

**Revealing atomic resolution structural insights
into membrane proteins in near-native
environments by proton detected solid-state NMR**

Dissertation

for the award of the degree

"Doctor of Philosophy" (Ph.D.)

of the Georg-August-Universität Göttingen

within the doctoral program **Chemistry**

of the Georg-August University School of Science (GAUSS)

submitted by

Tekwani Movellan Kumar

from

Andorra la vella, Andorra

Göttingen, 2020

Thesis Committee members

Professor Dr. Christian Griesinger

Dept. of NMR based structural biology, Max Planck Institute for biophysical chemistry

Professor Dr. Marina Bennati

Electron-Spin Resonance Spectroscopy, Max Planck Institute for biophysical chemistry

Loren B. Andreas, Ph.D.

Dept. of NMR based structural biology, Max Planck Institute for biophysical chemistry

Members of the Examination Board

Reviewer: Professor Dr. Christian Griesinger

Dept. of NMR based structural biology, Max Planck Institute for biophysical chemistry

Second Reviewer: Professor Dr. Marina Bennati

Electron-Spin Resonance Spectroscopy, Max Planck Institute for biophysical chemistry

Further members of the Examination Board

Loren B. Andreas, Ph.D.

Dept. of NMR based structural biology, Max Planck Institute for biophysical chemistry

Professor Dr. Franziska Thomas

Institute of organic chemistry, Heidelberg University

Professor Dr. Markus Zweckstetter

Structure Determination of Proteins Using NMR, Max Planck Institute for Biophysical Chemistry, Goettingen

Translation Structural Biology in Dementia, German Center for Neurodegenerative Disease, Goettingen

Professor Dr. Timothy A. Cross

Department of Chemistry, Florida State University, USA

Date of the oral examination: December 21st 2020

Affidativ

I hereby declare that this doctoral thesis entitles “Revealing atomic resolution structural insights into membrane proteins in near-native environments by proton detected solid-state NMR” is based on my own scientific work, except for the sources and aids quoted.

Göttingen, November 16th 2020

Acknowledgment

I would like to express my sincere gratitude to all the people who gave me the opportunity to do my Ph.D. in this excellent research environment. Especially to my supervisor, **Ph.D. Loren Andreas**, who gave me the opportunity to work on stimulating and exciting research topics during my doctorate. And, also for his support, guidance and mentoring throughout these years. I would also like to show my gratitude to **Prof. Dr. Christian Griesinger**, head of the NMRII Department for NMR-Based Structural Biology at the Max Planck Institute for Chemistry and Biophysics. I will be eternally grateful to both of you for this opportunity and for all the resources that have been made available to me during my time here.

I would like to thank **Junior Prof. Dr. Franziska Thomas** for being part of my thesis committee and the fruitful advice and support. I would like to further extend my appreciation to the extended committee board for the evaluation of the thesis.

This work would not have been possible without the support of **Dr. Stefan Becker and his group**. For that, I would like to acknowledge him and his group especially **Karin Giller, Melanie Wegstroth and Kerstin Overkamp**. Thank you all for your support and for everything you have taught me over the past years in the lab.

I would also like to thank all my colleagues, the former as well as the current members of the NMR-based structural biology department, for their useful advice and discussion especially to **Pablo, Michele, Juan and Vrinda**. And, I must also give especially acknowledgement to the solid-state NMR members **Brigitta, Eszter, Riza, Marcel, Cecily, Kai and Evgeny**.

I would like to thank Christian, Adriana and Filippo for the nice scientific discussions we had over these years.

I am indebted to **Dr. Virginie Gervais, Dr. Isabelle Muller and Prof. Dr. Ricardo Louro** for introducing protein NMR to me and having hosted during my internships at the

National Center for Scientific Research (CNRS) in Toulouse and at ITQB in Oeiras, Portugal.
Thank you.

I finally wish to thank my friends and family. Over these years, they have been those who advised, supported and encouraged me to go over these long and sometimes difficult years. I must especially mention my family, **Lidia, Ramesh, Nerea, Eric, Marc**. And, all the friends who encouraged me during this time, in particular **Gregory, Leo, Sonia, Tati and Daryna**.

And my more special gratitude to my wife, **Claudia**, for her support and comprehension along these years. She has never stopped motivating and encouraging me in the most difficult moments through this path.

Thank you, without all of you this would never have been possible.

Abstract

Membrane proteins are an essential part of any living organism. They are cellular switches that trigger metabolic cascades regulating the cell metabolism, including cell division, gene expression and others functions as needed. Despite the tremendous efforts put in place to unveil the structural details of membrane proteins, there are a few techniques available to inform on the structure-function relationship in a physiological context. Proton-detected solid-state nuclear magnetic resonance spectroscopy (ssNMR) provides unique abilities to address these challenging systems in near-native environments. However, there are still fundamental limitations such as the strong proton-proton dipolar couplings limiting high proton resolution. Although per-deuteration increases resolution, it comes with loss of structural information since only exchangeable amide protons are observed in the NMR spectra. Full protonation demands the fastest magic-angle spinning probes and the highest magnetic fields.

