[293-295] and was applied to p7 reconstituted into lipid bilayers of varying membrane composition for the first time by Gan *et al.*[171]. The assay records the change in pH over time and thus provides information about proton membrane permeability. For a schematic of the PFA see below. The selectivity of p7 of protons over cations such as potassium or sodium is yet to be analysed, but the PFA conducted by Gan *et al.*[171] supports the possibility that p7 proton conductance can be used as a quick measure to check for channel activity.

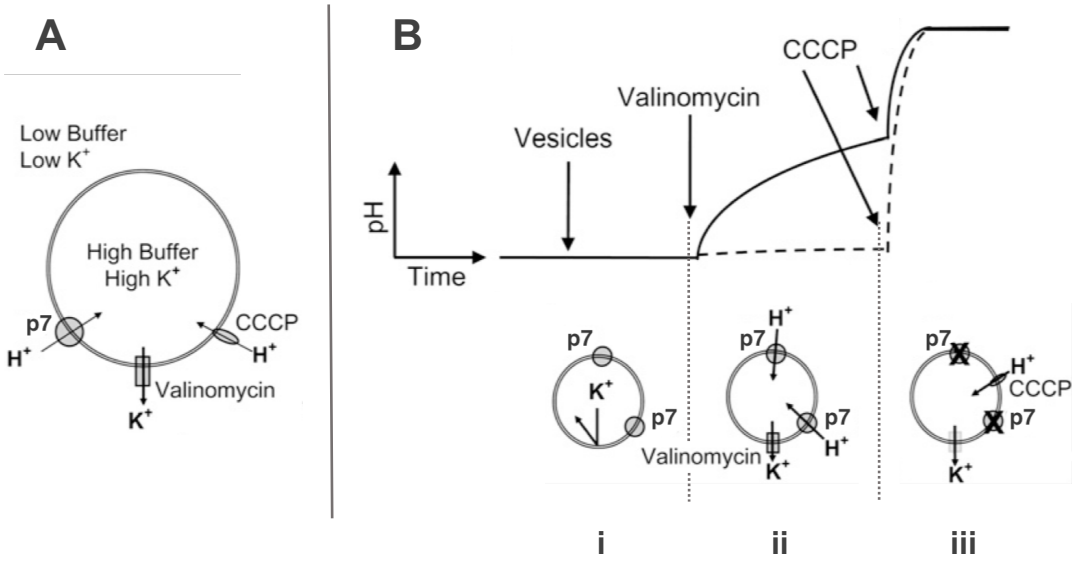


Figure 46.: Schematic of proton flux assay carried out on p7 in unilamellar liposomes. (A) Illustration of proteins and compounds involved in ion- and proton conductance. (B) *Top* – shown is the change in pH over time of a standard proton flux assay. *Bottom* – schematic of proteoliposome composition with i) starting condition liposomes contain p7, ii) addition of valinomycin leads to potassium release from the liposomes and thus a change in electrochemical potential across the membrane. The latter causes increased proton conductance of p7, iii) equilibration of proton concentration of external and internal buffer by addition of CCCP, end of measurement. Figure adapted from Moffat *et al.* 2008[294].

1.3.6.1 Preliminary proton flux measurements of p7 in DOPC:POPE membranes

In order to investigate the p7 ion-channel activity after reconstitution into lipid bilayers following the method described in sections 3.3.2 and 3.3.6 a proton flux assay was performed. P7 mutant 0 was reconstituted into DOPC:POPC membranes, in a similar way to samples prepared for CD and EPR analysis (see section 3.3.2), but with some

differences specifically required for the PFA. The main difference is that after dialysis, in order to remove any remaining detergent, proteoliposomes were extruded to obtain 100 μm diameter unilamellar vesicles. The p7 mutant 0 was chosen to avoid unspecific dimerization of p7 monomers due to disulphate bond formation of the natural occurring cysteine at position 27. Measurements took place as described in Gan *et al.* 2014[171] with minor adaptations in external proteoliposome buffer exchange (see sections 3.3.2 and 3.3.6).

PFA activity measurements of p7 mutant 0 reconstituted into DOPC:POPE (4:1, w/w) unilamellar liposomes (Figure 47A and B) show weak proton release (ΔpH) after changing the membrane's electrochemical potential by addition of the potassium ionophore valinomycin (A and B, at ~ 100 seconds). The proton release at LPR 1000:1 (ΔpH ~ 0.04) is about 4-fold higher than that at LPR 250:1 (ΔpH ~ 0.01) at the given time (Figure 47B). The end of the measurement, induced by the addition of the proton ionophore CCCP (A, at ~ 420 seconds), indicates a stronger change in ΔpH for LPR 1000:1, with an increase in pH of 0.2 with respect to the baseline. At LPR 250:1 the change in pH at termination of the measurement is less pronounced (ΔpH ~ 0.045). The results suggest that samples reconstituted at low LPRs have an overall lower proton capacity. Comparison to other PFA measurements in the literature[171,293-295] indicate a low recovery potential of the proteoliposomes prepared in this study, pointing towards a potential liposome proton leakage. Furthermore, the discrepancy in ΔpH of high and low LPR could be a result of the very high protein content of the latter, which could make the liposomes leakier.

1.3.6.2 Influence of the lipid composition on vesicle ion leakage for the PFA

To investigate the influence of the lipid composition on vesicle ion leakage, DOPC:POPE liposomes and *E.coli* polar lipid liposomes were prepared as in section 1.3.6.1, but in the absence of protein. *E. coli* polar lipids were chosen because the study by Gan *et al*. on p7 showed these lipids were suitable for the PFA[171]. For a detailed summary of the lipid compositions see Table 16. Figure 47C and D indicate that for DOPC:POPE membranes after valinomycin addition the pH remains stable until addition of CCCP (at ~300-400 s), but again showing a very low proton capacity of the DOPC:POPE liposomes formed under the experimental conditions. This suggests that from the moment DOPC:POPE vesicles are formed until the assay is started (0.5-1h time lapse), proton leak occurs and the gradient necessary for the assay is mostly lost.

For liposomes prepared from *E. coli* polar lipids the observed proton capacity of the liposomes seems to be about 10times higher than for DOPC:POPE mixtures (Figure 47C), as shown by a ΔpH of about 0.3 after addition of CCCP. The overall proton capacity appears to be the highest recorded so far. The addition of valinomycin (at ~100 s) causes a small increase of ΔpH over time with similar values to the one described by p7 in DOPC:POCE at high LPRs (Figure 47B).

Dynamic light scattering (DLS) analysis of PFA lipid-only samples show that unilamellar liposomes comprising of DOPC:POPE (4:1, w/w) are of larger nature and wider radius distribution than *E. coli* polar lipid liposomes prepared under same conditions (Table 15). The DOPC:POPE vesicles have on average double the radius (~200 nm) of the intended 100 nm (use of 0.2µm extrusion membrane), which could reflect vesicle fusion events after extrusion.

The preparation of p7 containing *E. coli* polar lipid vesicles shows a change in particle size from an on average 50 nm radius distribution in the absence of protein to about 105-125 nm with p7 reconstituted into the lipid bilayer. The findings support the assumption that DOPC:POPE membranes are not suitable for the analysis of proton membrane permeation under the assay conditions used in this thesis.

Table 15: Dynamic light scattering (DLS) analysis of different lipid and lipid-protein samples. All measurements are done post SEC buffer exchange of the PFA protocol. Shown are the observed particle hydrodynamic radii and the overall particle radii distribution in nanometres. All samples were extruded using a syringe coupler in combination with a 0.2 μm pore sized membrane.

Sample	**Particles**	
	radius main peak (nm)	**radius distribution (nm)**
DOPC:POPE (4:1, w/w)	50	4-400
E. coli polar lipids	36	10-100
E. coli polar lipids, p7 mutant 0, LPR 250:1	92	18-220
E. coli polar lipids, p7 mutant 0, LPR 1000:1	100	15-260

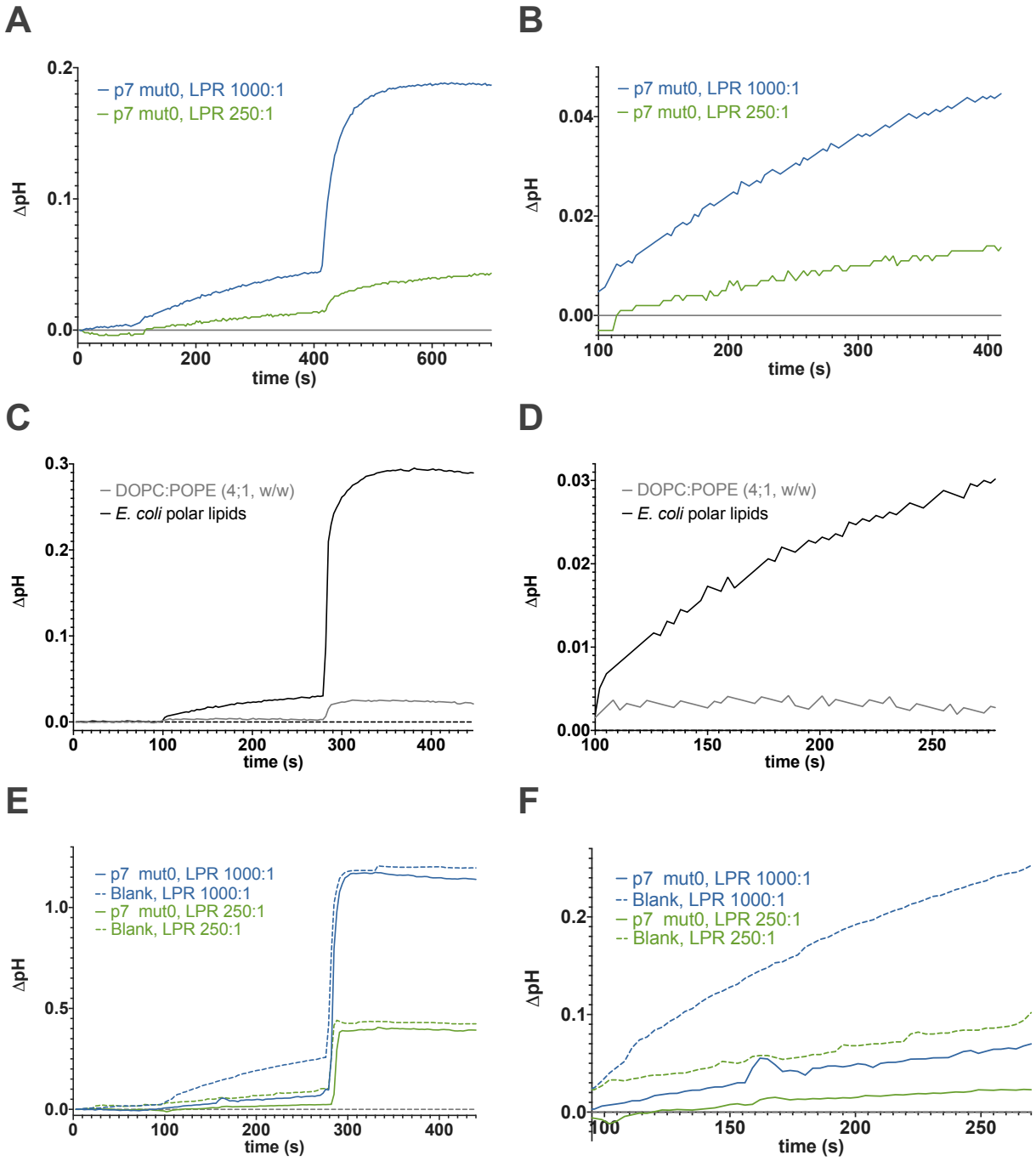


Figure 47.: Proton flux assay using unilamellar liposomes of varying lipid compositions, in the presence and absence of p7 mutant 0 (corresponding LPR is stated in figure legend). The change in pH (ΔpH) over time (s) is shown. Lipid compostion; DOPC:POPE 4:1 w/w (A/B), DOPC:POPE 4:1 w/w and *E. coli* polar lipids (C/D, grey and black graph, respectivily), only *E. coli* polar lipids (E/F). After ~100 s post start of recording a defined amount of valinomycin is added followed by a ~200-300 s delay until the addition of CCCP (A/C/E).

Table 16: List of lipid compositions used in this study. Shown are the percentages of single components included. *Asterix* – Percentages according to the manufacturer's website (https://avantilipids.com/product/100600). *Double asterix* – Study according to Swift *et al*. 1995[296] based on gas chromatography analysis of rat liver rough ER membranes.

Component	Content (%)
DOPC : POPE	
DOPC	78.5
POPE	21.5
E. coli* polar lipid extract	
PE	67.0
PG	23.2
CA	9.8
PAESC	
PA	50.0
PE	20.0
PS	20.0
PC	10.0
Rough ER membrane composition**	
PC	57.1
PE	23.9
PI	12.9
PS	3.2
SM	2.8

In Gan *et al*.[171] the authors describe the use of three different lipid compositions for p7 reconstitution purposes (Table 16): *E. coli* polar lipids, PAESC and DOPC:POPE (4:1, w/w, as used in this study), but do not report PFA results for the latter. PFA measurements in this work still need to be finalised, but preliminary results highlight that the choice of the lipid composition is crucial in order to successfully measure p7 activity in lipid bilayers. Lipid membranes have complex physicochemical properties that depend on the characteristics (molecular shape, headgroup charge, acyl chain saturations, etc) and interplay of their lipid components[297]. In the present study, the

comparison of DOPC:POPE vs *E. coli* polar lipid membranes indicate that the more complex lipid mixture of the latter is better suited for the PFA. This is surprising considering the high content of PE in the *E. coli* membrane, which would be expected to introduce instability in vesicles due to its tendency to introduce positive curvature due to its conical molecular shape[297-299]. For comparison, Table 16 includes the ER lipid composition of rat liver cells. Rat liver ER membranes show a high content of PC and PE with about 57% and 24%, respectively, which is more similar to the 78 and 22% DOPC:POPE model membrane used in this study. An interesting possibility would be to test the PFA using a bilayer composition closer to that of the ER membrane, as the presence of other components could favour creating stable vesicles. Another approach could be based on changing the extrusion protocol for the PFA proteoliposome preparation. In this study samples were extruded 21times through a 0.2 µm membrane (see section 3.3.6). The use of fewer extrusion cycles and bigger membrane pores of up to 1 µm have been used successfully for citrate permeation assays of the mitochondrial citrate carrier SLC25A1 in *E. coli* polar lipids[300]. Furthermore, the use of a bigger extrusion pore size might allow the use of the otherwise excluded DOPC:POPE lipid composition described in this study. A change in extrusion cycles or generally in the way the protein is reconstituted into lipid bilayer could have implications on the protein's secondary structural features as well[171]. This could influence the oligomerisation behaviour of p7 and therefore also its ion channel activity.

In the establishment of a p7 activity readout based on a PFA it would be of great interest to test, besides the normal channel activity, ion channel inhibition as well. The assay may be potentially used to test p7 inhibition by amantadine, rimantadine and N-

alkyl DNJs to confirm earlier findings by BLM[57,70,73] and support potential future findings using DEER methodology.

5. Materials and Methods – Alpha-Glucosidase II

5.1. GluII constructs

All GluII constructs used in this study are based on the ER alpha-glucosidase II heterodimer from *Mus musculus* (Mm, α-subunit UniProt: Q8BHN3, *GANAB_mouse*, β-subunit UniProt: O08795, *PRKCSH_mouse*). Further details and DNA sequences can be found in appendix 8.2.

For the identification of interface disruptors (see section 5.6.2 and 6.7) *Mm*GluII was 'humanised' within the alpha-beta interface based on the *Mm*GluII PDB structure (PDB-ID: 5F0E) and the UniProt entries *GANAB_human* (α-subunit, Q14697) and *PRKCSH_human* (β-subunit, P14314), respectively. For this the following amino acids were changed from the mouse to the human residues: αF724G, βL88P and βS90N on the α- and β-subunit, respectively. This was done with drug development in mind, that is aimed eventually for use in humans.

5.1.1. Cloning

*Mm*GluII in pHLsec expression vectors were a gift from Dr Alessandro Caputo (former Zitzmann-lab member). They were used for protein expression in mammalian cells (see section 5.2.1), and also served as starting points for further GluII mutagenesis.

5.1.2. *Mm*GluII subcloning into destination vector

Amplification of the target genes was done by polymerase chain reaction (PCR) using the primer pairs and Tms stated in Table 17. Solutions were mixed according to the 'Q5 High-Fidelity DNA Polymerase' (M0491) – protocol provided by New England Biolabs (NEB) and the reaction mix incubated in a Veriti PCR machine (ABI Biosystems). Thermocycles were adjusted as stated in Q5-protocol (NEB).

Table 17: DNA-primers (Gibson assembly ready) and temperatures used for target gene amplification. *Capital letters* – DNA sequence of flanking vector, *Lower case letters* – terminal sequences of gene, *highlighted in red* – introduced Stop-codon, *Asterix* – primer can be used for mPRKCSH, mLDLRa and mβΔMRH, respectively. Primers can be used for subcloning purposes of mouse- as well as humanised GluII constructs

Construct	Primer	Sequence (5'-3')	Tm [°C]	Ta [°C]	Destination vector
mGANAB	For.	agcttcccatggtgacgtcaggcCTTGCTGTGGATAGAAGC	60	59	Litmus 28i
	Rev.	aggccttgactagagggtacTCGAAGATGAATACTCCAG	58	59	
mPRKCSH	For.	agggcagatcttcgaatgcaggcGCTGTAGAAGTTAAGAGACCCC	64	65	Litmus 28i
	Rev.	agatatcctgcaggaattcc*ttatca*GTCCCCATCACTGGGTGC	69	65	
mLDLRa	For.	gatgggttgcgtagctgaaaccggtTTCTATGAAGAATCTAAACCTTTCACCTG	64	65	Litmus 28i
	Rev.	gtgatggtggtgcttggtacc*cta*TCTGCAGGTGTTCTCGCAG	68	65	

mβΔMRH	For.	agctgaaaccggtggtaccaATGGGCCTGCTGCTGCTG	72	69	Litmus 28i
	Rev.	cgccccgaagaacgttttcc**cta**ACCGAAATCAAAGGAGATCTCTTGTTCC	68	69	
mGANAB	For.	gatgggttgcgtagctgaaaccggtGGCCTTGCTGTGGATAGAAG	66	62	pHLsec
	Rev.	gtgatggtggtgcttggtaccTCGAAGATGAATACTCCAGTC	61	62	
mPRKCSH	For.	gatgggttgcgtagctgaaaccggtGGCGCTGTAGAAGTTAAGAG	63	64	pHLsec
	Rev.	gaattatcgcgatactagtc**tca**TCAGTCCCCATCACTGGG	67	64	
mPRKCSH nFLAG *	For.	gaattatcgcgatactagtcTTATCAGTCCCCATCACTG	56	57	pHLsec
	Rev.	See mPRKCSH into pHLsec			
mPRKCSH n3xFLAG *	For.	gatgggttgcgtagctgaaaccggtGACTACAAGGACCACGAC	62	62	pHLsec
	Rev.	See mPRKCSH into pHLsec			
mLDLRa	For.	See mPRKCSH into pHLsec			pHLsec
	Rev.	gaattatcgcgatactagtc**tca**TCAAAACTCTCCAGAGGG	61	59	
mβΔMRH	For.	See mPRKCSH into pHLsec			pHLsec
	Rev.	gaattatcgcgatactagtc**tca**TCATCTGCAGGTGTTCTC	58	59	

5.1.3. Gibson assembly

Amplified genes mentioned in section 5.1.1 could be directly used for Gibson assembly (GA) into linearised destination vectors. Prior to GA, vectors were double digested using the restriction enzymes AgeI/KpnI for α-subunit and XhoI/NsiI for β-subunit constructs, respectively. For GA see manufacturer's protocol 'NEBuilder® HiFi DNA Assembly Master Mix' (NEB). Site-directed mutagenesis (see section 5.1.4) was carried out in Litmus 28i (NEB) before subcloning of the target gene into the mammalian expression vector pHLsec (gift from Prof Radu Aricescu[301]). The GA product was transformed into *E. coli* DH5a cells (NEB) following a standard heat-

shock protocol. Positive clones were selected, the DNA isolated (QIAprep® Spin Miniprep Kit, standard protocol) and the identity of the DNA verified by sequencing.

5.1.4. Site-directed mutagenesis

Site-directed mutagenesis was exclusively carried out in Litmus 28i (NEB). For subcloning of the *Mm*GluII genes from other donor vectors see section 5.1.. Primers for nucleotide modification as well as nucleotide insertion were designed using the NEBaseChanger™ software tool (NEB, primers see Table 18). The two-step mutagenesis protocol includes the amplification of the gene as described in section 5.1.4 using the primer pairs from Table 18, followed by the circularisation of the linear vector. For this, the PCR product was mixed with NEB – KLD reaction mix (Table 19) and incubated 5 min at RT followed by 30-60 min at 37˚C. The final product was used for heat shock transformation into *E. coli* DH5α cells for vector amplification and subsequent DNA sequencing.

Table 18: DNA-primers and temperatures used for site directed mutagenesis. *Capital letters* – DNA sequence of target gene, *Lower case* – mutations introduced into gene, *Asterix* – primer can be used for mPRKCSH, mLDLRa and mβΔMRH, respectively

Construct	Primer	Sequence (5'-3')	Tm [°C]	Ta [°C]
mGANAB F724G	For.	AGATGCCTTGggcCAGCGATATTC	61	62
	Rev.	CGGATTGCATCTTGGTATTG	61	62
mPRKCSH L88P_S90N *	For.	caacCGGGTCAATGATGGGGTA	64	62
	Rev.	gagggGATGTACAAGGGCTTGTAC	61	62
mPRKCSH nFLAG *	For.	gatgacgacaagGGCGCTGTAGAAGTTAAG	61	62
	Rev.	gtctttgtagtcTGCATTCGAAGATCTGCC	62	62
mPRKCSH n3xFLAG *	For.	cacgacatcgactacaaggacgacgacgacaagGGCGCTGTAGAAGTTAAG	61	62
	Rev.	gtccttgtagtcaccgtcgtggtccttgtagtcTGCATTCGAAGATCTGCC	62	62

Table 19: KLD reaction mix used for vector circularisation. Sample was incubated 5 min at RT, followed by 30-60 min at 37˚C

Component [-]	Volume [µl]
PCR product	1
2x KLD reaction buffer	5
10x KLD enzyme mix	1
dH_2O	3

5.2. Expression and Purification

5.2.1. Transfection of mammalian cells

The pHLsec-GluII constructs designed as described in section 5.1 were used to co-/transfect FreeStyle HEK 293F cells (Thermo Fisher Scientific) for protein expression. A HEK 293F cell culture with a cell density of $1.2x10^6$ cells/ml was transfected with a total of 1 µg/ml plasmid DNA. For co-transfection an α-subunit to β-subunit molar ratio of 1:1.1 was used. Cell transfection was carried out using the transfection reagent FreeStyle MAX (Thermo Fisher Scientific) and OptiPRO SFM medium (Thermo Fisher Scientific) following the manufacturer's protocol. Cells were incubated in polycarbonate vented Erlenmeyer flasks, shaking at 37°C and 8% CO_2 for four days. For the cell harvest see section 5.2.2.

5.2.2. Processing of mammalian cell culture supernatant

Cells were harvested by centrifugation at 3000 g for 30 min and the supernatant decanted. While stirring the sample phosphate buffered saline (PBS) was added to a final concentration of 1x PBS followed by imidazole at a final concentration of 2 mM and the pH adjusted to 7.4 using NaOH. A 0.45 µm followed by a 0.22 µm filter membrane was used to remove any impurities.

5.2.3. Immobilised metal affinity chromatography

A 5 ml HisTrap excel column (General Electrics -Healthcare, GE) was equilibrated with 1% buffer B (for buffer composition see Table 20) and the filtered supernatant (see section 5.2.2) applied at low flow rate. Subsequently, the column was washed

with 10 column volumes (CV) of 6% buffer B followed by a step-elution with 50% buffer B over 10 CVs. The eluate was collected in 1 ml fractions and the purification monitored by recording the OD_{280}. Fractions showing an increased OD_{280} were analysed by SDS-PAGE (see section 5.4.1) to identify peaks containing GluII.

Table 20: Nickel affinity chromatography buffers.

Buffer	Buffer composition
A	1x PBS, pH 7.4
B	1x PBS, pH 7.4, 500 mM imidazole

After SDS-PAGE the desired fractions were pooled and dialysed in a 10 kDa MWCO dialysing tube against one litre of 20 mM HEPES, pH 7.4, 120 mM NaCl with a total of three changes within a minimum of 2 h. The dialysed sample was concentrated to about 2.5 ml in an Ultra-15 spin concentrator (Amicon) at 3200 g, 15°C.

5.2.4. Trypsin digestion for protein crystallography

Full length protein was trypsin digested before further purification by size exclusion chromatography (SEC), as required for protein crystallisation of *Mm*GluII.

The dialysed and concentrated sample for the metal affinity chromatography (section 5.2.3) was prepared for trypsin digestion as described in the table below. The reaction mix was incubated at 4˚C overnight and the reaction stopped by the addition of 1 mM (final concentration) of the protease inhibitor phenylmethylsulfonyl fluoride (PMSF). Subsequently the sample was purified by SEC (see section 5.2.5).

Table 21: Reaction mix for *Mm*GluII trypsin digestion

Component [-]	Amount
Protein	X mg
10x TBS	1x
$CaCl_2$	2 mM
Trypsin (13k U/mg)	X/100 mg
dH_2O	Y µl
Final volume	3000 µl

5.2.5. Size exclusion chromatography

The concentrated sample from section 5.2.3 or 5.2.4 was applied onto a Superdex 200 10/300 column (GE-Healthcare) equilibrated with a buffer suitable for subsequent experiments (see Table 22). Peak fractions containing the target protein, as determined by SDS-PAGE, were collected and pooled. The sample was concentrated as described before and protein aliquots snap-frozen in liquid nitrogen.

Table 22: Buffers used for size exclusion chromatography.

Subsequent experiment	Buffer composition
FRET, activity assays	20 mM HEPES pH 7.4, 150 mM NaCl
Crystallography	20 mM HEPES pH 7.4, 25 mM NaCl

5.2.6. Purification by ion exchange chromatography

For the purification of untagged GluII β-subunit, which was needed as a positive control for TR-FRET (see section 5.2.6), the supernatant of β-subunit transfected cells was harvested as described in section 5.2.2. The sample was transferred into 10 kDa MWCO dialysis tubes and dialysed against buffer A (30 x the sample volume) (see Table 23) at 4˚C overnight.

An anion exchange HiPrep Q XL 16/10 column (GE-Healthcare) was equilibrated with buffer A and the dialysed sample applied at low flow rate. The column was washed with 5 CVs of buffer A, followed by a 10 CVs gradient elution from 0-100% buffer B. Peaks at OD_{280} were analysed by SDS-PAGE and β-subunit containing fractions pooled and concentrated as described in section 5.2.4. To decrease the salt content and further improve sample purity SEC was carried out (see section 5.2.5).

Table 23: Buffers used for anion exchange chromatography

Buffer	Composition
A	20 mM Bis-Tris propane pH 7.0
B	20 mM Bis-Tris propane pH 7.0, 1 M NaCl

5.3. Protein concentration determination

A Nanodrop® 1000 (Thermo Fisher Scientific) was used for protein concentration determination. The device was set to 280 nm (A_{280}) and the absorbance measured by applying 2 µl of sample. The protein concentration in mg/ml was determined by using the specific molar absorption coefficient (ε, $M^{-1}*cm^{-1}$) and molecular weight (Mw, $g*mol^{-1}$) of the protein construct analysed (see table below). Molar absorption coefficient and molecular weight were calculated by utilising the software tool ProtParam (ExPASy[302]).

Table 24: Molar absorption coefficient (ε) assuming that all cysteines are in a disulphate conformation and molecular weight (Mw) of the different *Mm*GluII constructs.

GluII constructs	**ε ($M^{-1}*cm^{-1}$)**	**Mw ($g*mol^{-1}$)**
*Mm*GluII α-subunit	191640	109461
*Mm*GluII β-subunit	56360	56762
*Mm*GluII heterodimer	240875	162850
*Mm*GluII α-subunit + LDLRa	194215	117080
*Mm*GluII α-subunit + $\beta\Delta$MHR	220685	151637

5.4. Gel electrophoresis and Blotting

5.4.1. SDS-PAGE

SDS-PAGE was performed as described in section 3.2.3 with the following minor adjustments: i) Gel loading buffer: 4x LDS loading dye with or without appropriate amounts of reducing agent, ii) Electrophoresis buffer: 1x NuPage MOPS buffer (50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7), iii) Electrophoresis run time: 50 min.

5.4.2. Western blotting

SDS-gels, prepared as described in section 5.4.1, were transferred onto polyvinylidene difluoride (PVDF) membranes using the iBlot 2 Dry Blotting system (Thermo Fisher Scientific) following the manufacture's protocol.

After the transfer the membrane was incubated in a 5% w/v milk powder solution (1x TBS, 0.1% Tween 20) for 1 h on a rocking table, followed by three washes with 1x TBS, 0.1% Tween 20, 10 min each. Anti-His HRP antibody (Novex^{TM}) was diluted 1:5000 in 1x TBS, 0.1% Tween 20 and added to the membrane. Antibody binding took place for 1 h on a rocking table before unbound antibodies were washed off as described before. Blots were imaged on a LAS-1000 Luminescence Image Analyser (Fuji Medical Systems).

5.4.3. Agarose gels

See section 3.1.3 of "Methods and Materials – HCV p7"

5.5. GluII activity measurements

5.5.1. Continuous 4-MUG assay

GluII hydrolase activity was determined by measuring the cleavage of 4-methylumbelliferyl α-d-glucopyranoside (4-MUG) into the fluorescent product 4-methylumbelliferone.

For the measurement a 10-fold dilution series of the protein was prepared using 20 mM potassium phosphate pH 7.4 and a starting protein concentration of 0.1 mg/ml. 50 µl were transferred into black, non-binding surface treated 96-well plates (Corning) and mixed with 50 µl 0.5 mM 4-MUG (final concentration). The plate was incubated at 37˚C for 15 min before fluorescence was measured using a SpectraMax M5 pre-heated to 37˚C. The excitation and emission wavelengths were 355 nm and 460 nm, respectively. Fluorescence was recorded every minute over a time period of 30 min.

5.6. Stability measurement

5.6.1. Differential scanning fluorimetry

GluII crystal soaks were carried out in the presence of DMSO. Therefore, to investigate the DMSO tolerance of the GluII protein produced differential scanning fluorimetry (DSF) was performed first. For this method protein solutions were mixed with SYPRO Orange dye (Thermo Fisher Scientific) and the change in fluorescence over the change in temperature was recorded. Solutions stated in Table 25 were mixed and transferred to MicroAmp™ Fast Optical 96-well reaction plates (Thermo Fisher Scientific). Plates were sealed and spun to remove air bubbles. The plates were incubated in a 7500 Fast Real-Time PCR System (Applied Biosystems®) at 25°C for 15 min before the temperature was increased to 95˚C using a 1˚C/minute increment. Fluorescence was recorded every minute using the ROX® reporter settings supplied by the manufacturer. The Tm was evaluated by fitting a Boltzmann sigmoidal curve to the data recorded.

Table 25: Reaction mix used for DSF. Solutions were mixed and the volume adjusted to 50 µl with 20 mM HEPES pH 7.4. DMSO percentage range from 0.15 – 15% (v/v). All reaction mixes were prepared in quadruplicates.

Component [-]	**Amount**
Protein	50 ng/µl
250x SYPRO Orange dye	5x
DMSO	X% (v/v)
Final volume	50 µl

5.6.2. Time resolved fluorescent resonance energy transfer

Structural changes of the GluII α-subunit / β-subunit interface due to the incubation with different compounds were analysed by time resolved fluorescent resonance energy transfer (TR-FRET).

Unless otherwise indicated 20 µl of 7.5 nM GluII, specifically designed for TR-FRET (hexahistidine-tagged α-subunit and FLAG-tagged / untagged β-subunit, see section 5.1), were mixed with 0.2 µl of a ten-point logarithmic dilution series of compound starting from 10 mM in DMSO. For a counter assay GluII was replaced by 20 µl of 7.5 nM control peptide (HHHHHHAQDYKDDDDK-$CONH_2$ in 10% BSA (w/v), 10% TritonX-100 (w/v), 1x PBS). All samples were prepared in triplicates, in white 384-well plates (Corning). For details about the plate layout see appendix Figure B. 6. The plate was incubated for 60 min at 37˚C before the addition of 10 µl detection mix (7.5 nM mAb Anti FLAG M2-XL665 - cisbio, 0.75 nM mAb Anti 6HIS-Tb cryptate Gold - cisbio, 10% BSA (w/v), 10% TritonX-100 (w/v), 1x PBS), followed by 20 min incubation at RT. Detection took place in an 2104 Envision plate reader (PerkinElmer®) using the protocol settings stated in Table 26. Results are plotted in HTRF (homogeneous time resolved fluorescent) ratio over log compound concentration with typical IC_{50} concentrations ranging between 2.5-7.5 nM.

Table 26: TR-FRET settings. *HTRF* – homogeneous time resolved fluorescensce, *Exc*. – excitation, *Ems*. – emission.

Variable	**Setting**
Detection type	HTRF, ratio ($A_{655\ nm}/A_{620\ nm}$) * 10,000
Top mirror	LANCE/DELFIA Dual Enh (D400/D630)
Exc. filter	UV2 (TRF) 320
Exc. wavelength	320 nm
Ems. filter	APC 665
Ems. wavelength	665 nm
2nd Ems. filter	Europium 615
2nd Ems. wavelength	615 nm
Measurement height	3 mm
Cycle	2000
Delay	150 µsec
No. of flashes	200
No. of flashes for 2nd detector	20
No. of sequence windows	1
Total time of windows	800
Exc. light strength	100%

5.7. X-ray crystallography

5.7.1. Crystallisation

Prior to protein crystallisation protein solutions were spun for 10 min at 13k rpm to remove impurities introduced during sample purification or thawing. If not stated otherwise, all experiments were carried out at a 1:1 and/or 1:2 crystallisation solution to protein ratio (v/v, total volume 200-300 nl). Reliable dispersion in the nanolitre range was achieved by using a Mosquito Crystal robot (TTP Labtech) in combination with 96-well MRC 2-well crystallisation plates (Swissci). Reservoir volumes normally ranged from 40-50 µl. All plates were stored in a Rigaku Gallery HT (Rigaku) plate imager at 18˚C and images were taken every week over a time period of up to 2.5 months.

5.7.1.1. *Mm*GluII$_{tryps}$ crystallography

In this study orthorhombic as well as trigonal crystal forms of *Mm*GluII$_{tryps}$ were obtained. The former were a result of crystal seeding (see below). In general, about 5.5 mg/ml trypsin digested *Mm*GluII was mixed with crystallisation solution at the ratios stated above. For more information about solutions used see Table 28.

5.7.1.2. Crystal seeding

Crystal seeding resulted in the growth of trigonal crystals. For this, about ten different conditions of small *Mm*GluII$_{tryps}$ crystals, found in varying conditions of a Morpheus® HT-96 screen (Molecular Dimensions), were crushed using a small glass rod. The broken crystals were taken up in 1 µl crystallisation solution (see Table 27) and pooled in a 500 µl Eppendorf tube prior to 4 min of vortexing. Dilutions at a 1/10 and 1/100 ratio were prepared using the same crystallisation solution as before and samples

vortexed again. 96-well MRC 2-well crystallisation plates were set up at a 2:1:1 ratio (protein: crystallisation solution: seeding solution) and stored as described above.

Table 27: Crystallisation conditions for *Mm*GluII$_{tryps}$. All solutions were purchased form Molecular Dimensions.

Crystal form	Crystallisation solution
Orthorhombic	32% ethylene glycol/PEG 8000 mix, 0.1 M ethylene glycol mix, 0.1 M buffer system 2 pH 7.25
Trigonal	30% glycerol/PEG 4000 mix, 0.12 M alcohol mix, 0.1 M buffer system 1 pH 6.5

5.7.1.3. GluII crystal compound soaking

*Mm*GluII$_{tryps}$ crystals were soaked with different active site- as well as GluII α-/ β-subunit interface inhibitors to validate binding and mechanism of interaction.

All compounds were solubilised in DMSO at 100 mM and subsequently diluted down to 10 mM with soaking solution (16% PEG 8000, 20% PEG 400, 0.05 M Morpheus carboxylic acids mix and 0.1 M Morpheus buffer system 2 pH 7.2). Depending on the crystal form used, the compound containing soaking solution was diluted with the corresponding crystallisation solution matching the concentrations stated in Table 28. 0.2-0.5 µl of the solution were then transferred into a 96-well MRC 2-well crystallisation plate and up to five *Mm*GluII$_{tryps}$ crystals placed into the solution. Soaks were incubated at RT for a minimum of one hour before cryocooling for X-ray radiation.

Table 28: *Mm*GluII$_{tryps}$ crystal soaks carried out in this study. Long alkyloxy iminosugars were provided by Dr J. L. Kiappes at 100 mM in DMSO.

Compound	Crystal form soaked	Concentration of compound soaked
N-(9-methoxynonyl)-DNJ (M*O*N-DNJ)	orthorhombic	2.5 mM
N-(8-ethoxyoctyl)-DNJ (E*O*O-DNJ)	orthorhombic	2.5 mM
N-(7-propoxyheptyl)-DNJ (R*O*H-DNJ)	orthorhombic	2.5 mM
N-(6-butoxyhexyl)-DNJ (B*O*X-DNJ)	orthorhombic	2.5 mM
N-(5-pentyloxypentyl)-DNJ (P*O*P-DNJ)	orthorhombic	2.5 mM
N-(4-hexyloxybutyl)-DNJ (X*O*B-DNJ)	orthorhombic	2.5 mM
N-(3-heptyloxypropyl)-DNJ (H*O*R-DNJ)	orthorhombic	2.5 mM
N-(2-octyloxyethyl)-DNJ (O*O*E-DNJ)	orthorhombic	2.5 mM

5.7.1.4. *Mm*GluII$_{LDLRa}$ crystallography

In this study a monoclinic crystal of *Mm*GluII$_{LDLRa}$ was obtained. The former was a result of screening the chemical space using the commercial screen Morpheus HT-96 (Molecular Dimensions). In general, about 6 mg/ml *Mm*GluII$_{LDLRa}$ was put onto plates at the ratios stated above. Growth of the monoclinic crystal form was observed in the C1-well of the screen.

5.7.2. Data collection

GluII crystals were fished using nylon loops and immediately cryo-cooled in liquid nitrogen. The cooling was done to reduce radiation damage[303]. Crystals were screened at one of the molecular crystallography beamlines I02, I03, I04, I04-1 or I23 at Diamond Light Source (DLS, Harwell, UK) and data were collected under a flow of nitrogen, keeping the sample frozen at ~100 K. Determination of the crystal lattice as well as crystal orientation followed by the computation of the data collection strategy was done by the computer program EDNA[304], run in automated fashion by the DLS

IspyB data collection suite. For more information about data collection refer to settings listed in Table 29.

Table 29: Data collection parameters used for X-ray radiation of GluII crystal soaks. For abbriviations of the different *N*-alkyloxy deoxynojirimycin analogues (e.g. E*O*O-DNJ) see Table 28. *Asterisk* – GluII crystal soaking and data collection was performed by Benoit Darlot (former member of the Zitzmann laboratory). *DLS* – Diamond Light Source, Harwell, UK. *ESRF* – European Synchrotron Radiation Facility, Grenoble, France

	E*O*O-DNJ	**R*O*H-DNJ***	**B*O*X-DNJ***	**P*O*P-DNJ**	**X*O*B-DNJ***
Beamline	DLS I04	DLS I04	DLS I04	DLS I04	DLS I02
Detector	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M
Wavelength [Å]	0.9660	0.9281	0.9281	0.9660	0.9794
Beam size [µm]	63 x 50	63 x 50	63 x 50	63 x 50	60 x 20
Transmission [%]	100	100	100	100	100
Exposure time [ms]	282	98	98	57	68
Oscillation / image [º]	0.15	0.1	0.1	0.05	0.1
No. of images	840	1500	1500	3140	1800

	H*O*R-DNJ*	**O*O*E-DNJ***
Beamline	DLS I04-1	DLS I04-1
Detector	PILATUS 6M	PILATUS 6M
Wavelength [Å]	0.9281	0.9281
Beam size [µm]	70 x 70	70 x 70
Transmission [%]	69	100
Exposure [ms]	87	147
Oscillation / image [º]	0.1	0.2
No. of images	1500	900

5.7.3. Data processing

Diffraction images recorded during data collection were indexed, integrated, scaled and merged using the autoPROC program suit[305] (Global phasing). Molecular replacement was done by using DIMPLE[306] (DLS, Harwell, UK) or Phaser[307] in combination with the toolbox CCP4i (Harwell, UK). Model building was carried out in Coot[308] and the models refined using autoBUSTER-TNT[309] (Globel Phasing) or REFMAC5[310]. Validation took place in Coot and MolProbity[311] combined with additional rounds of refinement when needed. Structure validation for publication has yet to be performed.

6. Results and Discussion ER alpha-glucosidase II

6.1. *Mm*GluII cloning

In order to simplify *Mm*GluII purification for X-ray crystallography (see section 5.2.4) and to address the requirement for different GluII mutants for the TR-FRET assay development (see section 6.7), up to ten different *Mm*GluII constructs were designed (see Table 30).

Originally, *Mm*GluII crystallisation trials were carried out subjecting full length *Mm*GluII to partial trypsinolysis (for X-ray crystallography, see section 6.6). To simplify protein purification and improve reproducibility, a construct was designed based on the trypsinised *Mm*GluII N-terminal β-subunit LDLRa domain as observed in the crystal structures (e.g. PDB ID: 5IEG). The amplification of the *Mm*GluII β-subunit LDLRa N-terminal domain from the *Mm*GluII β-subunit DNA (UniProt: O08795, *PRKCSH_mouse*) was achieved without complications. However, ligation into the new destination vector repeatedly failed. Trouble-shooting revealed that the chosen primer was inadequate and not suitable for the formation of 3'-overhangs

('sticky ends') by endonuclease treatment prior ligation. Therefore the cloning strategy was changed from a ligation-based protocol to using new primers suitable for Gibson assembly (for primers see Table 17). Subcloning of the *Mm*GluII β-subunit LDLRa construct into a mammalian expression vector (pHLsec) was achieved and confirmed as correct by DNA-sequencing. For more information regarding the designed construct, see Table 30, construct *Mm* β_LDLRa.

All subsequent *Mm*GluII cloning, including the *Mm* β_ΔMRH construct designed for crystallisation attempts, and the humanised/tagged *Mm*GluII constructs needed for TR-FRET assay development, was carried out following the GA-protocol described in section 5.1.3.

Table 30: *Mm*GluII DNA constructs designed for the experiments described in this thesis. All mutants are based on the UniProt entry Q8BHN3 (*GANAB_mouse*, α-subunit) and O08795 (*PRKCSH_mouse*, β-subunit), respectively. *Mm* – *Mus musculus*, C-/N- – Position of protein tag attached (N- or C-terminus). F – flag tag (DYKDDDDK), 3F – triple flag tag repetition, H – C-terminal 6xHis tag, h – humanised sequence, Secretion signal – MGILPSPGMPALLS, part of the pHLsec vector system

Construct	AA sequence	Mutation	Terminal tag	Secretion signal
Crystallography				
*Mm*GANAB	32-967	-	C- His^6	YES
*Mm*PRKCSH	15-518	-	-	-
Mm β_LDLRa	31-118	-	-	-
Mm β_ΔMRH	31-397	-	-	-
TR-FRET				
Mm α_hH	32-967	F724G	C- His^6	YES
Mm β_h	15-518	L88P_S90N	-	YES
Mm β_hF	15-518	L88P_S90N	N- FLAG	-

Mm β_h3F	15-518	L88P_S90N	N- 3xFLAG	-
Mm β_LDLRa_hF	31-118	L88P_S90N	N- FLAG	-
Mm β_LDLRa_h3F	31-118	L88P_S90N	N- 3xFLAG	-
Mm β_ΔMRH_hF	31-397	L88P_S90N	N- FLAG	-
Mm β_ΔMRH_h3F	31-397	L88P_S90N	N- 3xFLAG	-

Attempts to express human GluII failed and thus - for the TR-FRET fragment based lead discovery screen (section 6.7) – decisions were made to 'humanise' murine GluII at the α-/ β-interface, the human GluII being the ultimate intended drug target for any hits in the screen. Sequence alignment of the human and murine GluII heterodimer shows an 86% amino acid sequence identity (see appendix Figure B. 5) with only three amino acids differences at the α-/ β-interface: F724G (mouse to human) on the α-subunit and L88P/S90N on the β-subunit. Site directed mutagenesis of these point mutations as well as the insertion of the desired FLAG-/3xFLAG- tags for FRET antibody recognition were carried out in the cloning vector Litmus 28i (NEB) before subcloning into the pHLsec expression vector as described in "Materials and Methods – GluII".

6.2. GluII expression and purification

Co-transfection of HEK 293F cells with *Mm*GluII α-/ β-subunit mutants (Table 30) was carried out in culture volumes of up to 900 ml. The secretion signal of pHLsec, which was fused to the N-terminus of three of the GluII mutants, showed peak overexpression and secretion of the target protein at about four days post cell transfection, as confirmed by SDS-PAGE (data not shown). Nine out of ten GluII mutants could be successfully expressed by α-/ β-subunit co-transfection and purified from mammalian cell culture media using a combination of metal-affinity- and size exclusion chromatography (see Figure 48, for methods see section 5.2). Figure 48A-D show a sample purification of full length *Mm*GluII expressed in HEK 293F cells. By affinity binding of the protein's α-subunit hexahistidine tag to an immobilised nickel chromatography resin, the target protein could be separated from the majority of undesired proteins contained in the cell media (Figure 48A and C). The protein was successfully eluted using an imidazole gradient and showed *Mm*GluII-containing fractions eluting at about 110 mM imidazole (Figure 48A and C). The subsequent SEC resulted in a monodisperse peak of the *Mm*GluII heterodimer (Figure 48B and D). The elution volume varied slightly depending on the *Mm*GluII mutant expressed, but was always in agreement with the protein molecular weight. In general, up to 7 mg target protein could be purified from one litre of transfected mammalian cell culture.

As shown in Table 30 the majority of the mutations localise to the GluII α-/ β-interface. The so-called 'humanised mutants' had no influence on protein purification or stability and could be purified using the protocol developed for the *Mm*GluII wild-type protein.

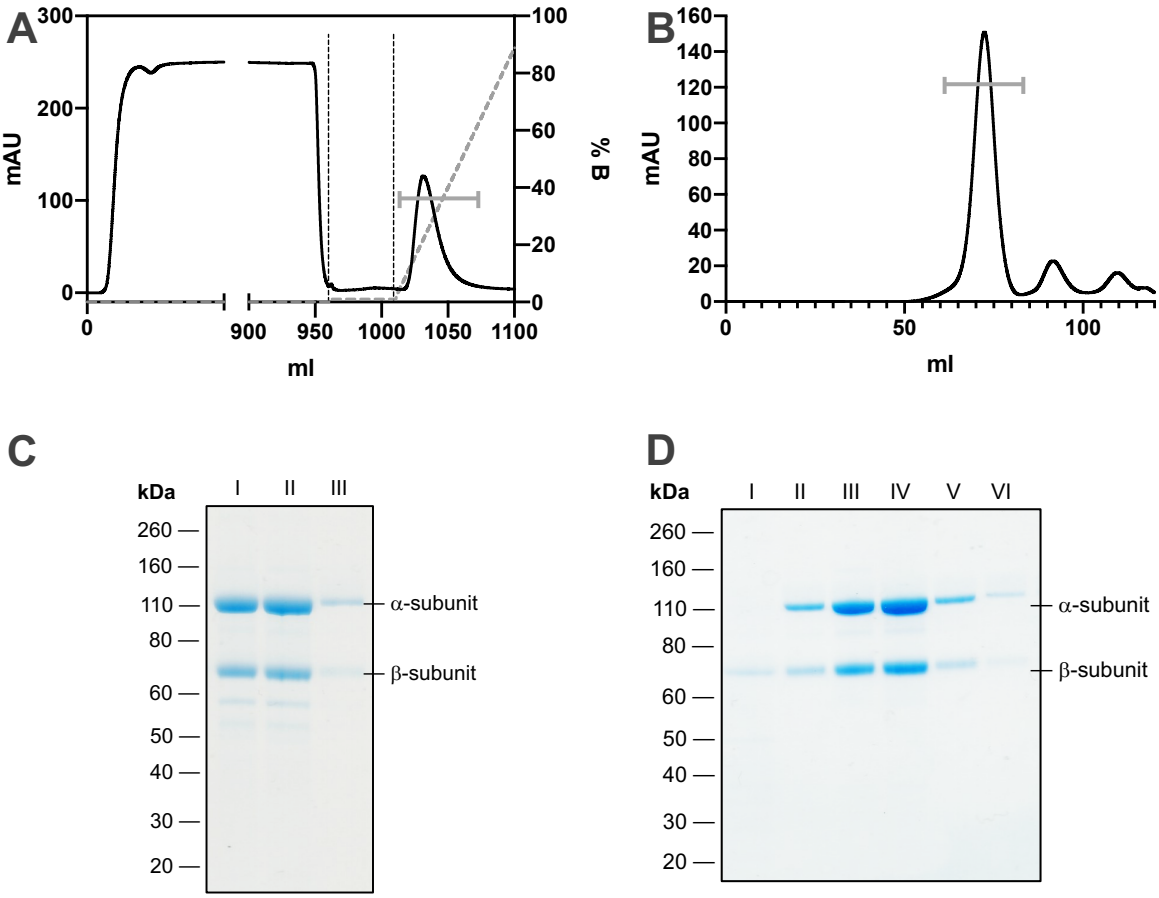


Figure 48: Purification of full length *Mm*GluII. (A) IMAC chromatogram of 900 ml of cell culture supernatant purified using a 5 ml HisTrap Excel column. Chromatography steps separated by *vertical dotted lines*: sample application, wash and gradient elution, respectively. Percentage of buffer B indicated by *grey dashed line*. Fractions pooled marked by grey *solid line.* (B) Elution profile from a Superdex 200 HP 16/60 size exclusion chromatography (SEC) column. (C) SDS-PAGE of IMAC elution. Fractions loaded (lane I-III) are taken across elution peak in A) indicated by *grey solid lines.* (D) SDS-PAGE of SEC fractions (lane I-VI, indicated by *grey solid lines* in B)) chosen for further processing.

The untagged *Mm*GluII *β*-subunit was needed as a positive control for the TR-FRET assay development (for results see section 6.7): the untagged *β*-subunit competes with its tagged counterpart for binding to the tagged α-subunit in the *Mm*GluII heterodimer and upon replacing the tagged *β*-subunit, it prevents FRET from occurring. For the purification of the *Mm*GluII *β*-subunit without FLAG- or hexahistidine-tag (Table 30, *Mm β*_h) a new purification method had to be established. The protein was also expressed in mammalian cell culture, similarly to the other *Mm*GluII heterodimer

mutants, but the requirements of the FRET assay ruled out co-transfection with the *Mm*GluII α-subunit. After calculating the protein's theoretical isoelectric point (pI 4.41) using the ProtParam software tool (ExPASy) anion-exchange chromatography (IEC) was chosen for purification. The method relies on electrostatic interactions between positively charged proteins and the negatively charged functional groups of the chromatography resin. The whole cell culture supernatant was applied to a HiPrep Q XL column at a low flow rate. In order to elute proteins bound to the resin, a salt gradient was applied reaching 1 M NaCl over 10 CV. As shown in Figure 49A, the elution profile showed three main elution peaks at about 380, 410 and 430 ml. SDS-PAGE of the protein-containing fractions confirmed the presence of the *Mm*GluII β-subunit in the middle peak (Figure 49, C). The peaks were well resolved and the SDS-PAGE analysis confirmed good purity. A subsequent SEC step was performed to remove remaining impurities and to exchange the protein into milder buffer conditions (Figure 49, B and D). This resulted in typical yields of about 12.4 mg/l purified protein.

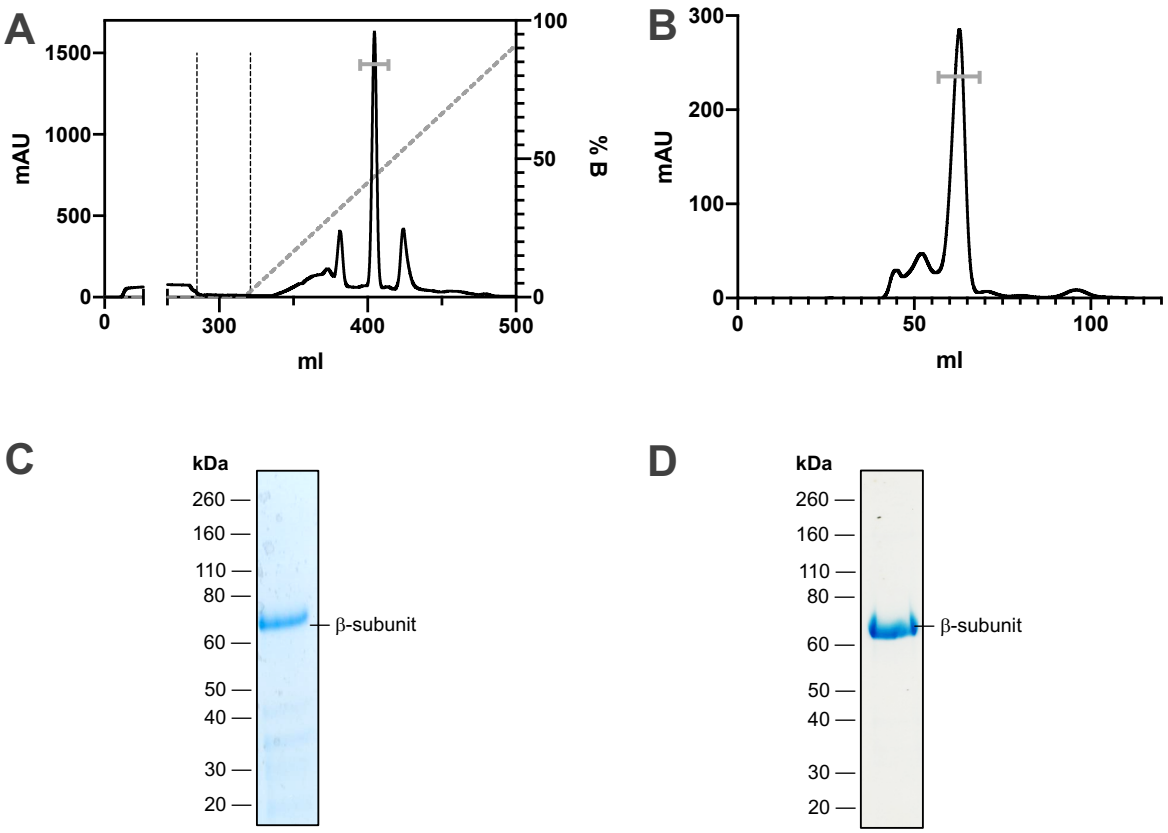


Figure 49: Purification of untagged humanised *Mm*GluII *β*-subunit (Table 30, *Mm β*_h). (A) Ion exchange chromatogram of 280 ml mammalian cell culture supernatant using a HiTrep Q XL 16/10. The *grey dashed line* indicates percentage of buffer B used. Vertical *dashed black lines* separate different chromatography steps: sample application, wash and gradient elution, respectively. Final pooled fractions are indicated by a *solid grey line*. (B) Gel filtration chromatography elution profile of concentrated IEC fractions injected and run on a Superdex 75 HP 16/60 SEC column. Fractions pooled are indicated by a *solid grey line*. (C) SDS-PAGE of the collected IEC peak. (D) SDS-PAGE of final pooled SEC fractions.

6.3. Activity determination by continuous 4MUG assay

After expression and purification of the *Mm*GluII heterodimer, the enzymatic activity of the protein was confirmed by the continuous 4-methylumbelliferyl α-D-glucopyranoside (4MUG) assay described in section 5.5.1. The use of 4MUG as a substrate for α-glucosidases was established several decades ago and the molecule is a very useful tool for *Mm*GluII activity assays[312,313]. The enzyme-catalysed hydrolysis of 4MUG produces α-D-glucose and the fluorescent compound 4-methylumbelliferone (4MU), the measured fluorescence of which correlates to α-glucosidase activity over time.

Figure 50 shows the increase in 4MU concentration over a time period of 30 min for five different *Mm*GluII batches. The assay demonstrates the presence of active enzyme and therefore confirms the successful purification of *Mm*GluII from mammalian cell culture. All five protein batches showed equivalent activity (Figure 50).

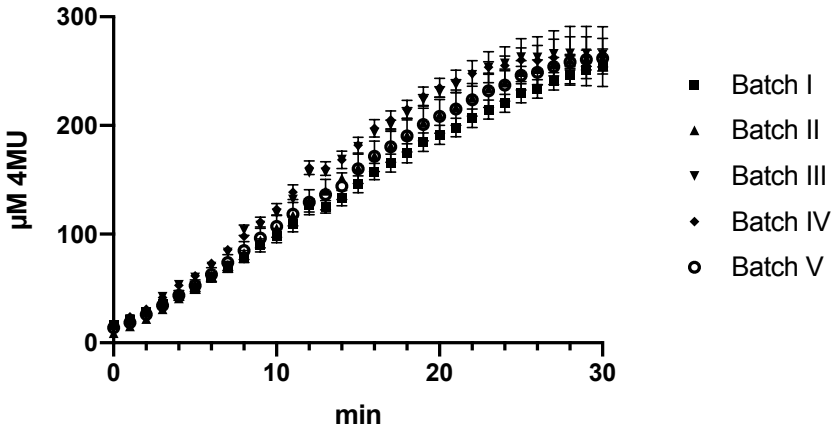


Figure 50: 4MUG activity assay of different *Mm*GluII batches. Shown is the increase in fluorescent hydrolysis product (4MU) over time. Assays were run with 0.1 mg/ml protein and 0.5 mM 4MUG at 37˚C.

6.4. X-ray crystallography of *Mm*GluII α-βLDLRa

No crystal structure of the full length GluII heterodimer has been published. Entries in the RCSB PDB database (http://www.rcsb.org) from three species; *Mus musculus* (PDBID: 5F0E, 5IED, 5H9O, 5HJO, 5HJR, 5IEE, 5IEG and 5IEF), *Schizosaccharomyces pombe* (PDBID: 2LVX and 4XQM), *Chaetomium thermophilum* (PDBID: 5DKX, 5DKY, 5DKZ and 5DL0) all lack the majority of the β subunit, either because they were obtained by limited trypsinolysis or because they used cloned truncated constructs of specific β subunit domains.

In particular, full length *Mm*GluII needed to undergo limited trypsinolysis in order for crystals to be obtained[201]. As a consequence of the trypsinolysis, the crystals contained only part of the β-subunit (the N-terminal two tandem LDLRa domains, corresponding to 18% of the protein) and do not inform on the structure of the C-terminal 82% of the protein. The *Mm*GluII α-subunit survives limited proteolysis largely intact, with only two loops missing (aa 186-243 and aa 351-369). The absence of the second loop could be a result of unspecific proteolysis because the N-terminal amino acid of the loop does not correspond to a lysine or arginine (needed for a trypsin cleavage site). Another explanation for the lack of electron density for this loop could be a high degree of disorder of the loop within the crystal structure.

The trypsin digestion step requires an extra SEC purification step which results in a significant decrease of final protein yields. To address the loss of protein and the unreliability of the trypsinolysis new β-subunit constructs were designed which would allow structural and functional studies of *Mm*GluII α-βLDLRa while at the same time saving resources and improving batch to batch protein reproducibility/consistency.

Based on the *Mm*GluII β-subunit sequence found in PDBID: 5IEG, a construct was designed spanning the two tandem LDLRa folds (in the following referred to as '*Mm*GluII β_LDLRa'). The sequence chosen starts at aa 15 of the original mouse β-subunit sequence (UniProt: O08795) and ends at aa 118. The construct was cloned into a mammalian expression vector. Co-expression trials with unmodified α-subunit resulted in good expression and is referred to as '*Mm*GluII α-β_LDLRa'. Subsequent purification resulted in up to 35% higher yields compared to those achieved after limited trypsinolysis[201].

6.4.1. Crystallography

Crystallisation trials of the *Mm*GluII mutant α-β_LDLRa were carried out under the same conditions described for the *Mm*GluII$_{tryps}$[201] making use of the departmental crystallography facilities. No crystal growth could be observed under the conditions described for *Mm*GluII$_{tryps}$, therefore a wider chemical space was screened using commercially available screens. After several days, the Morpheus HT-96 screen from Molecular Dimensions showed crystals in two different conditions (see Table 31). The Morpheus HT-96 screen allows the testing of commonly used precipitants, additives and pHs in protein crystallography in a simultaneous fashion**(-)**.

Table 31: Morpheus HT-96 screen crystallisation conditions for *Mm*GluII *α-β*_LDLRa. Formulation and stock solutions are all purchased from Molecular Dimensions. *NPS* - 0.3 M sodium nitrate, 0.3 M sodium phosphate dibasic, 0.3 M ammonium sulphate. *Buffer System 1* – 1 M imidazole/MES monohydrate (acid) pH 6.5. *Precipitant Mix 1* – 40% v/v PEG 500 MME, 20% w/v PEG 20000. *Carboxylic acids* - 0.2 M sodium format, 0.2 M ammonium acetate, 0.2 M sodium citrate tribasic dihydrate, 0.2 M potassium sodium tartrate tetrahydrate, 0.2 M sodium oxamate. *Buffer System 2* - 1.0 M sodium HEPES/MOPS (acid) pH 7.5

Well	Morpheus HT-96 crystallisation condition
C1	0.09 M NPS, 0.1M Buffer System 1 pH 6.5, 30% v/v Precipitant Mix 1
G5	0.1 M Carboxylic acids, 0.1 M Buffer System 2 pH 7.5, 30% v/v Precipitant Mix 1

Crystals found in wells C1 and G5 (Figure 51) were cooled down and shipped to the Diamond Light Source (DLS, Didcot, UK) for X-ray diffraction measurements. Only crystals found in crystallization condition C1 diffracted well. Data collection, structure determination and refinement are described in section 6.4.2. Fresh crystallisation trials were set up mimicking the conditions found in C1 of the Morpheus HT-96 screen, in order to prepare crystals for ligand soaking. However, the crystals could not be reproduced, even after several trials and the use of different protein batches. Attempts to screen around the original condition by varying the pH in the range 5.5 – 7.25 (0.25 pH/increment) and the precipitant concentration in the range 22.5 – 36.25% (1.25%/increment) were also unsuccessful. It is possible that the crystals had formed because of a small gap between plate and adhesive film, changing the evaporation rate or introducing nucleation agents from the air; conditions that could not be reproduced. In parallel to the optimization trials around condition C1, I supervised further crystallisation attempts around the G5 condition, carried out by Ms Lucy Britton, a part II student in the Zitzmann laboratory (see Table 31). As the original crystals showed absorbance under UV-light but no X-ray diffraction, we hoped that variation in crystallisation conditions would favour improved packing or a novel polymorph

altogether. A two-dimensional screen designed around the pH and precipitant concentration of the original condition resulted in over-nucleation. To reduce nucleation, the screen was repeated with the addition of 0 – 8.75% glycerol (1.25%/increment). The crystals decreased in number but remained too small to be fished for X-ray diffraction trials.

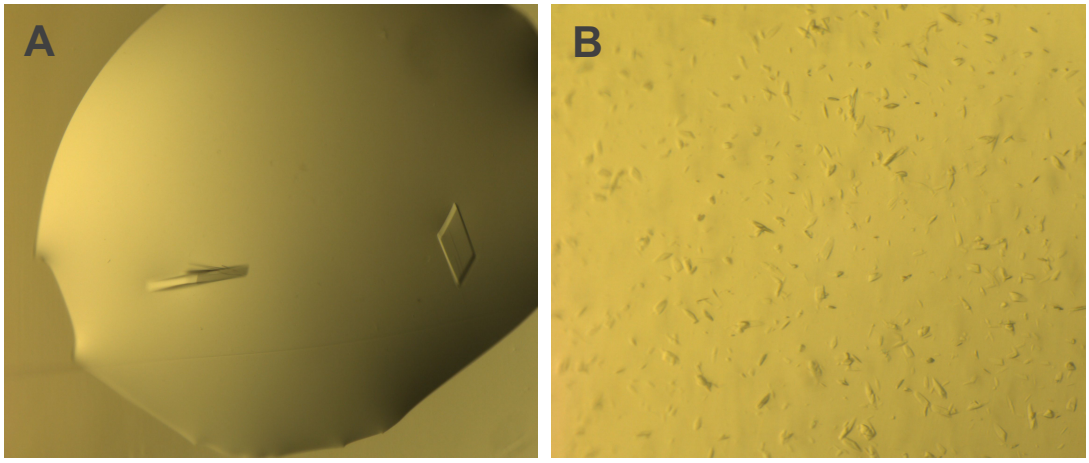


Figure 51: *Mm*GluII *α-β*_LDLRa crystals found in the Morpheus HT-96 screen (Molecular Dimensions). (A) Monoclinic *Mm*GluII *α-β*_LDLRa crystals found in well C1. (B) No*n*-diffracting *Mm*GluII *α-β*LDLR crystals in well G5. Protein and crystallisation solution were mixed at a 2:1 ratio (v/v) and incubated at 18˚C.

6.4.2. Data processing and refinement

Crystals mounted in nylon loops were exposed to X-ray radiation under a flow of nitrogen gas at T~100 K. For each sample, after crystal centering, three diffraction images were collected, each separated by Δφ= 45˚. Determination of the crystal lattice as well as crystal orientation followed by the computation of the data collection strategy was done by the computer program EDNA[304], run in automated fashion by the DLS IspyB data collection suite.

Recorded diffraction images of *Mm*GluII *α-β*LDLRa were processed using the automated data processing pipeline autoPROC (Global phasing)[305,315,316]. In short, the pipeline indexes, performs inter-frame scaling and integrates the diffraction maxima (XDS[315]); it estimates the space group from the observed point symmetry and the systematic extinctions (POINTLESS[305]); it performs summation over partials and merges the integrated intensities under the most likely point group symmetry (AIMLESS[305]). A single reflection file is produced containing a unique set of integrated intensities $\{I_{hkl}\}$ and their associated estimated errors $\{\sigma(I_{hkl})\}$ for indices in the reciprocal space asymmetric unit (asu).

Data processing statistics are listed in Table 32. The $\kappa = 180°$ section of the self-rotation function of the crystal calculated from the data processed in spacegroup $P2_1$ is shown in Figure 52.

Table 32: Diffraction data processing statistics for *Mm*GluII *α-β*LDLRa. Values in parentheses represent high resolution data shell.

	***α-β*LDLRa**
Resolution range (Å)	100.92 - 2.50 (2.69-2.50)
Cell dimensions	
a, b, c (Å)	87.09, 126.98, 103.64
α, β, γ (˚)	90, 103.07, 90
Spacegroup	$P2_1$
Total reflections	196188 (7268)
Unique reflections	59432 (2973)
R_{meas}	0.186 (0.883)
R_{pim}	0.101 (0.524)
I/□(I)	5.5 (1.4)
$CC_{1/2}$	0.991 (0.697)
Completeness (%)	91.6 (45.5)
Multiplicity	3.3 (2.4)

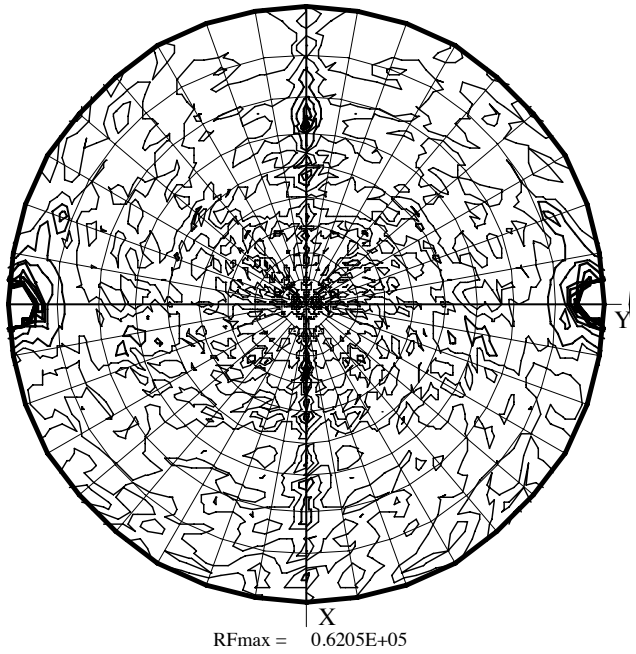


Figure 52: The κ =180° section of the self-rotation function of the *Mm*GluII *α*-*β*LDLRa crystal, calculated from the data processed in spacegroup $P2_1$. Figure generated using Molrep (CCP4 program suite).

Molecular replacement, followed by a first round of refinement, was carried out with DIMPLE[306] starting from the published *Mm*GluII$_{tryps}$ structure (PDBID: 5F0E). If the polymorph is the same, DIMPLE simply re-indexes the data to match the reference structure and runs a first rigid-body refinement (REFMAC5[310], CCP4[316]). If the polymorph is a novel one (as is the case for *Mm*GluII *α*-*β*LDLRa), DIMPLE runs Molecular Replacement (Phaser[307]) before running the first refinement. Subsequent iterative model building and refinement were carried out alternating model building in COOT[308] (CCP4) with refinement in autoBUSTER[309].

Table 33 shows the *Mm*GluII *α*-*β*LDLRa model refinement statistics at the time of writing (the structure has not been deposited in the Protein Data Bank yet).

Table 33: Data refinement statistics of the *Mm*GluII *α-β*_LDLRa mutant.

	***α-β*LDLRa at 2.48 Å**
Resolution range (Å)	84.74 – 2.48
Observations	
total	59482
in R_{free} set	1190
R_{work}	0.203
R_{free}	0.244
RMSD bond lengths (Å)	0.01
RMSD bond angles (˚)	1.1
Ramachandran outliers (%)	0.16
Ramachandran favoured (%)	96.46
MolProbity clash score	0.52
MolProbity score	1.09
Good Rotamers (%)	93.65
B-factor (Wilson plot)	41.87
Residues modelled	
A	33-182, 253-967
B	33-116
C	33-182, 253-968
D	31-116
Non-protein molecules	44
Waters total	322

The *Mm*GluII *α-β*_LDLRa structure is equivalent to the *Mm*GluII$_{tryps}$ one described by Caputo *et al.*[201]. In both structures, the lattice shows crystal contacts in close proximity to the *Mm*GluII active site (Figure 53), suggesting that soaking small molecules into crystals obtained of *Mm*GluII *α-β*_LDLRa is still feasible, while access to the active site on the part of bigger molecules such as high-mannose glycans may be sterically hindered.

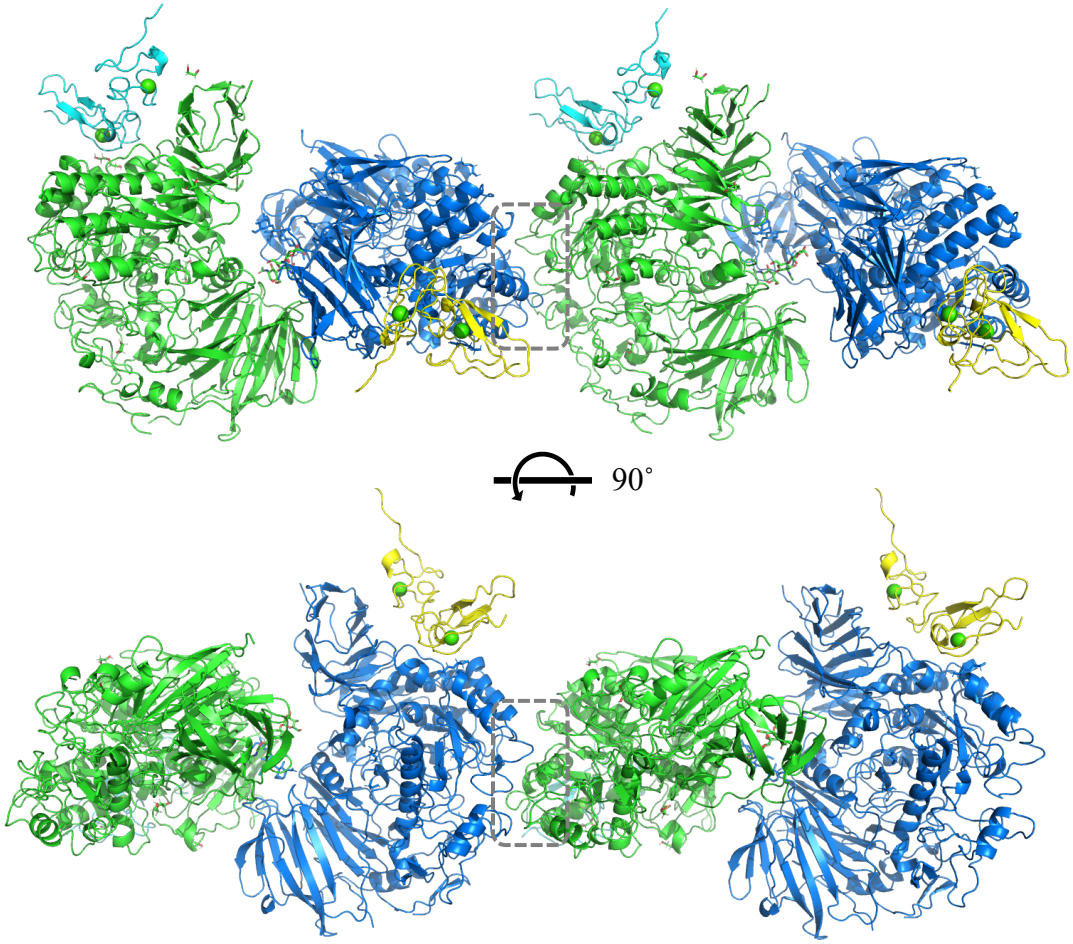
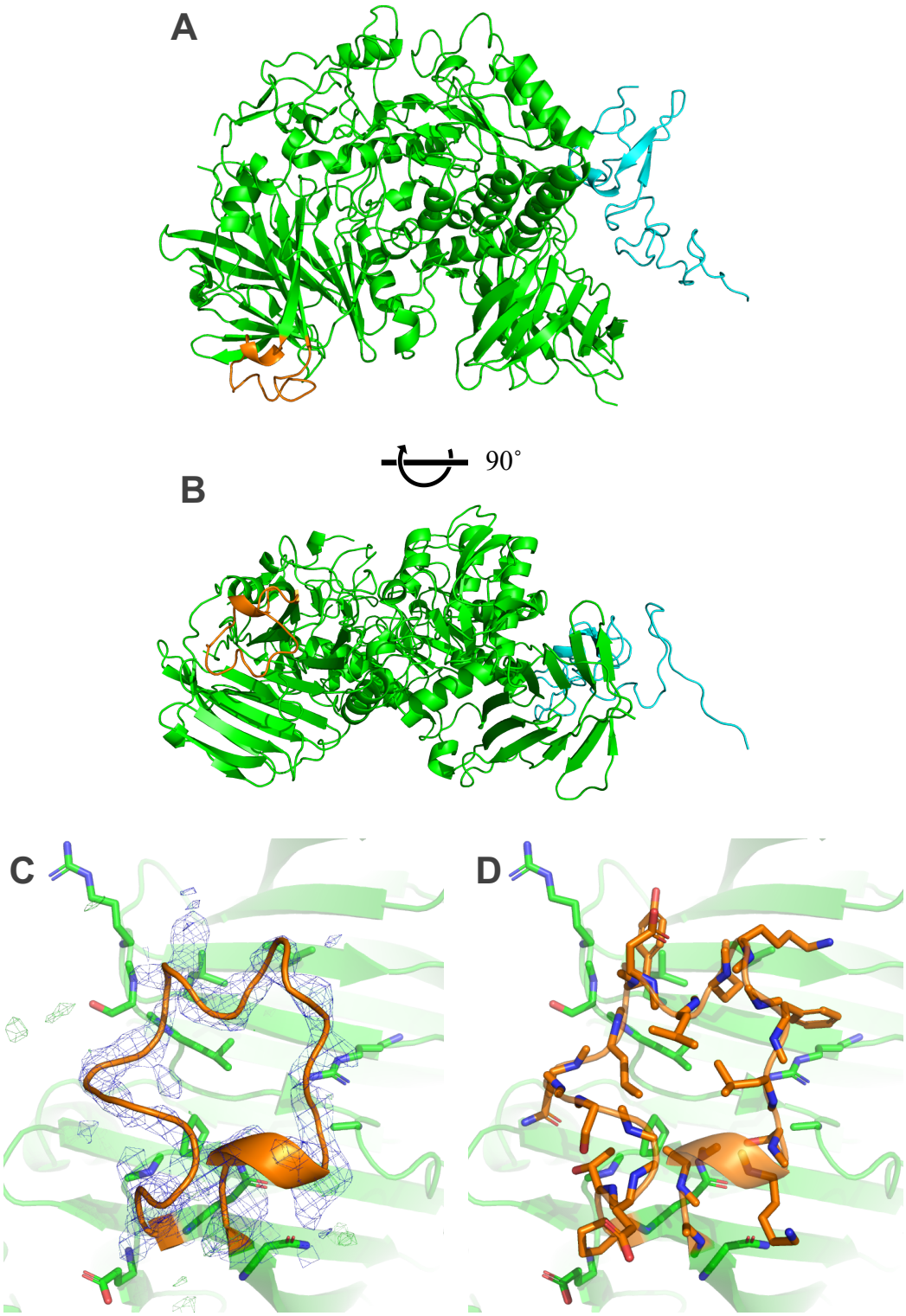


Figure 53: Two orthogonal views of two consecutive cells of the $P2_1$. *Mm*GluII α-β_LDLRa crystal. Blue and green – GluII α subunits. Cyan and yellow – GluII β-subunits. Green dots – Calcium ions. *Dashed square* –crystal contact between regions close to the active sites of two GluII α-subunits.

The main difference between the published *Mm*GluII$_{tryps}$ structures and the structure obtained of *Mm*GluII α-β_LDLRa in this study is that the latter shows ordered electron density for the majority of the main chain of the α subunit loop (aa 351-369) which was absent in the structure published by Caputo *et al.*[201] (Figure 54). Ordered density for a few side chains is visible, too (Figure 54C and D). The inner core of the loop is formed by hydrophobic residues (A352, L356, M360, L361 and L364, Figure 54D) which seem to be located above a hydrophobic patch formed by I150, L166 and P273 of the same chain. Residue Y363 of the loop is also stacked between L161 of the same

chain and P901 of a different protein copy across a crystal contact (Figure 54E). The electrostatic potential[317] of the interface between aa 351-369 and the rest of the α-subunit backbone further supports the assumption of uncharged interactions (Figure 55). In addition, charged residues can be found on the outer surface of the loop, contributing to the overall surface charge of *Mm*GluII.

As to the second loop that was absent in the *Mm*GluII$_{tryps}$ structures (aa 186-243) no electron density for it is visible in the *Mm*GluII α-β_LDLRa structure either. Either this region is unstructured or it has been cleaved by intra- or extra-cellular proteases, as no protease inhibitors were used during purification.


A
90°
B
C
D

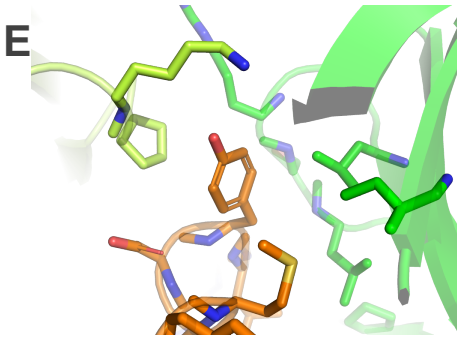


Figure 54: *Mm*GluII *α-β*_LDLRa *α*-subunit loop region aa 351-369. (A-B) Cartoon representation of the *Mm*GluII *α-β*_LDLRa structure obtained by X-ray crystallography. *Green* – *α*-subunit, *Cyan* – LDLRa domain (*β*-subunit), *Orange* – loop region aa 351-369 (*α*-subunit). (C) Close up of loop region. *Blue mesh* – electron density map contourd at 1 σ (2Fo-Fc). (D) Stick representation of loop region (orange) with residues in 3 Å distance highlighted as sticks (green). (E) Tyr363 is stacked between Leu161 of the same chain and *α*Pro901 of a neighbouring protein copy in the lattice (light green).

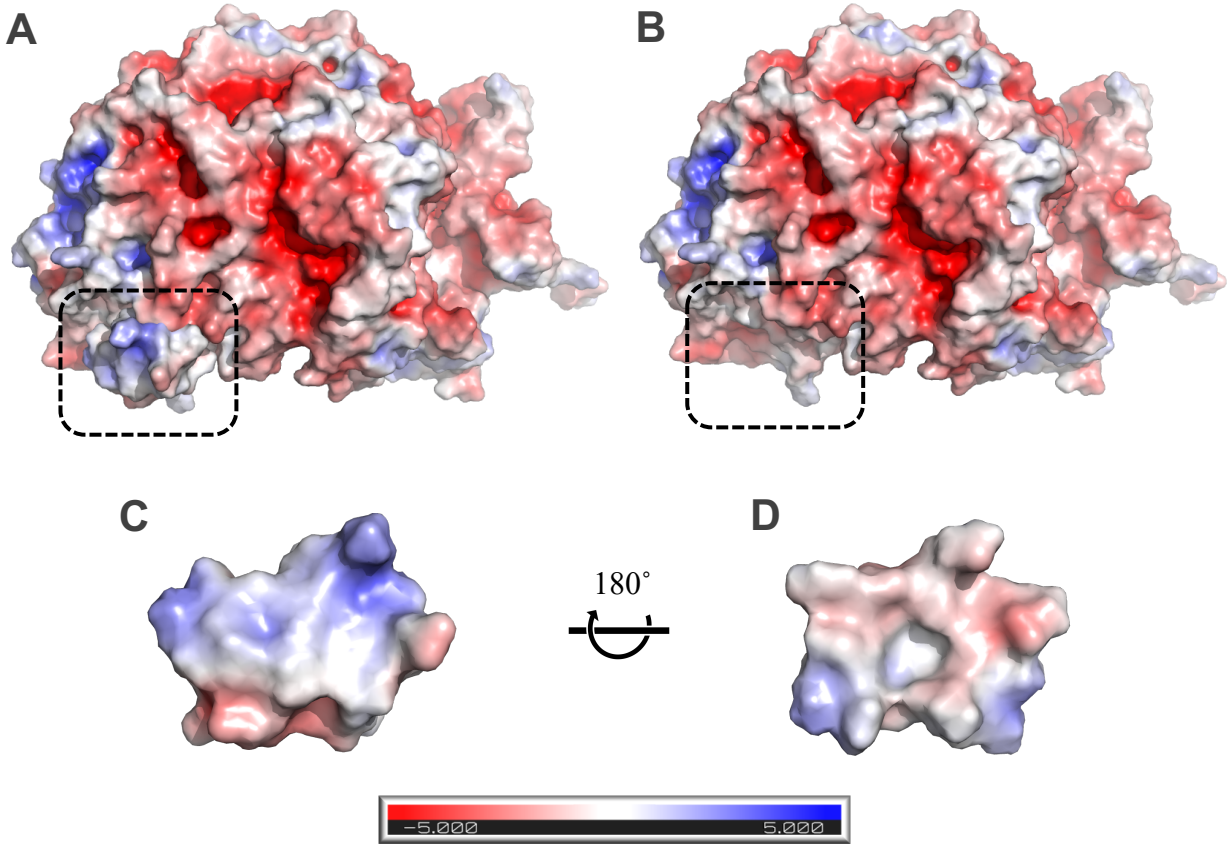


Figure 55: Electrostatic potential[317] contoured from -5.0 to 5.0 k_bT/e (red to blue) on the surface of *Mm*GluII *α-β*_LDLRa in the presence (A) and absence (B) of loop region aa 351-369 (*α*-subunit). *Dashed square* –loop region. (C-D) Loop region aa 351-369 only.

6.4.3. Discussion

By cloning and expressing the *Mm*GluII β_LDLRa domains, I was successful in improving yield and batch-to-batch reproducibility of the *Mm*GluII α-β_LDLRa construct expression and purification, disposing of the need for trypsinolysis. From this protein construct, I obtained a 2.5 Å resolution structure of the *Mm*GluII α-β_LDLRa in a novel monoclinic polymorph. The structure shows minor differences to previously published *Mm*GluII$_{tryps}$ structures, and ordered electron density for the loop region aa 351-369. The new crystal form is a suitable platform for structure determination and interaction studies of small binding partners.

To be able to use the *Mm*GluII α-β_LDLRa new crystal form for future experiments I attempted to reproduce the crystals found in C1 of the Morpheus screen. Even after several rounds of optimisation, including screening of pH and precipitant conditions, as well as co-crystallisation with *N*B-DNJ, no crystals were obtained. This could be due to the fact that the *Mm*GluII α-β_LDLRa crystals grew in well C1, close to the edge of the plate, possibly experiencing different equilibration rates or receiving nucleants from the air in case the plate was not sealed properly. When I used the commercially available Morpheus screen in an attempt to reproduce the original *Mm*GluII$_{tryps}$ crystals, crystals were found in well G5 (Figure 51), which did not diffract and resisted extensive optimisation attempts. As the *Mm*GluII α-β_LDLRa crystals could not be reproduced, subsequent soaking experiments of active site inhibitors were carried out using trypsin digested *Mm*GluII instead (see section 6.6).

6.5. *Mm*GluII biophysical characterisation for protein crystallography

6.5.1. Differential scanning fluorimetry

Gathering information about the ligand binding motifs within the active site or close to the GluII α-/β-subunit interface is one of the main aims of this thesis. In order to find more potent and selective drugs, protein X-ray crystallography (section 6.6 and 6.8) as well as time-resolved fluorescence resonance energy transfer (TR-FRET, section 6.7) were carried out on *Mm*GluII. Both techniques rely on the integrity of the protein's secondary structure and would be highly influenced by protein degradation. Compound libraries supplied by Evotec as well as novel DNJ-based active site inhibitors synthesised in the Zitzmann laboratory were supplied in DMSO. Therefore, I tested the maximum DMSO concentration tolerated by *Mm*GluII *in vitro* without altering the protein's secondary structure, giving an initial indication of the feasibility of DMSO-solubilised ligand soaks into *Mm*GluII$_{tryps}$ crystals.

Differential scanning fluorimetry (DSF) in combination with the dye SYPRO Organge^{TM} shows shifts in protein melting temperatures upon buffer changes or ligand binding. The technique was used to determine the influence of increasing DMSO concentrations on *Mm*GluII stability.

The protein melting temperature decreases with increasing DMSO concentration (Figure 56). This can be correlated with a loss in protein secondary structure. At DMSO concentrations in the range 1 - 2.5% the stability of *Mm*GluII is not affected. At concentrations > 2.5% DMSO, the protein starts to unfold and as a consequence hydrophobic core residues are exposed which favour dye binding and thus result in an increase in fluorescence. At the highest DMSO concentration tested the protein shows a decrease in melting temperature of about 3.5˚C.

Based on the DSF results I carried out all crystal soaking experiments of competitive as well as non-competitive inhibitors (section 6.6 and 6.8, respectively) at final DMSO concentrations of up to 3%.

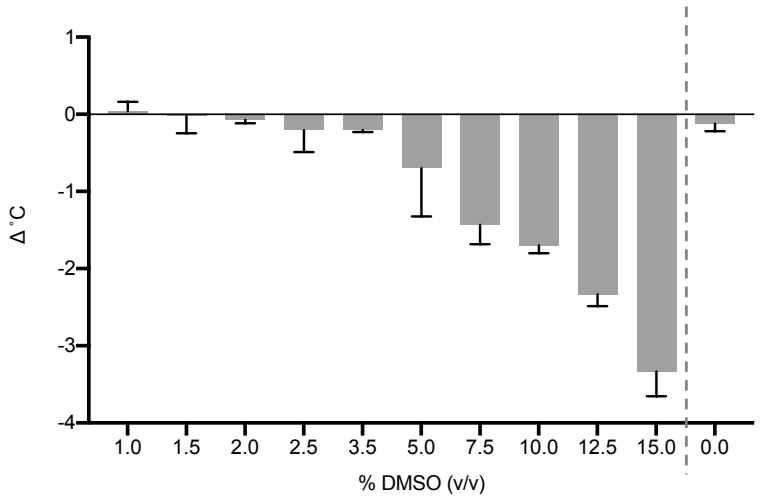


Figure 56: Differential scanning fluorimetry of *Mm*GluII in 20 mM HEPES pH 7.4, 120 mM NaCl at increasing DMSO concentrations (%, v/v). Negative Δ˚C indicates a decrease in protein stability.

6.6. X-ray crystallography of *Mm*GluII active site inhibitors

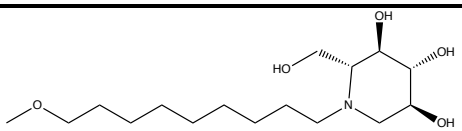
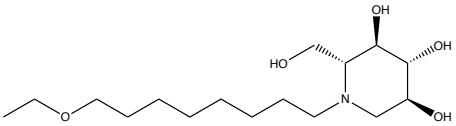
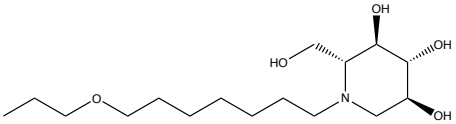
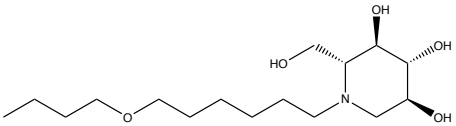
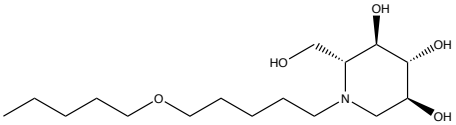
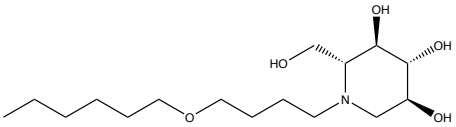
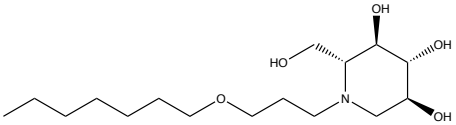
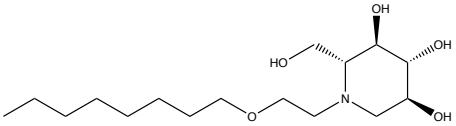
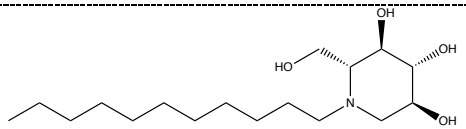
Iminosugars based on glucose mimetics reduce the release of infectious virus particles and virus infectivity[51,52,318-322] as described in section 1.1.2. Some of the viruses affected by iminosugar treatment rely on the host's glycoprotein folding machinery in order to secrete infectious virus. When the alpha-glucosidases in the glycoprotein folding machinery are inhibited by iminosugar treatment, misfolded glycoproteins accumulate in the ER. This can trigger the degradation of misfolded proteins by ERAD and virus envelope glycoproteins are unable to assemble into new virus particles.

An increase in the alkyl chain length of alkylated iminosugars was found to improve GluII inhibition, due to less favourable entropic contributions to the free energy of binding[201]. Unfortunately, long-alkyl chain iminosugars (e.g. *N*-nonyl-1-deoxynojirimycin (*N*N-DNJ)) show an increase in cell toxicity compared to short-alkyl chain derivatives (e.g. *N*-butyl-1-deoxynojirimycin (*N*B-DNJ)). The introduction of an oxygen could help decrease toxicity, as shown e.g. with the anti-diabetic drug Miglitol (*N*-hydroxyethyl-1-deoxynojirimycin)[323] or the HCV targeting drug UT-231B (United Therapeutics, stopped at clinical phase II stage)[324]. This was then addressed in a more systematic fashion by Dr J.L. Kiappes (Zitzmann laboratory) who designed a new group of alkyloxy iminosugars based on *N*-undecyl-DNJ (*N*Ud-DNJ, bottom of Table 34) in order to interrogate the importance of the exact oxygen position in the alkyl sidechain with respect to both compound activity as well as toxicity. Eight new derivatives were added to the alkyloxy DNJ family; for more information and a list of the iminosugars' IC_{50} values against *Mm*GluII see Table 34 and Table 37, respectively.