There are still many opportunities to improve solid-state NMR methods. We demonstrate that proton detected ssNMR under fast MAS can be used to **map the membrane-surroundings** in highly perdeuterated proteins in two independent systems. The membrane insertion of the alkane transporter, AlkL, from *Pseudomonas putida* and the human voltage dependent anionic channel, hVDAC, was investigated by proton detected solid-state NMR under 55 kHz MAS. This study showed that proton spin diffusion is sufficiently quenched using a perdeuterated protein with 100% back-exchange of amide protons at 55 kHz MAS. Having the dipolar network quenched allows for site specific information using proton-proton z-mixing experiments. And, we additionally introduce **alpha proton exchange by transamination (α -PET)**, a novel method which consists of re-introducing H α backbone protons while maintaining other protein sites highly deuterated. By applying α -PET, to both a microcrystalline α -spectrin Src-homology 3 (SH3) and a lipid reconstituted hVDAC samples, we showed an improvement in the NMR proton line widths with respect to the fully protonated samples by almost a factor of two. This allows for facile **H α assignment** as well as **unambiguous H α -H α long-range distances** adding restraints for structure calculation at 55 kHz MAS. In addition, α -PET allows protein expression in protonated media which **overcomes exchange limitations** for the amide sites often seen in membrane systems.

We applied proton detected ssNMR under fast and ultra-fast MAS to two distinct membrane proteins in membrane environments. To begin, we investigated Mic10, a double pass transmembrane helical protein that adopts a hairpin-like structure in the inner mitochondrial membrane (IMM) involved in the formation of the cristae junctions. Our data indicated that the second transmembrane domain of the protein undergoes a **conformational transition**, which we hypothesize to be, in part, involved in the **modulation of the IMM**. To terminate, we investigated matrix protein 2 (M2). M2 is a homo-tetrameric membrane proton channel from influenza A. Using a fully protonated M2 protein in combination with ultra-fast MAS (105 kHz), we for the first time obtain complete assignment of the important residue tryptophan 41 and histidine 37 (H37). The proton assignment allowed us to unambiguously assign the τ tautomer in DPhPC membranes at pH 7.8 and to identify the hydrogen bonded arrangement of the key residue **H37 by detecting a ${}^2J_{\text{NN}}$ inter imidazole-imidazole J coupling**. In addition, real-time NMR measurement at low temperature allows determination of a **high energy barrier of ~130 kcal/mol for rimantadine (rmt) pore binding** which is consistent with the structural rearrangement of M2 upon pore binding previously proposed. Furthermore, we found that **rmt pore binding disturbs the imidazole-imidazole H-bond**. Finally, the location of a structured **pore water hydrogen bonded to N δ 1 of H37** was identified at high pH by combining low temperature NMR, DNP and DFT calculation.

Altogether, in this thesis, we first introduce **two novel methods for studying membrane proteins by proton detected ssNMR under fast MAS (55 kHz)**. Secondly, on our more applied approaches, we **unveil specific structural characteristics of both membrane proteins, Mic10 and M2, in the relevant context of lipid bilayers**. These pieces of evidence further support the necessity of carrying out structural studies in close to near-native conditions since the environment might play an important role in the structure and the function as well.

Table of contents

Acknowledgment	i
----------------------	---

Abstract	iii
----------------	-----

Chapter 1

General introduction and motivations

1. Proteins	1
1.1 Proteins in biology	1
1.2 Available methods for structural characterization of proteins	3
1.3. Studies of proteins by NMR	4
1.3.1 Solution NMR and spin-spin magnetization transfer mechanisms	5
1.3.2 Solid-state NMR	8
2. Brief theoretical NMR introduction	11
2.1 Basics of nuclear magnetic resonance	11
2.1.1 NMR relaxation processes	14
2.1.2 Quantum mechanics treatment of spin interactions	15
2.2. Labeling strategies in NMR	19
3. Aims and Thesis structure	21

Chapter 2

Alpha protons as NMR probes in deuterated proteins

1. Abstract	23
2. Introduction	24
3. Material and methods	27
3.1 L-amino acid oxidase stock	27
3.2 Preparation of keto acid mix	27
3.3 Protein expression	28
3.4 NMR measurements	29
4. Results and Discussion	30
4.1 Characterization of the labeling pattern	30
4.2 Glycine is labeled stereospecifically	33
4.3 Resolution and structural data under MAS conditions	33
4.4 α -PET labeling for solution NMR	38
5. Conclusion	39
Supplementary	40