To observe interaction of long chain alkyloxy iminosugars with their target *Mm*GluII, X-ray crystallography was carried out on crystals of the trypsin-treated protein soaked

with the compounds. The focus of this study was to obtain structural information about interactions of the protein with oxygen atoms introduced at different positions into the DNJ's *N*-alkyl chain. As shown for M*O*N-DNJ, hydrogen bonding of the tail-substituted oxygen to the protein as well as hydrophobic interactions of the alkyl chain itself with hydrophobic side chains can favour specific conformations within the binding pocket and thus improve binding[201].

Table 34: *N*-alyloxy deoxynojirimycin analouges used in this study (exaption: M*O*N-DNJ)

Compound name	Abbreviation	Structure
N-(9-methoxynonyl)-deoxynojirimycin	M*O*N-DNJ	
N-(8-ethoxyoctyl)-deoxynojirimycin	E*O*O-DNJ	
N-(7-propoxyheptyl)-deoxynojirimycin	R*O*H-DNJ	
N-(6-butoxyhexyl)-deoxynojirimycin	B*O*X-DNJ	
N-(5-pentyloxynonyl)-deoxynojirimycin	P*O*P-DNJ	
N-(4-hexyloxybotyl)-deoxynojirimycin	X*O*B-DNJ	
N-(3-heptyloxypropyl)-deoxynojirimycin	H*O*R-DNJ	
N-(2-octyloxyethyl)-deoxynojirimycin	O*O*E-DNJ	
N-undecyl-deoxynojirimycin	*N*UD-DNJ	

6.6.1. Crystallisation

Following the crystallisation protocol published by Caputo *et al.*[201] crystal rods of *Mm*GluII$_{tryps}$ were obtained within two weeks after setting up crystallisation trials (Figure 57A). All *N*-alkyloxy iminosugars (excluding M*O*N-DNJ, see Table 34) were soaked into *Mm*GluII$_{tryps}$ crystals by transferring suitable crystals into a soaking solution (16% PEG 8000, 20% PEG 400, 0.05 M Morpheus carboxylic acids mix and 0.1 M Morpheus buffer system 2 pH 7.2) with the addition of up to 3 mM ligand. Soaks were carried out for up to 16 h prior to crystal preparation for synchrotron X-ray radiation. A maximum concentration of 3 mM ligand, from a 100 mM stock solution in 100% DMSO was chosen not to alter the protein secondary structure by the strong denaturation effects of DMSO. For more information see section 6.5.1.

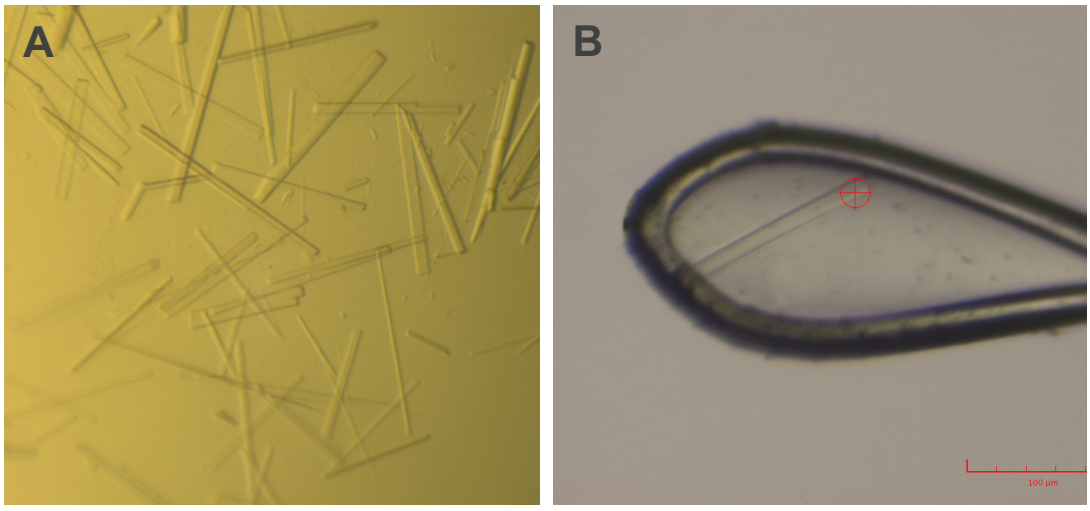


Figure 57: Orthorhombic *Mm*GluII$_{tryps}$ crystal rods. Protein and crystallisation solution were mixed at a 3:1 ratio (v/v) and incubated at 18˚C. (A) Crystals as found in crystallisation condition after a 2 week incubation time. (B) Single crystal from A) mounted in a nylon loop. Picture taken at Diamond Light Source, beamline I03. *Cross with circle* indicates X-ray beam size. *Scale at bottom right* marks 100 x 100 µm.

6.6.2. Data processing and refinement

Recorded diffraction images were processed and the structures determined and refined as described in section 6.4.2. I obtained high resolution coordinates for all seven *N*-alkyloxy iminosugars (again excluding M*O*N-DNJ) soaked into *Mm*GluII$_{tryps}$ crystals. Diffraction resolution ranged from 2.37 to 1.46 Angstrom (Å), with all structures in spacegroup $P2_12_12$. For more information regarding diffraction data processing statistics, see Table 35 below.

Table 35: Diffraction data processing statistics for *Mm*GluII$_{tryps}$ soaked with *N*-alkyloxy deoxynojirimycins. For abbriviations of the different *N*-alkyloxy deoxynojirimycin analogues (e.g. E*O*O-DNJ) seeTable 34. Values in parentheses represent high resolution data shell.

	E*O*O-DNJ	**RO*H*-DNJ**	**BO*X*-DNJ**
Resolution range (Å)	88.83 - 2.29 (2.32 - 2.29)	105.36 - 2.21 (2.25 - 2.21)	105.40 - 1.96 (1.99 – 1.96)
Cell dimensions			
a, b, c (Å)	103.2, 174.8, 63.8	105.4, 175.7, 63.9	105.4, 176.2, 63.8
α, β, γ (˚)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Spacegroup	$P2_12_12$	$P2_12_12$	$P2_12_12$
Total reflections	143869 (240)	268037 (1043)	321231 (143)
Unique reflections	30657 (66)	46736 (159)	56671 (35)
R_{meas}	0.099 (1.315)	0.348 (1.924)	0.237 (0.960)
R_{pim}	0.060 (0.817)	0.195 (1.038)	0.135 (0.634)
I/σ (I)	11.9 (1.1)	5.9 (1.4)	7.6 (1.6)
$CC_{1/2}$	0.998 (0.494)	0.981 (0.48)	0.990 (0.696)
Completeness (%)	88.7 (7.2)	89.9 (3.8)	93.7 (14.9)
Multiplicity	4.7 (3.6)	5.7 (6.6)	5.7 (4.1)

	P*O*P-DNJ	**X*O*B-DNJ**	**H*O*R-DNJ**
Resolution range (Å)	173.12 - 1.46 (1.48 - 1.46)	174.10 - 2.37 (2.41 – 2.37)	105.47 - 1.72 (1.75 – 1.72)
Cell dimensions			
a, b, c (Å)	103.7, 173.1, 62.8	103.9, 174.1, 63.15	105.5, 175.8, 63.8
α, β, γ (˚)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Spacegroup	$P2_12_12$	$P2_12_12$	$P2_12_12$
Total reflections	980504 (1748)	231048 (224)	535691 (941)
Unique reflections	170807 (322)	34879 (42)	104450 (243)
R_{meas}	0.071 (1.230)	0.330 (3.780)	0.122 (1.047)
R_{pim}	0.040 (0.690)	0.174 (1.927)	0.069 (0.651)
I/ σ (I)	13.7 (1.5)	6.2 (0.7)	10.4 (1.3)
$CC_{1/2}$	0.999 (0.483)	0.983 (0.488)	0.998 (0.546)
Completeness	95.8 (9.0)	92.5 (10.7)	93.8 (8.0)
Multiplicity	5.7 (5.4)	6.6 (5.3)	5.1 (3.9)

	O*O*E-DNJ
Resolution range (Å)	175.51 - 2.03 (2.07 - 2.03)
Cell dimensions	
a, b, c (Å)	105.1, 175.5, 63.8
α, β, γ (˚)	90.0, 90.0, 90.0
Spacegroup	$P2_12_12$
Total reflections	403926 (409)
Unique reflections	60660 (64)
R_{meas}	0.287 (1.484)
R_{pim}	0.152 (0.800)
I/ σ (I)	7.6 (1.7)
$CC_{1/2}$	0.989 (0.284)
Completeness	92.4 (12.0)
Multiplicity	6.7 (6.4)

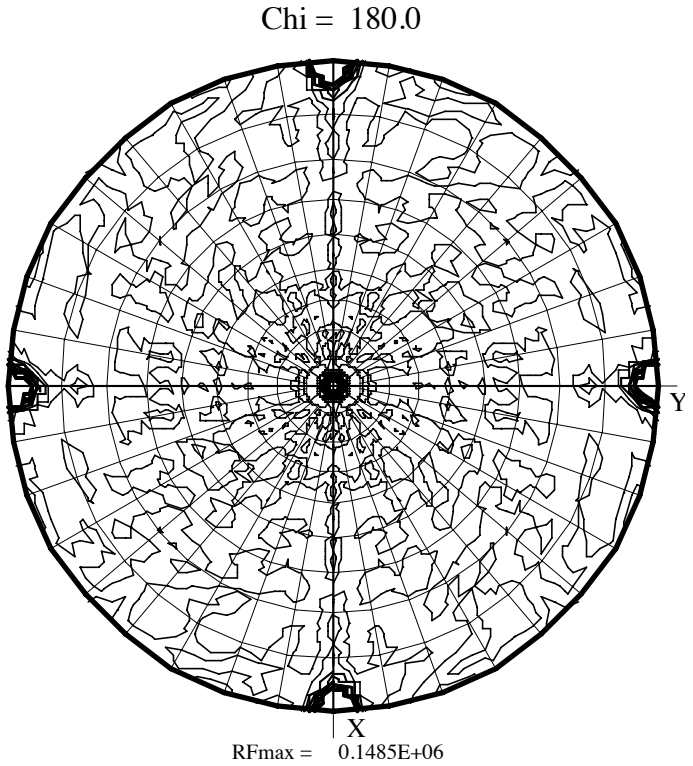


Figure 58: The κ =180° section of the self-rotation function of a *Mm*GluII$_{tryps}$ crystal calculated from data processed in $P2_1\,2_1\,2$. Figure generated using Molrep (CCP4 program suit).

Rigid body refinement of the published *Mm*GluII$_{tryps}$ structure PDBID: 5IEG was initially carried out against each of the datasets from the soaked crystals listed in Table 35. After several round of iterative apo protein model building and refinement, Fo-Fc difference electron densities were observed in the active side of the *Mm*GluII$_{tryps}$ α-subunit. Subsequent refinements resulted in a good fit for the DNJ ring for each of the seven datasets. The structures show the close proximity of the ring's hydroxy groups to hydrogen bond forming amino acid side chains (D451, R624, D640 and H698) of the *Mm*GluII α-subunit (Figure 59). Furthermore, a water molecule trapped between D640 and D669 forms an additional hydrogen bond and thus stabilises the inhibitors. The ring nitrogen is in close proximity to the active site catalytic residue D564 in an

orientation that is supposed to mimic the transition state of the glycosidic bond hydrolysis[201,325].

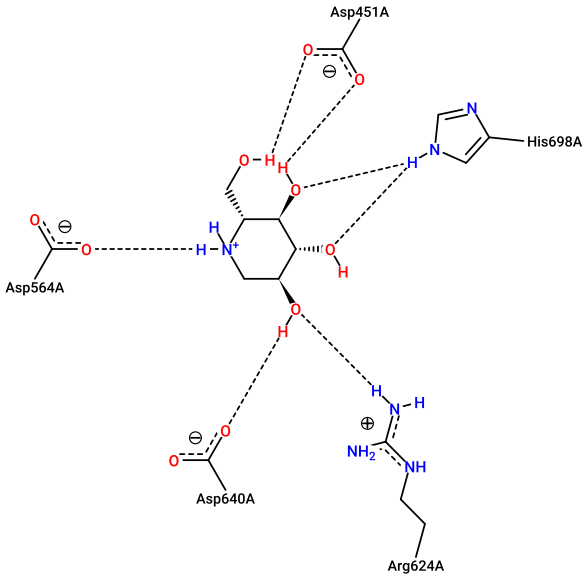


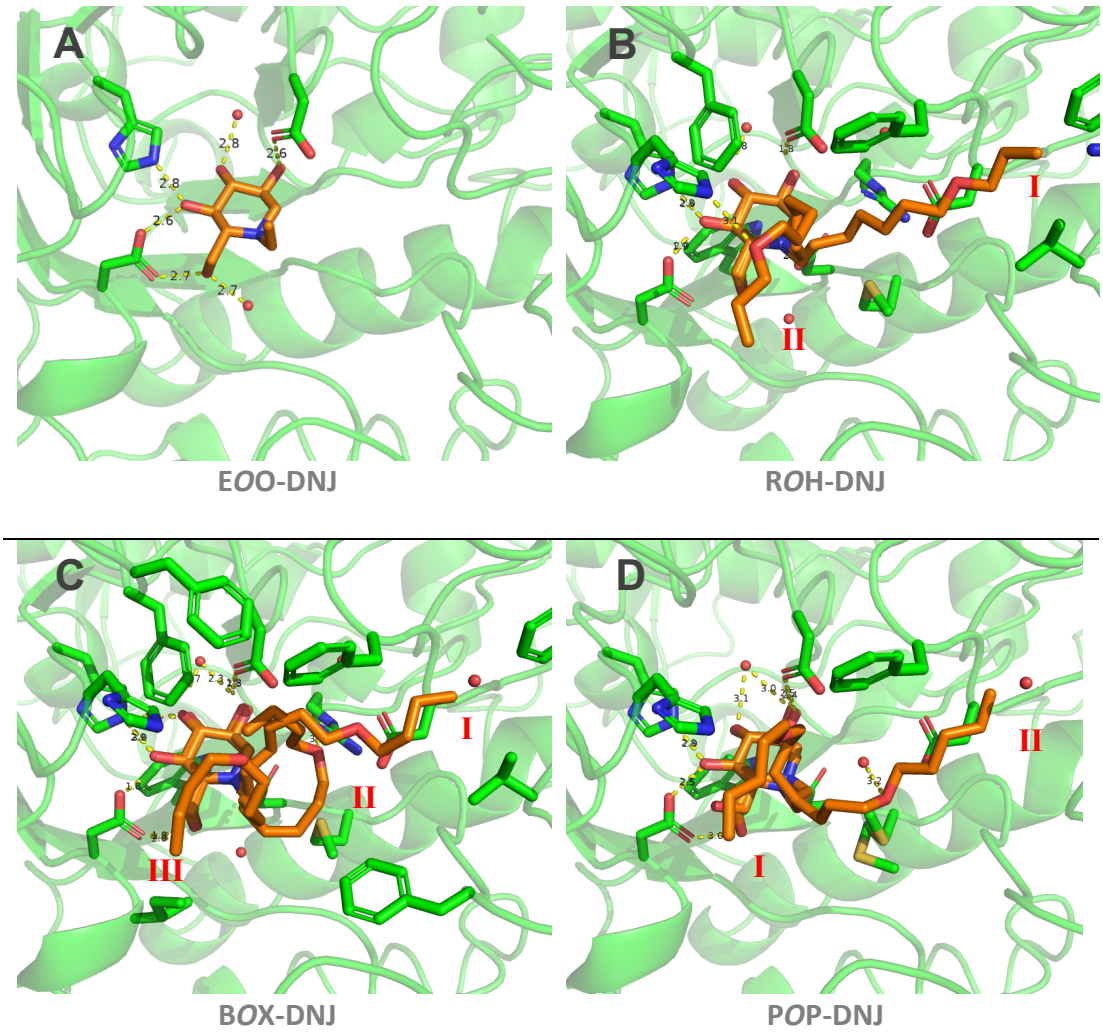
Figure 59: 2-dimensional diagram of deoxynojirimycin (DNJ) in the *Mm*GluII active site. Dashed lines indicate interactions between protein amino acid side chains and ligand. Diagram created using PoseView (Protein.Plus online program suit[326,327])

Subsequent refinement allowed the modelling of up to two carbon atoms of the *N*-alkyloxy chain for all DNJ-analogues. Similar to the M*O*N-DNJ molecule described by Caputo *et al.*[201], some of the compounds tested (Figure 60, P*O*P-, H*O*R-, O*O*E-DNJ) show a displacement of the α-subunit loop aa 524-526 as a result of steric hindrance. This effect is most pronounced in the data provided for the iminosugar analogue O*O*E-DNJ. No additional atoms of the E*O*O-DNJ *N*-alkyloxy chain were visible in the electron density.

Figure 60A-G show the models of the DNJ analogues: X*O*B-, H*O*R-, P*O*P-, O*O*E-, R*O*H and B*O*X-DNJ, respectively. These datasets enabled building of at least two different conformations for the side chain of each of the DNJ analogues. Across these datasets we observed no conserved interaction of the iminosugar's alkyl chain oxygen with the GluII active site amino acids. Moreover, the introduction of an oxygen atom

into the *N*-alkyl chain does not induce a single binding pose of the *N*-alkyloxy chain in any of the derivatives. The refined values of the occupancies for all conformers of each of the DNJ-analogues are shown in Table 36.

In the case of O*O*E-DNJ, the oxygen in position three of the alkyloxy chain is in close proximity to the aromatic ring of W525 and seems to direct the iminosugar's alkyloxy chain to one side of the GluII active site (Figure 60G).



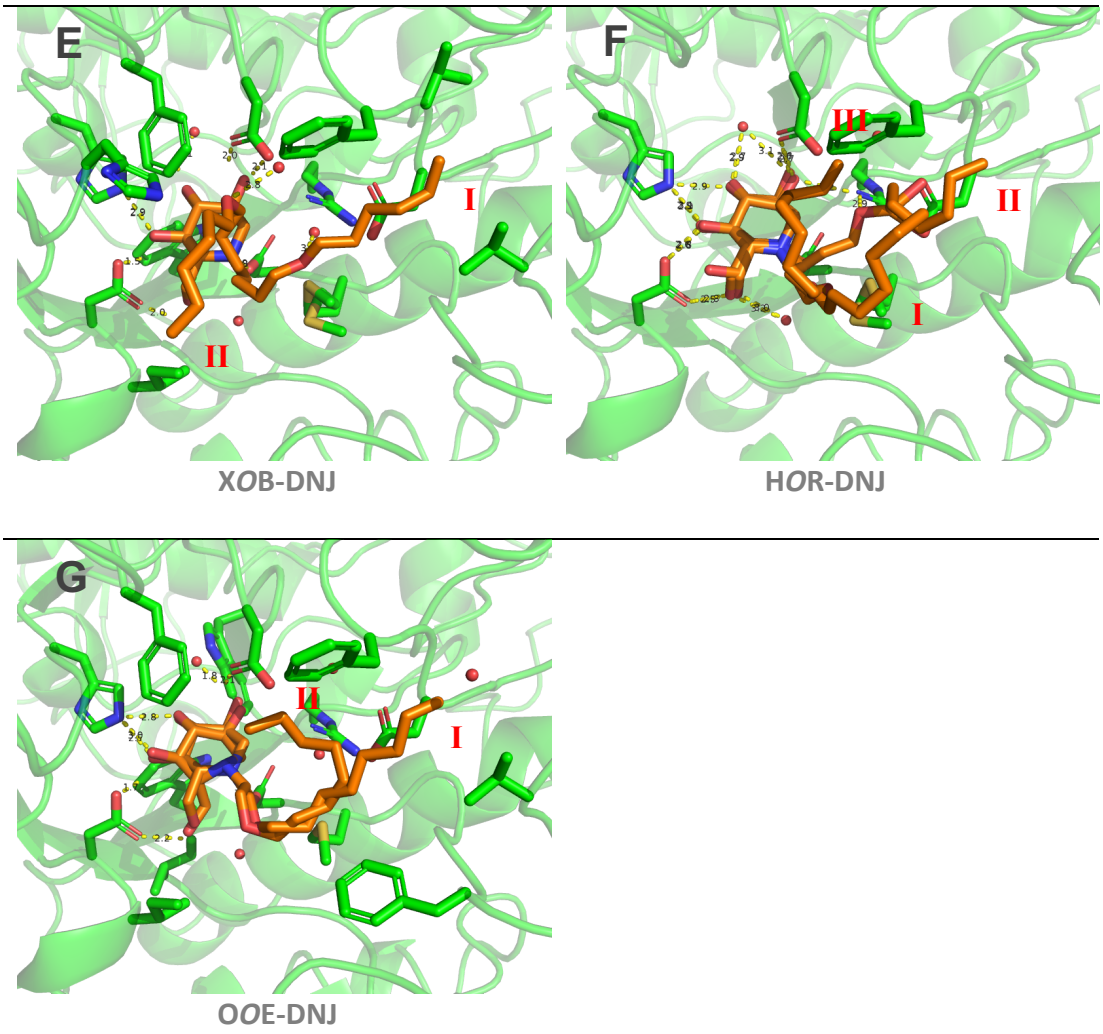


Figure 60: *Mm*GluII active site bound with different N-alkyloxy deoxynojirimycin analouges. Structures were obtained by soaking *Mm*GluII$_{tryps}$ crystals 16 h with ligands before exposure to X-ray radiation. *Green* – cartoon representation of the α-subunit. Polar interactions are highlighted with dashed, yellow lines. *Red spheres* – α-subunit located water molecules. *I/II/III* – different occupancies modelled for each of the DNJs, coressponing occupancy values found in Table 36. For more information and abbriviations of the different *N*-alkyloxy DNJ analogues (e.g. E*O*O-DNJ) see section 5.7 and Table 34, respectivly.

Table 36: *N*-alkyloxy DNJ analouges occupancy refinement according to Figure 60.

	Occupancy of conformation		
DNJ-analogue	**I**	**II**	**III**
E*O*O	na	na	na
O*O*E	0.47	0.53	na
H*O*R	0.46	0.34	0.19
X*O*B	0.57	0.43	na
P*O*P	0.65	0.33	na
B*O*X	0.40	0.24	0.37
R*O*H	0.46	0.54	na

Overall, I successfully obtained high resolution X-ray crystallography data for seven *N*-alkyloxy iminosugars soaked into *Mm*GluII$_{tryps}$ crystals, showing the iminosugar ring mimicking glucose binding within the active site of the protein. The binding is mediated by hydrogen bonding of the DNJ ring with charged amino acid side chains of the protein. The incorporation of a single oxygen atom in the DNJ's alkyl chain does not appear to influence compound binding as non-covalent interactions between protein and inhibitor are sporadic and bear little effect on inhibitory activity. The few interactions observed between the DNJ analogues chains and the protein (Figure 60, A-G) do not appear to induce any systematic trend in the compounds' IC$_{50}$s, all within the same magnitude of binding (μM) for all of the *N*-alkyloxy DNJs tested (Table 37). The structures of the *N*-alkyloxy DNJs molecules bound to *Mm*GluII$_{tryps}$ determined in this study support the conclusion that, due to weak non-covalent interactions between their *N*-alkyloxy side chains and the GluII active site lining residues, all compounds provide equivalent entropic contributions of the free energy of binding and do not influence displacement of certain amino acid side chains within the active site of the protein and thus do not show higher efficacy in inhibition than long *N*-alkyl DNJ's.

Table 37: IC_{50}- and Ki values of *N*-alkyloxy iminosugars in combination with *Mm*GluII. *Asterix* - IC_{50} evaluated at 95% confidence. *N*-alkyloxy data obtained from Benoit Darlot. *N*N – *N*-nonyl DNJ for comparison. IC_{50} value obtained from Dr Alessandro Caputo.

DNJ-analogue	IC_{50} (µM)*	K_i (µM)
O*O*E	3.34 ± 0.36	6.6 ± 0.5
H*O*R	5.18 ± 0.63	10.3 ± 0.5
X*O*B	4.05 ± 0.63	8.0 ± 0.5
P*O*P	6.22 ± 1.13	12.3 ± 0.5
B*O*X	1.15 ± 0.33	2.3 ± 0.6
R*O*H	2.09 ± 0.61	4.1 ± 0.6
E*O*O	1.01 ± 0.29	2.0 ± 0.6
D*O*	1.03 ± 0.23	2.0 ± 0.6
*N*UD	2.61 ± 1.27	5.2 ± 0.5
*N*N	3.11 ± 0.71	-

Once again, this study confirms the need for more specific GluII inhibitors as the main binding motif of the *N*-alkyloxy DNJs is the glucose mimetic headgroup. Glucose is involved in many cellular processes and thus is recognised by a vast variety of different proteins, resulting in off-target effects.

A new approach for GluII inhibition is described in section 6.7 and 6.8, respectively. The idea is based on the separation of the GluII α-/ β-subunit interface in order to affect the protein's activity levels within its natural location in the ER lumen.

6.7. Fluorescent resonance energy transfer

DNJ-derivatives are potential GluII inhibitors[328-331] binding to the active site of the GluII α-subunit[201]. DNJ-based inhibitors mimic glucose binding within the active site of the protein. The bound iminosugar prevents $Glc_2Man_9GlcNAc_2$-R oligosaccharides from entering the cleavage pocket whereby the iminosugar's ring nitrogen is suggested to facilitate the mimicry of the oxo-carbenium state of natural substrate[332,333]. These binding properties of DNJ-derivatives in combination with the high structure similarity as well as cleavage mechanisms of the GH31- and GH13 enzyme families cause many off-target side effects. My study of a new class of *N*-alkyloxy DNJs (see section 6.6) revealed similar kinetics compared to *N*-alkyl DNJs of similar chain length (Table 37, *N*UD-DNJ) and that binding was independent of the position of the oxygen in the side chain, while showing no significant improvement in binding.

In 2016, a crystallography and fragment-based lead discovery (FBLD) screen on *Mm*$GluII_{tryps}$ was carried out by Dr Alessandro Caputo and Dr Pietro Roversi (former members of the Zitzmann laboratory) in collaboration with the XChem facility at Diamond light source, Harwell, UK. The screen, of about 1100 compounds, resulted in twelve potential fragment hits, clustered in four different locations one the α-subunit surface. Two fragments were bound at the opening of the enzyme's active site. Another five bound to the opposite face of the molecule, to sites scattered on the backside of the α-subunit in a T-shaped pattern. The remaining five fragments were located within two small binding pockets of the α-subunit in close proximity to the GluII α-/ β-interface (data unpublished, for more information regarding the ligands see section 6.7.2). The aim of the FBLD screen was to identify allosteric site inhibitors of ER alpha-glucosidase II, which would be enzyme specific and not bind to other

carbohydrate-binding enzymes. In particular the study was focused on compounds with the potential of GluII α-/ β-subunit separation. With the discovery of the four hits in close proximity to the α-/ β-interface a basis was established for the experiments conducted in this study.

The *Mm*GluII α-/ β-interface spans an area of about 706 Å^2, stabilised by the existence of 16 salt bridges and up to 13 hydrogen bonds (Figure 61). The main protagonists on the β-subunit are N-terminal, acidic residues coordinating calcium ions in the two LDLRa tandem repeats: D53, D54, D57, D94 and D98. These salt bridges are established with basic amino acids on the α-subunit: R837, R839, R840, R951 and K952. *In planta* co-expression of fluorescently tagged GluII α- and β-subunit showed localisation of the two subunits in the ER of the cell. The αR787E mutant showed no influence on protein localisation; however, the introduction of a second mutation αR784E led to a loss of localisation of the α-subunit double mutant from the ER lumen. The mutagenesis experiments in plants suggest functional importance of residues R787 and R784 (corresponding to R837 and R840 in the mouse sequence) in localising the enzyme to the ER via α-/ β-interface stabilisation[201].

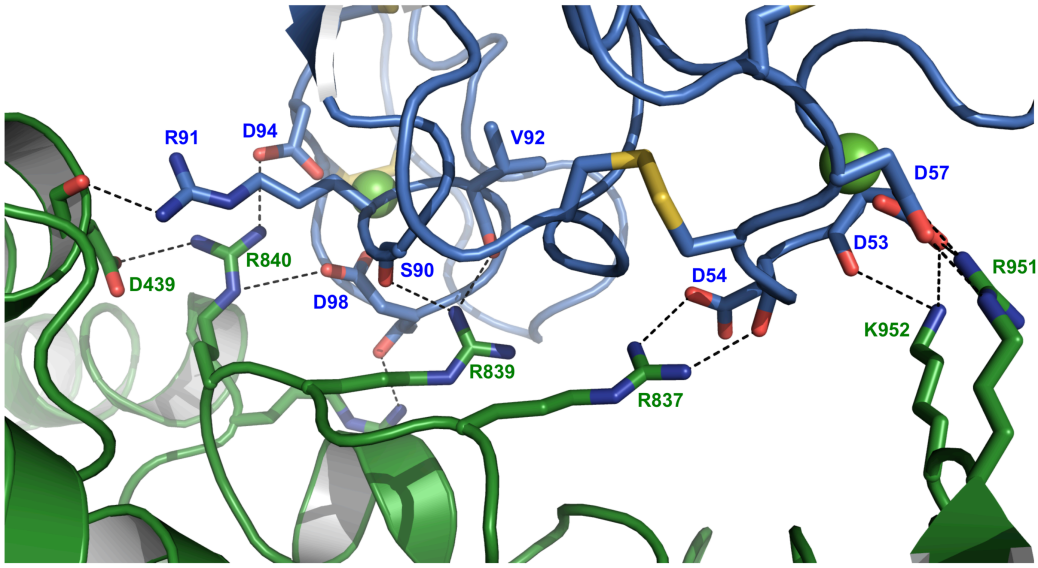


Figure 61: Mouse GluII α-/ β-interface polar interactions. *Green* – cartoon representation of the α-subunit. Polar interactions are highlighted with dashed lines. *Blue* – β-subunit. *Green spheres* – β-subunit located calcium ions. Residues involved in β-subunit binding are highlighted as sticks. Figure adopted from Caputo *et al.*[201].

The GluII active site is located within the α-subunit of the protein. The β-subunit importantly carries a small C-terminal ER retrieval signal (HDEL) which keeps the whole heterodimer retained within the ER lumen, where it can act as the ERQC usher, mediating entry/exit of a glycoprotein into/from the calnexin cycle (see section 1.3). Were the interactions between the α-/ β-interface disrupted by suitable compounds, the α-subunit would detach from the β-subunit, and upon secretion it could no longer be returned to the ER lumen by the KDEL receptor associated retrieval machinery[216] (Figure 62A-D), eventually leading to loss of ER GluII activity in the ER. The enzymatically inactive β-subunit on the other hand, carrying an HDEL C-terminal signal, would continue to be retrieved back to the ER. The loss of the α-subunit from the ER would result in an accumulation of unfolded glycoproteins carrying $Glc_2Man_9GlcNAc_2$- or $Glc_1Man_9GlcNAc_2$ N-glycans, and the newly expressed

glycoproteins would be prevented to enter the calnexin cycle of the ERQC for assisted folding.

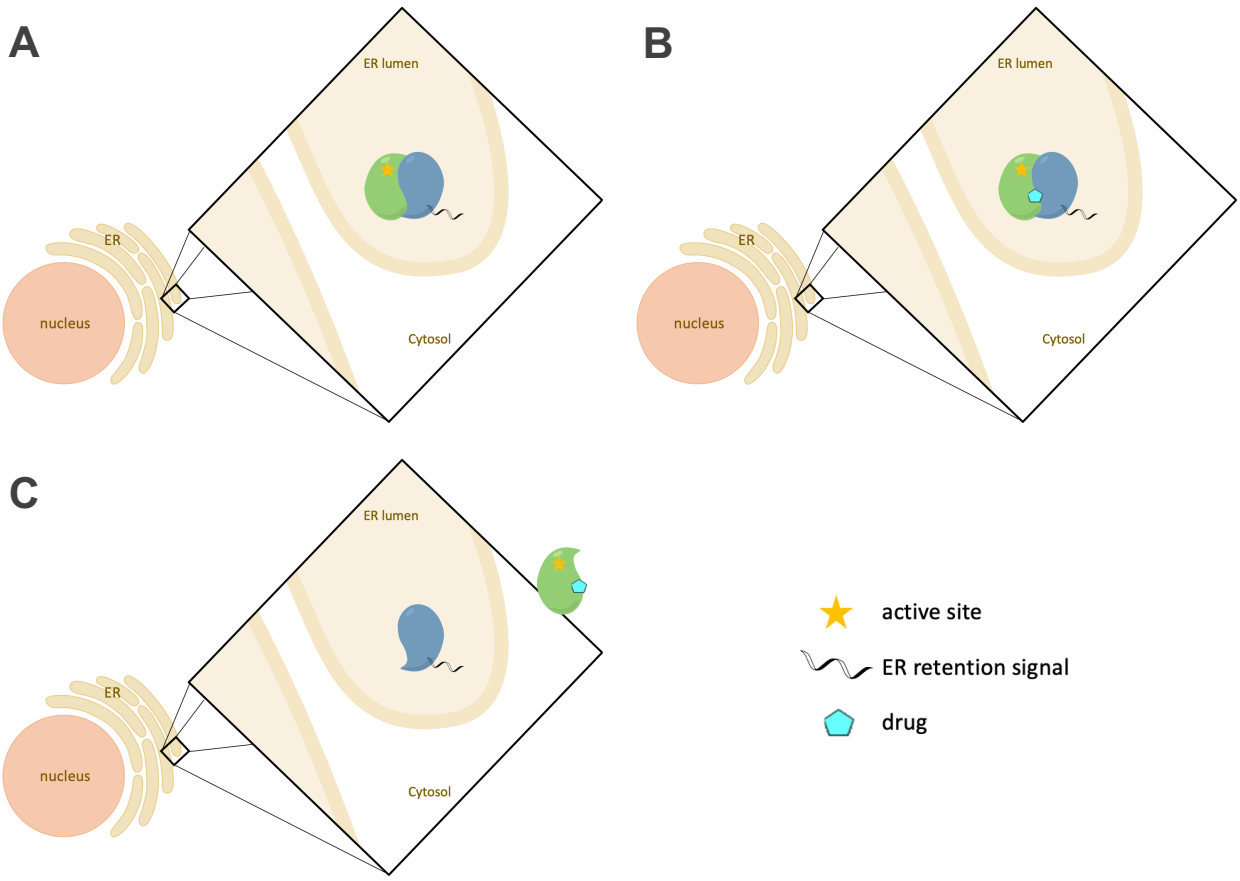


Figure 62: GluII α-/ β-interface interference concept. (A) GluII located in its natural environment in the ER lumen. (B) Drug binding close to the protein's α-/ β-interface. (C) After GluII heterodimer dissociation initiated by the drug, the α-subunit remains without ER retrieval signal and – unlike the β-subunit – cannot be retrieved to the ER via the KDEL receptor associated machinery[216]. *Green sphere* – GluII α-subunit, *Blue sphere* – GluII β-subunit.

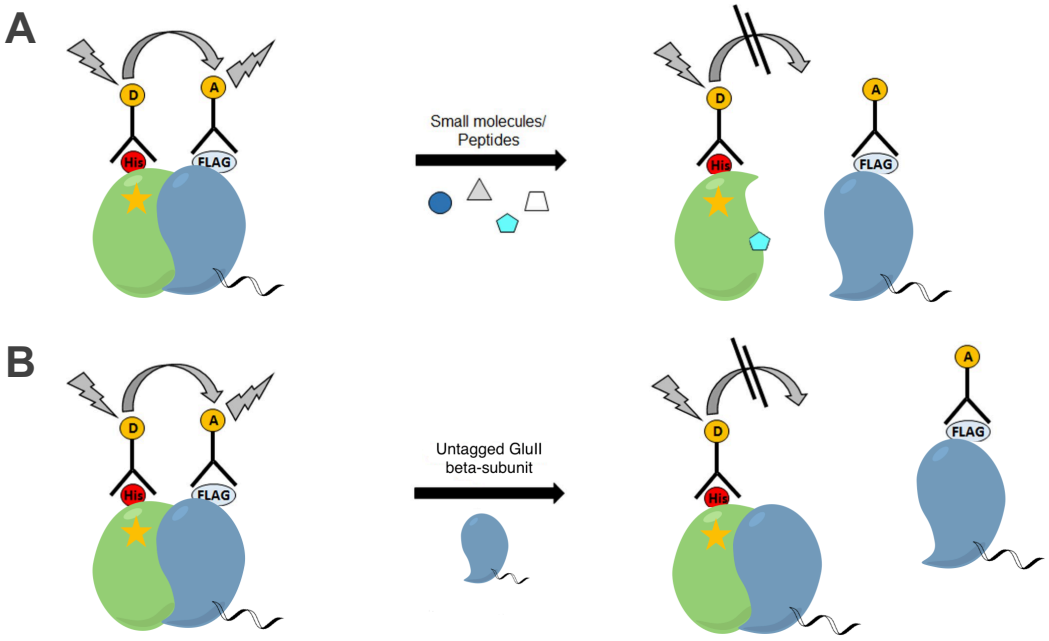


Figure 63: TR-FRET high throughput screening concept. The α-subunit carries a His tag and the β-subunit carries a FLAG tag. A TR-FRET donor is conjugated to an anti-His tag antibody, and a TR-FRET acceptor is conjugated to an anti-FLAG tag antibody. (A) TR-FRET of GluII in the presence of interface splitting compounds: no TR-FRET can take place. (B) Competition assay TR-FRET positive control (counter screen), same principle as A) but utilising untagged GluII β-subunit, to compete off its FLAG-tagged counterpart.

The discovery of new chemotypes able to induce a dissociation of the GluII α-/ β-subunits was one of the main challenges of our inhibition approach. The development of the required high-throughput assay, to monitor interface disruption by several compounds in a fast timescale, was carried out in collaboration with Evotec under a LAB282 grant.

To monitor the dissociation of the GluII subunits we established an *in vitro* TR-FRET assay (for FRET principle see section 2.3) using a hexahistidine-labelled α-subunit and a FLAG-labelled β-subunit. Monoclonal antibodies (mAbs) capable of recognising hexahistidine- or FLAG-tags were purchased with a donor or acceptor fluorophore, respectively. When binding to the proteins' tags the mAbs and thus the fluorophores come into close proximity, allowing energy transfer between the fluorophores. As long as the heterodimer stays intact, FRET is occurring. In the event

of subunit separation, the FRET efficacy will decrease until the distance exceeds the Förster limit.

As the overall aim is the development of new molecules that interfere with GluII α-/ β-subunit binding, to be tested as broad-spectrum antivirals, human GluII should be used in the TR-FRET high-throughput screen. As previous attempts to express and purify full length human GluII failed, I first wanted to 'humanise' at least the murine GluII α-/ β-subunit interface, by mutating mouse-specific amino acids to their human counterparts in this area of the protein. Sequence alignments of murine (*Mus musculus)* and human (*Homo sapiens)* GluII α- and β-subunit sequences show three differences only within the interface area in either subunit (mouse to human): F724G (α-subunit), L88P and S90N (β-subunit). For more information see Figure 64 and the sequence alignments with corresponding mutations in the appendix Figure B. 5.

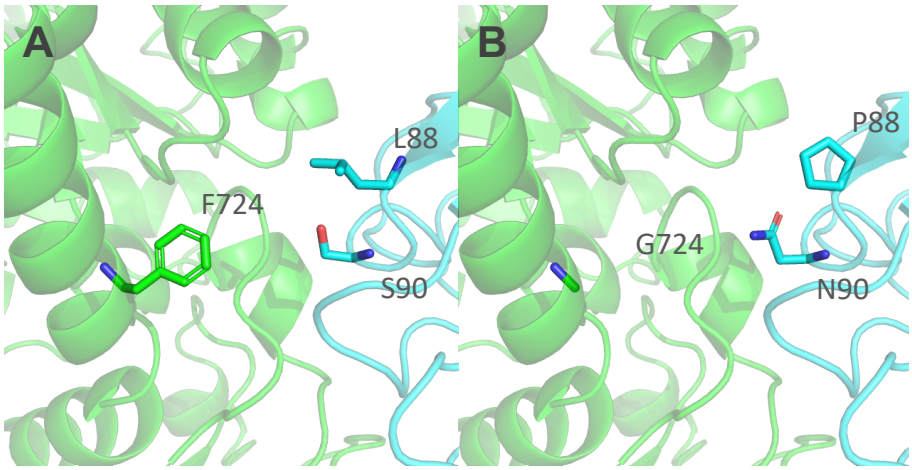


Figure 64: Humanisation of the mouse GluII heterodimer interface. (A) Native mouse GluII interface. Murine residues mutated to match the human GluII interface are highlighted in stick representation. (B) Humanised mouse GluII interface as used for the TR-FRET assay. Green – GluII α-subunit, Cyan – GluII β-subunit.

6.7.1. GluII TR-FRET assay development

To be able to distinguish changes in fluorescence introduced by the separation of the two GluII subunits and not by an unspecific interference of the antibody and/or fluorophore with the compounds screened, a control peptide was designed for the TR-FRET assay developed in this study. The control dipeptide has a N-terminal His- and a C-terminal FLAG-tag (His_6-AQ-FLAG).

In order to optimise the assay conditions, I first tested the behaviour of a control dipeptide (see above) in TR-FRET. The peptide was used to measure changes in homogeneous time resolved fluorescence (HTRF) ratio of the two fluorophores terbium (Tb, donor) and XL665 (acceptor[334]), covalently linked to either an anti-His or an anti-FLAG Ab. As the specific protein tag affinities between the two mAbs might differ, and thus the bound donor-/ bound acceptor ratio, measurements were performed sampling both mAb-fluorophore combinations. Figure 65 shows an increase in HTRF ratio when using a combination of anti-His-Tb and anti-FLAG-XL665 compared to anti-FLAG-Tb and anti-His-XL665 at all concentrations tested. A change in mAb concentrations (in the range of 0.1, 1.0 and 5 nM Tb and in the range of 1.0, 10 and 50 nM XL665, respectively) resulted in a small signal increase at 1:10 Tb:XL665 ratio over 1.0:1 and 5:50 Tb:XL665 ratios. The decrease in HTRF ratio at higher donor/acceptor concentrations could be due to cross interference of different fluorophore pairs in close proximity to each other. The titration of control peptide also revealed a correlation between peptide concentration and fluorescence detected. At the concentrations tested the HTRF ratio increases with increasing peptide concentration at all mAb concentrations tested. Measurements at 1:10 and 5:50 mAb Tb:XL665 ratios show a step-wise fluorescence increase between 1 and 10 nM of control peptide.

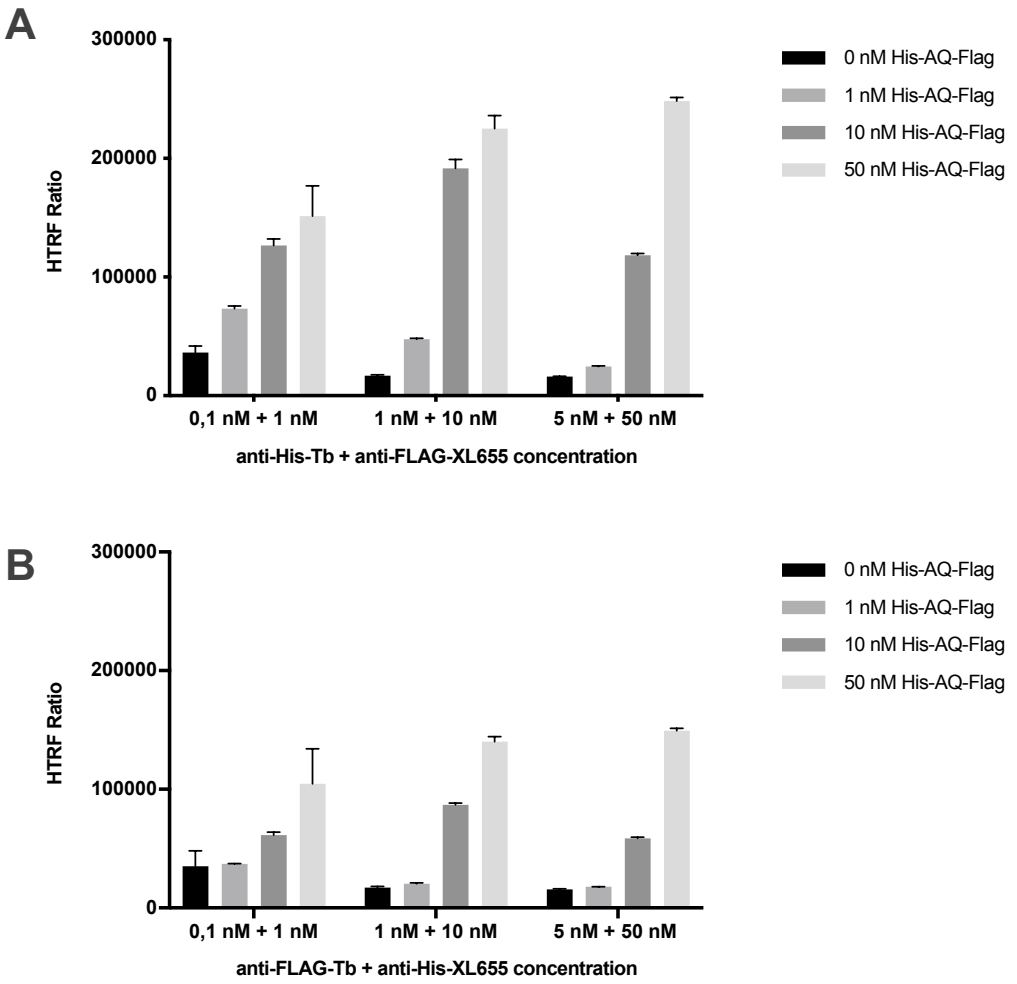


Figure 65: TR-FRET antibody titration over increasing concentrations of control peptide (His-AQ-FLAG). (A) anti-His-Tb and anti-FLAG-XL655 antibody with varying His-AQ-FLAG concentration. (B) anti-FLAG-Tb and anti-His-XL655 antibody with varying His-AQ-FLAG concentration. In both bar charts a significant amount of noise can be observed in the absence of His-AQ-FLAG.

Based on the results from the first TR-FRET experiments (above), mAb concentrations of 1 nM anti-His-Tb and 10 nM anti-FLAG-XL665 were chosen to further validate the HTRF performance of three different GluII mutants. During construct design, the decision was made to clone a version of the β-subunit with a triple repetition of the FLAG-tag (3xFLAG, Table 30) to potentially favour anti-FLAG mAb recognition. Experiments measuring TR-FRET at different GluII mutant heterodimer concentrations showed a qualitatively similar concentration dependence for the HTRF ratios of FLAG- and 3xFLAG-tagged protein (Figure 66). Both mutants show an

increase in HTRF signal strength up to their maximum signal at 10 nM, past which concentration the fluorescence decreases. Comparison at the highest HTRF signal strength shows a 2-fold higher fluorescence of FLAG-labelled GluII compared to 3xFLAG-labelled enzyme. This could be due to each of the following reasons, singularly or in combination: i) interaction of the 3xFLAG-tag with other amino acid side chains of the protein and therefore loss in tag mobility; ii) lower binding affinity of the Ab to the tag; and/or iii) higher flexibility of the tag due to its extension into the media and thus higher flexibility of the Ab/fluorophore. The latter would result in a distance increase between the fluorophores and therefore in a decrease of FRET efficiency.

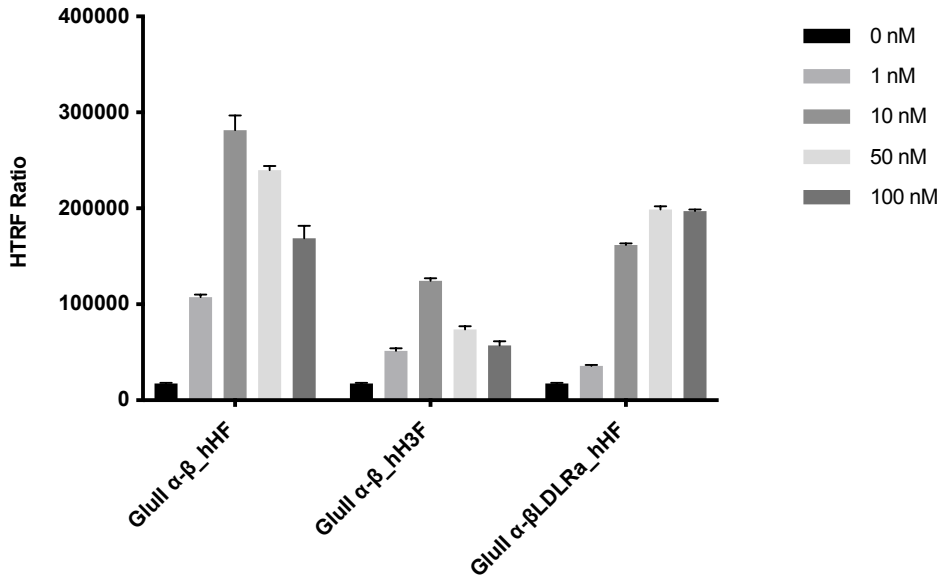


Figure 66: TR-FRET of three different GluII heterodimer mutants at varrying concentrations at constant mAb mix (1 nM anti-His-Tb and 10 nM anti-FLAG-XL665). Legend: *h* – humanised within the *α*-/ *β*-interface. *H* – C-terminal hexahistidine tag on *α*-subunit. *F* – N-terminal FLAG tag on *β*-subunit. *3F* – N-terminal 3xFLAG tag on *β*-subunit. *βLDLRa* – truncated *β*-subunit.

In addition to the two mutants described in Figure 66 the HTRF ratios for *Mm*GluII α-βLDLRa_hHF are shown. This mutant has a truncated β-subunit, i.e. contains only the N-terminal LDLRa domains, while the GluII α-β_hHF construct contains the full-length β-subunit (for more information see section 6.1, Table 30). The differences observed between the TR-FRET signal in presence of full-length vs. truncated β-subunit imply that the α-β interaction depends on β-subunit domains other than the N-terminal LDLRa domains. A hydrogen-deuterium exchange (HDX) study suggested an additional interface between the α- and β-subunit in close proximity of the α-subunit's active site, without reporting on whether this would involve the β-subunit MRH C-terminal domain[201]. In the mutant tested that retains only the two tandem repeats of the LDLRa domains (GluII α-βLDLRa_hHF), no additional interfaces are present other than the one with the LDLRa domains, so it is possible that in the heterodimer retaining the full length β-subunit the disruption of the interface may be more difficult to achieve. By comparing the HTRF ratios of GluII α-β_hHF and truncated α-βLDLRa_hHF (Figure 66) a drop in fluorescence can be observed for GluII α-βLDLRa_hHF at almost all concentrations tested. In contrast to the full length construct, truncated GluII shows a different TR-FRET behaviour: the HTRF ratio levels off at its maximum HTRF ratio with increasing concentrations used, which however is a third lower than the one for GluII α-β_hHF at its maximum HTRF ratio. The reason for the drop in HTRF when using the full-length construct is not known. As the binding specificity of the anti-FLAG Ab should not be different for the mutants tested – both are carrying a single FLAG-tag – the decrease in fluorescence may be a result of a change in the tag's molecular environment. By deleting 75% of the β-subunit's amino acid sequence charged amino acids close to the N-terminus could now

be exposed and in close proximity to the FLAG-tag and thus change its flexibility/accessibility.

After optimising the fluorophores-antibody combinations and concentrations (from here on referred to as 'detection mix'), as well as the use of different GluII mutants (Figure 65 and Figure 66, respectively) we proceeded with the TR-FRET assay development. Due to the overall aim of this study – finding compounds capable of disrupting the native GluII α-/ β-interface *in vivo* – the ideal model assay would use human full length GluII. As human GluII was not available at the time of assay development, the humanised GluII mutant α-β_hHF was the closest mimic candidate and used in our model assay. Furthermore, full-length GluII α-β_hHF shows a stronger HTRF ratio than the truncated GluII α-βLDLRa_hHF mutant at lower concentrations (10 nM, Figure 66). The use of GluII α-βLDLRa_hHF could have been beneficial for the detection of weaker binding fragment candidates, however potential hits would still require testing against full length GluII α-β_hHF. By not using GluII α-βLDLRa_hHF in addition to GluII α-β_hHF we saved time and resources, which we used to screen compounds of a wider chemical space within the given time.

In order to minimise the amount of protein used for compound screening, a titration of GluII α-β_hHF in small incremental steps under the conditions described above was carried out. Figure 67 shows the titration of 0.5 up to 25 nM of protein and the corresponding HTRF ratios. We observed the steep increase in FRET from 0.5 to 3 nM and a subsequent levelling off at higher concentrations. For further TR-FRET experiments a GluII α-β_hHF concentration of 3 nM was chosen as the best compromise between HTRF ratio signal strength and protein amount necessary.

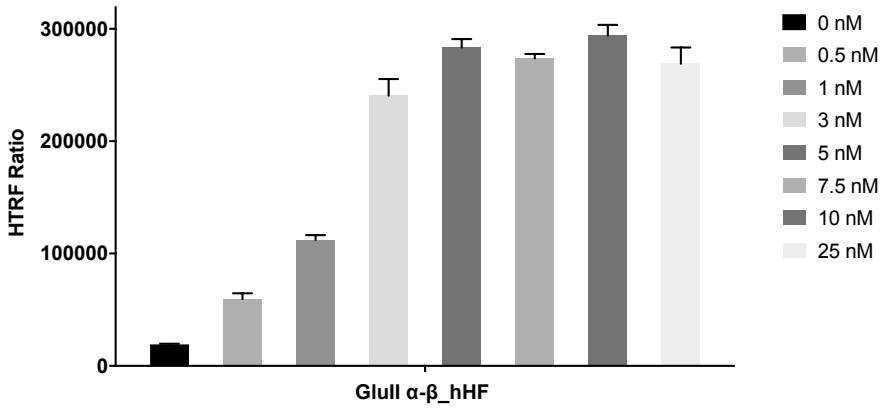


Figure 67: TR-FRET titration experiment of humanized full length GluII (GluII α-β_hHF), using 1 nM anti-His-Tb and 10 nM anti-FLAG-XL665 and a 1h incubation time.

After several rounds of optimisation, the final step of the TR-FRET assay development was to validate suitable FRET-antibody concentrations for the GluII mutant used. Figure 68 shows a TR-FRET titration experiment where the GluII concentration is kept stable and the donor- and acceptor-Ab mixture is varied as indicated. From 0 to 0.25 nM anti-His-Tb the overall TR-FRET signal is increasing with higher concentrations of anti-His-Tb. Between 0.5 and 1.0 nM anti-His-Tb the signal intensity stays about constant with some minor loss of signal at lower anti-FLAG-XL665 concentrations. When going from 1.0 to 5.0 nM anti-His-Tb, the drop in TR-FRET signal intensity is the strongest and reaches almost the same levels as for samples without the donor-fluorophore terbium. The titration of anti-FLAG-XL665 over the different anti-His-Tb concentrations shows a similar pattern as found for the control peptide (Figure 65).

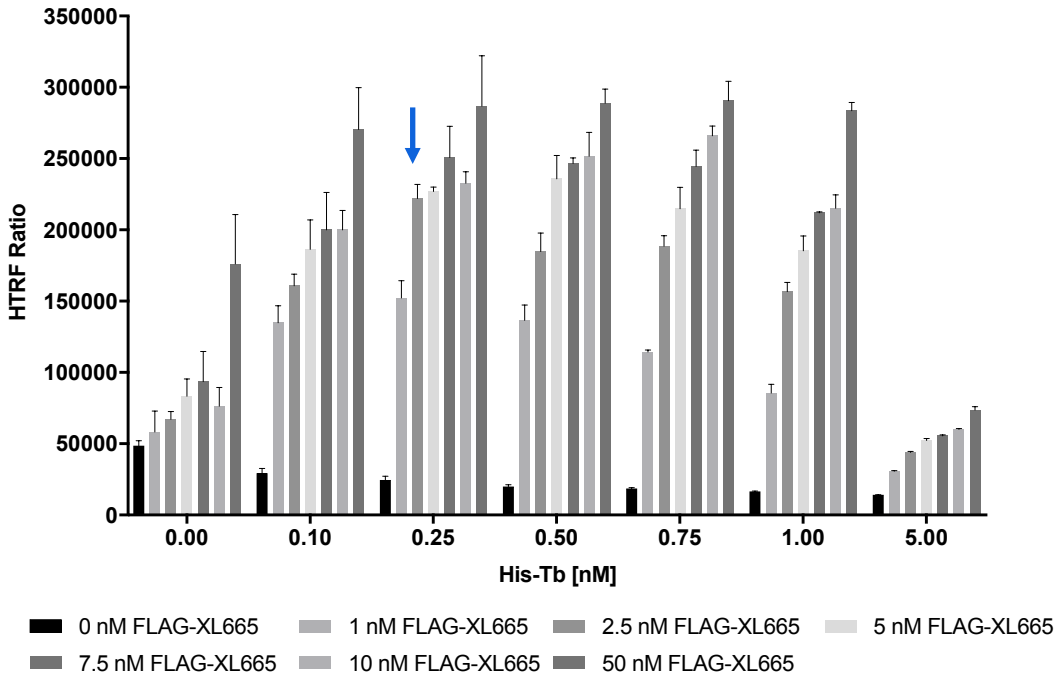


Figure 68: TR-FRET assay with double titration of anti-His-Tb (0.0-5.0 nM) and anti-FLAG-XL655 antibody (0.0-50 nM) using 3 nM GluII α-β_hHF. Reaction time 1h at 37˚C. The arow indicates the concentration of His-TB mAb and FLAG-XL665 mAb that was settled upon for subsequent compound testing.

In order to test for onset of protein degradation, a time course experiment was carried out in the presence of GluII (Figure 69). The experiment is intended to show time related changes in FRET intensity due to fluorophore-Ab and/or mAb and/or GluII degradation. In the absence of GluII no time dependent change of the HTRF ratios is detectable between zero and 180 min, with HTRF intensities around 100,000. This experiment having provided a baseline for the signal, measurements are repeated in the presence of 3 nM GluII heterodimer. The results show about four-fold higher HTRF values, with a slight increase over time and a minor dip at 75 min. The slightly higher HTRF ratios at later time points could be a result of weak interactions between the Abs and the protein tags, resulting in more Ab-protein interactions with longer timeframes. Another reason could be that the binding of the Abs lowers the heterodimer's diffusion coefficient due to an overall bigger size of the complex. The

additional mass would slow down the on-and-off rates of the α- and β-subunits with the effect that the fluorophores stay longer within the Förster distance. In addition, mAb aggregation cannot be excluded.

When comparing the HTRF signal strength at minutes 75 and 180 the difference between the two is about 20,000. Based on an average HTRF ratio of about 100,000 and the overall error margin of about +/- 10,000 an incubation time of 60 min was chosen for future experiments, allowing to perform repeat measurements in the time given.

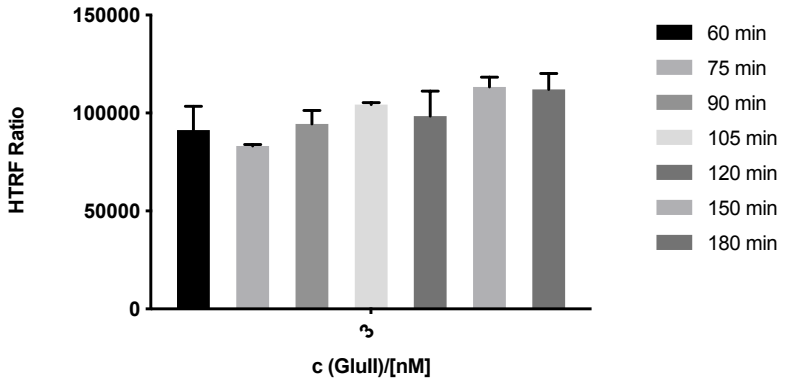


Figure 69: TR-FRET of GluII α-β_hHF at varying incubation times. Samples were incubated at 37˚C for the times stated prior to additon of detection mix. Protein concentration: 3 nM, detection mix: 0.25 nM His-TB mAb and. 2.5 nM FLAG-XL665 mAb.

After the initial FRET assay development scouting experiments described so far, subsequent experiments were carried out under conditions stated in section 5.6.2, Table 26.

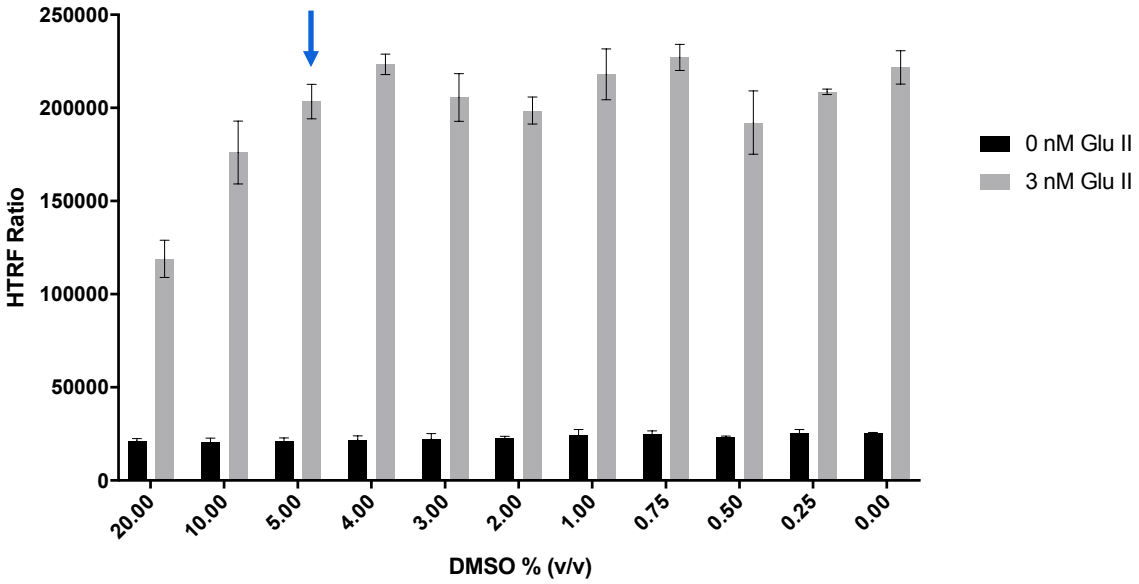


Figure 70: TR-FRET GluII α-β_hHF DMSO tolerance. Measruements were done at zero and 3 nM protein with DMSO concentrations ranging from 0-20% (v/v). Incubation time 1h at 37˚C. The arrow indicates the concentration of DMSO that was settled upon for subsequent compound testing.

Due to the design of the high throughput screens and the low water solubility of most of the compounds included, all screens were supplied with compounds solubilised in 100% DMSO. I first tested the DMSO tolerance of GluII in the TR-FRET assay. The results in Figure 70 show a TR-FRET experiment using the conditions outlined above with the addition of increasing DMSO concentrations (%, v/v). Besides minor assay-related fluctuations in the FRET intensities in the range from 0 to 5% DMSO, the results show a strong decay in fluorescence at DMSO concentrations > 5%. The findings agree with the DSF studies done previously, indicating protein unfolding at > 3% DMSO (see section 6.5.1). The DMSO tolerance of GluII would allow the testing of compounds of up to about 5 mM with a final DMSO concentration of about 5% under TR-FRET conditions. One of the main challenges of the TR-FRET assay development was the establishment of a positive control. The positive control should

be able to reliably interrupt the α-/ β-interface and would be run in parallel to the compounds tested.

First attempts to destabilise the α-/ β-interface using high concentrations of salt (up to 300 mM) showed no success. A second approach based on the removal of the two LDLRa-bound Ca^{2+}-ions with a chelating agent (up to 100 mM EDTA) showed no effect either (data not shown). A change in strategy finally delivered the desired results. Instead of chemically weakening the GluII α-/ β-interface, an untagged β-subunit mutant (GluII β_h, section 6.1, Table 30) was used to compete off its bound FLAG-tagged counterpart. Figure 71 shows the titration of 0.01 to 310 nM untagged β-subunit onto 3 nM GluII α-β_hHF. ('primary assay', two pairs of technical duplicates on subsequent days, red and orange graph, respectively). To be able to distinguish changes in the HTRF ratio introduced by the separation of the two GluII subunits and not by an unspecific interference of the antibody and/or fluorophore with the compounds screened, the control peptide described before ('counter assay', technical duplicate, blue graph) was run in parallel. Figure 71 shows the first measurement taken with the newly developed GluII TR-FRET assay. Instead of the HTRF ratio, the y-axis gives the fluorescent readout in percentage activity plotted against increasing compound concentration on the x-axis. The percentage of activity is a quantitative measure and results from the maximum decrease in fluorescence of the positive control which was run on each plate in parallel to the compounds tested. For a detailed plate layout see appendix Figure B. 6.

At about 0.3 nM GluII β_h the curves in Figure 71 indicate a change in FRET activity for the GluII heterodimer (red and orange curves, representing independent series of

measurements), while the counter assay (blue graph) stays at baseline level. The trend continues in a sigmoidal fashion and a maximum is reached at about 310 nM GluII β_h (end of titration). At all GluII β_h concentrations tested the control assay shows no changes in fluorescence. This indicates that the observed decrease in fluorescence is a result of the untagged GluII β_h competing away the tagged version of the protein, and not due to unspecific interactions between Glu II and the detection mix.

The experiment confirms the use of TR-FRET as a technique suitable to analyse dissociation of the GluII α-/ β-subunits and supports the use of untagged β-subunit as a positive control for future experiments.

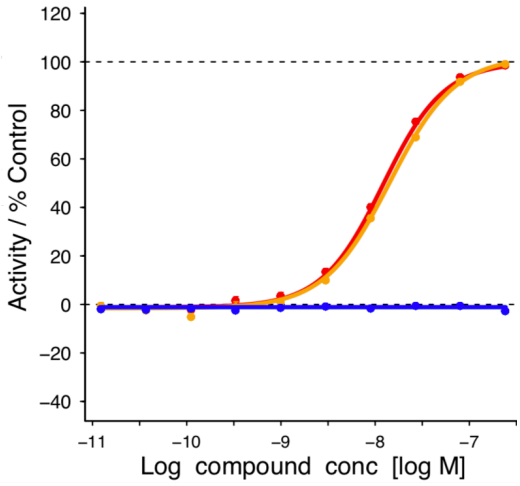
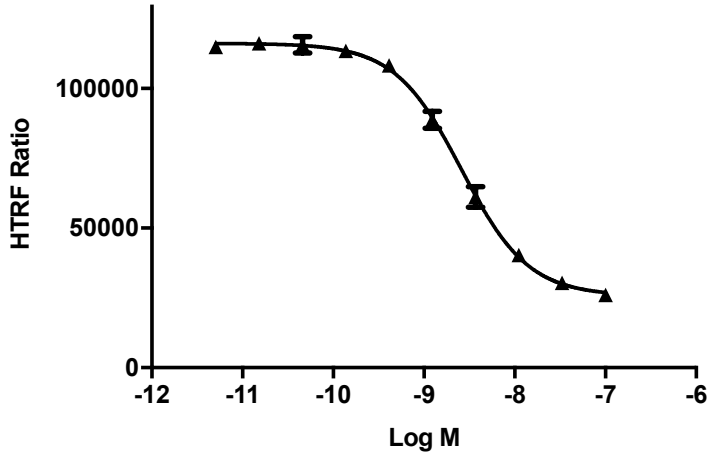


Figure 71: TR-FRET assay positive control. Shown is the change in activity (decrease in flourescence) over the titration of GluII β_h into GluII α-β_hHF loaded wells. *Red and orange graph* – Duplicates under same assay condition carried out on subsequent days. *Blue graph* – counter assay

A dissociation constant (K_d) for the GluII α-/ β-subunits has never before been reported. By fitting a sigmoidal function to the data points shown in Figure 72 and evaluating its inflexion point (1st derivative peak concentration), a K_d of 2.59 nM was calculated under the conditions used. The low K_d verifies the strong affinity between the GluII subunits and the difficulties faced separating them using chemical

approaches. Interferometric scattering mass spectrometry (ISCAM) carried out by Dr Weston Struwe on GluII α-β_hHF resulted in a K_d of 2.3 ± 0.7 nM, which is in a good agreement with the TR-FRET experiments carried out in this study. For more information regarding ISCAM see appendix Figure B. 7.



HTRF ratio		
Top	116170	[-]
Bottom	25723	[-]
Log IC_{50}	-8.58	[log M]
Hill Slope	-1.17	[HTRF/log M]
K_d	$2.59e^{-9}$	[M]

Figure 72: TR-FRET titration experiment of increasing GluII β_h concentration over 3 nM GluII α-β_hHF. Experimental conditions are as stated above. Table below graph summarises the values of the K_d validation. *M* – GluII β_h concentration in moles.

6.7.2. GluII TR-FRET compound screening

The initial results of the X-ray crystallography based FBLD screen, performed by Drs Alessandro Caputo and Pietro Roversi in 2016, served as a starting point for the GluII TR-FRET assay compound selection. Overall, selection was designed around three main strategies:

i) X-ray crystallography FBLD hits

From the FBLD screen comprising around 1000 fragments, 12 fragments were identified that bound to GluII, of which three were clustered in the same location on the GluII α-subunit in close proximity to the α-/ β interface (Figure 73, A - C). The fragments shown in Figure 73 bind in a pocket on the surface of the α-subunit and are too small to reach the actual interface but formed the basis for the design of new compounds extending towards it. NZ_IF-3 is the most promising fragment, with its pyrazole ring sitting closest to the α-/ β-interface. Synthesis and purchasing of NZ_IF derivatives was done by Dr J.L. Kiappes in collaboration with Evotec A.G.. In total 41 NZ_IF derivatives were taken forward to be tested by the GluII TR-FRET assay.

ii) Crystal breaking compounds

Besides the 12 hits found during the FBLD screen, another observation was made. About 6.5% of the total ~1000 GluII crystals did not survive the soaking procedure and either dissolved or did not diffract X-rays. The observation caught our attention, because all crystals originated form the same batch, and soaking DMSO concentrations were kept below 3% (v/v). The latter therefore cannot explain loss of diffraction by some crystals and not others. This suggested that the compounds that either destroyed the crystals or caused loss of diffraction were disrupting the crystal lattice. Disruption of the crystal lattice may have occurred at generic crystal contacts, but we did not want to rule out the possibility that some of the fragments may be able to dissociate the GluII subunits. Therefore we purchased and tested the 66 fragments of the original FBLD screen that either destroyed the crystals or induced loss of diffraction. As the original fragment-based lead discovery approach was designed to

cover the largest chemical space possible, the fragments belong to many different chemical scaffolds.

iii) D-mannose virtual simulations

More recently and not as part of the crystallography-based FBLD screen, a D-mannose moiety was discovered by Dr Pietro Roversi to bind at the GluII heterodimer interface (Figure 73, D). To further investigate the area around this mannose binding patch computer aided drug design (CADD) was carried out by Dr Someina Khor and Dr Tim James (Evotec A.G., Abingdon, UK). During virtual compound modification αC844, which is in close proximity to the D-mannose binding site, was thought to be a good candidate for the formation of a covalent linkage between compound and protein. The accessibility and reactivity of the cysteine side chain has yet to be confirmed, but by using the known fragment-protein interactions of the D-mannose binding pocket as a first anchor, a functional group, such as an acrylamide group, could be added to molecules binding in this pocket to favour the formation of a covalent linkage to αC844. The formation of a linkage in combination with the D-mannose binding side could add to the interface-disrupting efficacy of new compounds.

For GluII TR-FRET assay experiments 86 commercially available acrylamide fragments as well as 82 compounds synthesised based on the results of the D-mannose virtual screen were tested.

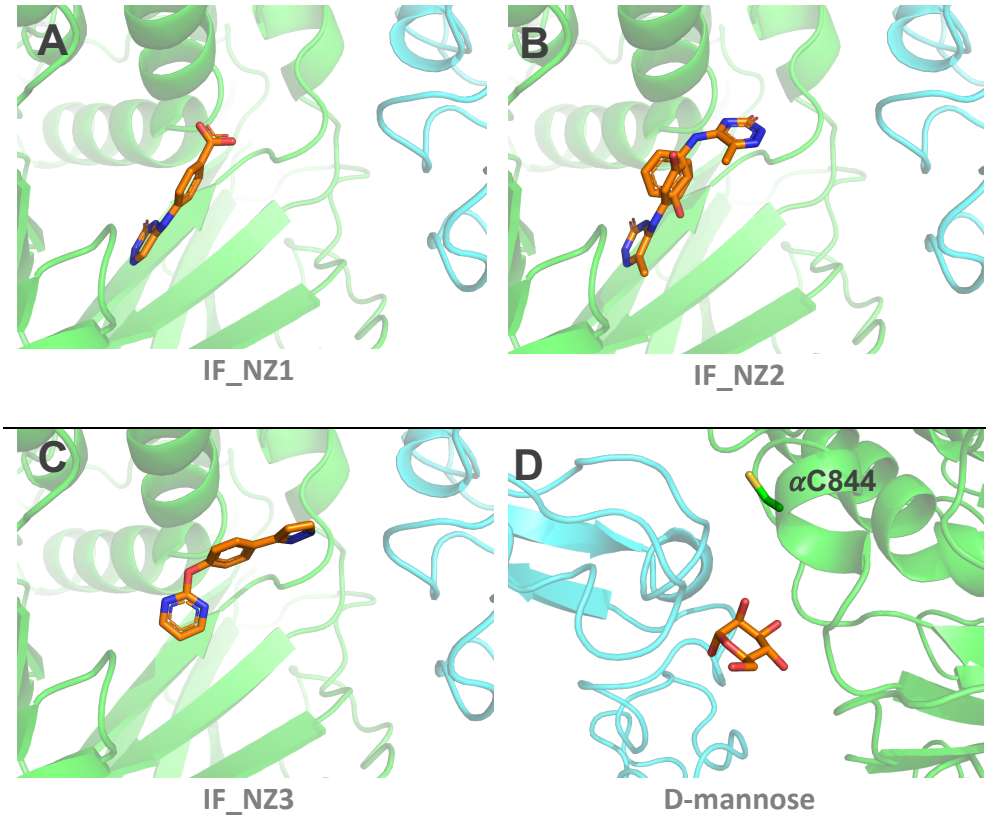


Figure 73: Structures of GluII bound ligands found in previous, X-ray crystallography based studies. (A-C) FBLD hit NZ_IF1 (A), IF_ZN2 (B), IF_NZ3 (C) and mannose binding side (D) in close proximity to the GluII α-/β-interface and αC844 (stick representation) as found by Drs Alessandro Caputo and Pietro Roversi.

6.7.3. Assay evaluation

In total 275 compounds were screened using the GluII TR-FRET assay designed for the analysis of compounds capable of disrupting the GluII α-/ β-subunit interface. The assay as well as the compounds indicate robust and reproducible readouts, also across different protein batches. The Z' values – a measure for the raw score position in regard to sample mean – are typically high (0.76 – 0.84) and repetitions of the same experiment intra- and inter-days resulted in XC50 variabilities within acceptable bounds (~ 3-fold).

The evaluation of the GluII TR-FRET assay shows a high diversity in response to the compounds tested. Figure 74A-D demonstrates the four most common types of TR-FRET readouts observed.

Figure 74A shows a typical case where none of the compound concentrations tested show any influence on primary or counter assay readout, indicating no change in GluII quaternary structure.

The graphs in Figure 74B describe an ideal case of an assay readout for the positive control described in section 6.7.1. The primary assays (day1 and 2, blue and red, respectively) are at baseline level for lower compound concentrations up to about 0.3 nM, but show a sigmoidal change in activity at higher concentrations. The counter assay stays at baseline level throughout the whole concentration range, indicating no cross reaction between compound screened and detection mix.

In close relation to the graphs described in Figure 74B, but observed more frequently, is a primary assay activity increase at higher compound concentrations in combination with an associated in- or decrease in counter assay activity (Figure 74C). Depending on the compound screened the shift occurs at similar concentration as the primary assay or is delayed by a certain concentration. Changes in the counter assay's fluorescence cannot be addressed to fluorescent properties of the compound exclusively, as intermolecular interactions between protein, compound and/or Ab might also influence the assay outcome, and we have not established the reasons behind the behaviour observed.

A small number of compounds tested showed a negative response in assay activity for both primary as well as counter assay (Figure 74D). A negative response indicates an increased energy transfer caused be a shorter Förster distance, i.e. fluorophores are in closer proximity to each other, or interference of the compound tested with the

fluorophore system used. The former can either be introduced by a GluII interface stabilisation or the compound-mediated Ab-protein interactions.

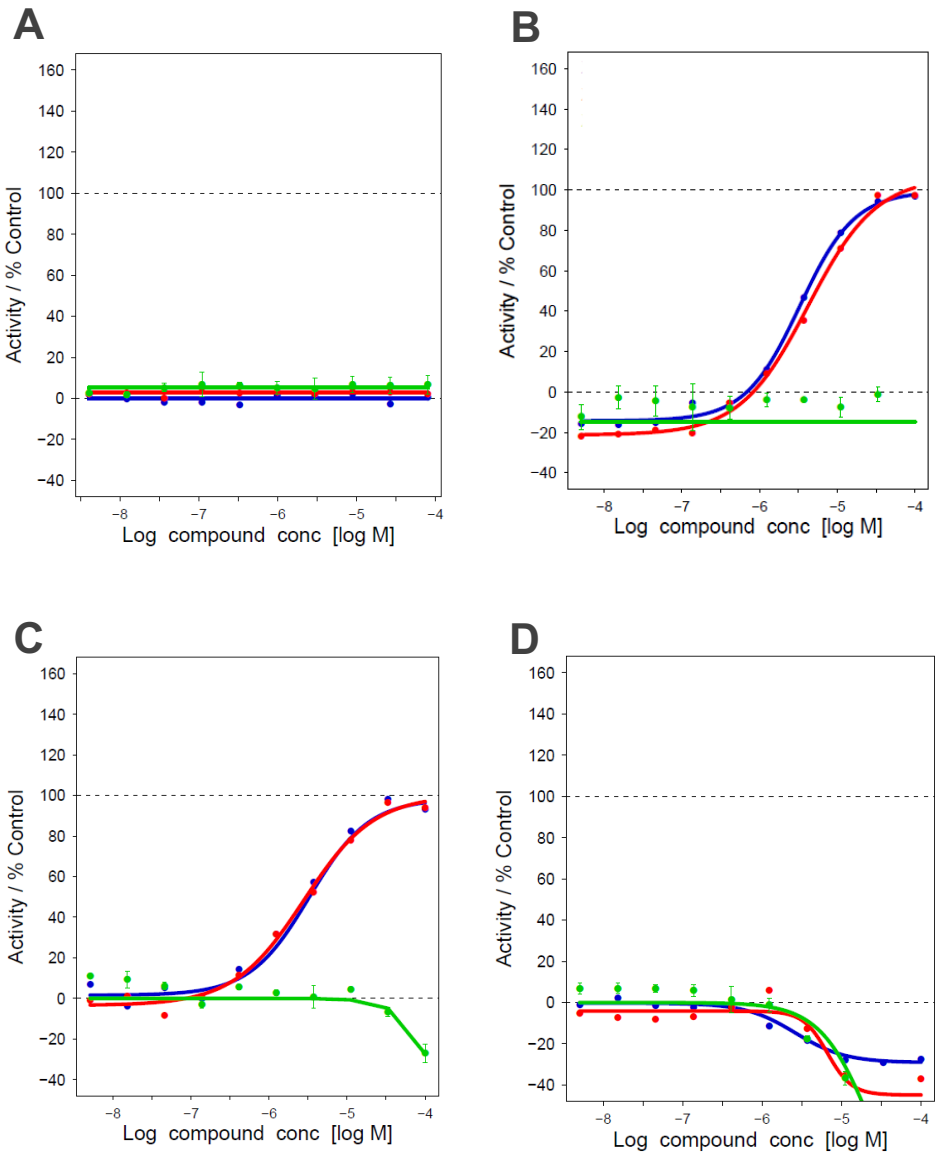


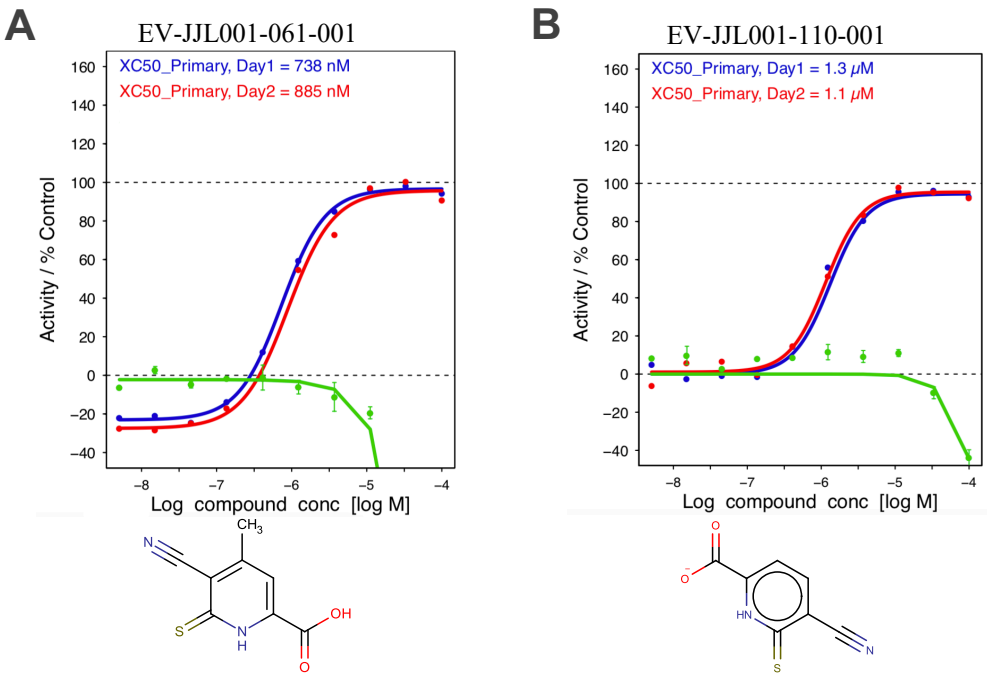
Figure 74: Illustration of the four most common types of TR-FRET readouts observed in this study. Shown is the change in activity (decrease in flourescence) over the compound titration into GluII α-β_hHF loaded wells. *Blue and red graphs* – primary assay of day 1 and 2, *Green graph* – counter assay.

Figure 75 shows a list of TR-FRET hits with a positive primary response and XC50 values below 5.0 µM. In total eight compounds were identified as potential GluII α-/β-interface disruptors. The list includes active compounds with negative responses in

the counter assay as to date we do not know what causes this phenomenon, and our data are too limited to draw a conclusion about which samples to in- or exclude in the list of TR-FRET hits at this stage.

The most potent GluII interface splitting compound was EV-JJL001-061-001 from the CADD-based compound screen. EV-JJL001-061-001 is the only TR-FRET hit showing XC50 values in the upper nM range. Within the TR-FRET hits all compounds are from one chemical scaffold (thiopyridones).

Eleven out of the 275 compounds had some sort of stabilising effect on the protein system used with negative responses in both primary assays (data not shown).

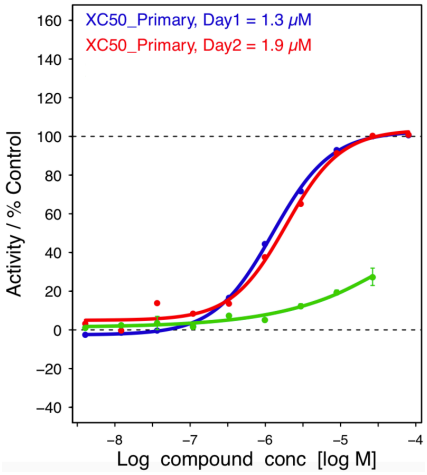


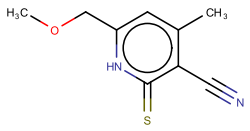
XC50: 0.8 µM

XC50: 1.2 µM

C

EV-ZC0151001

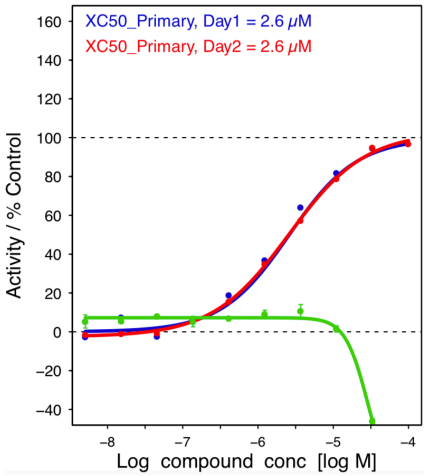


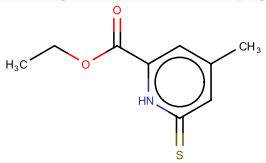


XC50: 1.6 µM

D

EV-JJL001-118-001

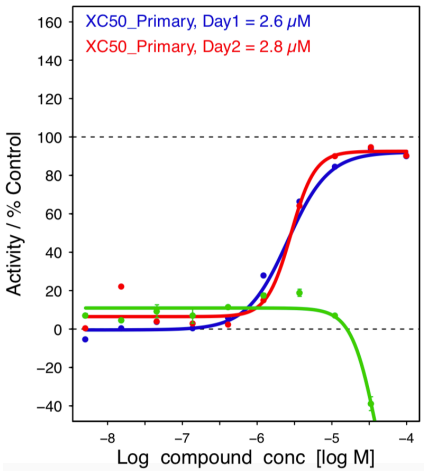


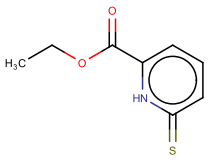


XC50: 2.6 µM

E

EV-JJL001-114-001

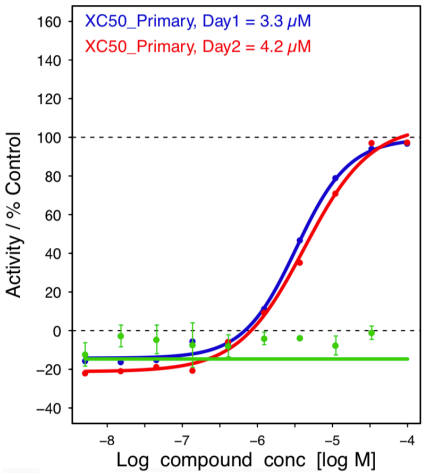


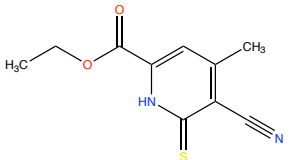


XC50: 2.7 µM

F

EV-JJL001-060-001





XC50: 3.8 µM

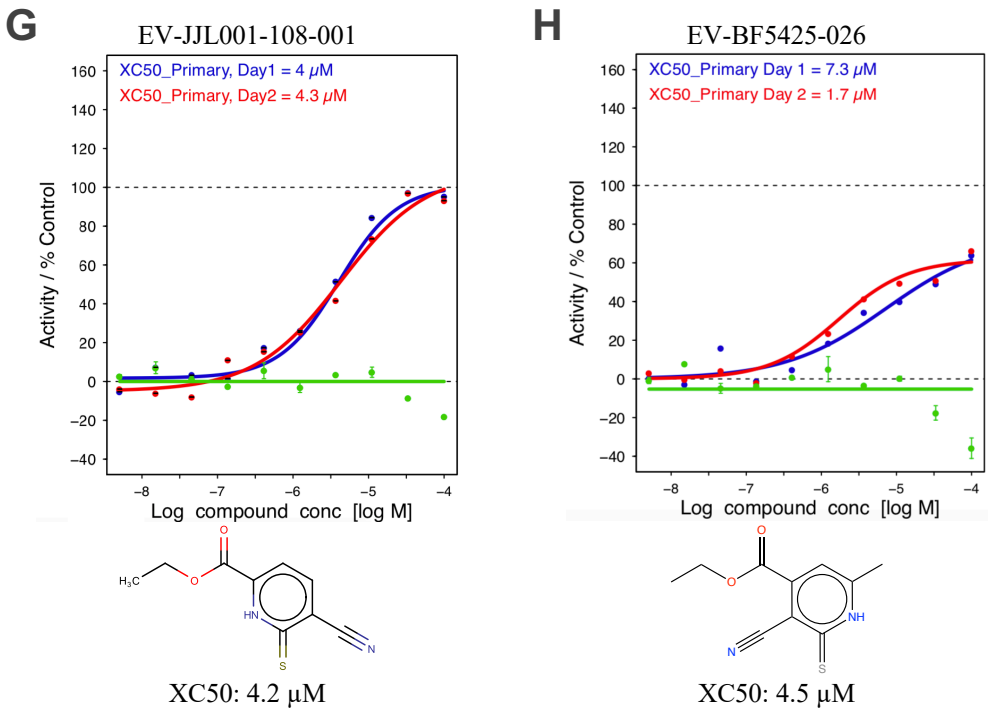


Figure 75: TR-FRET assay hits. Compounds are ordered by decreasing GluII interface splitting efficiency. Y-axis shows the change in activity (decrease in fluorescence) over the compound titration into GluII α-β_hHF loaded wells. *Blue and red curves* – primary assay of day 1 and 2, *Green curve* – counter assay.

Compound EV-JJL001-061-001 – originating from the D-mannose virtual simulations – shows good TR-FRET activity and only causes a small drop in the counter assay readout at higher concentrations (Figure 76A). The proximity of a cysteine to the predicted binding site led us to speculate whether the compound's sulphur group may be involved in a covalent linkage to the protein, and so we decided to perform some limited structure activity relationship (SAR) studies, aimed at figuring out the importance of the sulphur atom. The addition of a methyl group to the ring secondary amine resulted in a total loss of compound activity, with a negative response in primary as well as counter assay performance at about 1 µM (Figure 76B). The loss could be explained by a steric hindrance towards the sulphur group in ring position 5. The TR-FRET results suggest the need of a chemically accessible sulphur group for the

compound to be active as a GluII α-/ β-interface disruptor. By trapping the sulphur between methyl- and nitrile groups the compound seems to be less reactive. Even though, the activity cannot assign to the sulphur group exclusively because the exact binding pose is still unknown. Mass spectrometry experiments are planned to check if this sulphur atom is involved in forming a covalent link to the target protein.

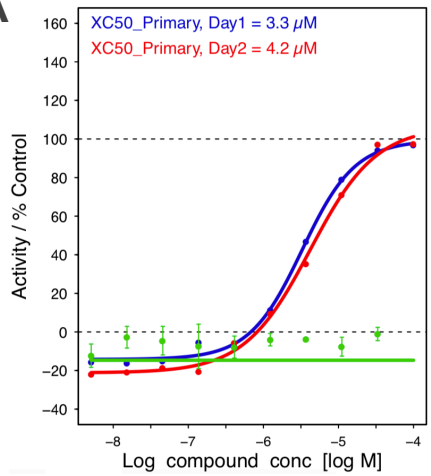
Modification of the compound's ethyl group shows no effect in the primary assay but causes a clear shift in the response of the counter assay, which changes from a previously negative to a positive response (Figure 76C). The shift in counter assay activity could be caused by the newly formed carboxylic acid and its interactions with the detection mix.

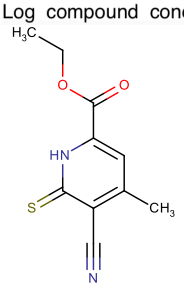
By moving the methyl group from position +3 onto the sulphur group and thus forming a thioether shows similar results as described in Figure 76B and C, respectively (Figure 76D-E).

The synthesis of the four compounds described in Figure 76B-E was the first approach to increase our understanding of the compound's binding behaviour with the target protein GluII. The modifications chosen gave some insights into potential atom(s)/-groups involved in binding but fail to provide the full picture of the compound's binding motif for which we have planned a mass spectrometry based approach in future.