Chapter 3

Probing membrane protein insertion into lipid bilayers by solid-state NMR

1. Abstract.....	59
2. Introduction.....	60
3. Materials and methods.....	64
4. Results and discussion.....	65
5. Conclusion.....	76
Supplementary.....	78

Chapter 4

Structural and functional characterization of Mic10 from the MICOS complex

1. Introduction.....	81
1.1 Cellular organization.....	81
1.2 Mitochondria.....	82
1.2.1 Mitochondrial membrane.....	83
1.2.2 The inner mitochondrial membrane (IMM) and cristae membrane (CM).....	85
1.3 Mitochondrial contact site and cristae organizing system (MICOS).....	87
1.3.1 Mic60.....	87
1.3.2 Mic10.....	88
2. Materials and Methods.....	90
2.1 Protein design.....	90
2.1.1 Constructs used for this work.....	90
2.1.2 Constructs production.....	90
2.1.3 Test Expression of the different constructs in minimal media.....	92
2.2 Peptide synthesis.....	93
2.3 Protein purification, refolding and reconstitution.....	94
2.3.1 Protein purification.....	94
2.3.2 Refolding and reconstitution protocol.....	97
2.4 Structural and functional data acquisition.....	98
2.4.1 Circular dichroism (CD).....	98
2.4.2 Nuclear magnetic resonance (NMR).....	100
2.4.3 Enzymatic treatment.....	103
3. Results.....	104
3.1 Increasing Mic10His yields in minimal media for structural studies.....	104
3.2 Biophysical evidence for Mic10 GxxxG motifs adopting a beta sheet and/or alpha-helical conformation in detergent environment.....	105

3.2.1 Bioinformatic studies of Mic10 - Predicted domains	105
3.2.2 Biophysical evidence for the secondary helical conformation adopted by Mic10	108
3.3 Inhomogeneous NMR line shape from Mic10 reconstituted in native-like lipid environment.....	114
3.4 The glycine rich motifs in Mic10's might not be the only site driving the oligomerization	117
3.4.1 TM2 domain homo-oligomerization is highly likely by bioinformatics predictions.....	117
3.4.2 Soluble domains might have a role in the stabilization of oligomers	119
3.5 Both transmembrane domains of Mic10 show helical propensity in detergent	122
3.5.1 The presence of detergent is required for driving the TM2 towards a helical conformation.....	122
3.5.2 TM2 domain forms a helical domain of at least 19 residues	124
3.5.3 Structure of TM2.....	127
3.5.4 Studies of TM2 in lipid membranes by solid-state NMR.....	129
3.5.5 No evidence of interaction between TM1 and TM2 peptides in detergent.....	131
4. Discussion.....	132
4.1 Mic10 forms an inhomogeneous sample not suitable for ssNMR studies	132
4.2 Most of the Glycines are found in a helical conformation	133
4.3 Evidence of the implication of the soluble domains in the membrane bending mechanism	134
4.4 The solid-state NMR data suggests structural plasticity of TM2 in membrane bilayers	135
5. Conclusion	137
Appendix.....	138

Chapter 5

Studies of M2 from influenza A in a membrane environment using proton detected solid-state NMR, DNP and DFT calculations

1. Introduction.....	160
1.1 Viruses - origin and socio-economic implications	160
1.2 Viral diseases and their treatments	161
1.3 Evolution and classification of viruses.....	162
1.4 Influenza virus.....	164
1.4.1 Influenza: Origin, genome and proteome	164
1.4.2 Mechanism of Influenza infection and treatments.....	165
1.4.3 The proton channel of influenza A: Matrix 2 (M2) protein.....	167
1.4.3.1 Role of the HxxxW motif in proton conduction	168
1.4.3.2 Structural differences in the M2 tetrad	169
2. Materials and Methods	174
2.1 Expression and purification.....	174
2.2 Refolding and reconstitution	174
2.3 Drug binding assays	175
2.4 Re-Solubilization of membrane inserted M2 in DHPC.....	175

2.5 Solution NMR, solid-state NMR and DNP	175
2.5.1 Solution NMR.....	175
2.5.2 Solid-state NMR (ssNMR)	176
2.5.3 Dynamic nuclear polarization (DNP)	179
2.6 Structure calculations of fully protonated M2.....	181
2.7 Density functional theory calculations	182
3. Identification and characterization of an imidazole – imidazole hydrogen bond in M2 from influenza A	183
3.1 Abstract.....	183
3.2 Results and discussion.....	183
3.3 Conclusion.....	187
4. Structure determination of M2 in membranes by proton detected NMR under ultra fast MAS	188
4.1 Abstract.....	188
4.2 Results and discussion.....	189
4.3 Conclusion.....	195
5. Non-specific binding of rimantadine	195
5.1 Abstract.....	195
5.2 Results and discussion.....	196
5.3 Conclusion.....	203
6. Pore bound water to histidine 37 sidechain in influenza A M2	203
6.1 Abstract.....	203
6.2 Results and discussion.....	204
6.3 Conclusion.....	211
7. Conclusion	213
Appendix.....	216
Bibliography	240