A

EV-JJL001-060-001

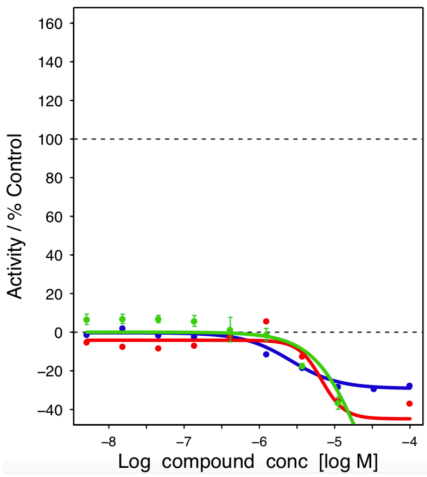


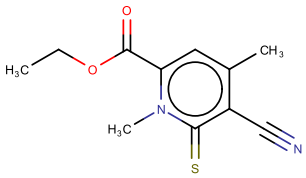


XC50: 3.8 µM

B

EV-JJL001-121-001

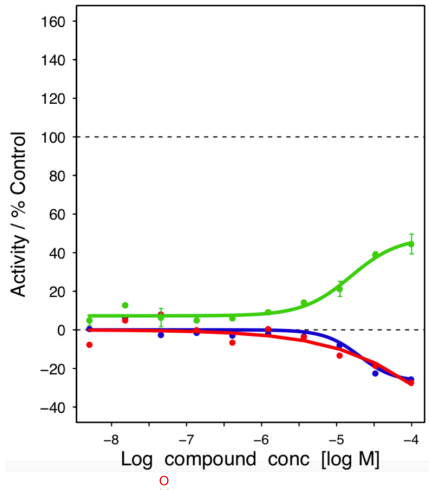


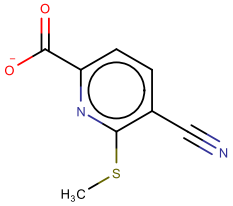


XC50: -na-

C

EV-JJL001-123-001

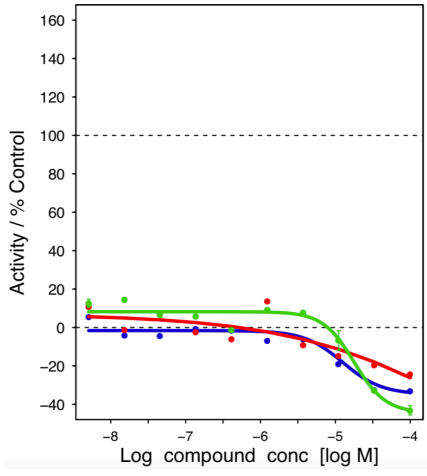


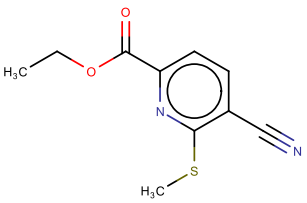


XC50: -na-

D

EV-JJL001-113-001





XC50: -na-

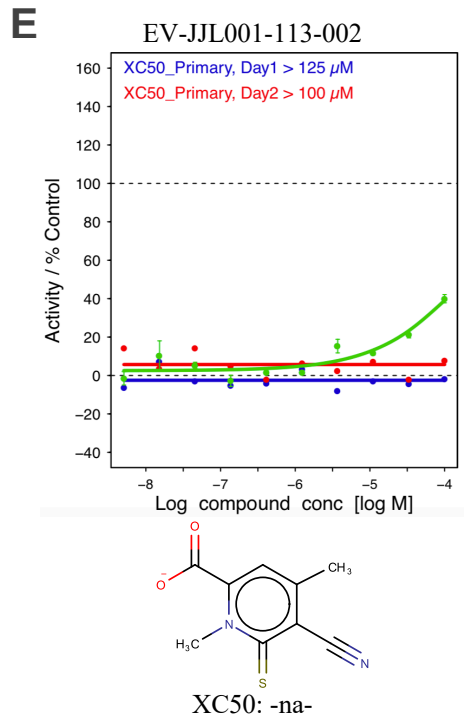


Figure 76: TR-FRET assay-based structure activity relationship study of EV-JJL001-060-001 derivatives. *Blue and red curves* – primary assay of day 1 and 2, *Green curve* – counter assay.

In addition to the compounds of the thiopyridone family, NZ_IF2 shows FRET activity in the primary assay with an additional signal rise in counter assay (Figure 77A). The compound has an XC50 of about 25 µM. The concentration separation between the two responses – primary and counter assay – is not as pronounced as shown for compounds of the thiopyridone family (Figure 75), leading to the hypothesis that the compound is acting via a non-specific FRET interference mechanism. To test this hypothesis 19 additional analogues of IF_NZ2 were acquired of which most showed no FRET activity and only one an improved separation of primary and counter assay response with a five-fold increase in XC50 compared to IF_NZ2 (Figure 77B).

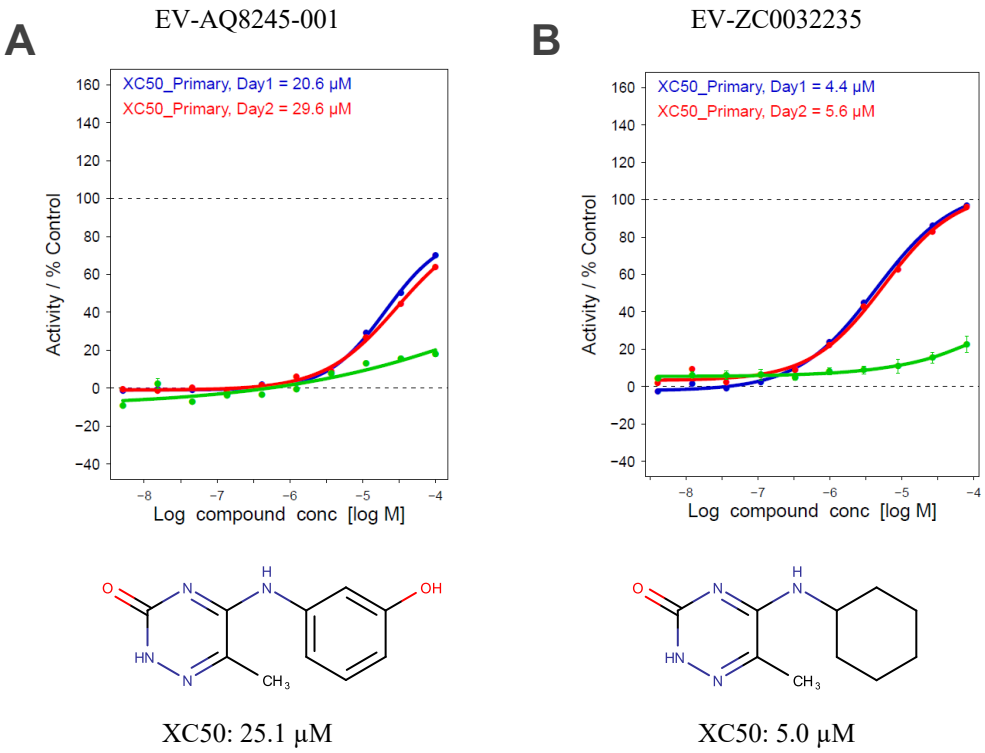


Figure 77: TR-FRET assay-based structure activity relationship study of IF_ZN2 analogues. *Blue and red curves* – primary assay of day 1 and 2, *Green curve* – counter assay.

6.7.4. GluII TR-FRET assay conclusion

Overall the development of the GluII TR-FRET assay for the detection of GluII α-/ β-interface interfering compounds was a success. The process included the production of humanised murine GluII, assay development and the first round of compound screening with the establishment of a positive control and the use of a counter assay for the detection of cross-reactions between compound and reaction mix. The assay delivers robust and reproducible results under compound screening conditions (section 6.7.1) and the first kinetics determination for the GluII α-/ β-subunit confirmed the expected high affinity binding between α- and β-subunits, with a K_d of about 2.6 nM. As the compounds chosen originated from different sources and are based on four

different strategies of potential interface disruptor design, their chemical properties cover a wide chemical space (see section 6.7.2).

Results in this thesis describe the first round of compound screening for the detection of GluII α-/ β-interface disruptors and in addition to identifying a total of nine potential interface disrupting compounds in the first screening round, we have conducting a limited SAR study around compound EV-JJL001-061-001 (section 6.7.3).

Overall, the GluII TR-FRET assay developed seems to be a good platform for the detection of future GluII inhibitors based on the disruption of the α-/ β-interface.

To support the TR-FRET results and to further screen for strong interface disruptors the Zitzmann lab also invested in the development of a GluII-based ELISA with an immobilised GluII α-subunit. Future studies will show whether the way the subunits are presented in the two assays delivers similar results, and how much either system is influenced by the properties of the compounds tested.

Of the total 275 compounds screened, nine were identified as potential GluII α-/ β-interface disruptors, with varying responses in their counter assay activity (section 6.7.3). The majority of the counter assay shifts observed occur at higher compound concentrations (> 300 nM) and appear after a response in primary assay. The reason for this phenomenon is still unknown, but investigations are underway to analyse the emission/absorbance characteristics of each compound, in order to rule out UV-absorbance interference. As UV absorbance interference seems to be a risk for some compounds, we have ruled out a thermal shift based approach (such as DSF).

A shortcoming of the GluII TR-FRET assay is the lack of structural information provided regarding compound binding. The TR-FRET activity indicates a change in protein quaternary structure and cannot be translated into binding motifs, neither is it able to determine the site of binding on the GluII heterodimer. We only know the intended binding site on the GluII α-subunit of a few of the compounds tested (i.e. NZ_IF derivatives and compounds designed around the accessory D-mannose binding side, section 6.7.2), all other compounds tested could bind on either the α- or β-subunit. The design of new inhibitors based on initial TR-FRET hits is very limited due to the uncertainty of binding partners involved. To increase the inhibitor-protein affinity structural information is needed in order to steer compound design into right direction.

Saturation-transfer difference (STD)-NMR spectroscopy on GluII TR-FRET active compounds (see section 6.7.3) was carried out by Ms Zala Sekne and Dr J.L Kiappes (Zitzmann lab) in order to gain structural information about compound binding. The technique reports on the magnetization transfer from excited H nuclei on the protein to the H nuclei of a binding compound. Comparing 'binding epitopes' indicates compound H nuclei that are in closer proximity to the protein and therefore show a greater magnetisation transfer compared to compound H nuclei at a greater distance from the protein[335,336].

STD-NMR helps to get a first impression of compound binding, even though the technique cannot report on potential binding motifs on the protein, which would help to further understand compound binding. This shows the need for other methods providing structural information about compound-protein interactions.

Another technique that could provide more structural information about the amino acids involved in compound binding is protein X-ray crystallography. For this thesis the method has previously been used to successfully determine the binding poses of *N*-alkyloxy-iminosugars within the GluII active site (section 6.6). For protein X-ray crystallography of putative GluII α-/ β-interface disruptors see section 6.8.

6.8. Crystallisation of *Mm*GluII with α-/β-interface disruptors

X-ray crystallography is a powerful tool for structure determination of macromolecules[337]. The technique was used to determine the structure of the truncated *Mm*GluII α-βLDLRa (section 6.4) and it helped to understand the binding pose of seven different *N*-alkyloxy-DNJs within the active site of *Mm*GluII (for more information section 6.6).

Soaking *Mm*GluII (trypsin truncated) crystals with GluII interface disruptors we attempted to gain information about their binding poses and binding sites, in an attempt to inform the design of compounds capable of more efficiently dissociating the GluII α-/ β-subunit interface.

Following the method described in section 5.7, *Mm*GluII$_{tryps}$ crystals were soaked with different concentrations of the interface disruptors found by TR-FRET (section 6.7.3). The growth of fresh *Mm*GluII$_{tryps}$ crystals suitable for compound soaking failed after several attempts under the conditions described in section 5.7. Attempts included the use of different protein batches of *Mm*GluII$_{tryps}$ (fresh and thawed) and the two-dimensional screening around the initial orthorhombic *Mm*GluII$_{tryps}$ crystal growth condition by variation in pH and precipitant concentration. As none of the above resulted in crystals, crystallisation attempts were widened to the conditions for growth of monoclinic and trigonal *Mm*GluII$_{tryps}$ crystal forms previously found by Dr A. Caputo. The setup of new plates with conditions favouring monoclinic and trigonal crystal growth resulted in small bipyramidal (trigonal) crystals and little rods with split ends (star-like, monoclinic) crystals. Some of the trigonal crystals were soaked with compounds and diffraction measured (Figure 79A and B). As the majority of the new crystals observed were not suitable for compound soaking, crystal containing wells

were pooled and ground up for crystal seeding experiments, in combination with the growth solutions prepared initially for monoclinic, orthorhombic and trigonal crystals, as well as the Morpheus HT-96 three-dimensional screen (Molecular Dimensions). The crystal seeding method resulted in bigger crystals which could be taken further for compound soaking. Figure 78 shows the comparison of crystal growth attempts of *Mm*GluII$_{tryps}$ in a commercially available screen (Morpheus HT-96, Molecular Dimensions), with and without crystal seeding.

Due to time limitations and restricted access to synchrotron beamtime towards the end of my DPhil, different *Mm*GluII$_{tryps}$ crystals, grown under a variety of conditions, were used for the soaking experiments described in section 5.7.



Figure 78: *Mm*GluII$_{tryps}$ crystal growing attempts using the Morpheus HT-96 screen (Molecular Dimensions). (A) Screen setup with protein and crystallisation condition mixed at a 2:1 ratio (v/v). (B) Protein, crystallisation condition and seeding solution mixed at a 2:0.5:0.5 ratio (v/v). Both plates were incubated at 18˚C for several days. *Orange squares* – wells containing unsuitable crystals for planned compound soaking experiments. *Red squares* – wells containing suitable crystals for compound soaking experiments.

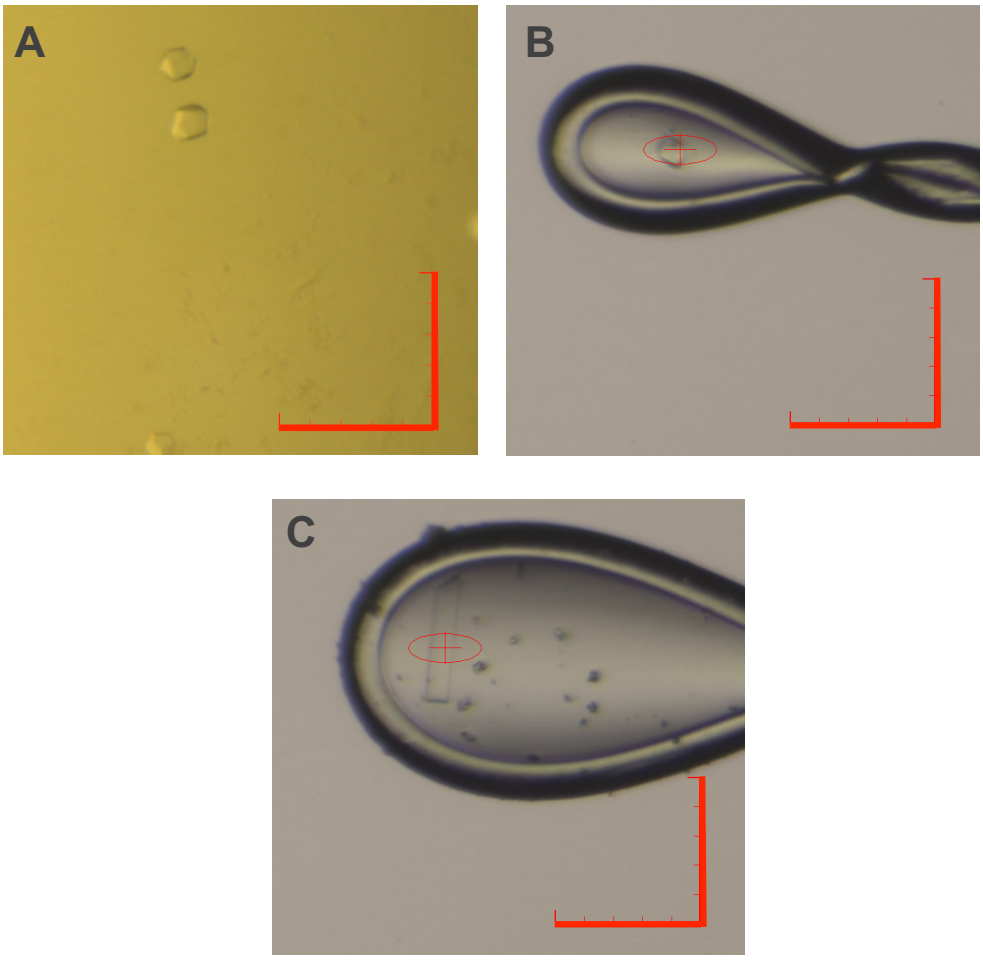


Figure 79: Different morphologies of *Mm*GluII$_{tryps}$ crystals. Protein, crystallisation condition and seeding solution were mixed at a 2:0.5:0.5 ratio (v/v) and incubated at 18˚C. (A) Trigonal bipyramidal crystals as found in crystallisation condition after several days of incubation. (B) Single crystal from A) mounted in nylon loop. (C) A monoclinic crystal rod found in a different crystallisation solution. (B and C) *Red cross with circle* indicates X-ray beam size. *Red scale at bottom right* marks 100 x 100 µm. Loop pictures were taken at Diamond Light Source, beamline i03.

Table 38 is an attempt to summarise the different soaks carried out using different compounds and compound concentrations. According to the results of the GluII TR-FRET assay (section 6.7.3 and 6.7.4) the identified compounds have the potential to disrupt the GluII α-/ β-interface. GluII crystal soaks with these compounds therefore entail the risk that the disruption of the α-/ β-interface introduces changes in the crystal lattice. As a consequence, the crystals might lose their ability to diffract X-rays, or get destroyed completely. This would support the findings of the GluII TR-FRET assay,

but would not provide the desired additional structural information about how these compounds bind.

As the TR-FRET compound soaks may introduce changes in the crystal lattice, a crystal was first shot without any soaked compound, and crystals from the same drop where then soaked with DMSO to final concentrations of 5, 2 and 1% DMSO (Table 38, sample number 0.0 – 0.1) and also subjected to X-ray diffraction. The majority of the untreated crystals diffracted to about 3.8 – 2.2 Å and showed a DMSO tolerance of up to 2%. Subsequently, crystal soaks were carried out on crystals grown in the same drop as highly diffracting crystals, using 1 and 2 mM compound (i.e. at 1 and 2% DMSO) for each of the TR-FRET active compounds listed in Table 38.

For the compounds listed in Table 38, datasets could be obtained from crystals soaked with 1 or 2 mM compound (corresponding to 1 or 2% DMSO), but none of the datasets shows any electron difference maps which could be assigned to successfully bound compound.

The trigonal crystals of *Mm*GluII$_{tryps}$ were detected as twinned by Phenix Xtriage[338]. After a first round of data processing, utilising automated space group assignment by AutoPROC[305], molecular replacement using Phaser[307] and subsequent refinement carried out in REFMAC5[310], as autoBUSTER[309] cannot refine models against twinned data, refinement could not lower R-factors below 0.35-0.40, so the space group assignment was questioned.

Previous attempts by Dr Alexandro Caputo and Dr Pietro Roversi to solve the structure of a twinned trigonal *Mm*GluII$_{tryps}$ data set resulted in good refinement statistics in space group $P3_2$. Reprocessing of the fourteen data sets described above in $P3_2$ and subsequent rigid-body refinement with REFMAC5[310], taking into account possible twin operators, proved to be correct with good refinement statistics and R-factors

between 0.18-0.25. Evaluation in coot[308] for potential different density maps corresponding to the soaked compounds remained unsuccessful.

The reason for their absence is yet to be analysed. Potential causes for this could be a found in the low affinity of the compounds tested to the target protein GluII. Even though the evaluation of the TR-FRET assay experiments (section 6.7) show XC50 values in the low millimolar range. To increase compound soaking concentration new stock solutions in DMSO could be prepared aiming for compound concentrations > 100 mM.

Another influence could be found in the low water solubility of some of the TR-FRET active compounds tested. While adding some compound stock solutions to the crystal soaking solution coloured precipitation occurred in some cases corresponding to the compounds colour solubilised in DMSO. Thus it might be, that the concentrations reached in aqueous solution might not be sufficient for crystal soaking experiments.

As the binding location of GluII TR-FRET active compounds on the target protein are still unknown it has to be considered that for some of the compounds allosteric inhibition takes place away from the GluII α-/ β-interface. Changes in the β-subunit's secondary structure could be introduced by binding compounds which then leads to the dissociation of the two subunits. If the location of binding is not present in the $GluII_{tryps}$ crystals, due to the processing of the protein pre crystallisation, the compounds cannot bind an lose there α-/ β-interface splitting activity.

Table 38: Summary of GluII crystal soaks with TR-FRET active compounds. If not indicated otherwise, soaking times ranged from 16-18h prior crystal cooling. *Asterix* – apo-version of crystal form. *Plus symbol* – crystals diffracted also post compound soak. *Minus symbol* – Crystals showed no diffraction after compound soak. Percentage DMSO in v/v.

Number	**DMSO**	**Compound**		**Crystal survived**
[-]	[%]	[-]	[mM]	[-]
0.0	-*	-	-	+
0.5	5	-	-	-
0.2	2	-	-	+
0.1	1	-	-	+
60.2	2	EV-JJL001-060-001	2	+
60.1	1	EV-JJL001-060-001	1	+
58.2	2	EV-JJL001-058-001	2	-
58.1	1	EV-JJL001-058-001	1	n.a.
54.2	2	EV-JJL001-054-001	2	-
54.1	1	EV-JJL001-054-001	1	+
52.2	2	EV-JJL001-052-001	2	+
52.1	1	EV-JJL001-052-001	1	n.a.

6.8.1. Conclusion of X-ray crystallography carried out on TR-FRET GluII active compounds

Overall, the attempts to crystallise some of the TR-FRET active compounds soaked into protein crystals did not deliver the structural information desired, and we could not gain additional information about the binding sites of any of the compounds.

The main issue of the X-ray crystallography experiments carried out on $GluII_{tryps}$ crystals soaked with TR-FRET active compounds can be found in the lack of the ability to grow uniform crystals diffracting at high resolution. The use of multiple crystal growth condition resulted in at least three different crystal forms which made processing and refinement difficult and very time consuming. In order to carry out

compound soaks a more stable and controlled crystal form has to be established for future experiments. This could be achieved by modifying some of the protein's secondary structure features in order to generate a more rigid and stable initial product to carry out crystal growth attempts. Those modifications could include the substitution of the α-subunit loop regions aa 186-243 and aa 351-369 with short linkers, as these loops are of higher disorder than the rest of the subunit. For the same reason mutagenesis of the $GluII_{tryps}$ termini should be considered. Only if a stable and reproducible crystal form is established, additional compound soaks of TR-FRET active compounds can be carried out in a reliable fashion. To further improve the success rate of soaking TR-FRET compounds into GluII crystals commercially available kits like the CryoSol™ from Molecular Dimension could be used. The kit allows to solubilise hydrophobic compounds in pre-designed soaking solutions helping to improve compound solubility as well as crystal integrity.

In case a stable crystal form suitable for the use in soaking experiments can be established, TR-FRET active compounds could still have the ability to alter the tertiary structure of GluII and thus the crystal lattice, resulting in weak diffraction or even dissolving crystals. Thus, a co-crystallisation strategy of the different GluII heterodimer constructs with TR-FRET active compounds should be carried out. Upon compound binding and α-/ β-subunit dissociation the separated subunits might be able to crystallise on their own, even though this could never be achieved with individually purified subunits, due to both the lack of ability to express and purify the murine GluII α-subunit on its own and unsuccessful attempts to crystallise varying constructs of the murine GluII β-subunit.

7. Conclusion and future directions

The research presented in this study was carried out in the last three years and provides a strong platform for the analysis of iminosugar-based inhibitors with their two target proteins alpha-glucosidase II and HCV p7. In addition, the thesis describes the development and first employment of a high throughput screen for novel α -glucosidase II inhibitors as well as a system to introduce p7 oligomerisation in lipid environments.

7.1. HCV p7

The parallel production of many different p7 single and double cysteine mutants was necessary and invaluable for the structural analysis of the membrane protein by NMR and EPR-based techniques.

Over the years many different studies provided evidence for the importance of the chemical environment for a protein's structural and functional properties. Several

solution NMR studies of p7 carried out in solvents or detergents agree on the protein being monomeric under the conditions used[77,164-166]. In this dissertation and the Nature Brief Communication Arising (BCA) from Ouyang *et al.* 2013 by Oestringer *et al.* 2018[84] it is shown that p7 solubilised in the detergent fos-chline-12 is also monomeric, different to the authors' conclusion made in their 2013 paper.

EPR being one of a few techniques able to provide structural information about proteins embedded in lipid bilayers helped to gain evidence about p7's structural arrangement in lipid mixtures of DOPC:POPE – two main constituents of the ER membrane, p7's physiologically relevant environment. Furthermore, experiments conducted with samples prepared at different lipid-to-protein molar ratios helped to understand the oligomerisation state of the protein and thus contribute to the very limited information available to date about p7 in lipid mixtures.

The information gained, especially those related to the oligomeric state of p7 in lipid bilayers, form the basis to further determine the protein's interaction with iminosugar-derived inhibitors in the most natural environment currently possible to test this in. The effect of iminosugars on the oligomeric state of the p7 viroporin could be assessed by pulsed EPR techniques based on the information gained in this study.

Nevertheless, high resolution data is needed to fully understand structural and thus functional aspects of the p7 mono- as well as oligomer. Here membrane protein X-ray crystallography carried out in the liquid cubic- or sponge phase of certain lipids could provide a powerful tool for structure determination. Experiments utilising this technique are ongoing in collaboration with the Membrane Protein Laboratory (MPL) within the Research Complex at Harwell (RCaH), UK.

An additional approach to be considered is the newly developed Time Proportional Phase Incremented (TPPI) EPR detection method for so called hub-and-spoke

systems. These systems rely on metal as well as organic spin systems to be present in the nanoobject to be analysed. Therefore it could be applied to putative copper-p7 oligomer chelates[167,339]. Measurements on model systems are currently under development by the group of Dr Alice Bowen (Department of Inorganic Chemistry, University of Oxford) and could be carried out on p7 embedded in lipid bilayers soon. The technique utilises two spin systems of different nature to validate their interactions similar to the DEER methodology carried out in this study, but without the distortion of multi-spin effects.

7.2. alpha-glucosidase II

High resolution structural data provide one of the most efficient ways to understand drug binding at an atomic level and give insights into binding motifs of each of the involved molecules.

For the study presented here several high resolution X-ray crystallography datasets of *Mm*GluII$_{tryps}$ in a ligand bound conformation were obtained. The datasets inform on the binding interactions of a new class of *N*-alkyloxy iminosugars within the active site of the enzyme. Furthermore, one of the datasets resulted in the highest resolution structure of protease digested *Mus musculus* alpha-glucosidase II recorded so far. Although defined interactions of the iminosugars' *N*-alkyloxy chain could not be described, the data provided may aide the design of new inhibitors with more specific binding and fewer off-target effects than current iminosugar-based drugs.

This thesis is also investigating the novel approach of inhibiting GluII by disrupting its heterodimeric interface which should result in the α-subunit leaving the ER. This

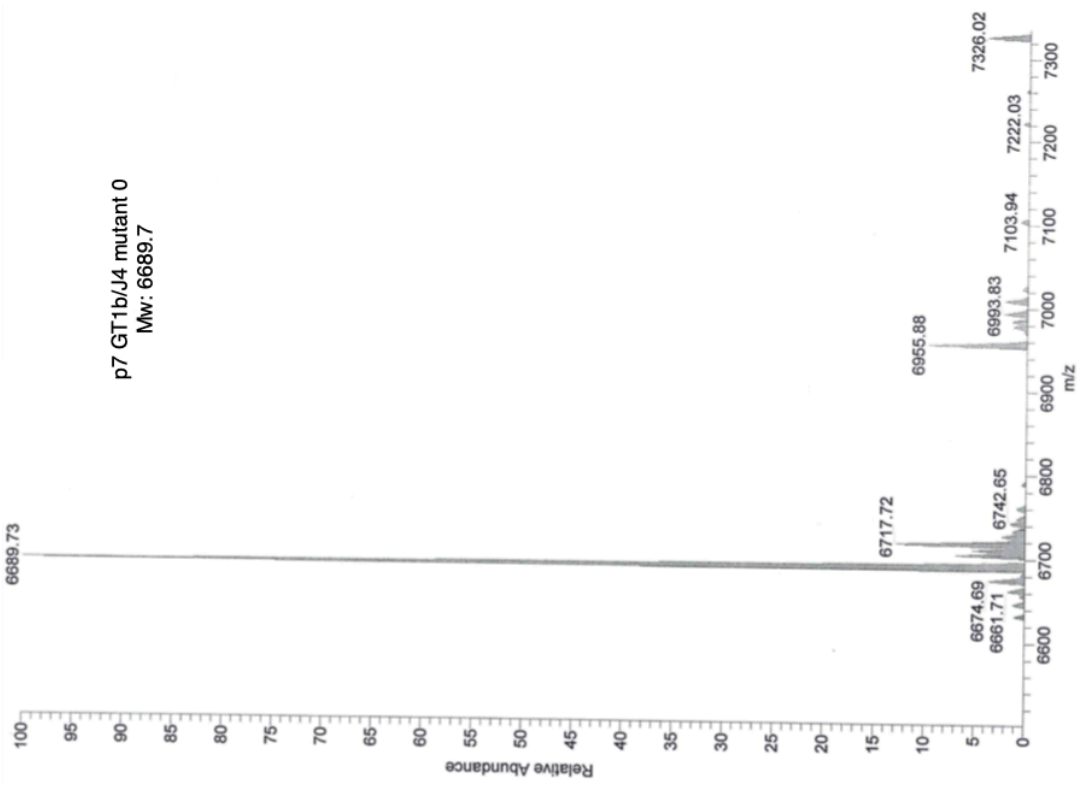
new approach is based on the dissociation of the GluII α-/ β-interface upon compound binding, and was tested using high-throughput screens *in vitro*, using the isolated target enzyme. I describe here the development and first use of a TR-FRET-based screen of GluII, identifying several compounds that show interface disrupting activity.

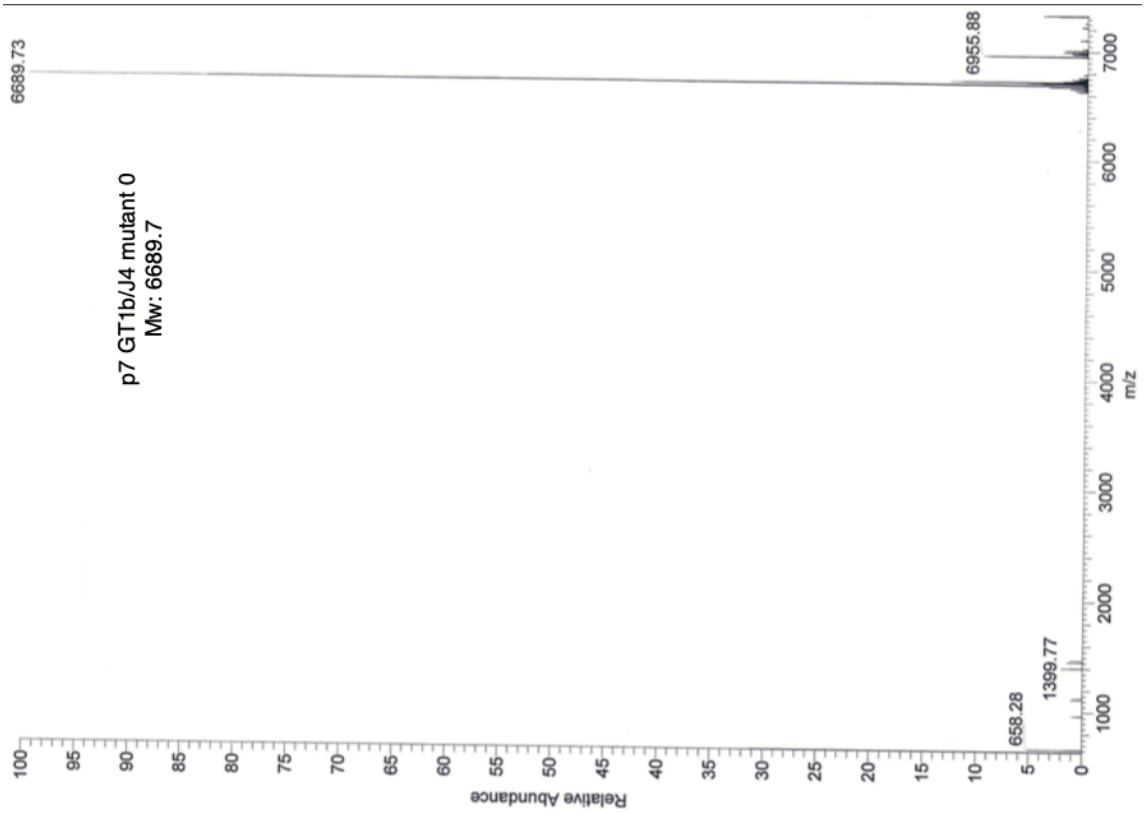
The establishment of a GluII specific high-throughput screen allows future scans of new compound libraries based on the information provided in this study. The nature of the interactions of GluII with interface disrupting compounds is yet to be identified and cannot be provided by the screen. Future structural analyses such as NMR, mass spectroscopy and/or protein X-ray crystallography will help to draw a more complete picture of the binding interactions involved.

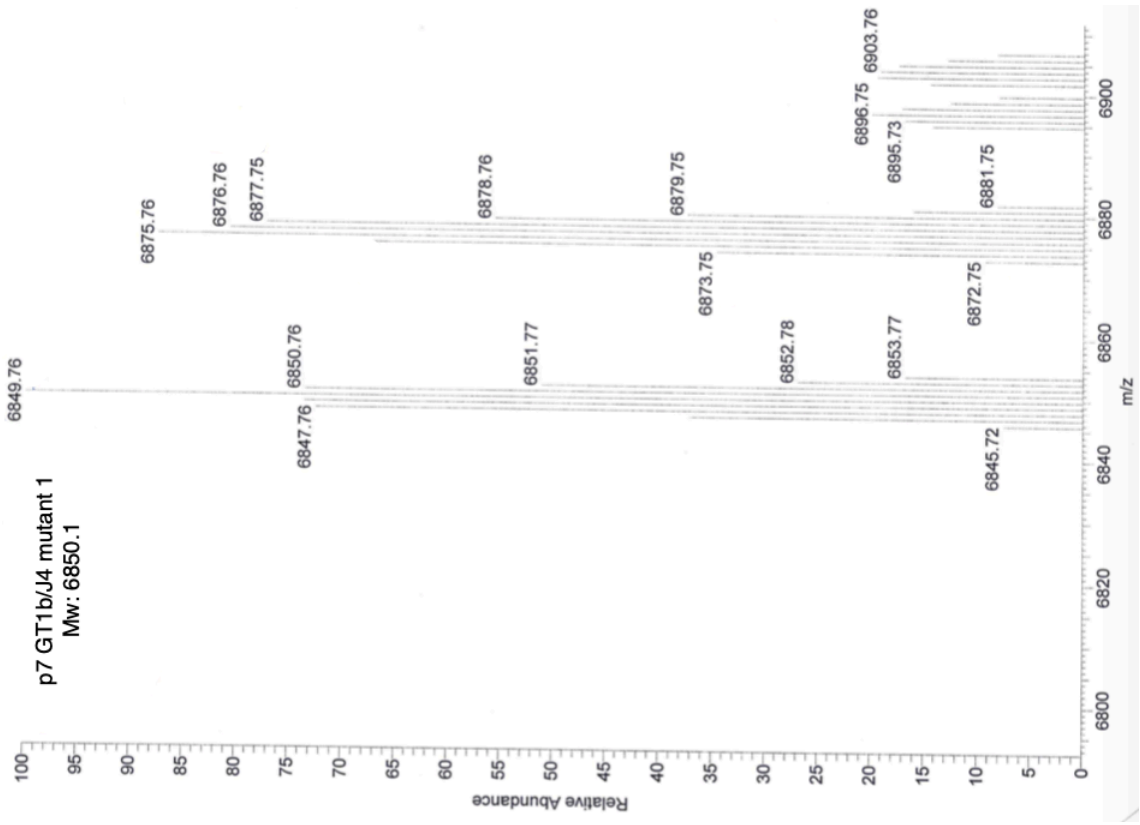
Due to the absence of high resolution structural data for the majority of the GluII β-subunit, and the as yet unsuccessful attempts to grow full-length GluII heterodimer, or β-subunit only crystals, it might be difficult to answer the question of where these compounds bind as some binding motifs might be located within an as yet unknown part of the β-subunit. Furthermore, understanding the involvement of the β-subunit's C-terminal MRH domain in glycan binding might be crucial to further understand the interaction of GluII with di- and monoglucosylated N-linked glycoproteins.

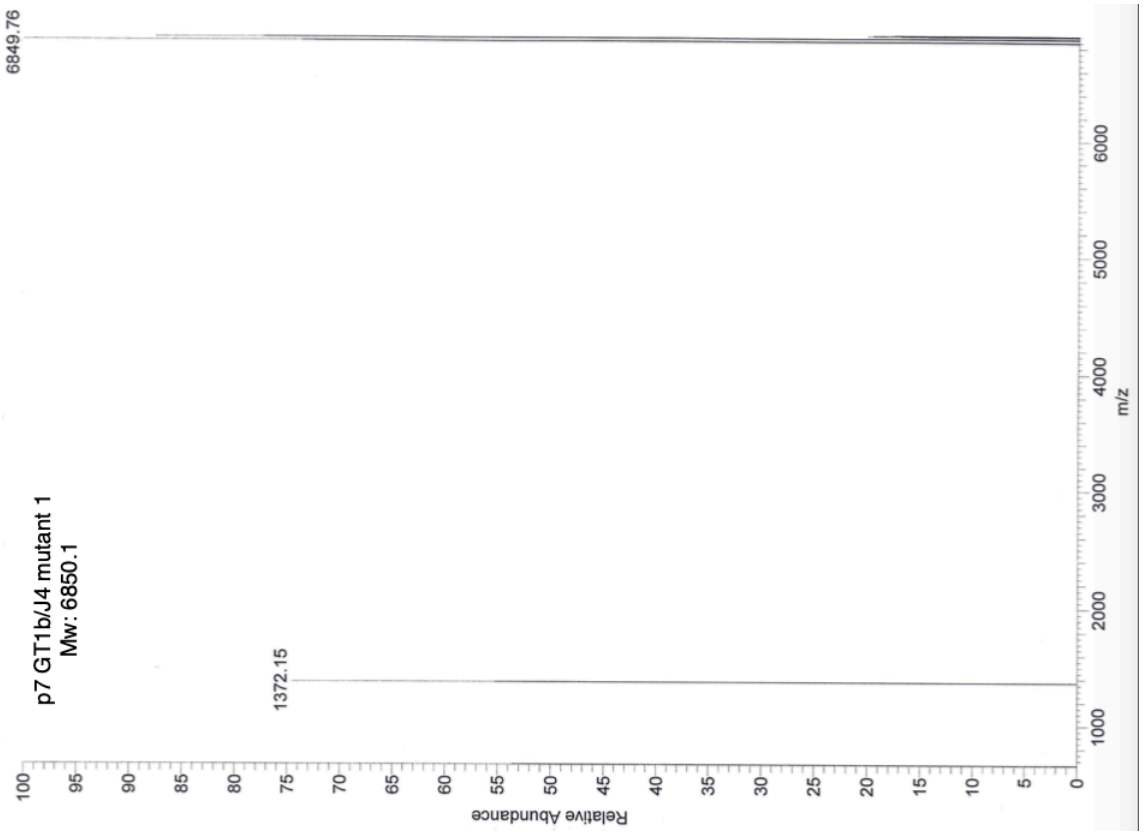
8. Appendix

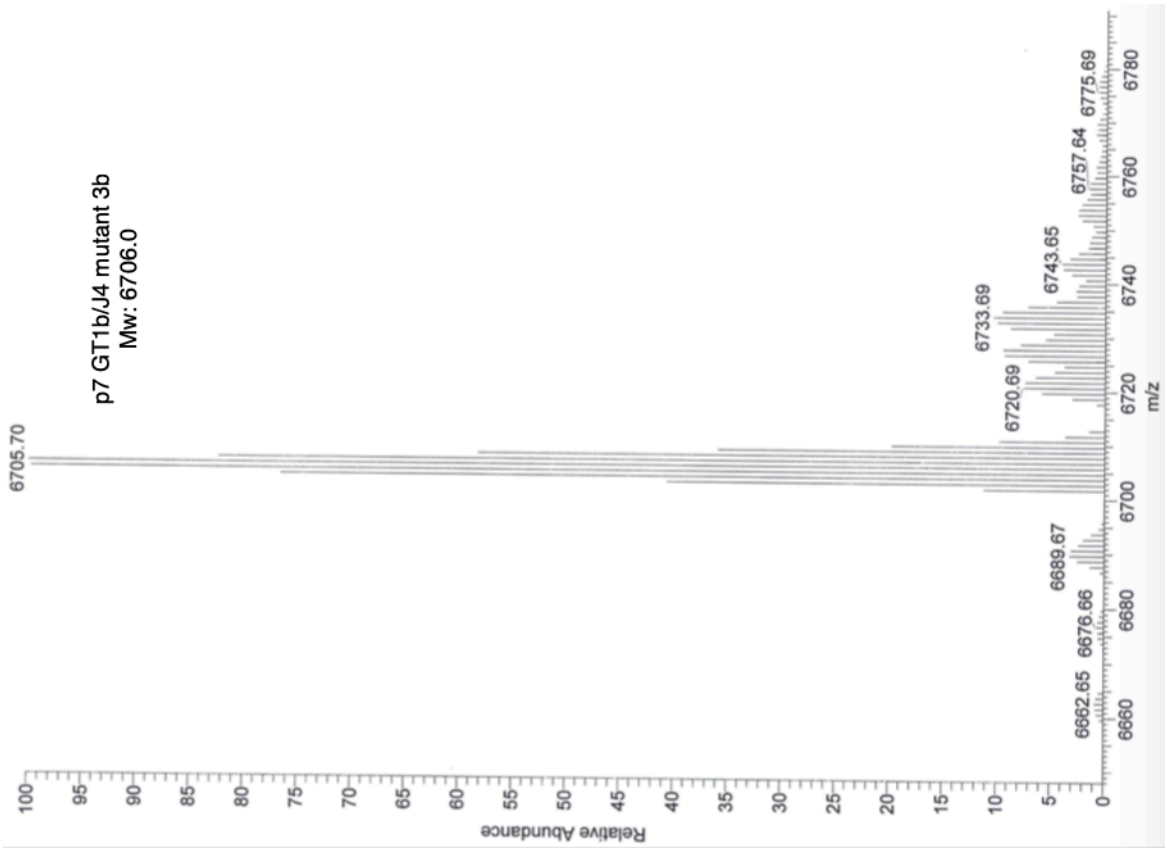
8.1. **Appendix A** – HCV p7

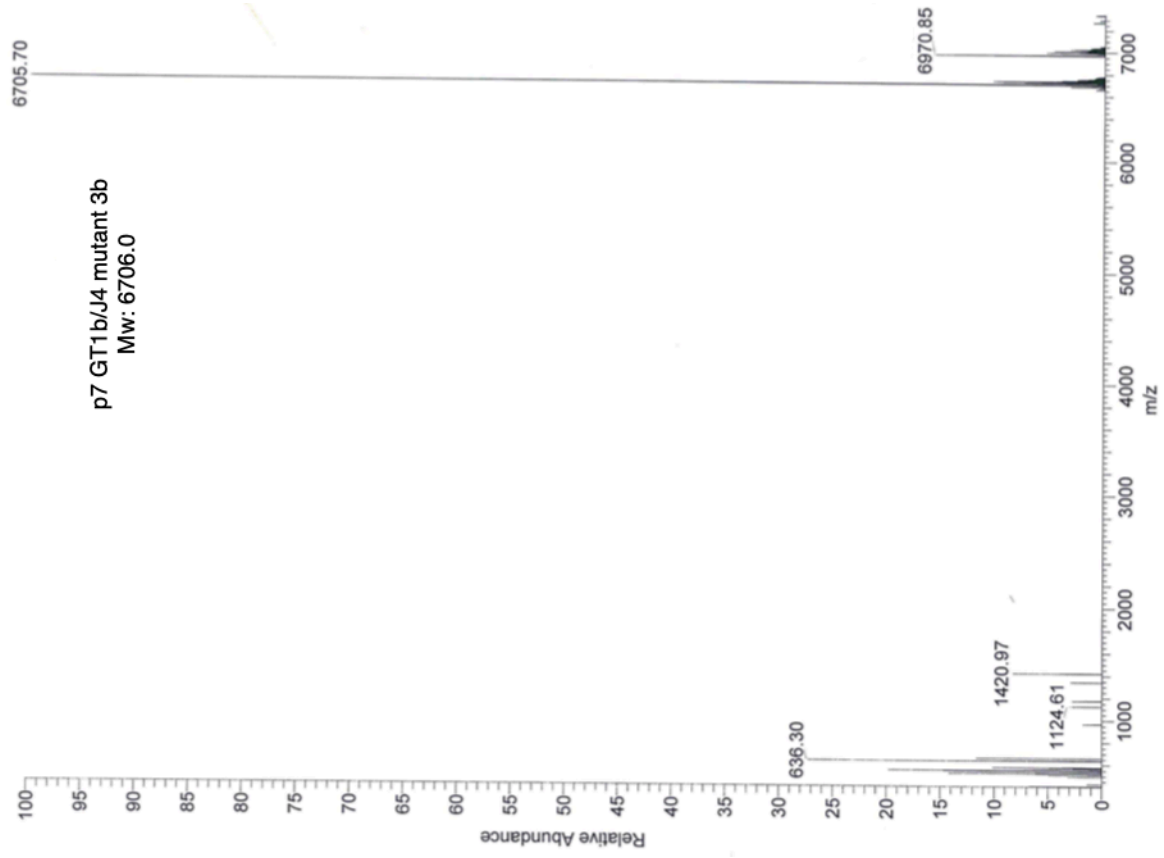


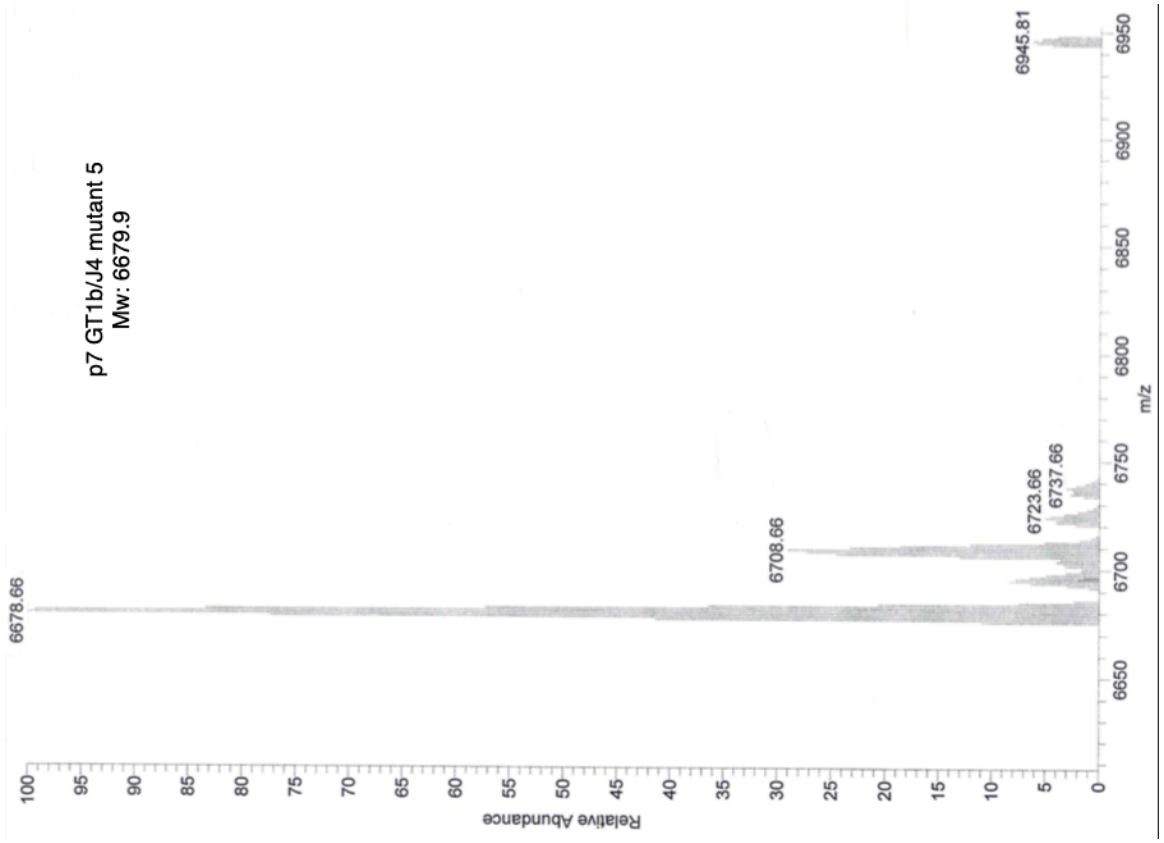

p7 GT1b/J4 mutant 0
Mw: 6689.7
6689.73
6955.88
1399.77
658.28
Relative Abundance
m/z

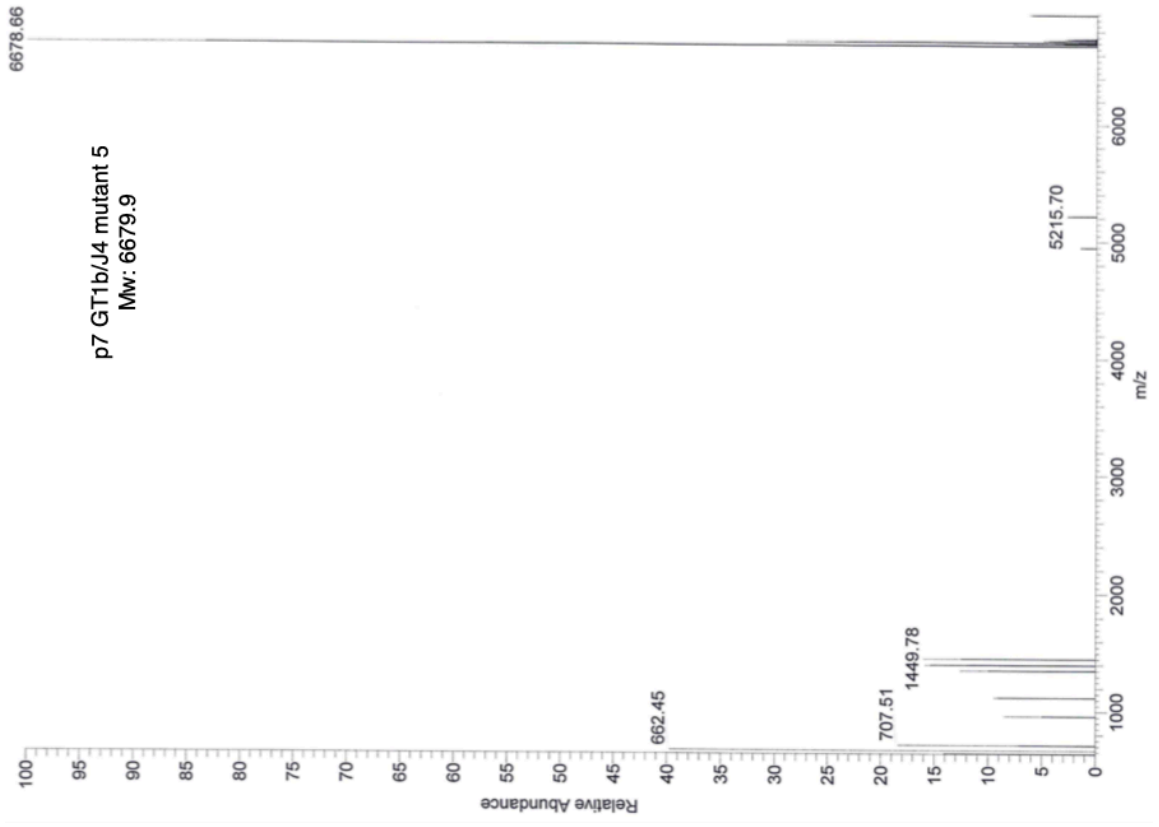

p7 GT1b/J4 mutant 1
Mw: 6850.1
6849.76
6903.76
6896.75
6895.73
6881.75
6878.76
6879.75
6876.76
6877.75
6875.76
6873.75
6872.75
6853.77
6852.78
6851.77
6850.76
6847.76
6845.72
Relative Abundance
m/z

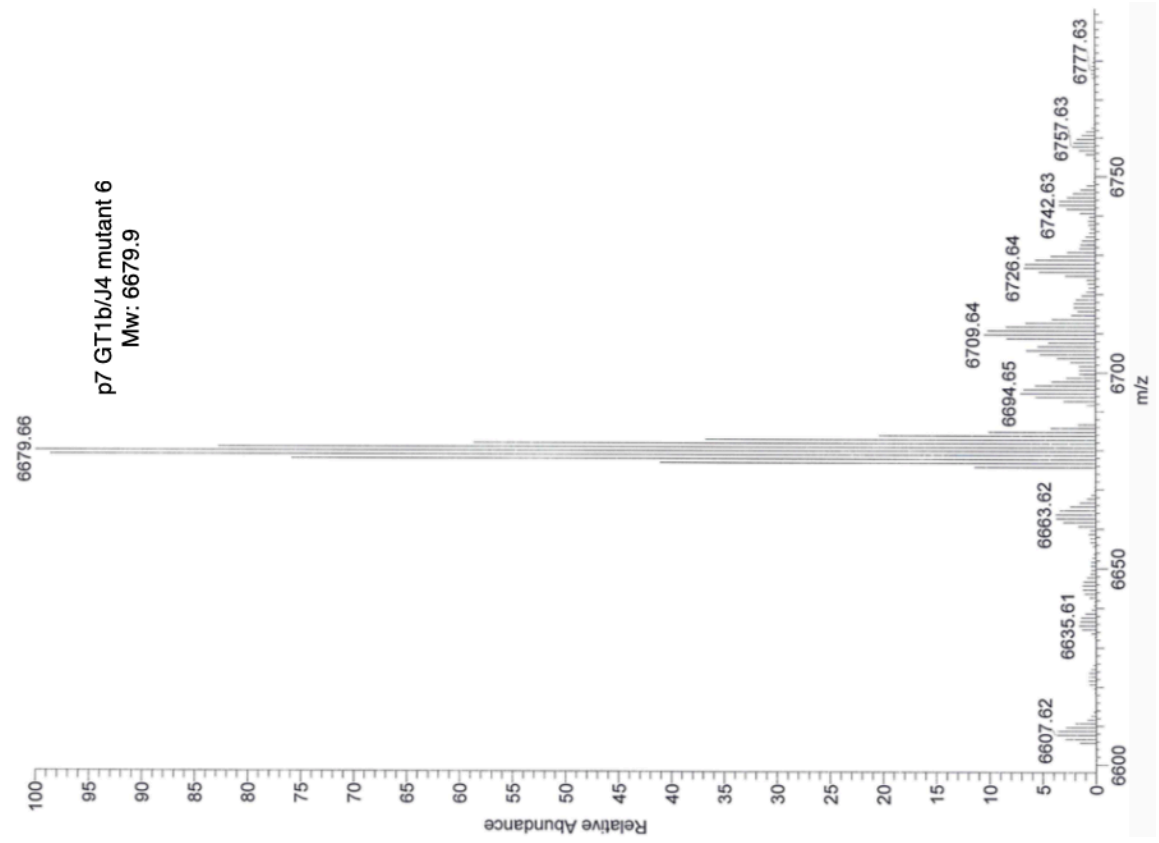

p7 GT1b/J4 mutant 1
Mw: 6850.1
6849.76
1372.15
Relative Abundance
m/z

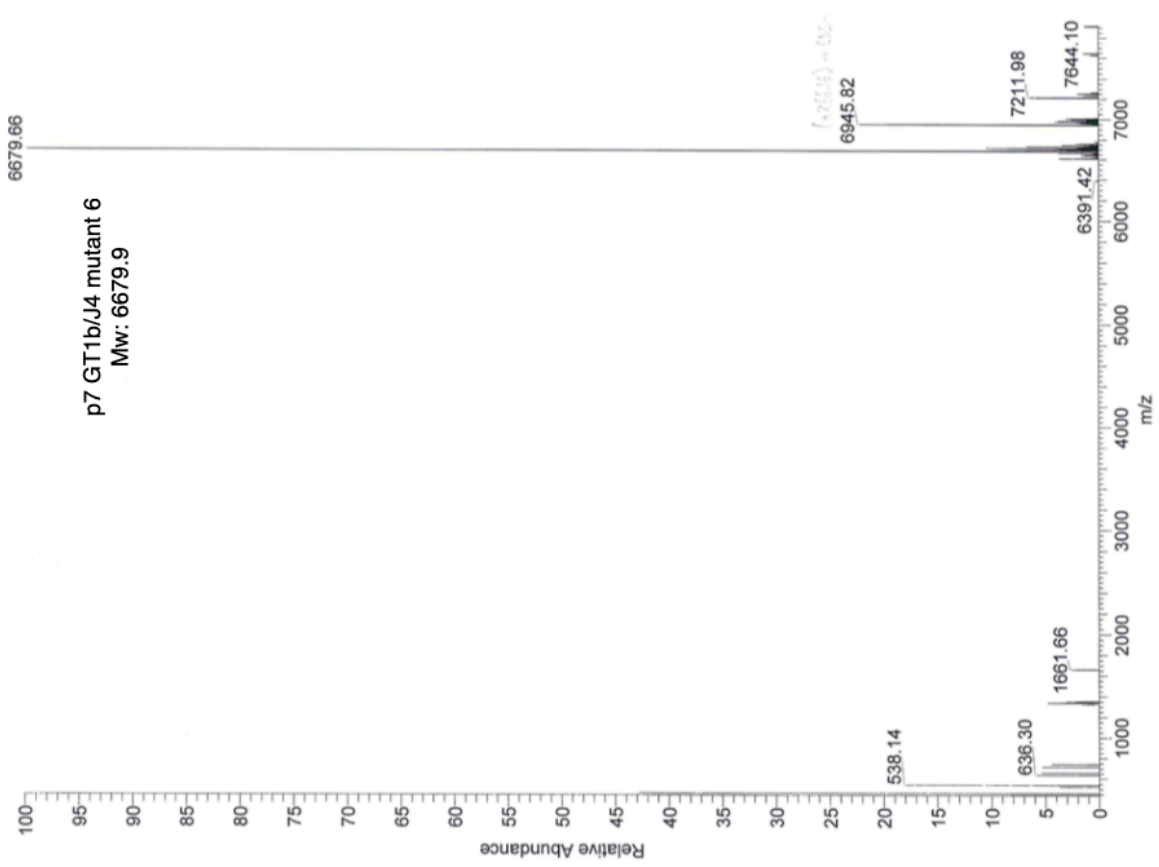

p7 GT1b/J4 mutant 3b
Mw: 6706.0
6705.70
6733.69
6720.69
6743.65
6757.64
6775.69
6689.67
6676.66
6662.65
Relative Abundance
m/z

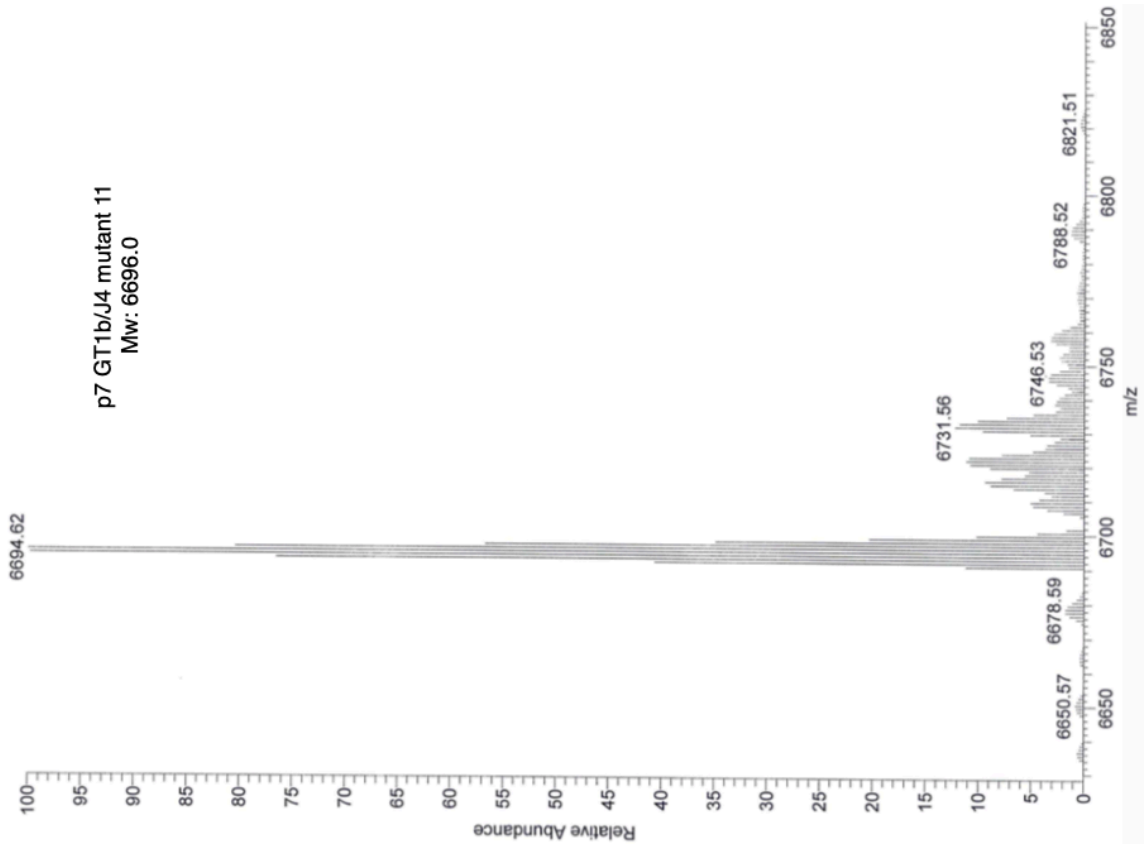

p7 GT1b/J4 mutant 3b
Mw: 6706.0
6705.70
6970.85
1420.97
1124.61
636.30
Relative Abundance
m/z


p7 GT1b/J4 mutant 5
Mw: 6679.9
6678.66
6945.81
6708.66
6723.66
6737.66
Relative Abundance
m/z


p7 GT1b/J4 mutant 5
Mw: 6679.9
6678.66
5215.70
1449.78
707.51
662.45
Relative Abundance
m/z


p7 GT1b/J4 mutant 6
Mw: 6679.9
6679.66
6709.64
6726.64
6742.63
6757.63
6777.63
6694.65
6663.62
6635.61
6607.62
Relative Abundance
m/z


p7 GT1b/J4 mutant 6
Mw: 6679.9
6679.66
6945.82
7211.98
7644.10
6391.42
1661.66
636.30
538.14
Relative Abundance
m/z


p7 GT1b/J4 mutant 11
Mw: 6696.0
6694.62
6731.56
6746.53
6788.52
6821.51
6678.59
6650.57
Relative Abundance
m/z

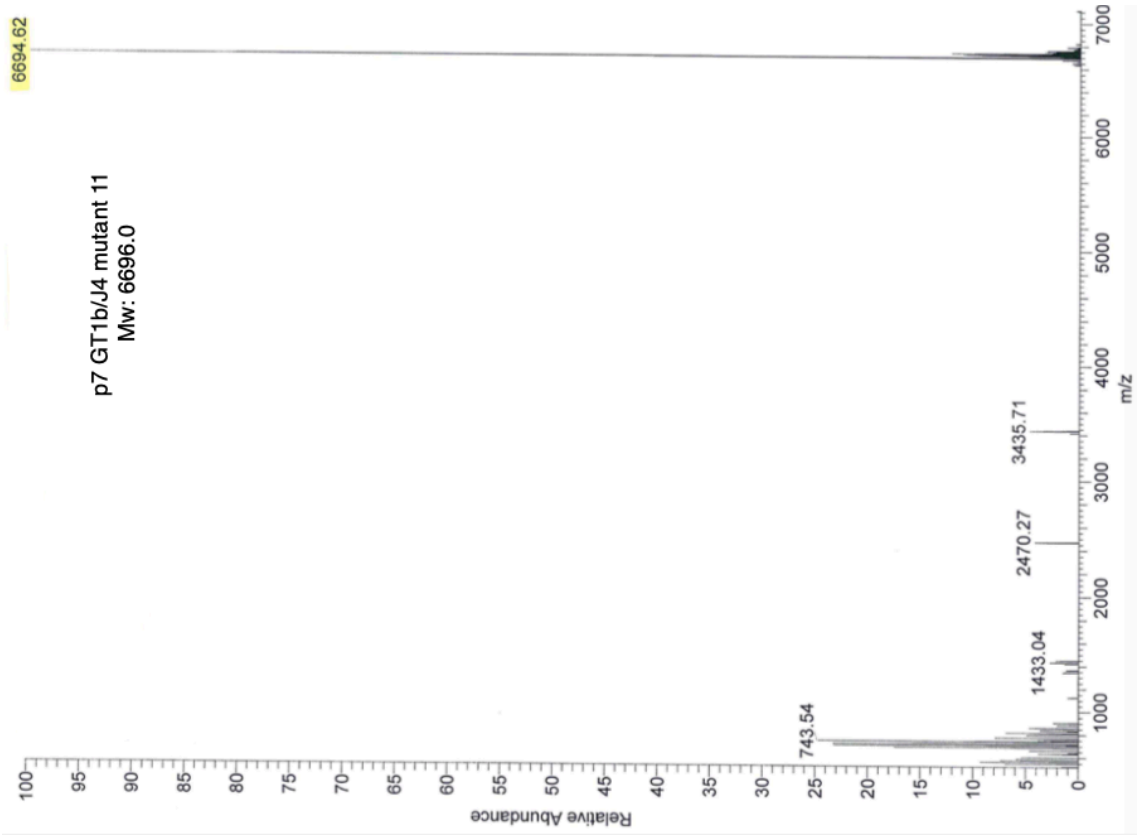


Figure A. 1. Mass spectrometry of different p7 single and double cysteine mutants (see Table 2). Mutants as well as calculated protein masses are indicated in graphs. P7 was solubilised in 70% (v/v) formic acid prior injection and spun 20 min at 13k rpm to remove impurities. The peaks < 1400 m/z are general impurities caused by the device used and thus can be neglected.

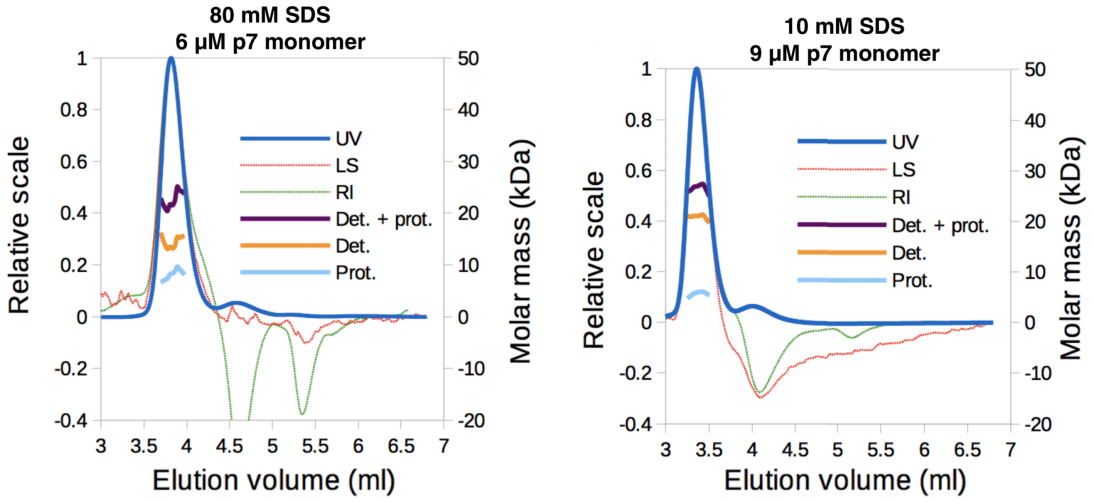


Figure A. 2. Size exclusion – multi angle laser light scattering of p7 (GT5a/JFH-1) mutant 0a solubilised in 10 and 80 mM sodium lauryl-sulphate. Sample detergent concentrations and maximum concentation of p7 monomer detected at peak elution are indicated above the graphs. As indicated by the mobility of p7 on SDS-PAGE the protein is found to run in a monomeric conformation (at about 7 kDa, data not shown), which is in aggrement with the all SEC-MALLS molar mass analysis descirbed above in this study. The evaluation also confirms the trend found with DPC (see section 4.3), showing that the amount of protein associated detergent and is consistensly higher in samples with a overall lower detergent concententration. For more information see section 4.3.

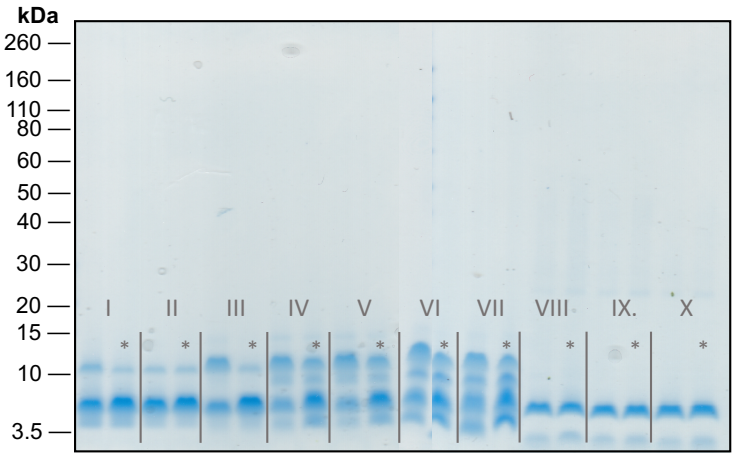


Figure A. 3. MTSL labelling test of different p7 single and double cysteine mutants in the presence and absence of spin label. Lyophilised p7 powder was resuspended in 1.25 M urea, 11 mM SDS, 20 mM sodium phosphate and incubated for 24 h. *Asterisk* – Resuspension took place in the presence of 10 mM MTSL. The sample was mixed with gel loading dye and equal protein amounts were put on each lane. (I) p7 mutent 1, (II) p7 mutant 3b, (III) p7 mutant 5 (IV-V) p7 mutant 6, (VI-VII) two batches of the signle cysteine mutant 8 (not used in this study). Samples show lower weight impurities. (VIII-X) p7 double cysteine mutant 11 from three different purification batches. The protein shows no sign of dimer formation, that even in the absence of MTSL, indicating the formation of intra-monomer disulphide bonds of cysteines in close proximity.

```
% Basic deconvolution of two component Nitroxide cwEPR spectrum (2 -components)
% ================================================================================
% Initial parameters set for 150 Gauss experimental sweep and 1500 data points.
% Initial hyperfine tensor values and anisotropy taken from:
% http://www.sciencedirect.com/science/article/pii/S109078070600406X
% 2 step fitting - Genetic global search followed by Levenberg-Marquardt local
% Data input: text file with two columns: (I) # points/magnetic field (II) data
% Requires easyspin package
% Plotted as function of data points (x - axis) not magnetic field
% 11/9/2016    Marc Pfeil (Watt's Group)
%%

N = maxNumCompThreads
LASTN = maxNumCompThreads(8)
LASTN = maxNumCompThreads('automatic')

clear, clc, clf

% Some experimental parameters

sweep = 150;        % Sweep width in Gauss
nump = 1500;        % Number of points
temp = 308;         % (in Kelvin)
freq = 9.5;         % in GHz

% Specify data path

spc = load('/Users/mario/Documents/Oxford/DPhil/p7/cwESR/160920/2000deconvulation.txt');
xf = spc(:,1); yf=spc(:,2);
xf = 1:1:(numel(xf));

%%

% set baseline to 0 based on first data point and normalize to maximum of
% data set

n = yf(1);

if n < 0
    yf = yf + abs(n);

else
    yf = yf - abs(n);

end

m = max(yf);
yf = yf./m;

% plot of experimental data

ax1 = subplot(2,2,1);
xf = xf;

figure(1)

plot(xf,yf); hold on
title('experimental data');
xlabel('data points');

% Definition of Spin systems (2 nitroxide spins)

% Spin System 1

Nitroxide1.g = [2.0063,2.0085,2.0025];
Nitroxide1.Nucs = '14N';
Nitroxide1.A = mt2mhz([5.8, 5.8, 30.8]/10);
Nitroxide1.weight = 1.369;

% Spin System 2

Nitroxide2.g = [2.0063,2.0085,2.0025];
```

```
Nitroxide2.Nucs = '14N';
Nitroxide2.A = mt2mhz([5.8, 5.8, 30.8]/10);
Nitroxide2.weight = 1.817;

% Dynamic parameters for both spin systems. Correlation time in -log base 10

Nitroxide1.lw =  [0 0.1];
Nitroxide1.logtcorr = -8.40;
Nitroxide1.Exchange = 4;

Nitroxide2.lw =  [0 0.1];
Nitroxide2.logtcorr = -8.00;
Nitroxide2.Exchange = 4;

% Experimental parameters defining sweep and frequency bandwidth

Exp.mwFreq = freq;
swx = sweep/10;
Exp.CenterSweep = [338 swx];
%Exp.CenterSweep = [338 swx]; bigger vailue moves exp. spectra to the right
Exp.Temperature = temp;

% create spin system and plot simulation starting points normalised to
% maximum of simulated spectra

[x2, spc] = chili(Nitroxide1,Exp);
[x3, spc1] = chili(Nitroxide2,Exp);

spc = spc./max(spc);
spc1 = spc1./max(spc1);

ax2 = subplot(2,2,2);
plot(x2,spc); hold on
title('simulated fitting starting point #1');
xlabel('magnetic field (mT)');

ax3 = subplot(2,2,3);
plot(x3,spc1);
title('simulated fitting starting point #2');
xlabel('magnetic field (mT)');

% Parameters to be varied during fitting process for both spin systems

Vary1.logtcorr = 0.5;
Vary2.logtcorr = 0.1;
Vary1.lw = [0 0.1];
Vary2.lw = [0 0.1];
Vary2.g = [0.003 0.003 0.003];
Vary1.g = [0.003 0.003 0.003];

Vary2.weight = 2;
Vary1.weight = 2;
Vary1.A = [6 6 10];
Vary2.A = [6 6 10];
Vary1.Exchange = 5;
Vary2.Exchange = 5;

% Overlay the simulation starting points and the experimental data

xscale = [1:nump/1024:nump];
subplot(2,2,4);

% Data shifted to so that maxima align

[pks, locs] = max(spc);
[pks1, locs1] = max(yf);

xx = locs./1024;
xx1 = locs1./nump;

if xx < xx1

    shift = (locs./1024 - locs1./nump).*nump;
```

```
    shift = round(shift);

    yf = circshift(yf,[shift shift]);

else

    shift = (locs1./nump - locs./1024).*nump;
    shift = round(shift);
    shift = abs(shift);

    yf = circshift(yf,[shift shift]);

end

plot(xf,yf); hold on
plot(xscale,spc); hold on
plot(xscale,spc1); hold on
title('overlay: experimental and starting points');
xlabel('Number of points');
legend('experimental', 'Starting point #1', 'Starting point #2', 'Location', 'Southwest');

% Calling the fitting function for global fitting

SimOpt.Method = 'perturb';
FitOpt1.Method = 'genetic int';
% FitOpt1.TolEdgeLength = 1e-1    % simplex algorithm, integrals of spectra
FitOpt1.maxGenerations = 20;
FitOpt1.PopulationSize = 45;

[BestSys BestSpc] = esfit('chili',yf,{Nitroxide1 Nitroxide2},{Vary1 Vary2},Exp,SimOpt,FitOpt1);

figure(2)

plot(xf,yf); hold on
plot(xf,BestSpc)
title('starting point from global fitting');
legend('experimental', 'global starting point');

% New spinsystems from global fitting

Nitroxide3.Nucs = '14N';
Nitroxide4.Nucs = '14N';

Nitroxide3.g = BestSys{1,1}.g;
Nitroxide4.g = BestSys{1,2}.g;

Nitroxide3.logtcorr = BestSys{1,1}.logtcorr;
Nitroxide4.logtcorr = BestSys{1,2}.logtcorr;

Nitroxide3.weight = BestSys{1,1}.weight;
Nitroxide4.weight = BestSys{1,2}.weight;

Nitroxide3.Exchange = BestSys{1,1}.Exchange;
Nitroxide4.Exchange = BestSys{1,2}.Exchange;

Nitroxide3.A = BestSys{1,1}.A;
Nitroxide4.A = BestSys{1,2}.A;

Nitroxide3.lw = BestSys{1,1}.lw;
Nitroxide4.lw = BestSys{1,2}.lw;

% Local fitting

Vary3.logtcorr = 0.2;
Vary4.logtcorr = 0.2;
Vary3.lw = [0 0.05];
Vary4.lw = [0 0.05];
Vary3.g = [0.003 0.003 0.003];
Vary4.g = [0.003 0.003 0.003];

Vary3.weight = 2;
Vary4.weight = 2;
Vary3.A = [6 6 10];
```

```
Vary4.A = [6 6 10];
Vary3.Exchange = 5;
Vary4.Exchange = 5;

%FitOpt.Method = 'simplex int';
FitOpt.Method = 'levmar fcn';
%FitOpt.Method = 'simplex fcn';

[BestSys1 BestSpc] = esfit('chili',yf,{Nitroxide3 Nitroxide4},{Vary3 Vary4},Exp,SimOpt,FitOpt);

% plotting final result

Nitroxide5.Nucs = '14N';
Nitroxide6.Nucs = '14N';

Nitroxide5.g = BestSys1{1,1}.g;
Nitroxide6.g = BestSys1{1,2}.g;

Nitroxide5.logtcorr = BestSys1{1,1}.logtcorr;
Nitroxide6.logtcorr = BestSys1{1,2}.logtcorr;

Nitroxide5.weight = BestSys1{1,1}.weight;
Nitroxide6.weight = BestSys1{1,2}.weight;

Nitroxide5.Exchange = BestSys1{1,1}.Exchange;
Nitroxide6.Exchange = BestSys1{1,2}.Exchange;

Nitroxide5.A = BestSys1{1,1}.A;
Nitroxide6.A = BestSys1{1,2}.A;

Nitroxide5.lw = BestSys1{1,1}.lw;
Nitroxide6.lw = BestSys1{1,2}.lw;

[x2, spc] = chili(Nitroxide5,Exp,SimOpt);
[x3, spc1] = chili(Nitroxide6,Exp,SimOpt);

max3 = max(spc)+max(spc1);

spc = spc./max3;
spc1 = spc1./max3;

spc_w1 = (Nitroxide6.weight./(Nitroxide5.weight + Nitroxide6.weight));
spc_w2 = (Nitroxide5.weight./(Nitroxide5.weight + Nitroxide6.weight));

% Overlay the simulations and the experimental data

xscale = [1:nump/1024:nump];

% Data shifted so that maxima align for plotting (x-scale: number of points)

spc2 = spc1 + spc;
spc2 = spc2./max(spc2);
[pks, locs] = max(spc2);
[pks1, locs1] = max(yf);

xx = locs./1024;
xx1 = locs1./nump;

if xx < xx1

   shift = (locs./1024 - locs1./nump).*nump;
   shift = round(shift);

   yf = circshift(yf,[shift shift]);

else

   shift = (locs1./nump - locs./1024).*nump;
   shift = round(shift);
   shift = abs(shift);

   yf = circshift(yf,[shift shift]);
```

```
end

figure(3)

subplot(1,3,1);
plot(xf,yf); hold on
title('Experimental data');
xlabel('data points');

subplot(1,3,2);
plot(xf,yf); hold on

plot(xscale,spc); hold on
plot(xscale,spc1); hold on
title('Experimental data and two deconvoluted components');
xlabel('data points');
legend('Experimental data', strcat(num2str(round(spc_w2*100)), '% fast motion component'),
strcat(num2str(round(spc_w1*100)), '% slow motion component'), 'Location', 'Southwest');

subplot(1,3,3);
plot(xf,yf); hold on
xscale1 = [1:1:nump];
plot(xscale1, BestSpc)
title('Experimental data and Simulated two component spectrum');
xlabel('data points');
legend('experimental data', 'simulated spectrum', 'Location', 'Southwest')
```

Figure A. 4. Basic deconvolution of two component Nitroxide cwEPR spectrum (2 -components). 2 step fitting - Genetic global search followed by Levenberg-Marquardt local. Script by Dr Marc-Philipp Pfeil (Watts group, Department of Biochemistry, University of Oxford). Code adapted from http://www.sciencedirect.com/science/article/pii/S109078070600406X.

p7 mutant	LPR	Zero time [ns]	Phase [°]	Background beginning [ns]	Background cutoff [ns]	Background dimension [-]	Regression parameter [-]	Corresponding figure in this Study
1	150:1	76	-0.0	344	1784	2.27	3.98E+01	Figure 39
1	2000:1	203	-0.9	272	1864	2.00	2.00E+01	Figure 41
3b	250:1	74	0.1	74	1424	2.00	3.98E+01	Figure 39
3b	1000:1	72	4.7	240	2100	2.02	6.31E+01	Figure 41
5	250:1	74	0.1	176	1528	2.00	3.98E+01	Figure 39
6	250:1	72	-0.7	608	1752	2.00	3.98E+01	Figure 39
6	1000:1	72	-.02	160	1690	2.27	3.98E+01	Figure 41

Figure A. 5. DEER data processing parameters used in DeerAnalysis2019[241,282] according graphs and distance distributions shown in section 4.4.4 of this study.

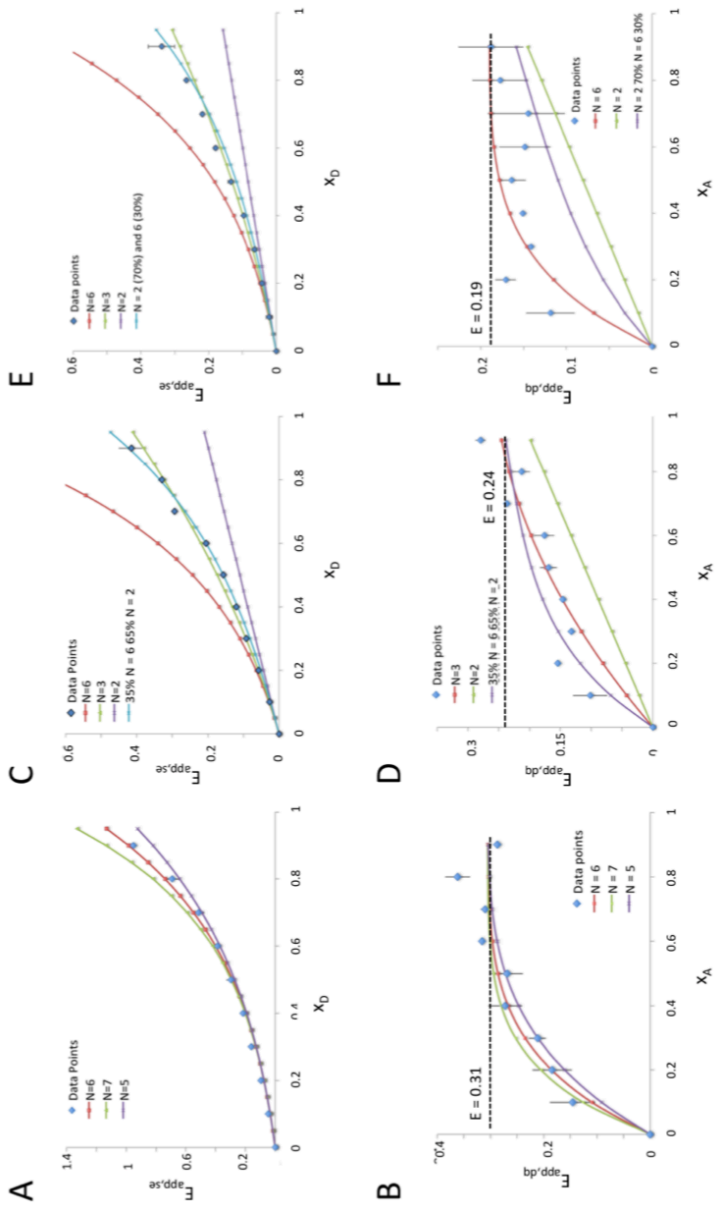


Figure A. 6. Stoichiometric analysis of FRET measurements carried out by Dr Marc-Philipp Pfeil (Professor Anthony Watts group, Department of Biochemistry, Oxford University) on fluorophore labelled p7 reconstituted into DOPC:POPE (4:1 w/w) at different LPRs: (A,B) 150:1, (C,D) 100:1 and (E,F) 2000:1. The apperent FRET efficiency, based on emission (top panels) and doner quenching (bottom panels), was calculate to validate the number of p7 protomers within an oligomeric arrangement. P7 mutant C27S_K33R was N-amine labeleld unsing Sulfo-Cyamine3 N-Hydroxysuccinimide ester (Cy3, donor) or Sulfo-Cyamine5 N-Hydroxysuccinimide ester (Cy5, acceptor) and labelled protein mixed at different donor-to-acceptor molar ratios (X-axis, molar doner fraction). Dr Pfeil's findings are in agreement with EPR data presented in this dissertation (see section 4.4.3), indicating that p7 formes oligomeric arrangments at lower LPR. *Solid lines* – theoretical FRET efficiencies calculated for different oligomeric arrangements of the p7 protomers (n). *Blue dots* – calculated FRET efficiencies for experimental data points. Figure adopted from Dr Marc-Philipp Pfeil's DPhil thesis.