List of Figures

Chapter 1

General introduction and motivations

Figure 1.1. Number of structures deposited to the protein data bank per year separated by technique.....	4
Figure 1.2. 1D and 2D NMR spectra of a 7 kDa protein in solution.....	6
Figure 1.3. Binding study of the membrane protein M2 from Influenza A.....	7
Figure 1.4. Different spin magnetization transfer mechanisms used in NMR.....	8
Figure 1.5. Comparison of 1D and 2D solution and solid-state NMR spectra of SH3.....	11
Figure 1.6: Representation of the spin states in a sample upon applying an external magnetic field (B0) and a pulse (B1).....	13

Chapter 2

Alpha protons as NMR probes in deuterated proteins

Figure 2.1. Amino acid metabolic pathways and the different enzymatic stages of the α -PET labeling method.....	26
Figure 2.2. Effective incorporation of H α protons in a ubiquitin sample, while suppressing many side-chain signals.....	30
Figure 2.3. Residue specific characterization of labeling from 1H-15N TOCSY-HSQC spectra of 1 mM ubiquitin using 75 ms MLEV-17 mixing.....	31
Figure 2.4. Cross-polarization based carbon-proton correlation spectra, (H)CH, of microcrystalline SH3 either uniformly α -PET labeled (red) and 13C,15N-labeled (black) crystallized from a protonated buffer.....	34
Figure 2.5. Long-range distance information is highlighted in a 3D H(H)CH spectrum of α -PET SH3 (pdb: 1aey) in D2O (blue) and in H2O (red).....	35
Figure 2.6. Identification of a cross beta strand contact (F99H α –I114H α) in the beta barrel membrane protein VDAC in lipid bilayers.....	37
Figure 2.7. Selected residues showing the reduction in proton (H α) R2 relaxation rates with α -PET labeling (red) as compared with full protonation (black). The correlation plot (right) shows a reduction for all residues.....	39
Figure S2.1. LAAO activity for different amino acids in solution.....	43
Figure S2.2. Side products from LAAO treatment on Leucine.....	44

Figure S2.3. LAAO treatment efficiency on a deuterated amino acid powder for different amino acids.	45
Figure S2.4. LAAO treatment efficiency on a deuterated amino acid powder for polar/charged amino acids.	46
Figure S2.5. LAAO treatment efficiency on a deuterated amino acid powder for hydrophobic amino acids.	47
Figure S2.6. Transaminase and other enzyme activity in <i>E. coli</i> is sufficient to exchange the amide position of many residues, with and without LAAO treatment.	48
Figure S2.7. Amino acid labelling pattern of α -PET Ubiquitin without LAAO treatment.	49
Figure S2.8. Comparison of spectral quality when labeling with deuterated glucose in otherwise protonated media.	50
Figure S2.9. Transverse relaxation (T_2') is compared for α -PET samples in 100% deuterated buffer and 100% protonated buffer.	51
Figure S2.10. Predicted $H\alpha$ - $H\alpha$ contacts accordingly to α -PET labeling mapped into the VDAC NMR structure.	52
Figure S2.11. Transverse relaxation rates (R_2) using CPMG for fully protonated (black) and α -PET labeled ubiquitin (red).	53
Figure S2.12. α -PET SH3 comparison using protonated buffer (red) or deuterated buffer (blue) for crystallization.	58

Chapter 3

Probing membrane protein insertion into lipid bilayers by solid-state NMR

Figure 3.1. Schematic representation of longitudinal magnetization transfer in a lipid bilayer sample.	62
Figure 3.3. Evolution of magnetization in the HhNH spectrum of AlkL and hVDAC for selected residues.	68
Figure 3.4. Lipid and water signal detected at the protein amides for AlkL and hVDAC.	73
Figure 3.5. Selected strips from the 4D spectra of hVDAC and AlkL using 50 ms proton-proton mixing.	74
Figure 3.6. Water and lipid contacts shown on the homology model of AlkL using OmpW (PDB 2F1T) as a template (A) and the solution NMR structure (PDB 5JDP) of hVDAC (B). ...	