8.2. **Appendix B** – ER alpha-glucosidase II

```
   1 ACCGGTGAGC TCCCATGGGC GCGGCAATAG CGGCAGTGGC GGCGCGTAGG AGGCGGTCTT GGTTAAGTTT GGTGCTGGCA
  81 TACTTAGGGG TCTGTCTGGG GATTACACTT GCTGTGGATA GAAGCAACTT TAAGACCTGT GATGAGAGTT CCTTTTGCAA
 161 ACGGCAGCGA AGCATTCGGC CAGGCCTCTC TCCTTACCGT GCCTTGCTGG ACACTCTGCA GCTTGGTCCT GATGCTCTTA
 241 CAGTCCATCT GATCCATGAA GTCACCAAGG TGCTGCTTGT GCTGGAACTC CAGGGCCTTC AGAAGAACAT GACTCGGATC
 321 AGGATCGATG AGCTAGAGCC CCGGCGGCCT CGATACCGAG TGCCAGATGT TTTAGTGGCT GACCCCCCCA CAGCTAGGCT
 401 TTCAGTCTCT GGCCGTGATG ACAACAGTGT GGAGCTAACA GTGGCTGAGG GACCCTACAA AATCATTTTG ACAGCACAGC
 481 CATTCCGCCT TGACCTGCTA GAAGATCGAA GCCTCCTGCT CAGTGTCAAT GCCCGAGGAC TTATGGCCTT TGAGCACCAG
 561 AGGGCCCCCA GGGTCCCTTT CTCGGATAAA GTTAGTCTCG CGCTCGGTAG CGTGTGGGAT AAGATCAAGA ACCTTTTCTC
 641 TAGGCAAGAA TCAAAAGACC CAGCTGAAGG CAATGGAGCC CAGCCTGAAG CAACACCTGG GGATGGTGAC AAGCCAGAGG
 721 AGACCCAGGA AAAGGCTGAG AAGGATGAGC CAGGAGCCTG GGAGGAGACA TTCAAAACAC ATTCTGACAG CAAGCCTTAT
 801 GGCCCCACGT CTGTAGGTTT GGACTTTTCT CTGCCAGGAA TGGAACATGT GTATGGGATC CCTGAGCATG CTGACAGCCT
 881 GAGACTGAAG GTCACTGAGG GCGGTGAGCC GTACCGCCTG TACAATTTGG ATGTGTTCCA GTATGAGCTG AACAACCCAA
 961 TGGCTCTATA TGGGTCTGTG CCTGTGCTCC TGGCACACAG CTTTCATCGA GACCTGGGCA TCTTCTGGCT TAATGCTGCT
1041 GAGACTTGGG TTGATATATC CTCCAACACG GCTGGGAAGA CCCTGTTTGG GAAGATGCTT GATTACCTGC AGGGCTCTGG
1121 GGAGACTCCA CAGACAGACA TTCGTTGGAT GTCAGAGAGT GGCATTATTG ATGTTTTCCT AATGCTTGGC CCTTCGGTTT
1201 TTGATGTCTT TAGGCAGTAT GCTAGTCTCA CAGGGACCCA GGCATTGCCC CCACTCTTCT CCCTCGGCTA TCACCAGAGT
1281 CGCTGGAACT ACCGGGATGA GGCTGATGTT TTGGAAGTGG ATCAGGGTTT TGATGATCAC AACATGCCTT GTGATGTCAT
1361 TTGGTTGGAC ATTGAACATG CTGATGGCAA GCGGTACTTC ACTTGGGACC CCACCCGATT TCCTCAGCCC CTCAATATGC
1441 TTGAGCACTT GGCTTCCAAG AGGCGGAAGC TGGTGGCCAT TGTGGACCCC CACATCAAGG TAGACTCTGG CTACCGAGTT
1521 CACGAAGAAT TGCGAAACCA CGGGCTGTAT GTTAAAACTC GGGATGGCTC TGATTACGAG GGCTGGTGCT GGCCAGGCTC
1601 AGCTAGTTAC CCTGACTTCA CTAATCCAAG GATGAGGGCC TGGTGGTCTA ACATGTTCAG CTTTGACAAT TATGAGGGTT
1681 CAGCTCCTAA TCTTTATGTT TGGAATGACA TGAATGAACC GTCTGTGTTC AATGGTCCTG AGGTCACCAT GTTGAAGGAT
1761 GCTGTGCATT ATGGAGGCTG GGAGCACCGG GACATCCATA ACATCTATGG CTTATATGTG CACATGGCGA CTGCTGATGG
1841 GCTAATACAG CGCTCTGGGG GCATAGAGCG TCCCTTTGTC CTGAGTAGGG CTTTCTTCTC AGGCTCCCAG CGCTTTGGAG
1921 CTGTGTGGAC AGGGGACAAC ACTGCCGAAT GGGATCATTT GAAGATCTCT ATCCCTATGT GTCTCAGCCT GGCACTGGTG
2001 GGGCTTTCCT TCTGTGGAGC GGATGTGGGT GGCTTCTTCA AGAACCCAGA GCCAGAGCTG CTTGTGCGCT GGTACCAAAT
2081 GGGTGCCTAC CAGCCGTTCT TTCGGGCTCA TGCCCACTTG GACACTGGGC GGCGAGAGCC CTGGCTGTTA GCGTCTCAAT
2161 ACCAAGATGC AATCCGAGAT GCCTTGTTCC AGCGATATTC TTTGCTGCCC TTCTGGTATA CCCTCTTCTA TCAAGCTCAC
2241 AAGGAAGGGT TTCCTGTCAT GAGGCCCCTC TGGGTACAGT ATCCTGAGGA TATGTCTACC TTCAGTATAG AGGATCAGTT
2321 CATGCTTGGT GATGCACTCC TTATTCACCC TGTATCGGAT GCTGGGGCCC ACGGAGTGCA GGTCTATTTG CCTGGCCAAG
2401 AAGAAGTGTG GTATGACATT CAGAGCTATC AGAAGCATCA TGGGCCCCAG ACCTTGTATC TGCCAGTAAC TTTGAGCAGT
2481 ATCCCCGTGT TCCAGCGTGG CGGAACCATT GTGCCTCGAT GGATGCGTGT GAGGCGCTCT TCAGACTGTA TGAAGGACGA
2561 TCCTATCACT CTCTTTGTTG CTCTCAGTCC CCAGGGTACT GCCCAAGGAG AACTCTTTCT AGATGATGGA CACACCTTTA
2641 ACTATCAGAC TCGCCATGAG TTCCTGTTGC GGCGGTTCTC TTTCTCTGGC AGCACACTGG TCTCTAGTTC AGCAGACCCC
2721 AAAGGCCACC TTGAGACACC TATTTGGATT GAGCGAGTAG TCATCATGGG GGCTGGAAAG CCAGCAGCTG TGGTGCTCCA
2801 GACCAAAGGA TCCCCTGAAA GTCGCCTGTC CTTCCAGCAT GACCCTGAGA CCTCAGTGCT GATATTGCGT AAACCTGGTG
2881 TCAGCGTGGC ATCCGACTGG AGTATTCATC TTCGATAAGC GGCCGC
```

Figure B. 1. DNA sequence of *mus musculus* alpha-glucosidaseII α-subunit used in this study. Uniprot: Q8BHN3, *GANAB_mouse.*

```
   1 ACCGGTGAGC TCCCATGGGC CTGCTGCTGC TGCTACTACT ACTACCCCTC TGTTGGGCTG TAGAAGTTAA GAGACCCCGG
  81 GGCGTTTCCC TCAGCAACCA TCACTTCTAT GAAGAATCTA AACCTTTCAC CTGTTTGGAC GGCACAGCCA CCATCCCATT
 161 CGATCAGGTG AACGACGACT ACTGCGACTG TAAGGACGGT TCAGATGAGC CTGGCACAGC TGCTTGTCCC AATGGCAGCT
 241 TTCACTGCAC CAACACTGGG TACAAGCCCT TGTACATCCT CTCCAGCCGG GTCAATGATG GGGTATGTGA CTGCTGTGAT
 321 GGCACAGACG AGTACAACAG CGGCACGGTC TGCGAGAACA CCTGCAGAGA GAAGGGTCGC AAGGAAAAAG AGTCCCTGCA
 401 GCAGCTGGCG GAAGTCACCC GTGAAGGGTT CCGCCTGAAG AAGATTCTCA TTGAGGAGTG GAAGACAGCC CGGGAAGAAA
 481 AGCAGAGTAA GCTTCTTGAG CTTCAGGCTG GAAAGAAGTC TCTGGAAGAC CAGGTAGAAA CACTGCGGGC AGCGAAAGAA
 561 GAAGCAGAGA GGCCAGAGAA GGAGGCCAAG GACCAGCACC GGAAGCTGTG GGAAGAGCAG CAAGCTGCTG CCAAGGCCCG
 641 GCGGGAACAG GAGCGGGCAG CCAGTGCCTT CCAGGAACTT GACGACAACA TGGATGGGAT GGTCTCGCTG GCTGAGTTAC
 721 AGACTCACCC GGAGCTGGAC ACAGATGGAG ATGGAGCGCT GTCTGAGGAG GAGGCCCAGG CCCTTCTCAG TGGAGACACA
 801 CAGACTGACA CCACCTCCTT CTATGACCGT GTCTGGGCTG CCATCAGGGA CAAGTACCGC TCTGAGGTCC CGCCCACTGA
 881 CATACCTGTT CCGGAGGAGA CTGAGCCCAA AGAGGAAAAG CCACCAGTGT TGCCACCCAC AGAGGAGGAG GAAGAGGAGG
 961 AGGAGGAGCC AGAAGAAGAG GAGGAGGAAG AGGAAGAGGA GGAGGAGGCT CCGCCCCCAC TGCAGCCCCC ACAGCCTCCC
1041 AGCCCCACAG AGGATGAGAA GATGCCGCCC TATGATGAGG AGACCCAGGC CATCATCGAT GCTGCACAGG AGGCCCGGAG
1121 TAAGTTTGAG GAAGTCGAAC GGTCCTTGAA AGAGATGGAA GAGTCCATCA GGAGTTTGGA ACAAGAGATC TCCTTTGATT
1201 TCGGTCCCTC TGGAGAGTTT GCATATCTCT ACAGCCAATG CTACGAACTC ACCACCAATG AGTACGTCTA CCGGCTTTGC
1281 CCCTTCAAAC TGGTCTCCCA GAAACCCAAA CATGGGGGCT CCCCGACCAG CCTGGGCACA TGGGGCTCCT GGGCTGGCCC
1361 TGATCATGAC AAGTTCAGTG CGATGAAGTA CGAGCAGGGC ACGGGCTGTT GGCAGGGCCC CAACCGATCC ACCACAGTGC
1441 GCCTGCTGTG TGGCAAAGAG ACTGTGGTGA CCAGCACCAC GGAGCCCAGT CGCTGTGAGT ACCTCATGGA GCTGATGACA
1521 CCAGCAGCCT GCCCAGAGCC GCCACCAGAA GCACCCAGTG ATGGGGACCA TGACGAGCTG TAGGCGGCCG C
```

Figure B. 2. DNA sequence of *mus musculus* alpha-glucosidaseII β-subunit used in this study. Uniprot: O08795, *PRKCSH_mouse.*

```
   1 AAGGAGCAAA CTCTGCACAA GATGGCGGCG GTAGCGGCAG TGGCGGCGCG TAGGAGGCGG TCTTGGGCGT CTTTGGTACT
  81 GGCTTTTTTA GGGGTCTGCC TGGGGATTAC CCTTGCTGTG GATAGAAGCA ACTTTAAGAC CTGTGAAGAG AGTTCTTTCT
 161 GCAAGCGACA GAGAAGCATA CGGCCAGGCC TCTCTCCATA CCGAGCCTTG CTGGACTCTC TACAGCTTGG TCCTGATTCC
 241 CTCACGGTCC ATCTGATCCA TGAGGTCACC AAGGTGTTGC TGGTGCTAGA GCTTCAGGGG CTTCAAAAGA ACATGACTCG
 321 GTTCAGGATT GATGAGCTGG AGCCTCGGCG ACCCCGATAC CGTGTACCAG ATGTTTTGGT GGCTGATCCA CCAATAGCCC
 401 GGCTTTCTGT CTCTGGTCGT GATGAGAACA GTGTGGAGTT AACCATGGCT GAGGGACCCT ACAAGATCAT CTTGACAGCA
 481 CGGCCATTCC GCCTTGACCT ACTAGAGGAC CGAAGTCTTT TGCTTAGTGT CAATGCCCGA GGACTCTTGG AGTTTGAGCA
 561 TCAGAGGGCC CCTAGGGTCT CCCCCTCGTC CCTGCCCCCT CTGGATTGGA GCAGACAGCT CTCCTACCTT CCAGGCAAGG
 641 ATCAAAAGAC CCAGCTGAGG GCGATGGGGC CCAGCCTGAG GAAACACCCA GGGATGGCGA CAAGCCAGAG GAGACTCAGG
 721 GGAAGGCAGA GAAAGATGAG CCAGGAGCCT GGGAGGAGAC ATTCAAAACT CACTCTGACA ATGTCTGTGG GTTTGGACTT
 801 CTCTCTGCCA GGCATGGAGC ATGTCTATGG GATCCCTGAG CATGCAGACA ACCTGAGGCT GAAGGTCACT GAGGGTGGGG
 881 AGCCATATCG CCTCTACAAT TTGGATGTGT TCCAGTATGA GCTGTACAAC CCAATGGCCT TGTATGGGTC TGTGCCTGTG
 961 CTCCTGGCAC ACAACCCTCA TCGCGACTTG GGCATCTTCT GGCTCAATGC TGCAGAGACC TGGGTTGATA TATCTTCCAA
1041 CACTGCCGGG AAGACCCTGT TTGGGAAGAT GATGGACTAC CTGCAGGGCT CTGGGGAGAC CCCACAGACA GATGTTCGCT
1121 GGATGTCAGA GACTGGCATC ATTGACGTCT TCCTGCTGCT GGGGCCCTCC ATCTCTGATG TTTTCCGGCA ATATGCTAGT
1201 CTCACAGGAA CCCAGGCGTT GCCCCCACTC TTCTCCCTCG GCTACCACCA GAGCCGTTGG AACTACCGGG ACGAGGCTGA
1281 TGTGCTGGAA GTGGATCAGG GCTTTGATGA TCACAACCTG CCCTGTGATG TCATCTGGCT AGACATTGAA CATGCTGATG
1361 GCAAGCGGTA TTTCACCTGG GACCCCAGTC GCTTCCCTCA GCCCCGCACC ATGCTTGAGC GCTTGGCTTC TAAGAGGCGG
1441 AAGCTGGTGG CCATCGTAGA CCCCCACATC AAGGTGGACT CCGGCTACCG AGTTCACGAG GAGCTGCGGA ACCTGGGGCT
1521 GTATGTTAAA ACCCGGGATG GCTCTGACTA TGAGGGCTGG TGCTGGCCAG GCTCAGCTGG TTACCCTGAC TTCACTAATC
1601 CCACGATGAG GGCCTGGTGG GCTAACATGT TCAGCTATGA CAATTATGAG GGCTCAGCTC CCAACCTCTT TGTCTGGAAT
1681 GACATGAACG AACCATCTGT GTTCAATGGT CCTGAGGTCA CCATGCTCAA GGATGCCCAG CATTATGGGG GCTGGGAGCA
1761 CCGGGATGTG CATAACATCT ATGGCCTTTA TGTGCACATG GCGACTGCTG ATGGGCTGAG ACAGCGCTCT GGGGGCATGG
1841 AACGCCCCTT TGTCCTGGCC AGGGCCTTCT TCGCTGGCTC CCAGCGCTTT GGAGCCGTGT GGACAGGGGA CAACACTGCC
1921 GAGTGGGACC ATTTGAAGAT CTCTATTCCT ATGTGTCTCA GCTTGGGGCT GGTGGGACTT TCCTTCTGTG GGGCGGATGT
2001 GGGTGGCTTC TTCAAAAACC CAGAGCCAGA GCTGCTTGTG CGCTGGTACC AGATGGGTGC TTACCAGCCA TTCTTCCGGG
2081 CACATGCCCA CTTGGACACT GGGCGACGAG AGCCATGGCT GTTACCATCT CAGCACAATG ATATAATCCG AGATGCCTTG
2161 GGCCAGCGAT ATTCTTTGCT GCCCTTCTGG TACACCCTCT TATATCAGGC CCATCGGGAA GGCATTCCTG TCATGAGGCC
2241 CCTGTGGGTG CAGTACCCTC AGGATGTGAC TACCTTCAAT ATAGATGATC AGTACTTGCT TGGGGATGCG TTGCTGGTTC
2321 ACCCTGTATC AGACTCTGGA GCCCATGGTG TCCAGGTCTA TCTGCCTGGC CAAGGGGAGG TGTGGTATGA CATTCAAAGC
2401 TACCAGAAGC ATCATGGTCC CCAGACCCTG TACCTGCCTG TAACTCTAAG CAGTATCCCT GTGTTCCAGC GTGGAGGGAC
2481 AATCGTGCCT CGATGGATGC GAGTGCGGCG GTCTTCAGAA TGTATGAAGG ATGACCCCAT CACTCTCTTT GTTGCACTTA
2561 GCCCTCAGGG TACAGCTCAA GGAGAGCTCT TTCTGGATGA TGGGCACACG TTCAACTATC AGACTCGCCA AGAGTTCCTG
2641 CTGCGTCGAT TCTCATTCTC TGGCAACACC CTTGTCTCCA GCTCAGCAGA CCCTGAAGGA CACTTTGAGA CACCAATCTG
2721 GATTGAGCGG GTGGTGATAA TAGGGGCTGG AAAGCCAGCA GCTGTGGTAC TCCAGACAAA AGGATCTCCA GAAAGCCGCC
2801 TGTCCTTCCA GCATGACCCT GAGACCTCTG TGTTGGTCCT GCGCAAGCCT GGCATCAATG TGGCATCTGA TTGGAGTATT
2881 CACCTGCGAT AA
```

Figure B. 3. DNA sequence of *homo sapiens* alpha-glucosidaseII α-subunit used in this study. Uniprot: Q14697, *GANAB_human*.

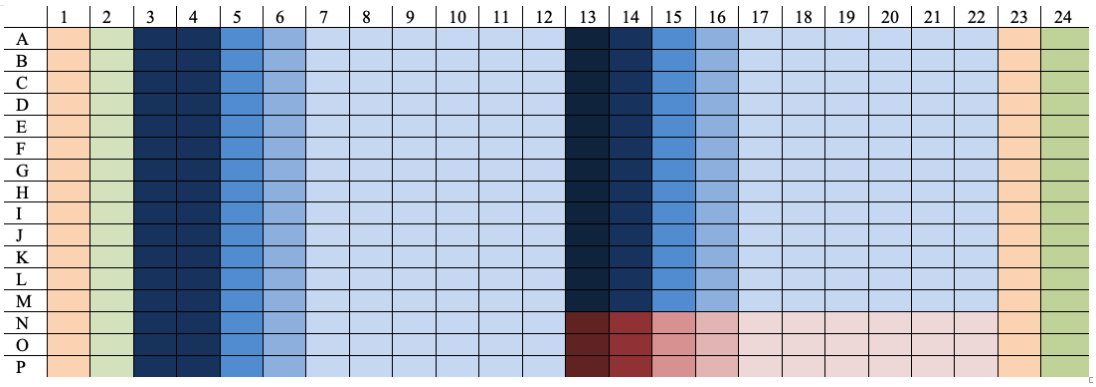
```
   1 ATGCTGTTGC CGCTGCTGCT GCTGCTACCC ATGTGCTGGG CCGTGGAGGT CAAGAGGCCC CGGGGCGTCT CCCTCACCAA
  81 TCATCACTTC TACGATGAGT CCAAGCCTTT CACCTGCCTG GACGGTTCGG CCACCATCCC ATTTGATCAG GTCAACGATG
 161 ACTATTGCGA CTGCAAAGAT GGCTCTGACG AGCCAGGCAC GGCTGCCTGT CCTAATGGCA GCTTCCACTG CACCAACACT
 241 GGCTATAAGC CCCTGTATAT CCCCTCCAAC CGGGTCAACG ATGGTGTTTG TGACTGCTGC GATGGAACAG ACGAGTACAA
 321 CAGCGGCGTC ATCTGTGAGA ACACCTGCAA AGAGAAGGGC CGTAAGGAGA GAGAGTCCCT GCAGCAGATG GCCGAGGTCA
 401 CCCGCGAAGG GTTCCGTCTG AAGAAGATCC TTATTGAGGA CTGGAAGAAG GCACGGGAGG AGAAGCAGAA AAAGCTCATT
 481 GAGCTACAGG CTGGGAAGAA GTCTCTGGAA GACCAGGTGG AGATGCTGCG GACAGTGAAG GAGGAAGCTG AGAAGCCAGA
 561 GAGAGAGGCC AAAGAGCAGC ACCAGAAGCT GTGGGAAGAG CAGCTGGCTG CTGCCAAGGC CCAACAGGAG CAGGAGCTGG
 641 CGGCTGATGC CTTCAAGGAG CTGGATGATG ACATGGACGG GACGGTCTCG GTGACTGAGC TGCAGACTCA CCCGGAGCTG
 721 GACACAGATG GGGATGGGGC GTTGTCAGAA GCGGAAGCTC AGGCCCTCCT CAGTGGGGAC ACACAGACAG ACGCCACCTC
 801 TTTCTACGAC CGCGTCTGGG CCGCCATCAG GGACAAGTAC CGGTCCGAGG CACTGCCCAC CGACCTTCCA GCACCTTCTG
 881 CCCCTGACTT GACGGAGCCC AAGGAGGAGC AGCCGCCAGT GCCCTCGTCG CCCACAGAGG AGGAGGAGGA GGAGGAGGAG
 961 GAGGAAGAAG AGGCTGAAGA AGAGGAGGAG GAGGAGGATT CCGAGGAGGC CCCACCGCCA CTGTCACCCC CGCAGCCGGC
1041 CAGCCCTGCT GAGGAAGACA AAATGCCGCC CTACGACGAG CAGACGCAGG CCTTCATCGA TGCTGCCCAG GAGGCCCGCA
1121 ACAAGTTCGA GGAGGCCGAG CGGTCGCTGA AGGACATGGA GGAGTCCATC AGGAACCTGG AGCAAGAGAT TTCTTTTGAC
1201 TTTGGCCCCA ACGGGGAGTT TGCTTACCTG TACAGCCAGT GCTACGAGCT CACCACCAAC GAATACGTCT ACCGCCTCTG
1281 CCCCTTCAAG CTTGTCTCGC AGAAACCCAA ACTCGGGGGC TCTCCCACCA GCCTTGGCAC CTGGGGCTCA TGGATTGGCC
1361 CCGACCACGA CAAGTTCAGT GCCATGAAGT ATGAGCAAGG CACGGGCTGC TGGCAGGGCC CCAACCGCTC CACCACCGTG
1441 CGCCTCCTGT GCGGGAAAGA GACCATGGTG ACCAGCACCA CAGAGCCCAG TCGCTGCGAG TACCTCATGG AGCTGATGAC
1521 GCCAGCCGCC TGCCCGGAGC CACCGCCTGA AGCACCCACC GAAGACGACC ATGACGAGCT CTAG
```

Figure B. 4. DNA sequence of *homo sapiens* alpha-glucosidaseII β-subunit used in this study. Uniprot: P14314, *GLU2B_HUMAN*.

```
GANAB_mouse      MGAAIAAVAARRRRSWLSLVLAYLGVCLGITLAVDRSNFKTCDESSFCKRQRSIRPGLSP   60
GANAB_human      -MAAVAAVAARRRRSWASLVLAFLGVCLGITLAVDRSNFKTCEESSFCKRQRSIRPGLSP   59
                   **:*********** *****:******************:*****************

GANAB_mouse      YRALLDTLQLGPDALTVHLIHEVTKVLLVLELQGLQKNMTRIRIDELEPRRPRYRVPDVL   120
GANAB_human      YRALLDSLQLGPDSLTVHLIHEVTKVLLVLELQGLQKNMTRFRIDELEPRRPRYRVPDVL   119
                 ******:******:****************************:******************

GANAB_mouse      VADPPTARLSVSGRDDNSVELTVAEGPYKIILTAQPFRLDLLEDRSLLLSVNARGLMAFE   180
GANAB_human      VADPPIARLSVSGRDENSVELTMAEGPYKIILTARPFRLDLLEDRSLLLSVNARGLLEFE   179
                 ***** *********:******:***********:*********************: **

GANAB_mouse      HQRAPRVPFSDKVSLALGSVWDKIKNLFSRQESKDPAEGNGAQPEATPGDGDKPEET--Q   238
GANAB_human      HQRAPRVSPSSLPPLD----WSRQLSYLPGKDQKTQLRAMGPSLRKHPGMATSQRRLRGR   235
                 *******  *.   *     *.:  . :  ::.*   .. * . .  ** . . ..   :

GANAB_mouse      EKAEKDEPGAWEETFKTHSDSKPYGPTSVGLDFSLPGMEHVYGIPEHADSLRLKVTEGGE   298
GANAB_human      QRKMSQEPGR--------RHSKLTLTMSVGLDFSLPGMEHVYGIPEHADNLRLKVTEGGE   287
                 ::  .:***          .**     **********************.**********

GANAB_mouse      PYRLYNLDVFQYELNNPMALYGSVPVLLAHSFHRDLGIFWLNAAETWVDISSNTAGKTLF   358
GANAB_human      PYRLYNLDVFQYELYNPMALYGSVPVLLAHNPHRDLGIFWLNAAETWVDISSNTAGKTLF   347
                 ************** ***************. ****************************

GANAB_mouse      GKMLDYLQGSGETPQTDIRWMSESGIIDVFLMLGPSVFDVFRQYASLTGTQALPPLFSLG   418
GANAB_human      GKMMDYLQGSGETPQTDVRWMSETGIIDVFLLLGPSISDVFRQYASLTGTQALPPLFSLG   407
                 ***:**************:*****:*******:****: **********************

GANAB_mouse      YHQSRWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDPTRFPQPLNMLEH   478
GANAB_human      YHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSRFPQPRTMLER   467
                 *************************:**********************:***** .***:

GANAB_mouse      LASKRRKLVAIVDPHIKVDSGYRVHEELRNHGLYVKTRDGSDYEGWCWPGSASYPDFTNP   538
GANAB_human      LASKRRKLVAIVDPHIKVDSGYRVHEELRNLGLYVKTRDGSDYEGWCWPGSAGYPDFTNP   527
                 ****************************** ********************.*******

GANAB_mouse      RMRAWWSNMFSFDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAVHYGGWEHRDIHNIY   598
GANAB_human      TMRAWWANMFSYDNYEGSAPNLFVWNDMNEPSVFNGPEVTMLKDAQHYGGWEHRDVHNIY   587
                  *****:****:**********:********************** *********:****

GANAB_mouse      GLYVHMATADGLIQRSGGIERPFVLSRAFFSGSQRFGAVWTGDNTAEWDHLKISIPMCLS   658
GANAB_human      GLYVHMATADGLRQRSGGMERPFVLARAFFAGSQRFGAVWTGDNTAEWDHLKISIPMCLS   647
                 ************ *****:******:****:******************************

GANAB_mouse      LALVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHAHLDTGRREPWLLASQYQD   718
GANAB_human      LGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHAHLDTGRREPWLLPSQHND   707
                 *.**************************************************** **:**

GANAB_mouse      AIRDALFQRYSLLPFWYTLFYQAHKEGFPVMRPLWVQYPEDMSTFSIEDQFMLGDALLIH   778
GANAB_human      IIRDALGQRYSLLPFWYTLLYQAHREGIPVMRPLWVQYPQDVTTFNIDDQYLLGDALLVH   767
                  *****:************:****:**:***********:*::**.*:**::******:*

GANAB_mouse      PVSDAGAHGVQVYLPGQEEVWYDIQSYQKHHGPQTLYLPVTLSSIPVFQRGGTIVPRWMR   838
GANAB_human      PVSDSGAHGVQVYLPGQGEVWYDIQSYQKHHGPQTLYLPVTLSSIPVFQRGGTIVPRWMR   827
                 ****:************ ******************************************

GANAB_mouse      VRRSSDCMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTRHEFLLRRFSFSGSTLVSS   898
GANAB_human      VRRSSECMKDDPITLFVALSPQGTAQGELFLDDGHTFNYQTRQEFLLRRFSFSGNTLVSS   887
                 *****:**************************************:***********.*****

GANAB_mouse      SADPKGHLETPIWIERVVIMGAGKPAAVVLQTKGSPESRLSFQHDPETSVLILRKPGVSV   958
GANAB_human      SADPEGHFETPIWIERVVIIGAGKPAAVVLQTKGSPESRLSFQHDPETSVLVLRKPGINV   947
                 ****:**:***********:*********************************:*****:.*

GANAB_mouse      ASDWSIHLR-     967
GANAB_human      ASDWSIHLR*     956
                 *********
```

```
PRKCSH_mouse      MGLLLLLLLLPLCWAVEVKRPRGVSLSNHHFYEESKPFTCLDGTATIPFDQVNDDYCDCK
PRKCSH_human      -MLLPLLLLLPMCWAVEVKRPRGVSLTNHHFYDESKPFTCLDGSATIPFDQVNDDYCDCK
                    ** ******:**************:******:**********:**********

PRKCSH_mouse      DGSDEPGTAACPNGSFHCTNTGYKPLYILSSRVNDGVCDCCDGTDEYNSGTVCENTCREK
PRKCSH_human      DGSDEPGTAACPNGSFHCTNTGYKPLYIPSNRVNDGVCDCCDGTDEYNSGVICENTCKEK
                  ****************************.*.*******************.:*****:**

PRKCSH_mouse      GRKEKESLQQLAEVTREGFRLKKILIEEWKTAREEKQSKLLELQAGKKSLEDQVETLRAA
PRKCSH_human      GRKERESLQQMAEVTREGFRLKKILIEDWKKAREEKQKKLIELQAGKKSLEDQVEMLRTV
                  ****:*****:****************:**.******.**:************** **:.

PRKCSH_mouse      KEEAERPEKEAKDQHRKLWEEQQAAAKARREQERAASAFQELDDNMDGMVSLAELQTHPE
PRKCSH_human      KEEAEKPEREAKEQHQKLWEEQLAAAKAQQEQELAADAFKELDDDMDGTVSVTELQTHPE
                  *****:**:***:**:****** *****::*** **.**:****:*** **::*******

PRKCSH_mouse      LDTDGDGALSEEEAQALLSGDTQTDTTSFYDRVWAAIRDKYRSEVPPTDIPVPE---ETE
PRKCSH_human      LDTDGDGALSEAEAQALLSGDTQTDATSFYDRVWAAIRDKYRSEALPTDLPAPSAPDLTE
                  *********** *************:******************. ***:*.*.    **

PRKCSH_mouse      PKEEKPPVLPPTEEEEEEE---EEPEEEEEEEEEEEEAPPPLQPPQPPSPTEDEKMPPYD
PRKCSH_human      PKEEQPPVPSSPTEEEEEEEEEEEEAEEEEEEEDSEEAPPPLSPPQPASPAEEDKMPPYD
                  ****:***     ******   **  *******:.*******.**** **:*::******

PRKCSH_mouse      EETQAIIDAAQEARSKFEEVERSLKEMEESIRSLEQEISFDFGPSGEFAYLYSQCYELTT
PRKCSH_human      EQTQAFIDAAQEARNKFEEAERSLKDMEESIRNLEQEISFDFGPNGEFAYLYSQCYELTT
                  *:***:********.****.*****:******.***********.***************

PRKCSH_mouse      NEYVYRLCPFKLVSQKPKHGGSPTSLGTWGSWAGPDHDKFSAMKYEQGTGCWQGPNRSTT
PRKCSH_human      NEYVYRLCPFKLVSQKPKLGGSPTSLGTWGSWIGPDHDKFSAMKYEQGTGCWQGPNRSTT
                  ****************** ************* ****************************

PRKCSH_mouse      VRLLCGKETVVTSTTEPSRCEYLMELMTPAACPEPPPEAPSDGDHDEL
PRKCSH_human      VRLLCGKETMVTSTTEPSRCEYLMELMTPAACPEPPPEAPTEDDHDEL
                  *********:**********************************::.*****
'
```

Figure B. 5. Amino acid sequence alignment of full length mouse and human GluII (UniProt: Q8BHN3/P14314 and Q14697/P14314, respectivly). *Red boxes* – mutation sites used for *Mm*GluII interface 'humanisation' described in section 6.7.1 (αF724G, βL88P, βS90N).



	DMSO only
	300 nM untagged beta
	1:3 serial dilution of untagged beta (Top 243 nM, 10 points) onto 3 nM humanised GluII
	1:3 serial dilution of tested compound (Top 100 µM, 10 points) onto 3 nM humanised GluII

Figure B. 6. TR-FRET assay plate layout

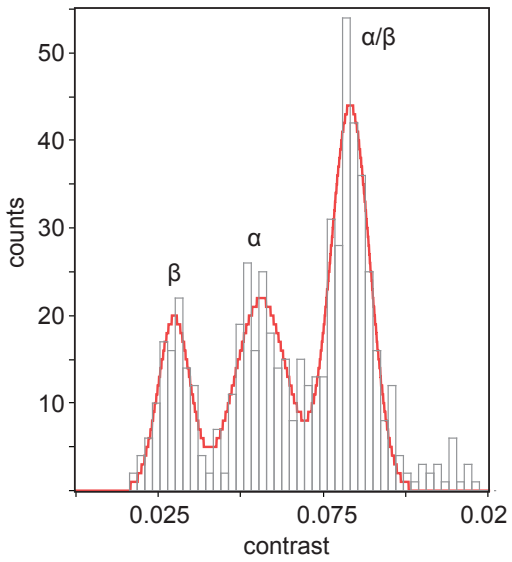


Figure B. 7. Interferometric scattering mass spectrometry (ISCAM)[340] conduced by Dr Weston Stuwe on *Mm*GluII in 20 mM HEPES, 120 mM NaCl, pH 7.4. Evaluation of the measurement gave a K_d value for the *Mm*GluII *α*-/ *β*-subunits of 2.3 ± 0.7 nM. The findings are in agreement with dta collected in this study by TR-FRET (K_d of 2.59 nM).

9. Bibliograph

1. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology 2008 9:2* **4,** 181–191 (2003).
2. Ferris, S. P., Kodali, V. K. & Kaufman, R. J. Glycoprotein folding and quality-control mechanisms in protein-folding diseases. *Disease Models & Mechanisms* **7,** 331–341 (2014).
3. Dwek, R. A., Butters, T. D., Platt, F. M. & Zitzmann, N. Targeting glycosylation as a therapeutic approach. *Nature Reviews Drug Discovery* **1,** 65–75 (2002).
4. Horne, G., Wilson, F. X., Tinsley, J., Williams, D. H. & Storer, R. Iminosugars past, present and future: medicines for tomorrow. *Drug Discovery Today* **16,** 107–118 (2011).
5. Alonzi, D. S., Scott, K. A., Dwek, R. A. & Zitzmann, N. Iminosugar antivirals: the therapeutic sweet spot. *Biochemical Society Transactions* **45,** 571–582 (2017).
6. Inouye, S., Tsuruoka, T. & Niida, T. The structure od nojirimycin, a piperidinose sugar antibiotic. **14,** 288–292 (1966).
7. Asano, N., Nash, R. J., Molyneux, R. J. & Fleet, G. W. J. Sugar-mimic glycosidase inhibitors: natural occurrence, biological activity and prospects for therapeutic application. *Tetrahedron: Asymmetry* **11,** 1645–1680 (2000).

8. Schmidt, D. D., Frommer, W. & Müller, L. Glucosidase-inhibitoren aus Bazillen. *Springer* (1979).
9. Murao, S. & Miyata, S. Isolation and Characterization New Trehalase Inhibitor, S-GI. *Agric. Biol. Chem.* **44,** 219–221 (1980).
10. EZURE, Y., Maruo, S., Miyazaki, K. & Kawamata, M. Moranoline (1-Deoxynojirimycin) Fermentation andIts Improvement. *Agricultural and Biological Chemistry* **49,** 1110–1125 (1985).
11. Mellor, H. R. *et al.* Cellular effects of deoxynojirimycin analogues: uptake, retention and inhibition of glycosphingolipid biosynthesis. 1–6 (2004).
12. Asano, N., Tomioka, E., Kizu, H. & Matsui, K. Sugars with nitrogen in the ring isolated from the leaves of Mona bombycis. *Carbohydrate Research* **253,** 235–245 (1994).
13. Asano, N., Oseki, K., Tomioka, E., Kizu, H. & Matsui, K. N-Containing sugars from *Morus alba* and their glycosidase inhibitory activities. *Carbohydrate Research* 243–255 (1994).
14. Hohenschutz, L. D. *et al.* Castanosperine, a 1,6,7,8-tretrahydroxyoctahydroindolizine alkaloid, from seeds of *Castanospermum Australe*. *Phytochemistry* **20,** 811–814 (1980).
15. Asano, N. Alkaloidal Sugar Mimetics: Biological Activities and Therapeutic Applications. *Journal of Enzyme Inhibition* **15,** 215–234 (2008).
16. Saunier, B., Kilker, R., Tkacz, J. S., Quaroni, A. & Herscovic, A. N-linked Complex Oligosaccharide Formation by 1-Deoxynojirimycin, an Inhibitor of Processing Glucosidases. *Journal of Biological Chemistry* **257,** 14155–14161 (1982).
17. Hettkamp, H., Bause, E. & Legler, G. Inhibition by nojirimycin and 1-deoxynojirimycin of microsomal glucosidases from calf liver acting on the glycoprotein oligosaccharides Glc1–3Man9GlcNAc2. *Bioscience Reports* **2,** 899–906 (1982).
18. Elbein, A. D., Tropea, O. E., Mike Mitchell & Kaushal, G. P. Kifunensine, a Potent Inhibitor of the Glycoprotein Processing Mannosidase I. *Journal of Biological Chemistry* **265,** 15599–15605 (1990).
19. Elbein, A. D. Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains. *Annual Review of Biochemistry* **56,** 497–534 (1987).
20. Andersson, U., Butters, T. D., Dwek, R. A. & Platt, F. M. N-Butyldeoxygalactonojirimycin: A More Selective Inhibitor of Glycosphingolipid Biosynthesis than N-Butyldeoxynojirimycin, In Vitro and In Vivo. *Biochemical Pharmacology* **59,** 821–829 (1999).
21. Rhinehart, B. L., Robinson, K. M., King, C.-H. R. & Liu, P. S. Castanospermine-glucosides as selective disaccharidase inhibitors. *Biochemical Pharmacology* **39,** 1537–1543 (1989).
22. Campbell, L. K., Baker, D. E. & Campbell, R. K. Miglitol: Assessment of Its Role in the Treatment of Patients with Diabetes Mellitus. *The Annals of Pharmacotherapy* **34,** 1291–1301 (2000).

23. Mitrakou, A. *et al.* Long-term Effectiveness of a New alpha-Glucosidase Inhibitor (BAY m1099-Miglitol) in Insulin-treated Type 2 Diabetes Mellitus. *Diabetic Medicine* **15,** 657–660 (1998).
24. Kingma, P. J., Sels, J. P., Nieuwenhuijzen Kruseman, A. C.Paulus P. C. A. Menheere. a-Glucosidase Inhibition by Miglitol in NIPPM Patients. *Diabetes Care* **15,** 478–483 (1992).
25. Cox, T. *et al.* Novel oral treatment of Gaucher's disease with N-butyldeoxynojirimycin (OGT 918) to decrease substrate biosynthesis. *The Lancet* **355,** 1481–1485 (2000).
26. Vunnam, R. R. & Radin, N. S. Analogs of the ceramide that inhibit glucocerebroside synthetase in mouse brain. *Chemistry and Physics of Lipids* **26,** 265–278 (1980).
27. Radin, N. S. Treatment of Gaucher diseasewith an enzyme inhibitor. **13,** 153–157 (1995).
28. Patterson, M. C., Vecchio, D., Prady, H., Abel, L. & Wraith, J. E. Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study. *The Lancet Neurology* **6,** 765–772 (2007).
29. Norez, C. *et al.* Rescue of functional delF508-CFTR channels in cystic fibrosis epithelial cells by the α-glucosidase inhibitor miglustat. *FEBS Letters* **580,** 2081–2086 (2006).
30. Hosokawa, N. *et al.* A novel ER α-mannosidase-like protein accelerates ER-associated degradation. *European Molecular Biology Organisation EMBO reports* **2,** 415–422 (2001).
31. Lubamba, B. *et al.* Airway Delivery of Low-Dose Miglustat Normalizes Nasal Potential Difference in F508del Cystic Fibrosis Mice. *Am J Respir Crit Care Med* **179,** 1022–1028 (2009).
32. Hadwiger, A., Niemann, H., Käbisch, A., Bauer, H. & Tamura, T. Appropriate glycosylation of the fms gene product is a prerequisite for its transforming potency. *The EMBO Journal* **5,** 689–694 (1986).
33. You, N. *et al.* Swainsonine inhibits growth and potentiates the cytotoxic effect of paclitaxel in hepatocellular carcinoma in vitro and in vivo. *Oncology Reports* **28,** 2091–2100 (2012).
34. Allan, G. *et al.* New Castanospermine Glycoside Analogues Inhibit Breast Cancer Cell Proliferation and Induce Apoptosis without Affecting Normal Cells. *PLoS ONE* **8,** e76411 (2013).
35. Wojtowicz, K., Januchowski, R., Sosińska, P., Nowicki, M. & Zabel, M. Effect of brefeldin A and castanospermine on resistant cell lines as supplements in anticancer therapy. *Oncology Reports* **35,** 2896–2906 (2016).
36. Shaheen, P. E. *et al.* Phase II study of the efficacy and safety of oral GD0039 in patients with locally advanced or metastatic renal cell carcinoma. *Invest New Drugs* **23,** 577–581 (2005).
37. Li, Z. *et al.* Swainsonine activates mitochondria-mediated apoptotic pathway in human lung cancer A549 cells and retards the growth of lung cancer xenografts. *International Journal of Biological Sciences* **8,** 394–405 (2012).
38. Aerts, J. M. *et al.* Pharmacological inhibition of glucosylceramide synthase enhances insulin sensitivity. *Diabetes* **56,** 1341–1349 (2007).

39. Chao, D. H. M. *et al.* The Iminosugar AMP-DNM Improves Satiety and Activates Brown Adipose Tissue Through GLP1. *Diabetes* **68,** 2223–2234 (2019).
40. Elena, S. F. & Sanjuán, R. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *Journal of Virology* **79,** 11555–11558 (2005).
41. Datema, R., Olofsson, S. & Romero, P. A. Inhibitors of protein glycosilation and glycoprotein processing in viral ststems. *Pharmaceutic Therapy* **33,** 221–286 (1987).
42. Tyrrell, B. E., Sayce, A. C., Warfield, K. L., Miller, J. L. & Zitzmann, N. Iminosugars: Promising therapeutics for influenza infection. *Critical Reviews in Microbiology* **43,** 521–545 (2016).
43. Block, T. M. *et al.* Secretion of human hepatitis B virus is inhibited by the iminosugar N-butyldeoxynojirimycin. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 2235–2239 (1994).
44. Taylor, D. L. *et al.* Loss of cytomegalovirus infectivity after treatment with castanospermine or related plant alkaloids correlates with aberrant glycoprotein synthesis. *Antiviral Research* **10,** 11–26 (1988).
45. Schlesinger, S., Koyama, A. H., Malfer, C., Gee, S. L. & Schlesinger, M. J. The effects of inhibitors of glucosidase I on the formation of Sindbis virus. *Virus Research* **2,** 139–149 (1984).
46. Schlesinger, S., Malfer, C. & Schlesinger, M. J. The Formation of Vesicular Stomatitis Virus (Sa Jnuan Strain) Becomes Temperature-sensitive When Glucose Residues Are Retained on the Oligosaccharides of the Glycoprotein. *J. Biol. Chem.* **259,** 7597–7601 (1984).
47. Wu, S.-F. *et al.* Antiviral effects of an iminosugar derivative on flavivirus infections. *Journal of Virology* **76,** 3596–3604 (2002).
48. Courageot, M.-P., Frenkiel, M.-P., Santos, Dos, C. D., Deubel, V. & Despres, P. Alpha-Glucosidase Inhibitors Reduce Dengue Virus Production by Affecting the Initial Steps of Virion Morphogenesis in the Endoplasmic Reticulum. *Journal of Virology* **74,** 564–572 (1999).
49. Sayce, A. C. *et al.* Iminosugars Inhibit Dengue Virus Production via Inhibition of ER Alpha-Glucosidases—Not Glycolipid Processing Enzymes. *PLOS Neglected Tropical Diseases* **10,** e0004524–22 (2016).
50. Vasudevan, S. G. & Hilgenfeld, R. *Dengue and Zika: Control and Antiviral Treatment Strategies*. **1062,** 1–375 (Springer Singapore, 2018).
51. Repp, R. *et al.* The Effectsof Processing Inhibitorsof N-Linked Oligosaccharides on the Intracellular Migrationof Glycoprotein E2 of Mouse Hepatitis Virus and the Maturationf Coronavirus Particles. *Journal of Biological Chemistry* **260,** 15873–15879 (1985).
52. Pinter, A., Honnen, W. J. & Li, J. S. Studies with Inhibitors of Oligosaccharide Processing Indicate a Functional Role for Complex Sugars in the Transport and Proteolysis of Friend Mink Cell Focus-Inducing Murine Leukemia Virus Envelope Proteins. *Journal of Virology* **136,** 196–210 (1984).
53. Poss, M. L., Mullins, J. I. & Hoover, E. A. Posttranslational Modifications Distinguish the Envelope Glycoprotein of the

Immunodeficiency Disease-Inducing Feline Leukemia Virus Retrovirus. **63,** 189–195 (1988).
54. Ratner, L., Vander Heyden, N. & Dedera, D. Inhibition of HIV and SIV Infectivity by Blockade of alpha-Glucosidase Activity. **181,** 180–192 (1990).
55. Stein, B. S. & Engleman, E. G. Intracellular Processing of the gp160 HIV-1 Envelope Precursor. *Journal of Biological Chemistry* **265,** 2640–2649 (1989).
56. Elbein, A. D. Glycosidase inhibitors as antiviral and/or antitumor agents. *Seminars in cell biology* **2,** 309–317 (1991).
57. Pavlovic, D. & Zitzmann, N. The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. 1–5 (2003).
58. Steinmann, E. *et al.* Antiviral effects of amantadine and iminosugar derivatives against hepatitis C virus. *Hepatology* **46,** 330–338 (2007).
59. Durantel, D. *et al.* Study of the mechanism of antiviral action of iminosugar derivatives against bovine viral diarrhea virus. *Journal of Virology* **75,** 8987–8998 (2001).
60. Chapel, C. *et al.* Reduction of the infectivity of hepatitis C virus pseudoparticles by incorporation of misfolded glycoproteins induced by glucosidase inhibitors. *Journal of General Virology* **88,** 1133–1143 (2007).
61. Elbein, A. D., Legler, G., Tlusty, A., McDowell, W. & Schwarz, R. The Effect of Deoxymannojirimycin on the Processing of Influenza Viral Glycoproteins. **235,** 579–588 (1984).
62. Warfield, K. *et al.* A Novel Iminosugar UV-12 with Activity against the Diverse Viruses Influenza and Dengue (Novel Iminosugar Antiviral for Influenza and Dengue). *Viruses* **7,** 2404–2427 (2015).
63. Warfield, K. L. *et al.* Inhibition of endoplasmic reticulum glucosidases is required for in vitro and in vivo dengue antiviral activity by the iminosugar UV-4. *Antiviral Research* **129,** 93–98 (2016).
64. Treston, A. M. & Warfield, K. L. Methods of treating Zika virus infection. *United States Patent Application Publication* 1–36 (2019).
65. Gonzalez, M. E. & Carrasco, L. Viroporins. *FEBS Letters* **552,** 28–34 (2003).
66. Wang, K., Xie, S. & Sun, B. Viral proteins function as ion channels. *Biochim. Biophys. Acta* **1808,** 510–515 (2011).
67. Schnell, J. R. & Chou, J. J. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **451,** 591–595 (2008).
68. Duff, K. C. & Ashley, R. H. The transmembrane domain of influenza A M2 protein forms amantadine-sensitive proton channels in planar lipid bilayers. *Virology* **190,** 485–489 (1992).
69. Pinto, L. H., Holsinger, L. J. & Lamb, R. A. Influenza virus M2 protein has ion channel activity. *Cell* **69,** 517–528 (1992).
70. Griffin, S. D. C. *et al.* The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Letters* **535,** 34–38 (2002).

71. Clarke, D. S. *et al.* Evidence for the Formation of a Heptameric Ion Channel Complex by the Hepatitis C Virus P7 Protein in Vitro. *Journal of Biological Chemistry* 1–13 (2006).
72. Luik, P. *et al.* The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy. *Proceedings of the National Academy of Sciences* **106,** 12712–12716 (2009).
73. Premkumar, A., Wilson, L., Ewart, G. D. & Gage, P. W. Cation-selective ion channels formed by p7 of hepatitis C virus are blocked by hexamethylene amiloride. *FEBS Letters* **557,** 99–103 (2003).
74. Griffin, S. D. C. A conserved basic loop in hepatitis C virus p7 protein is required for amantadine-sensitive ion channel activity in mammalian cells but is dispensable for localization to mitochondria. *Journal of General Virology* **85,** 451–461 (2004).
75. Wozniak, A. L. *et al.* Intracellular Proton Conductance of the Hepatitis C Virus p7 Protein and Its Contribution to Infectious Virus Production. *PLoS Pathog* **6,** e1001087–17 (2010).
76. Whitfield, T. *et al.* The influence of different lipid environments on the structure and function of the hepatitis C virus p7 ion channel protein. *Molecular Membrane Biology* **28,** 254–264 (2011).
77. Montserret, R. *et al.* NMR Structure and Ion Channel Activity of the p7 Protein from Hepatitis C Virus. *Journal of Biological Chemistry* **285,** 31446–31461 (2010).
78. Fischer, W. B. & Hsu, H.-J. Viral channel forming proteins — Modeling the target. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808,** 561–571 (2011).
79. Suzuki, T. *et al.* The Human Polyoma JC Virus Agnoprotein Acts as a Viroporin. *PLoS Pathog* **6,** e1000801–17 (2010).
80. Raghava, S., Giorda, K. M., Romano, F. B., Heuck, A. P. & Hebert, D. N. SV40 Late Protein VP4 Forms Toroidal Pores To Disrupt Membranes for Viral Release. *Biochemistry* **52,** 3939–3948 (2013).
81. Viroporins: structure, function and potential as antiviral targets. *J. Gen. Virol.* **96,** 2000–2027 (2015).
82. Lemaitre, V., Willbold, D., Watts, A. & Fischer, W. B. Full Length Vpu from HIV-1: Combining Molecular Dynamics Simulations with NMR Spectroscopy. *Journal of Biomolecular Structure and Dynamics* **23,** 485–496 (2006).
83. Chandler, D. E., Penin, F., Schulten, K. & Chipot, C. The p7 Protein of Hepatitis C Virus Forms Structurally Plastic, Minimalist Ion Channels. *PLoS Comput Biol* **8,** e1002702–10 (2012).
84. Oestringer, B. P. *et al.* Re-evaluating the p7 viroporin structure. *Nature* **562,** E8–E18 (2018).
85. Nieva, J. L., Madan, V. & Carrasco, L. Viroporins: structure and biological functions. *Nature Publishing Group* **10,** 563–574 (2012).
86. Wetherill, L. F. *et al.* High-risk human papillomavirus E5 oncoprotein displays channel-forming activity sensitive to small-molecule inhibitors. *Journal of Virology* **86,** 5341–5351 (2012).