Figure S3.1. Proximity of exchanging residues has little or no effect on the observed buildup of the cross peak between a mobile species and an amide proton.	78
Figure S3.2. Cross-peak intensities for backbone amides, lipids, and water.	79
Figure S3.3. Measurement of the average protein-lipid cross-relaxation rate in hVDAC by tracking the initial protein signal decay (blue) and the buildup on lipid resonances (blue).	80

Chapter 4

Structural and functional characterization of Mic10 from the MICOS complex

Figure 4.1. Schematic representation of an eukaryotic cell.	82
Figure 4.2. Mitochondrial membrane architecture and composition.	85
Figure 4.3. Schematic representation of the sub-complexes of the MICOS found at the cristae junction and Mic10 amino acid sequence.	87
Figure 4.4. Comparison of Mic10His expression and growth in both BL21(DE3) and Lemo21(DE3).	105
Figure 4.5. Predicted transmembrane domains of Mic10 by TMpred.	106
Figure 4.6. Mic10 structural prediction based on homology models from two structure prediction servers: GalaxyTBM and SWISS-MODEL.	107
Figure 4.7. Heterogeneous contribution from alpha and beta sheet secondary structure in refolded Mic10.	108
Figure 4.8. Detergent screening by NMR.	110
Figure 4.9. Poor stability of Mic10His in SDS detergent micelles conditions.	112
Figure 4.10. Part of the glycine content in Mic10His adopts a helical conformation in SDS micelles.	113
Figure 4.11. (H)NH spectra of Mic10-His reconstituted in different lipids.	115
Figure 4.12. (H)NH spectra of Mic10-His using different reconstituted protocols.	116
Figure 4.13. Negatively charged lipids do not improve Mic10His spectral quality in ssNMR.	117
Figure 4.14. Predicted helical-helical interaction from the different Mic10 transmembrane domains.	119
Figure 4.15. ssNMR spectra of trypsin treated Mic10 reconstituted in DMPC.	120
Figure 4.16. Enzymatic treatments disrupt Mic10His oligomers.	121
Figure 4.17. TM1 and TM2 adopt a helical conformation in the presence of detergents.	123

Figure 4.18. HH-TOCSY assignments of TM2.	124
Figure 4.19. HH-NOESY assignments of the TM2-GA.	125
Figure 4.20. TM2 adopts an alpha helical conformation based on the chemical shift index (CSI).	126
Figure 4.21. Ten best structures of the TM2 calculated by CYANA.	128
Figure 4.22. TM2 reconstituted in lipid bilayers displays inhomogeneous broadening indicative of conformational heterogeneity.	130
Figure 4.23. TM2 adopts different conformations in lipid bilayers.	130
Figure 4.24. No chemical shift perturbations observed on a mixed TM1-TM2 sample.	132
Appendix 4.V. Western blots obtained from the Mic10His test expressions.	148
Appendix 4.VI. Mic10His ¹³ C-HSQC spectra obtained for different detergents.	149
Appendix 4.VII. Comparison of (H)NH spectra of the non treated and trypsin treated Mic10His.	150
Appendix 4.VIII. Identification of Mic10 after trypsin treatment by mass spectrometry.	151
Appendix 4.IX. Purity and mass control of the TM2 peptide synthesis.	152
Appendix 4.X. Helical secondary structure of TM2 domain is induced by the presence of detergents.	153
Appendix 4.XI. Alpha proton (left panel) and amide (right panel) chemical shift index of TM2 in DPC micelles.	154

Chapter 5

Studies of M2 from influenza A in a membrane environment using proton detected solid-state NMR, DNP and DFT calculations

Figure 5.1.1. Mechanisms leading to antiviral resistance.	163
Figure 5.1.3. Influenza A life-cycle.	167
Figure 5.1.4. Comparison of M2 structures and H37 sidechain arrangements.	171
Figure 5.3.1. Measurement of 2hJNHN hydrogen bonding in H37 imidazole dimers within influenza M2.	184
Figure 5.3.2. Quantification of the intermolecular 2hJNHN J-coupling.	185
Figure 5.3.3. Histidine-water contact and assignment of H37 tautomer state.	185
Figure 5.3.4. Chemical shift changes in the histidine side-chain upon addition of the drug rimantadine (rmt).	186

Figure 5.4.1. Backbone walking assignment spectra of a fully protonated $^{13}\text{C},^{15}\text{N}$ M2 at 100 kHz MAS.	189
Figure 5.4.2. Solid-state NMR assignments of a fully protonated M2.	191
Figure 5.4.3. De novo structure calculation of M2 from influenza A.	193
Figure 5.4.4. Comparison of the M2 structure built by homology or de novo calculation.	194
Figure 5.5.1. Chemical shift changes in M2 induced by non-specific binding with 40 mM rmt.	196
Figure 5.5.2. Non-specific CSP.	197
Figure 5.5.3. Individual chemical shift perturbation observed for rmt non-specific binding in M2.	199
Figure 5.5.4. Histidine side-chain CSP and hydrogen bonding.	200
Figure 5.5.5. Kinetics of rmt binding.	201
Figure 5.5.6. Pore binding results in channel restructuring.	202
Figure 5.6.1. Identification of water close to $\text{N}\delta 1\text{B}$ of His37.	206
Figure 5.6.2. Water - imidazole cluster revealed under DNP conditions.	209
Figure 5.6.3. Chemical shifts of the structured water in the core of M2 calculated by density functional theory.	211
Summary figure showing the solved and remaining questions of the proton channel M2 in lipid bilayers.	214
Table S5.5. Manual restraints for helping tetramer formation.	235
Appendix 5.IV. pH at the DNP conditions.	236
Appendix 5.III. 2D (H)NH DNP at 90 K using 24 kHz MAS.	237
Appendix 5.V. DFT structures after geometry optimization.	238
Appendix 5.VI. NMR calculations using solvation (water with PCM).	239

List of tables

Chapter 1

General introduction and motivations

Table 1.1. Dipolar coupling strength.....	9
Table 1.2. Isotopes, natural abundance and spin number.....	12

Chapter 2

Alpha protons as NMR probes in deuterated proteins

Table 2.1. LAAO activity in deuterated Silantes media, as determined by solution NMR.	32
Table S2.1. Amino acid composition of Silantes media as supplied by the manufacturer, with negligible concentrations of carbohydrates (less than 30 mg per liter).	40
Table S2.2. Incorporation level at the H α position by amino acid type as estimated from ¹³ C-HSQC spectra without correcting for differences in T2 relaxation.	41
Table S2.3. Incorporation level at the H α position by amino acid type as estimated from ¹³ C-HSQC spectra corrected by the H α T2.	42
Table S2.4. Transverse relaxation rates (R2), obtained from CPMG measurements at 277 K for resolved residues in uniform ¹⁵ N, ¹³ C labeled and α -PET labeled ubiquitin measured at a 600 MHz spectrometer.	54
Table S2.5. Transverse relaxation rate (R2), obtained from CPMG measurements at 308 K for resolved residues in uniform ¹⁵ N, ¹³ C labeled and α -PET labeled ubiquitin measured at a 600 MHz spectrometer.	55
Table S2.6. Spectrum acquisition parameters recorded for the different samples.	57

Chapter 3

Probing membrane protein insertion into lipid bilayers by solid-state NMR

Chapter 4

Structural and functional characterization of Mic10 from the MICOS complex

Table 4.1. Mic10, Mic12, and Mic13 constructs used.	90
Table 4.2. PCR buffer composition used for Mic δ s preparation.....	91
Table 4.3. Ligation protocol used.....	92
Table 4.4. PCR protocol used.....	92

Table 4.5. Transmembrane domain constructs.....	93
Table 4.6. Optimal expression and purification obtained conditions for the different constructs.	96
Table 4.7. Detergents and Lipids used for Mic10 and TM studies.	97
Table 4.8. Constructs and conditions used for the acquisition of the circular dichroism data.	99
Table 4.9. Mic10His sample conditions used for solution NMR.....	100
Table 4.10. Sample conditions and NMR experiments used for the TMs constructs.	101
Table 4.11. Mic10His sample conditions used for solid-state NMR.	102
Table 4.12. Detergent properties.	110
Table S4.1. Minimal media M9 composition (For 1 Litre of M9).....	144
Table S4.2. Mic10 δ s PCR primers and the expected number fo amino acids colony PCR protocol tables.	146
Table S4.3. PCR protocol used for insert amplification	146
Table S4.4. Bacterial strains properties.....	147
Table S4.5. Assignment table of TM2 ¹³ C, ¹⁵ N-GA labeled.....	155
Table S4.6. Assignment table of TM2 ¹³ C, ¹⁵ N- IGG labeled.	157

Chapter 5

Studies of M2 from influenza A in a membrane environment using proton detected solid-state NMR, DNP and DFT calculations

Table 5.2.1. Solid-state NMR experiments and parameters.....	176
Table 5.2.2. DNP spectra performed with a 3.2 mm Bruker DNP probe.....	180
Table 5.2.3 DNP spectra performed with a 2.5 mm Phoenix DNP probe.....	181
Table 5.4.1. Spectra used for the structure calculation.	192
Table S5.1. ¹ H, ¹³ C and ¹⁵ N chemical shift assignment Helix A.	216
Table S5.2. Upper and lower distance used in Cyana for Helix A.....	233
Table S5.3. Upper and lower distance used in Cyana for Helix B.....	234
Table S5.4. Entry of the imidazole-imidazole hydrogen bond for Cyana.....	235

Chapter 1

General introduction and motivations

1. PROTEINS

This part of the thesis is based on different articles (cited during the text) and books: Molecular Biology of the Cell (1); Biochemistry (2); Spin Dynamics: Basics of Nuclear Magnetic Resonance (3), Solid-state NMR Spectroscopy (4) and Understanding NMR Spectroscopy (5).