87. Agarkova, I. *et al.* Chlorovirus-mediated membrane depolarization of Chlorella alters secondary active transport of solutes. *Journal of Virology* **82,** 12181–12190 (2008).
88. Hyser, J. M., Collinson-Pautz, M. R., Utama, B. & Estes, M. K. Rotavirus Disrupts Calcium Homeostasis by NSP4 Viroporin Activity. *Journal of Virology* **1,** 1–12 (2010).
89. Aldabe, R., Irurzun, A. & Carrasco, L. Poliovirus Protein 2BC Increases Cytosolic Free Calcium Concentrations. *Journal of Virology* **71,** 6214–6217 (1997).
90. Madan, V., Castelló, A. & Carrasco, L. Viroporins from RNA viruses induce caspase-dependent apoptosis. *Cellular Microbiology* **10,** 437–451 (2008).
91. Chami, M., Ferrari, D., Nicotera, P., Paterlini-Bréchot, P. & Rizzuto, R. Caspase-dependent alterations of Ca2+ signaling in the induction of apoptosis by hepatitis B virus X protein. *Journal of Biological Chemistry* **278,** 31745–31755 (2003).
92. Davies, W. L. *et al.* Antiviral Activity of 1-Adamantanamine (Amantadine). *Science* **155,** 826–863 (1964).
93. Togo, Y., Hornick, R. B. & Dawkins, A. T. Studies on Influenza in Man - 1st Double-blind studies designed to assess prophylactic efficacy of amantadine hydrochloride against A2/Rockville/1/65 strain. *The Journal of the American Medical Association* **203,** 1–6 (1968).
94. Dawkins, A. T., Gallager, L. R., Togo, Y., Hornick, R. B. & Harris, B. A. Studies on Influenza in Man - 2st Double-blind studies designed to assess prophylactic efficacy of amantadine hydrochloride against A2/Rockville/1/65 strain. *The Journal of the American Medical Association* **203,** 1–5 (1968).
95. Wingfield, W. L., Pollack, D. & Grunert, R. R. Therapeutic Efficacy of Amantadine HCl and Rimantadine HCl in Naturally Occurring Influenza A2 Respiratory Illness in Man. **281,** 579–584 (1969).
96. Hay, A. J., Wolstenholme, A. J., Skehel, J. J. & Smith, M. H. The molecular basis of the specific anti-influenza action of amantadine. *The EMBO Journal* **4,** 3021–3024 (1985).
97. Cady, S. D. *et al.* Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* **463,** 689–693 (2010).
98. Cross, T. A., Murray, D. T. & Watts, A. Helical membrane protein conformations and their environment. *Eur Biophys J* **42,** 731–755 (2013).
99. Ewart, G. D., Mills, K., Cox, G. B. & Gage, P. W. Amiloride derivatives block ion channel activity and enhancement of virus-like particle budding caused by HIV-1 protein Vpu. *Eur Biophys J* **31,** 26–35 (2014).
100. Römer, W. *et al.* Channel Activity of a Viral Transmembrane Peptide in Micro-BLMs: Vpu 1-32from HIV-1. *J. Am. Chem. Soc.* **126,** 16267–16274 (2004).
101. Lemaitre, V., Ali, R., Kim, C. G., Watts, A. & Fischer, W. B. Interaction of amiloride and one of its derivatives with Vpu from HIV-1: a molecular dynamics simulation. *FEBS Letters* **563,** 75–81 (2004).

102. Kim, C. G., Lemaitre, V., Watts, A. & Fischer, W. B. Drug–protein interaction with Vpu from HIV-1: proposing binding sites for amiloride and one of its derivatives. *Anal Bioanal Chem* **386,** 2213–2217 (2006).
103. Willbold, D., HOFFMANN, S. & ROSCH, P. Secondary Structure and Tertiary Fold of the Human Immunodeficiency Virus Protein U (Vpu) Cytoplasmic Domain in Solution. *European Journal of Biochemistry* **245,** 581–588 (1997).
104. Wittlich, M., Koenig, B. W., Stoldt, M., Schmidt, H. & Willbold, D. NMR structural characterization of HIV-1 virus protein U cytoplasmic domain in the presence of dodecylphosphatidylcholine micelles. *FEBS Journal* **276,** 6560–6575 (2009).
105. Park, S. H. & Opella, S. J. Conformational changes induced by a single amino acid substitution in the trans-membrane domain of Vpu: Implications for HIV-1 susceptibility to channel blocking drugs. *Protein Science* **16,** 2205–2215 (2007).
106. Park, S. H., De Angelis, A. A., Nevzorov, A. A., Wu, C. H. & Opella, S. J. Three-Dimensional Structure of the Transmembrane Domain of Vpu from HIV-1 in Aligned Phospholipid Bicelles. *Biophys. J.* **91,** 3032–3042 (2006).
107. Op De Beeck, A. & Dubuisson, J. Topology of hepatitis C virus envelope glycoproteins. *Rev. Med. Virol.* **13,** 233–241 (2003).
108. Farci, P. Milestines in liver disease: Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Journal of Hepatology* **244,** 259–362 (1989).
109. Murphy, K., Travers, P. & Walport, M. Janeway Immunologie, 7 edn Heidelberg, Spektrum. (2009).
110. Borgia, S. M. *et al.* Identification of a Novel Hepatitis C Virus Genotype From Punjab, India: Expanding Classification of Hepatitis C Virus Into 8 Genotypes. *The Journal of Infectious Diseases* **218,** 1722–1729 (2018).
111. Nakano, T., Lau, G. M. G., Lau, G. M. L., Sugiyama, M. & Mizokami, M. An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. *Liver International* **32,** 339–345 (2012).
112. Bukh, J. The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. *Journal of Hepatology* **65,** S2–S21 (2016).
113. Simmonds, P. *et al.* Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42,** 962–973 (2005).
114. Geller, R. *et al.* Highly heterogeneous mutation rates in the hepatitis C virus genome. *Nat Microbiol* **1,** 1–7 (2016).
115. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatitis C virus infection. *Journal of Hepatology* **60,** 392–420 (2014).
116. World Health Organisation. *WHO Global Hepatitis Report 2017*. 1–83 (2017).

117. Kaito, M. *et al.* Morphology of hepatitis C and hepatitis B virus particles as detected by immunogold electron microscopy. *Med Mol Morphol* **39,** 63–71 (2006).
118. Fénéant, L., Levy, S. & Cocquerel, L. CD81 and Hepatitis C Virus (HCV) Infection. *Viruses* **6,** 535–572 (2014).
119. Dubuisson, J. & Cosset, F.-L. Virology and cell biology of the hepatitis C virus life cycle – An update. *Journal of Hepatology* **61,** S3–S13 (2014).
120. Castelli, M. *et al.* A Biologically-validated HCV E1E2 Heterodimer Structural Model. *Nature Publishing Group* **7,** 1–13 (2017).
121. Basu, A., Beyene, A., Meyer, K. & Ray, R. The hypervariable region 1 of the E2 glycoprotein of hepatitis C virus binds to glycosaminoglycans, but this binding does not lead to infection in a pseudotype system. *Journal of Virology* **78,** 4478–4486 (2004).
122. Popescu, C.-I., Rouillé, Y. & Dubuisson, J. Hepatitis C Virus Assembly Imaging. *Viruses* **3,** 2238–2254 (2011).
123. Ma, Y. *et al.* Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins. *Journal of Virology* **85,** 86–97 (2011).
124. Stapleford, K. A. & Lindenbach, B. D. Hepatitis C virus NS2 coordinates virus particle assembly through physical interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. *Journal of Virology* **85,** 1706–1717 (2011).
125. Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. A second hepatitis C virus -encoded proteinase. *Proceedings of the National Academy of Sciences* **90,** 10583–10587 (1993).
126. Hijikata, M., Akagi, T., Mizushima, H. & Shimotohno, K. Two Distinct Proteinase Activities Required for the Processing of a Putative Nonstructural Precursor Protein of Hepatitis C Virus. *Journal of Virology* **67,** 4665–4675 (1993).
127. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W. & Weber, P. C. Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease–helicase. *Structure* **7,** 1353–1363 (1999).
128. Wölk, B., Sansonno, D., Kräusslich, H.-G. & Moradpour, D. Subcellular Localization, Stability, and *trans*-Cleavage Competence of the Hepatitis C Virus NS3-NS4A Complex Expressed in Tetracycline-Regulated Cell Lines. *Journal of Virology* **74,** 2293–2304 (1999).
129. Egger, D. *et al.* Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *Journal of Virology* **76,** 5974–5984 (2002).
130. Quintavalle, M., Sambucini, S., Di Pietro, C., De Francesco, R. & Neddermann, P. The alpha isoform of protein kinase CKI is responsible for hepatitis C virus NS5A hyperphosphorylation. *Journal of Virology* **80,** 11305–11312 (2006).
131. Moradpour, D., Penin, F. & Rice, C. M. Replication of hepatitis C virus. *Nat Rev Micro* **5,** 453–463 (2007).

132. Neddermann, P. *et al.* Reduction of hepatitis C virus NS5A hyperphosphorylation by selective inhibition of cellular kinases activates viral RNA replication in cell culture. *Journal of Virology* **78,** 13306–13314 (2004).
133. Evans, M. J., Rice, C. M. & Goff, S. P. Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proceedings of the National Academy of Sciences* **101,** 13038–13043 (2004).
134. Appel, N., Pietschmann, T. & Bartenschlager, R. Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *Journal of Virology* **79,** 3187–3194 (2005).
135. Ago, H., Adachi, T., Yoshida, A. & Miyano, M. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure* **7,** 1417–1426 (1999).
136. Lesburg, C. A. *et al.* Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nature America Inc.* **6,** 1–7 (1999).
137. Bressanelli, S., Tomei, L., Rey, F. A. & De Francesco, R. Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *Journal of Virology* **76,** 3482–3492 (2002).
138. Bartenschlager, R. & Pietschmann, T. Efficient hepatitis C virus cell culture system: What a difference the host cell makes. *Proceedings of the National Academy of Sciences* **102,** 9739–9740 (2005).
139. Lindenbach, B. D. *et al.* Complete Replication of Hepatitis C Virus in Cell Culture. *Bioscience Reports* **309,** 623–626 (2005).
140. Manns, M. P. & Hahn, von, T. Novel therapies for hepatitis C — one pill fits all? *Nature Publishing Group* **12,** 595–610 (2013).
141. Bartenschlager, R. *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy*. 1–353 (2013).
142. Wiegand, J. *et al.* Early monotherapy with pegylated interferon alpha-2b for acute hepatitis C infection: The HEP-NET acute-HCV-II study. *Hepatology* **43,** 250–256 (2006).
143. Hussaini, T. Paritaprevir/ritonavir-ombitasvir and dasabuvir, the 3D regimen for the treatment of chronic hepatitis C virus infection: a concise review. *HMER* **8,** 61–8 (2016).
144. Wedemeyer, H. Towards interferon-free treatment for all HCV genotypes. *The Lancet* **385,** 2443–2445 (2015).
145. Zhang, J., Nguyen, D. & Hu, K.-Q. Chronic Hepatitis C Virus Infection: A Review of Current Direct-Acting Antiviral Treatment Strategies. *North American Journal of Medical Science* **9,** 47–54 (2016).
146. Paolucci, S. *et al.* Naturally occurring resistance mutations to inhibitors of HCV NS5A region and NS5B polymerase in DAA treatment-naïve patients. *Virology Journal* **10,** 1–7 (2013).
147. Sorbo, M. C. *et al.* Hepatitis C virus drug resistance associated substitutions and their clinical relevance_ Update 2018. *Drug Resistance Updates* **37,** 17–39 (2018).

148. Vieyres, G. *et al.* Subcellular Localization and Function of an Epitope-Tagged p7 Viroporin in Hepatitis C Virus-Producing Cells. *Journal of Virology* **87,** 1664–1678 (2013).
149. Carrere-Kremer, S. *et al.* Subcellular Localization and Topology of the p7 Polypeptide of Hepatitis C Virus. *Journal of Virology* **76,** 3720–3730 (2002).
150. Griffin, S., Clarke, D., McCormick, C., Rowlands, D. & Harris, M. Signal Peptide Cleavage and Internal Targeting Signals Direct the Hepatitis C Virus p7 Protein to Distinct Intracellular Membranes. *Journal of Virology* **79,** 15525–15536 (2005).
151. Isherwood, B. J. & Patel, A. H. Analysis of the processing and transmembrane topology of the E2p7 protein of hepatitis C virus. *Journal of General Virology* **86,** 667–676 (2005).
152. Lin, C., Lindenbach, B. D., Pragai, B. M., McCourt, D. W. & Rice, C. M. Processing in the Hepatitis C Virus E2-NS2 Region: Identification of p7 and Two Distinct E2-specific Products with Different C Termini. *Journal of Virology* **68,** 5063–5073 (1994).
153. Mizushima, H. *et al.* Two Hepatitis C Virus Glycoprotein E2 Products with Different C Termini. *Journal of Virology* **68,** 6215–6222 (1994).
154. Sakai, A. *et al.* The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. *Proc. Natl. Acad. Sci. U.S.A.* 1–6 (2003).
155. Brohm, C. *et al.* Characterization of Determinants Important for Hepatitis C Virus p7 Function in Morphogenesis by Using trans-Complementation. *Journal of Virology* **83,** 11682–11693 (2009).
156. Jones, C. T., Murray, C. L., Eastman, D. K., Tassello, J. & Rice, C. M. Hepatitis C Virus p7 and NS2 Proteins Are Essential for Production of Infectious Virus. *Journal of Virology* **81,** 8374–8383 (2007).
157. Boson, B., Granio, O., Bartenschlager, R. & Cosset, F.-L. A Concerted Action of Hepatitis C Virus P7 and Nonstructural Protein 2 Regulates Core Localization at the Endoplasmic Reticulum and Virus Assembly. *PLoS Pathog* **7,** e1002144–17 (2011).
158. Tedbury, P. *et al.* The subcellular localization of the hepatitis C virus non-structural protein NS2 is regulated by an ion channel-independent function of the p7 protein. *Journal of General Virology* **92,** 819–830 (2011).
159. Convery, O. *et al.* The hepatitis C virus (HCV) protein, p7, suppresses inflammatory responses to tumor necrosis factor (TNF)- a via signal transducer and activator of transcription (STAT)3 and extracellular signal-regulated kinase (ERK)–mediated induction of suppressor of cytokine signaling (SOCS)3. *The FASEB Journal* **33,** 8732–8744 (2019).
160. Bentham, M. J., Foster, T. L., McCormick, C. & Griffin, S. Mutations in hepatitis C virus p7 reduce both the egress and infectivity of assembled particles via impaired proton channel function. *Journal of General Virology* **94,** 2236–2248 (2013).
161. Jirasko, V. *et al.* Structural and Functional Characterization of Nonstructural Protein 2 for Its Role in Hepatitis C Virus Assembly. *J. Biol. Chem.* **283,** 28546–28562 (2008).

162. Hagen, N., Bayer, K., Rösch, K. & Schindler, M. The Intraviral Protein Interaction Network of Hepatitis C Virus. *The American Society of Biochemistry and Molecular Biology* 1–14 (2014).
163. Lee, G. Y., Lee, S., Lee, H.-R. & Yoo, Y. D. Hepatitis C virus p7 mediates membrane-to-membrane adhesion. *Biochim. Biophys. Acta* **1861,** 1096–1101 (2016).
164. Cook, G. A., Dawson, L. A., Tian, Y. & Opella, S. J. Three-Dimensional Structure and Interaction Studies of Hepatitis C Virus p7 in 1,2-Dihexanoyl- sn-glycero-3-phosphocholine by Solution Nuclear Magnetic Resonance. *Biochemistry* **52,** 5295–5303 (2013).
165. Foster, T. L. *et al.* Structure-guided design affirms inhibitors of hepatitis C virus p7 as a viable class of antivirals targeting virion release. *Hepatology* **59,** 408–422 (2014).
166. Oestringer, B. P. *et al.* Hepatitis C virus sequence divergence preserves p7 viroporin structural and dynamic features. *Nature Publishing Group* **9,** 8383 (2019).
167. Chew, C. F., Vijayan, R., Chang, J., Zitzmann, N. & Biggin, P. C. Determination of Pore-Lining Residues in the Hepatitis C Virus p7 Protein. *Biophysj* **96,** L10–L12 (2009).
168. OuYang, B. *et al.* Unusual architecture of the p7 channel from hepatitis C virus. *Nature* **498,** 521–525 (2013).
169. Holzmann, N., Chipot, C., Penin, F. & Dehez, F. Assessing the physiological relevance of alternate architectures of the p7 protein of hepatitis C virus in different environments. *Bioorganic & Medicinal Chemistry* **24,** 4920–4927 (2016).
170. Fischer, W. B. *et al.* Transmembrane Peptide NB of Influenza B: A Simulation, Structure, and Conductance Study †. *Biochemistry* **39,** 12708–12716 (2000).
171. Gan, S. W., Surya, W., Vararattanavech, A. & Torres, J. Two Different Conformations in Hepatitis C Virus p7 Protein Account for Proton Transport and Dye Release. *PLoS ONE* **9,** e78494 (2014).
172. Soranzo, T. *et al.* Functional characterization of p7 viroporin from hepatitis C virus produced in a cell-free expression system. *Protein Expression and Purification* **118,** 83–91 (2016).
173. Griffin, S. *et al.* Genotype-dependent sensitivity of hepatitis C virus to inhibitors of the p7 ion channel. *Hepatology* **48,** 1779–1790 (2008).
174. Foster, T. L. *et al.* Resistance mutations define specific antiviral effects for inhibitors of the hepatitis C virus p7 ion channel. *Hepatology* **54,** 79–90 (2011).
175. StGelais, C. *et al.* Determinants of Hepatitis C Virus p7 Ion Channel Function and Drug Sensitivity Identified In Vitro. *Journal of Virology* **83,** 7970–7981 (2009).
176. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1473,** 4–8 (1999).

177. Pfeffer, S. *et al.* Structure of the mammalian oligosaccharyl- transferase complex in the native ER protein translocon. *Nature Communications* **5,** 1–8 (2019).
178. Mellquist, J. L., Kasturi, L., Spitalnik, S. L. & Shakin-Eshleman, S. H. The Amino Acid Following an Asn-X-Ser/Thr Sequon Is an Important Determinant of N-Linked Core Glycosylation Efficiency. *Biochemistry* **37,** 6833–6837 (1998).
179. Palcic, M. M. *et al.* Processing alpha-glucosidase I is an inverting glycosidase. *Glycoconjugate Journal* **16,** 351–355 (1999).
180. Schallus, T. *et al.* Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein. *Molecular Biology of the Cell* **19,** 3404–3414 (2008).
181. Qin, S.-Y. *et al.* Malectin Forms a Complex with Ribophorin I for Enhanced Association with Misfolded Glycoproteins. *J. Biol. Chem.* **287,** 38080–38089 (2012).
182. Hebert, D. N., Foellmer, B. & Helenius, A. Glucose Trimming and Reglucosylation Determine Glycoprotein Association with Calnexin in the Endoplasmic Reticulum. *Cell* **81,** 425–433 (1995).
183. Nauseef, W. M., McCormick, S. J. & Clark, R. A. Calreticulin Functions as a Molecular Chaperone in the Biosynthesis of Myeloperoxidase. *J. Biol. Chem.* **270,** 1–7 (1994).
184. Bertozzi, C. R. & Rabuka, D. *Structural Basis of Glycan Diversity*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009).
185. Trombetta, S. E., Gañan, S. A. & Parodi, A. J. The UDP-Glc:glycoprotein glucosyltrasferase is a soluble protein of the endoplasmic reticulum. *Glycobiology* **1,** 155–161 (1991).
186. Sousa, M. C., Ferrero-Garcia, M. A. & Parodi, A. J. Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* **31,** 97–105 (2002).
187. Gonzalez, D. S., Karaveg, K., Vandersall-Nairn, A. S., Lal, A. & Moremen, K. W. Identification, Expression, and Characterization of a cDNA Encoding Human Endoplasmic Reticulum Mannosidase I, the Enzyme That Catalyzes the First Mannose Trimming Step in Mammalian Asn-linked Oligosaccharide Biosynthesis*. *J. Biol. Chem.* **274,** 21375–21386 (1999).
188. De Praeter, C. M. *et al.* A Novel Disorder Caused by Defective Biosynthesis of N-Linked Oligosaccharides Due to Glucosidase I Deficiency. *The American Journal of Human Genetics* **66,** 1744–1756 (2000).
189. Porath, B. *et al.* Mutations in GANAB, Encoding the Glucosidase IIα Subunit, Cause Autosomal-Dominant Polycystic Kidney and Liver Disease. *The American Journal of Human Genetics* **98,** 1193–1207 (2016).
190. Vasylyeva, T. L. GANAB and PKD1 Variations in a 12 Years Old Female Patient With Early Onset of Autosomal Dominant Polycystic Kidney Disease. *fgene-10-00044.tex* **10,** 1–6 (2019).

191. Sampson, H. M. & Thomas, D. Y. Protein Trafficking Diseases: Small Molecule Approaches. *Wiley Encyclopedia of Chemical Biology* **167,** 1–9 (2008).
192. Amara, J. F., Cheng, S. H. & Smith, A. E. Intracellular protein trafficking defects in human disease. *Trends in Cell Biology* **2,** 145–149 (1992).
193. Hammond, C., Braakman, I. & Helenius, A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proceedings of the National Academy of Sciences* **91,** 913–917 (1994).
194. Danilczyk, U. G. & Williams, D. B. The lectin chaperone calnexin utilizes polypeptide-based interactions to associate with many of its substrates in vivo. *Journal of Biological Chemistry* **276,** 25532–25540 (2001).
195. Kozlov, G. *et al.* Structural basis of carbohydrate recognition by calreticulin. *J. Biol. Chem.* **285,** 38612–38620 (2010).
196. Frickel, E.-M. *et al.* TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proceedings of the National Academy of Sciences* **99,** 1954–1959 (2002).
197. Kozlov, G. *et al.* Structural Basis of Cyclophilin B Binding by the Calnexin/Calreticulin P-domain. *J. Biol. Chem.* **285,** 35551–35557 (2010).
198. Roversi, P. *et al.* Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint. *Proc. Natl. Acad. Sci. U.S.A.* **114,** 8544–8549 (2017).
199. Shin, G.-C. *et al.* PRKCSH contributes to tumorigenesis by selective boosting of IRE1 signaling pathway. *Nature Communications* **10,** 1–16 (2019).
200. Trombetta, E. S., Simons, J. F. & Helenius, A. Endoplasmic Reticulum Glucosidase II Is Composed of a Catalytic Subunit, Conserved from Yeast to Mammals, and a Tightly Bound Noncatalytic HDEL-containing Subunit*. *J. Biol. Chem.* **271,** 27502–27516 (1996).
201. Caputo, A. T. *et al.* Structures of mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E4630–8 (2016).
202. Ardèvol, A. & Rovira, C. The Molecular Mechanism of Enzymatic Glycosyl Transfer with Retention of Configuration: Evidence for a Short-Lived Oxocarbenium-Like Species. *Angewandte Chemie International Edition* **50,** 10897–10901 (2011).
203. Frandsen, T. P. & Svensson, B. Plant alpha-glucosidases of the glycoside hydrolase family 31. Molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. *Plant Molecular Biology* **37,** 1–13 (1998).
204. Totani, K., Ihara, Y., Matsuo, I. & Ito, Y. Substrate specificity analysis of endoplasmic reticulum glucosidase II using synthetic high mannose-type glycans. *Journal of Biological Chemistry* **281,** 31502–31508 (2006).

205. Petrescu, A. J. *et al.* The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *European Molecular Biology Organisation EMBO reports* **16,** 4302–4310 (1997).
206. Jakob, C. A., Burda, P., Roth, J. & Aebi, M. Degradation of misfolded endoplasmic reticulum glycoproteins in Saccharomyces cerevisiae is determined by a specific oligosaccharide structure. *J Cell Biol* **142,** 1223–1233 (1998).
207. Hubbard, S. C. & Robbins, P. W. Synthesis and Processing of Protein-linked Oligosaccharides in Vivo. *J. Biol. Chem.* **254,** 4568–4576 (1978).
208. Caramelo, J. J. & Parodi, A. J. Getting in and out from calnexin/calreticulin cycles. *Journal of Biological Chemistry* **283,** 10221–10225 (2008).
209. Olson, L. J. *et al.* Structure of the lectin mannose 6-phosphate receptor homology (MRH) domain of glucosidase II, an enzyme that regulates glycoprotein folding quality control in the endoplasmic reticulum. *J. Biol. Chem.* **288,** 16460–16475 (2013).
210. Olson, L. J. *et al.* Crystal Structure and Functional Analyses of the Lectin Domain of Glucosidase II: Insights into Oligomannose Recognition. *Biochemistry* **54,** 4097–4111 (2015).
211. Pelletier, M. F. *et al.* The heterodimeric structure of glucosidase II is required for its activity, solubility, and localization in vivo. *Glycobiology* **10,** 815–827 (2000).
212. Feng, J. Processing enzyme glucosidase II: proposed catalytic residues and developmental regulation during the ontogeny of the mouse mammary gland. *Glycobiology* **14,** 909–921 (2004).
213. Treml, K., Meimaroglou, D., Hentges, A. & Bause, E. The alpha- and beta-subunits are required for expression of catalytic activity in the hetero-dimeric glucosidase II complex from human liver. *Glycobiology* 1–10 (2000).
214. Semenza, J. C., Hardwick, K. G., Dean, N. & Pelham, H. F. B. ERD2, a Yeast Gene Required for the Receptor-Mediated Retrieval of Luminal ER Proteins from the Secretory Pathway. *Cell* **61,** 1349–1357 (1990).
215. Deprez, P., Gautschi, M. & Helenius, A. More than one glycan is needed for ER glucosidase II to allow entry of glycoproteins into the calnexin/calreticulin cycle. *Mol. Cell* **19,** 183–195 (2005).
216. Brauer, P. *et al.* Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science* **363,** 1–6 (2019).
217. D'Alessio, C., Fernandez, F., Trombetta, E. S. & Parodi, A. J. Genetic Evidence for the Heterodimeric Structure of Glucosidase II. **274,** 25899–25905 (1999).
218. Li, D. & Caffrey, M. Structure and Functional Characterization of Membrane Integral Proteins in the Lipid Cubic Phase. *Journal of Molecular Biology* 1–20 (2020). doi:10.1016/j.jmb.2020.02.024
219. Baker, M. Making membrane proteins for structures: a trillion tiny tweaks. *Nature Publishing Group* **7,** 429–434 (2010).

220. Lluis, M. W., Godfroy, J. I. & Yin, H. Protein engineering methods applied to membrane protein targets. *Protein Engineering Design and Selection* **26,** 91–100 (2013).
221. Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science* **248,** 1088–1092 (1990).
222. Klare, J. P. & Steinhoff, H.-J. Spin labeling EPR. *Photosynth Res* **102,** 377–390 (2009).
223. Mchaourab, H. S., Lietzow, M. A., Hideg, K. & Hubbell, W. L. Motion of Spin-Labeled Side Chains in T4 Lysozyme. Correlation with Protein Structure and Dynamics †. *Biochemistry* **35,** 7692–7704 (1996).
224. Griffith, O. H., Cornell, D. W. & McConnell, M. Nitrogen Hyperfine Tensor and g Tensor of Nitroxide Radicals. *The Journal of Chemical Physics* **43,** (1965).
225. Livshits, V. A., Dzikovski, B. G. & Marsh, D. Anisotropic motion effects in CW non-linear EPR spectra: relaxation enhancement of lipid spin labels. *Journal of Magnetic Resonance* **162,** 429–442 (2003).
226. Weil, J. A. & Bolton, J. R. Electron paramagnetic resonance: elementary theory and practical applications. (2007).
227. Klug, C. S. & Feix, J. B. Methods and applications of site-directed spin labeling EPR spectroscopy. *Methods Cell Biol.* **84,** 617–658 (2008).
228. Traeubl, H. & Sackmann, E. Crystalline-liquid crystaline phase transition of lipid model membranes. III. Structure of a steroid-lecithin system below and above the lipid-phase transition. *J. Am. Chem. Soc.* **94,** 4499–4510 (1972).
229. Junk, M. J. Assessing the functional structure of molecular transporters by EPR spectroscopy. (2012).
230. Watts, M. T., Lu, M. L. & Eastman, M. P. Electron spin resonance studies of Heisenberg spin exchange. Effect of macrocyclic polyethers on the spin exchange rate for ion pairs. *ACS Publication* **77,** 625–628 (1973).
231. Jeschke, G. DEER Distance Measurements on Proteins. *Annu. Rev. Phys. Chem.* **63,** 419–446 (2012).
232. Milov, A. D., Ponomarev, A. B. & Tsvetkov, Y. D. Electron-electron double resonance in electron spin echo: Model biradical systems and the sensitized photolysis of decalin. *Chemical Physics Letters* **110,** 67–72 (1984).
233. Kaplan, D. E. & Hahn, E. L. Expériences de double irradiation en résonance magnétique par la méthode d'impulsions. *J. Phys. Radium* **19,** 821–825 (1958).
234. Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Physical Chemistry Chemical Physics* **9,** 1895–16 (2007).
235. Schiemann, O. & Prisner, T. F. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Quart. Rev. Biophys.* **40,** 1–53 (2007).
236. Hahn, E. L. Spin echos. *Physical Review* **80,** 580–602 (1950).

237. Rogers, R. Dead-Time Free Measurement of Dipole–Dipole Interactions between Electron Spins. *Process Safety and Environmental Protection* **142,** 1–10 (1998).
238. Jeschke, G., Sajid, M., Schulte, M. & Godt, A. Three-spin correlations in double electron-electron resonance. *Physical Chemistry Chemical Physics* **11,** 6580–6591 (2009).
239. Milov, A. D. & Tsvetkov, Y. D. Double Electron-Electron Resonance in Electron Spin Echo Conformations of Spin-Labeled Poly-4-Vinilpyridine in Glassy Solutions. *Appl Magn Reson* 495–504 (1997).
240. Jeschke, G., Koch, A., Jonas, U. & Godt, A. Direct Conversion of EPR Dipolar Time Evolution Data to Distance Distributions. *Journal of Magnetic Resonance* **155,** 72–82 (2002).
241. Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl Magn Reson* **30,** 473–498 (2006).
242. Chiang, Y.-W., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *Journal of Magnetic Resonance* **172,** 279–295 (2005).
243. Degoree, F. *et al.* HTRF: A Technology Tailored for Drug Discovery –A Review of Theoreti- cal Aspects and Recent Applications. *Current Chemical Genomices* 1–11 (2009).
244. Yan, Y. & Marriott, G. Analysis of protein interactions using fluorescence technologies. *Current Opinion in Chemical Biology* **7,** 635–640 (2003).
245. Leavesley, S. J. & Rich, T. C. Overcoming limitations of FRET measurements. *Cytometry* **89,** 325–327 (2016).
246. Miozzari, G. F. & Yanofsky, C. Translation of the Leader Region of the Escherichia coli Tryptophan Operon. *Journal of Bacteriology* **133,** 1457–1466 (1977).
247. Kleid, D. G. *et al.* Cloned Viral Protein Vaccine for Foot-and-Mouth Disease: Responses in Cattle ansd Swine. *Science* **214,** 1–6 (1981).
248. Estephan, R. *et al.* Biosynthesis and NMR Analysis of a 73-Residue Domain of a Saccharomyces cerevisiaeG Protein-Coupled Receptor. *Biochemistry* **44,** 11795–11810 (2005).
249. Cook, G. A., Stefer, S. & Opella, S. J. Expression and purification of the membrane protein p7 from hepatitis C virus. *Biopolymers* **96,** 32–40 (2011).
250. Dev, J., Brüschweiler, S., OuYang, B. & Chou, J. J. Transverse relaxation dispersion of the p7 membrane channel from hepatitis C virus reveals conformational breathing. *J. Biomol. NMR* **61,** 369–378 (2015).
251. Scull, M. A. *et al.* The N-terminal Helical Region of the Hepatitis C Virus p7 Ion Channel Protein Is Critical for Infectious Virus Production. *PLoS Pathog* **11,** e1005297 (2015).
252. Shiryaev, V. A. *et al.* Molecular design, synthesis and biological evaluation of cage compound-based inhibitors of hepatitis C virus p7 ion channels. *Eur J Med Chem* **158,** 214–235 (2018).

253. Walter, S. *et al.* Ion Channel Function and Cross-Species Determinants in Viral Assembly of Nonprimate Hepacivirus p7. *Journal of Virology* **90,** 5075–5089 (2016).
254. Chen, W. *et al.* The Unusual Transmembrane Partition of the Hexameric Channel of the Hepatitis C Virus. *Structure* **26,** 627–634.e4 (2018).
255. Montserret, R. *et al.* Structural Analysis of the Heparin-Binding Site of the NC1 Domain of Collagen XIV by CD and NMR †,‡. *Biochemistry* **38,** 6479–6488 (1999).
256. Zhao, L. *et al.* Structural basis of interaction between the hepatitis C virus p7 channel and its blocker hexamethylene amiloride. *Protein & Cell* **7,** 300–304 (2016).
257. Chen, W., OuYang, B. & Chou, J. J. Critical Effect of the Detergent:Protein Ratio on the Formation of the Hepatitis C Virus p7 Channel. *Biochemistry* **58,** 3834–3837 (2019).
258. Stansfeld, P. J. *et al.* MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure/Folding and Design* **23,** 1350–1361 (2015).
259. Kalita, M. M. & Fischer, W. B. Asymmetric dynamics of ion channel forming proteins - Hepatitis C virus (HCV) p7 bundles. *Biochim. Biophys. Acta* **1858,** 1462–1470 (2016).
260. Fernández, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr. Opin. Struct. Biol.* **13,** 570–580 (2003).
261. Gardner, K. H., Rosen, M. K. & Kay, L. E. Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* **36,** 1389–1401 (1997).
262. Korepanova, A. & Matayoshi, E. D. *HPLC-SEC Characterization of Membrane Protein-Detergent Complexes*. *Current Protocols in Protein Science* **328,** 29.5.1–29.5.12 (John Wiley & Sons, Inc., 2012).
263. Turro, N. J., Turro, N. J., Yekta, A. & Yekta, A. Luminescent probes for detergent solutions. A simple procedure for determination of the mean aggregation number of micelles. *J. Am. Chem. Soc.* **100,** 5951–5952 (1978).
264. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology 2008 9:2* **9,** 112–124 (2008).
265. Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research* **32,** 668–673
266. Abdul-Gader, A., Miles, A. J. & Wallace, B. A. A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics* **27,** 1630–1636 (2011).
267. Compton, L. A. & Johnson, W. C. Analysis of Protein Circular Dichroism Spectra for Secondary Structure Using a Simple Matrix Multiplication. *Anal. Biochem.* **155,** 155–167 (1985).

268. Manavalan, P. & Johnson, W. C. Variable Selection Method Improves the Prediction of Protein Secondary Structure from Circular Dichroism Spectra. *Anal. Biochem.* **167,** 76–85 (1987).
269. Sreerama, N. & Woody, R. W. Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Anal. Biochem.* **287,** 252–260 (2000).
270. Mao, D., Wachter, E. & Wallace, B. A. Folding of the mitochondrial proton adenosine triphosphatase proteolipid channel in phospholipid vesicles. *Biochemistry* **21,** 4960–4968 (1982).
271. Etienne, E., Le Breton, N., Martinho, M., Mileo, E. & Belle, V. SimLabel: a graphical user interface to simulate continuous wave EPR spectra from site-directed spin labeling experiments. *Magn. Reson. Chem.* **55,** 714–719 (2017).
272. Budil, D. E. E. A. Nonlinear-Least-Squares Analysis of Slow-Motion EPR Spectra in One and Two Dimensions Using a Modified Levenberg-Marquardt Algorithm. 1–35 (1996).
273. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *Journal of Magnetic Resonance* **178,** 42–55 (2006).
274. Zou, P. & Mchaourab, H. S. Increased Sensitivity and Extended Range of Distance Measurements in Spin-Labeled Membrane Proteins: Q-Band Double Electron-Electron Resonance and Nanoscale Bilayers. *Biophysj* **98,** L18–L20 (2010).
275. Endeward, B., Butterwick, J. A., MacKinnon, R. & Prisner, T. F. Pulsed Electron−Electron Double-Resonance Determination of Spin-Label Distances and Orientations on the Tetrameric Potassium Ion Channel KcsA. *J. Am. Chem. Soc.* **131,** 15246–15250 (2009).
276. Giannoulis, A., Ward, R., Branigan, E., Naismith, J. H. & Bode, B. E. PELDOR in rotationally symmetric homo-oligomers. *Molecular Physics* **111,** 2845–2854 (2013).
277. Pliotas, C. *et al.* Conformational state of the MscS mechanosensitive channel in solution revealed by pulsed electron– electron double resonance (PELDOR) spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 2675–2682 (2012). doi:10.1073/pnas.1202286109/-/DCSupplemental
278. Valera, S. *et al.* Accurate Extraction of Nanometer Distances in Multimers by Pulse EPR. *Chem. Eur. J.* **22,** 4700–4703 (2016).
279. Hemminga, M. A. & Berliner, L. ESR spectroscopy in membrane biophysics. **27,** (2007).
280. Polyhach, Y. *et al.* High sensitivity and versatility of the DEER experiment on nitroxide radical pairs at Q-band frequencies. *Physical Chemistry Chemical Physics* **14,** 10762–10773 (2012).
281. Tait, C. E. & Stoll, S. Coherent pump pulses in Double Electron Electron Resonance spectroscopy. *Physical Chemistry Chemical Physics* **18,** 18470–18485 (2016).
282. Jeschke, G. *DeerAnalysis2013.2 User Manual*. 1–43 (2013).
283. Banham, J. E. *et al.* Distance measurements in the borderline region of applicability of CW EPR and DEER: A model study on a homologous

series of spin-labelled peptides. *Journal of Magnetic Resonance* **191,** 202–218 (2008).
284. Raitsimring, A. M. *et al.* Gd 3+Complexes as Potential Spin Labels for High Field Pulsed EPR Distance Measurements. *J. Am. Chem. Soc.* **129,** 14138–14139 (2007).
285. Goldfarb, D. Gd3+ spin labeling for distance measurements by pulse EPR spectroscopy. *Phys Chem Chem Phys* **16,** 9685–9699 (2014).
286. Edwards, D. T. *et al.* Determining the Oligomeric Structure of Proteorhodopsin by Gd3+-Based Pulsed Dipolar Spectroscopy of Multiple Distances. *Structure/Folding and Design* **22,** 1677–1686 (2014).
287. Lovett, J. E., Lovett, B. W. & Harmer, J. DEER-Stitch: Combining three- and four-pulse DEER measurements for high sensitivity, deadtime free data. *Journal of Magnetic Resonance* **223,** 98–106 (2012).
288. Rodriguez, F. *et al.* Structural model for the protein-translocating element of the twin-arginine transport system. 1–10 (2013).
289. Ward, R. *et al.* Probing the Structure of the Mechanosensitive Channel of Small Conductance in Lipid Bilayers with Pulsed Electron-Electron Double Resonance. *Biophysj* **106,** 834–842 (2014).
290. Terstappen, G. C. & Reggiani, A. In silico research in drug discovery. *TRENDS in Pharmacologicla Science* **22,** 23–26 (2000).
291. Davey, J. G-Protein-Coupled Receptors: New Approaches to Maximise the Impact of GPCRs in Drug Discovery. *Expert Opinion on Therapeutic Targets* **8,** 165–170 (2005).
292. StGelais, C. *et al.* Inhibition of hepatitis C virus p7 membrane channels in a liposome-based assay system. *Antiviral Research* **76,** 48–58 (2007).
293. Nguitragool, W. & Miller, C. Uncoupling of a CLC Cl-/H+ exchange transporter by polyatomic anions. *Journal of Molecular Biology* **362,** 682–690 (2006).
294. Moffat, J. C. *et al.* Proton transport through influenza A virus M2 protein reconstituted in vesicles. *Biophys. J.* **94,** 434–445 (2008).
295. Schroeder, C., Ford, C. M., Wharton, S. A. & Hay, A. J. Functional reconstitution in lipid vesicles of influenza virus M2 protein expressed by baculovirus: evidence for proton transfer activity. *Journal of General Virology* **75,** 3477–3484 (1994).
296. Swift, L. L. Assembly of very low density lipoproteins in rat liver: a study of nascent particles recovered from the rough endoplasmic reticulum. *J. Lipid Res.* **36,** 395–406 (1995).
297. Mouritsen, O. G. Lipids, curvature, and nano-medicine. *Eur. J. Lipid Sci. Technol.* **113,** 1174–1187 (2011).
298. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666,** 105–117 (2004).
299. Kooijman, E. E. *et al.* Spontaneous curvature of phosphatidic acid and lysophosphatidic acid. *Biochemistry* **44,** 2097–2102 (2005).
300. Majd, H., King, M. S., Smith, A. C. & Kunji, E. R. S. Pathogenic mutations of the human mitochondrial citrate carrier SLC25A1 lead to

impaired citrate export required for lipid, dolichol, ubiquinone and sterol synthesis. *BBA - Bioenergetics* **1859,** 1–7 (2018).

301. Aricescu, A. R., Lu, W., Jones, E. Y.IUCr. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr Sect D Biol Crystallogr* **62,** 1243–1250 (2006).
302. Gasteiger, E. *et al. Protein Identification and Analysis Tools on the ExPASy Server*. 571–607 (2005).
303. Garman, E. F. Summary of lecture and practical session at Biophysics and Structural Biology at Synchrotrons workshop: cryo-cooling in macromolecular crystallography—why and how? *Biophys Rev* **11,** 1–3 (2019).
304. Incardona, M. F. *et al.* EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Rad (2009). 16, 872-879 [doi:10.1107/S0909049509036681]* **16,** 1–8 (2009).
305. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr Sect D Biol Crystallogr* **67,** 293–302 (2011).
306. Wojdyr, M., Keegan, R. & Winter, G. DIMPLE-a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands. *th European Crystallographic Meeting, UK*
307. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst (2007). 40, 658-674 [doi:10.1107/S0021889807021206]* **40,** 1–17 (2007).
308. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst (2010). D66, 486-501 [doi:10.1107/S0907444910007493]* **66,** 1–16 (2010).
309. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Cryst (2004). D60, 2210-2221 [doi:10.1107/S0907444904016427]* **60,** 1–12 (2004).
310. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst (2011). D67, 355-367 [doi:10.1107/S0907444911001314]* **67,** 1–13 (2011).
311. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst (2010). D66, 12-21 [doi:10.1107/S0907444909042073]* **66,** 1–10 (2009).
312. Brada, D. & Dubach, U. C. Isolation of a homogeneous glucosidase II from pig kidney microsomes. *Biochemistry* **141,** 149–156 (1984).
313. Saxena, S., Shailubhai, K., Dong-Yu, B. & Vijay, K. Purification and characterization of glucosidase II involved in N-linked glycoprotein processing in bovine mammary gland. **247,** 563–570 (1897).
314. Gorrec, F.IUCr. The MORPHEUS protein crystallization screen. *J Appl Crystallogr* **42,** 1035–1042 (2009).
315. Kabsch, W.IUCr. XDS. *Acta Crystallogr Sect D Biol Crystallogr* **66,** 125–132 (2010).
316. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* **67,** 235–242 (2011).

317. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* **32,** W665–W667 (2004).
318. Schlesinger, S., Malfer, C. & Schlesinger, M. J. The Formation of Vesicular Stomatitis Virus (SaJnuan Strain) Becomes Temperature-sensitive When Glucose Residues Are Retained on the Oligosaccharides of the Glycoprotein. *Journal of Biological Chemistry* **259,** 7597–7601 (1984).
319. Datema, R., Romero, P. A., Rott, R. & Schwarz, R. T. On the Role of Oligosaeeharide Trimmin 9 in the Maturation of Sindbis and Influenza Virus. *Archives of Virology* **81,** 25–39 (1984).
320. Gruters, R. A. *et al.* Interference with HIV-induced syncytium formation and viral infectivity by inhibitors of trimming glucosidase. *Nature* **330,** 74–76 (1987).
321. Karpas, A. *et al.* Aminosugar derivatives as potential anti-human immunodeficiency virus agents. *Proceedings of the National Academy of Sciences* **85,** 9229–9233 (1988).
322. Ahmed, S. P. *et al.* Antiviral Activity and Metabolism of the Castanospermine Derivative MDL 28, 574 in Cells Infected with Herpes Simplex Virus Type 2. *Biochemical and Biophysical Research* **208,** 267–273 (1995).
323. Joubert, P. H., Bam, W. J. & Manyane, N. Effect of an Alpha-Glucosidase Inhibitor (BAY m 1099) on Post-Prandial Blood Glucose and Insulin in Type II Diabetics. *European Journal of Pharmacology* **30,** 253–255 (1985).
324. OLeary, J. G. & Davis, G. L. Hepatitis C virus replication and potential targets for direct-acting agents. *Therapeutic Advances in Gastroenterology* 43–53 (2010).
325. Gloster, T. M. & Vocadlo, D. J. Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. *Nature Chemical Biology* **8,** 683–694 (2012).
326. Stierand, K., Maass, P. C. & Rarey, M. Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics* **22,** 1710–1716 (2006).
327. Fricker, P. C., Gastreich, M. & Rarey, M. Automated Drawing of Structural Molecular Formulas under Constraints. *J. Chem. Inf. Comput. Sci.* **44,** 1065–1078 (2004).
328. Hettkamp, H., Bause, E. & Legler, G. Inhibition by nojirimycin and 1-deoxynojirimycin of microsomal glucosidases from calf liver acting on the glycoprotein oligosaccaride Glc1-3Man9GlcNAc2. *Bioscience Reports* 1–8 (1982).
329. Kaushal, G. P., Pastuszak, I., Hatanaka, K.-I. & Elbein, A. D. Purification to Homogeneity and Properties of Glucosidase II from Mung Bean Seedlings and Suspension-cultured Soybean Cells. *Journal of Biological Chemistry* **265,** 16271–16279 (1990).

330. Pan, Y. T. *et al.* Castanospermine inhibits the processing of the oligosaccharide portion of the influenza viral hemagglutinin. *Biochemistry* **22,** 3975–3984 (1983).
331. Hentges, A. & Bause, E. Affinity Purification and Characterization of Glucosidase IIfrom Pig Liver. *J. Biol. Chem.* **378,** 1031–1038 (1997).
332. Alonso, J. M., Santa-Cecilia, A. & Calvo, P. Glucosidase II From rat liver microsomes. *Biochemistry* 721–727 (1991).
333. Sinnott, M. L. Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.* **90,** 1171–1202 (1990).
334. Mathis, G. Rare Earth Crytates and Homogenous Fluoroimmuniassays with Human Sera. 1–7 (1993).
335. Lepre, C. A., Moore, J. M. & Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *ACS Publication* **104,** 3641–3676 (2004).
336. Viegas, A., Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **88,** 990–994 (2011).
337. Spencer, R. K. & Nowick, J. S. A Newcomer′s Guide to Peptide Crystallography. *Israel Journal of Chemistry* **55,** 698–710 (2015).
338. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst (2010). D66, 213-221 [doi:10.1107/S0907444909052925]* **66,** 1–9 (2010).
339. Garbuio, L. *et al.* Orthogonal Spin Labeling and Gd(III)–Nitroxide Distance Measurements on Bacteriophage T4-Lysozyme. *The Journal of ...* **117,** 3145–3153 (2013).
340. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360,** 423–427 (2018).