1.1 PROTEINS IN BIOLOGY

Proteins are one of the most important pillars of biology. Proteins are long polypeptide chains made by the combination of the 20 different building blocks (amino acids). Over the years, the scientific interest in understanding the structure-function relationship of proteins has increased in view of their role as the "master regulators" of the cellular metabolic pathways. Their specificity and efficiency in carrying out cellular reactions such as protein folding, degradation and phosphorylation make the understanding of the structure-function relationship a key step in preventing, regulating and treating a wide range of cellular functions. It is now possible to obtain high expression yields of recombinant proteins at a relatively low-cost due to a better understanding of cell metabolism as well as the advances made in biochemistry and biotechnology concerning protein expression (6). These technological advances have allowed the use of recombinant proteins in many fields such as medicine (7) or industry in a wide range of applications such as diagnostics, food additives and others (8). To keep advancing in these areas, it is, therefore, essential to develop methods to study the structure-function relationship of proteins in order to optimize their applications in different fields.

Determination of the three-dimensional structure of a protein as well as the protein-protein interactions are crucial to understand protein function(s), regulation(s) and to search for potent and specific drugs with important pharmacological targets (9). The protein amino acid sequence is encoded by the mRNA. After mRNA translation by the ribosome, the chain of amino acids is often taken over by the chaperones, which assist the protein folding

pathways by ensuring the correct fold that leads to biological function. Indeed, protein function is not completely described by the amino acid sequence but also by the 3D conformation in space (10). Incorrect protein folding known as “misfolding” has been related to a plethora of severe human diseases as well as distinct protein function (11). In order to improve the specificity and efficacy at the pharmacological level, it is essential to study protein folding in a relevant biological context to establish a “correct” relationship between structure and function.

Expression and purification of proteins in their physiologically relevant native forms is today the major bottleneck for structural studies. In a cellular context, proteins are often modified with post-translation modifications (PTMs) which regulates activity. In addition, about 2/3 of the pharmacological targets are membrane proteins, which often give low expression yields and require special treatments due to their hydrophobic nature. Although, enormous improvement has been made there is still more to be accomplished in this field.

The expression system must be carefully chosen according to the study requirements. Each system has its own advantages and disadvantages regarding yields, cost, and protein maturation including PTMs. Protein expression can be achieved using a variety of methods and/or organisms from the higher Eukaryotic cells including insects (SF9), yeast (*Pichia pastoris*) or even mammalian cells lines (Chinese hamster ovary (CHO) and human embryonic kidney (HEK))(12), which give low yields but result in native or near native PTMs, to lesser complex systems such as prokaryotic organisms (bacterial), which often give higher expression yields but lack PTMs. Other methods have become available such as protein synthesis or cell-free approaches (13).

The characterization of PTMs are not the subject of this thesis, thus *E. coli*, which allows both perdeuteration media and high protein yields at a reasonable cost, was used as an expression system for the whole set of proteins; The α -spectrin Src-homology 3 domain (SH3); Mic10 found in the mitochondrial membrane as part of the mitochondrial contact site and organization system (MICOS); and the Matrix 2 (M2) protein found in the influenza viral membrane; the human voltage dependent anionic channel (hVDAC) found in the outer mitochondria membrane; and ubiquitin (UBQ).

1.2 AVAILABLE METHODS FOR STRUCTURAL CHARACTERIZATION OF PROTEINS

There are a limited number of techniques that are currently available for solving the three dimensional structure of proteins at the atomic level. The vast majority of structures are solved by X-ray crystallography, where high resolution structural information is obtained via the electron density map derived from an X-ray diffraction pattern. Although today, X-ray crystallography is less dominant, it is still the technique adding the highest number of structures in the protein database (pdb) per year (Fig. 1.1, blue). Several requirements arise from X-ray crystallography, the most common are: (i) the need for crystals diffracting to high resolution (they are not trivial to obtain and can be far from the biological context); (ii) protein structure artifacts due to the formation of the crystals leading to “non-native” protein-protein contacts. Many of the deposited structures have density corresponding to the additives used for crystallization such as, for example, detergents in the case of membrane proteins. More recently, X-ray crystallography has been applied to membrane proteins in lipid cubic phase, which to a certain extent mimics the membrane environmental conditions.

In the past decade, electron microscopy based techniques, mainly single particle cryo-EM, have gained popularity. The recent hardware revolution has allowed this technique to reach a resolution below 3 Å. In 2020, Stark and co-workers achieved 1.25 Å resolution on an Apoferritin sample, the best resolution reported today by Cryo-EM (14). Despite the “native environment of the samples” and the high resolution obtained, cryo-EM applications are limited to system with a molecular size larger than 50 kDa (15, 16).

Unlike Cryo-EM and X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) has unique capabilities to determine three dimensional structure of proteins near-native conditions regarding temperature and environment. The sample condition used for NMR allow dynamic and thermodynamic characterization of the protein and the protein-ligand binding properties near physiological conditions. Currently (mid 2020), NMR has added more than 13 000 structures to the pdb (Fig. 1.1, red). The major bottleneck to carry out structural studies by NMR is obtaining the complete resonance assignment of the protein. While the majority of these structures are solution NMR, solid-state NMR (ssNMR) has begun to add structures. SsNMR has the advantage for membrane proteins in lipid bilayers.

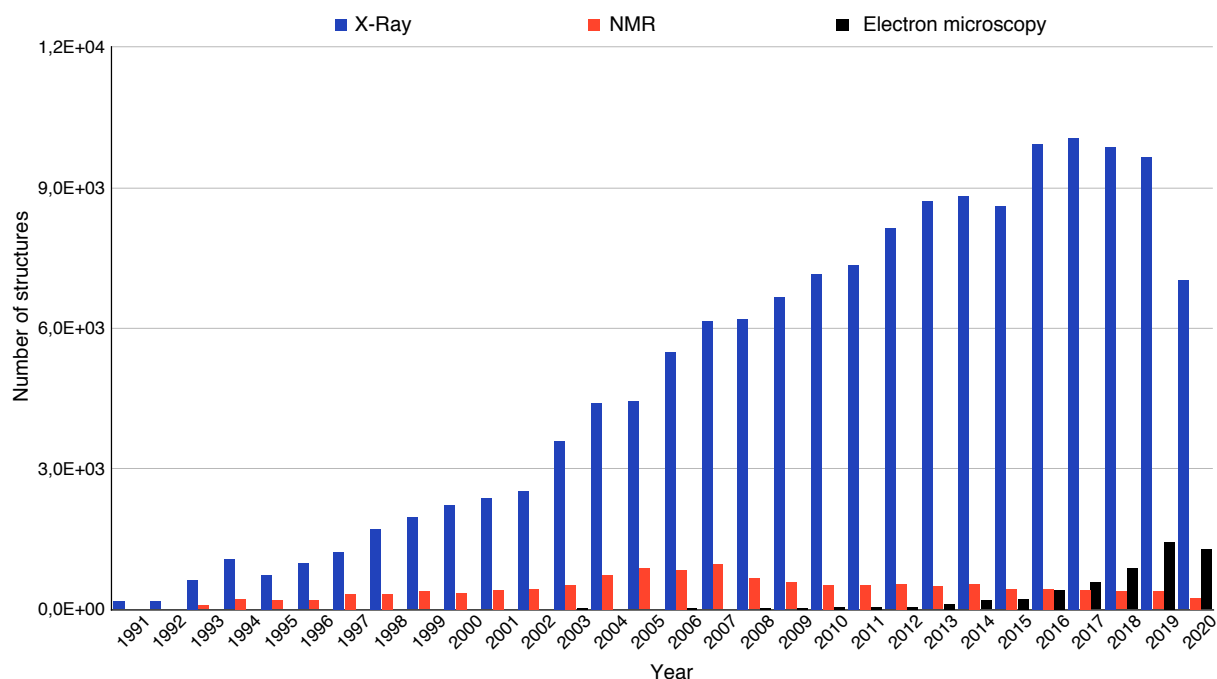


Figure 1.1. Number of structures deposited to the protein data bank per year separated by technique. The X-ray (blue), NMR (red) and electron microscopy (black) structures deposited.

1.3. STUDIES OF PROTEINS BY NMR

For the past decades, many methodological and hardware developments in NMR have been achieved leading to the structural elucidation of proteins in a wide range of environments. Two distinct approaches are nowadays widely used for structural purposes; (i) **solution NMR** and (ii) **solid-state NMR**. These will be introduced in sections 1.3.1 and 1.3.2, respectively. The low NMR sensitivity has led to the development of signal enhancement techniques such as dynamic nuclear polarization (DNP). Although DNP has increased in popularity since sensitivity enhancements of about 10 to 600 fold have been achieved, there is still a limited usage due to the need for expensive equipment, as well as poor resolution obtained for biological samples at the low temperature needed. The limited resolution obtained in DNP is mostly from sample heterogeneity. Unlike solution and ssNMR, DNP studies requires the use of polarizing agents with unpaired electrons (e.g. AMUPol) and cryogenic conditions (17). Under low temperature conditions the ensemble of the molecular orientations are retained leading to line broadening. In addition, the presence of electron spins in the sample from the polarizing agent decreases the T_2 relaxation which also contributes to the line broadening. The relaxation parameters will be discussed in more detail in section 2.1.1.

1.3.1 Solution NMR and spin-spin magnetization transfer mechanisms

In solution, small molecules or proteins experience fast tumbling with a correlation time on the order of nanoseconds. Under Brownian motion, anisotropic interactions such as the chemical shift anisotropy (\hat{H}_{CSA}) and dipolar couplings (\hat{H}_D) (discuss in section 2.1.2: “Quantum mechanics treatment of spin interactions”), which depend on the relative orientations of the molecules will be averaged out. The reduction of the spin interactions limit the spin relaxation leading to long coherence life time of the spin which results in narrow NMR lines below 1 Hz.

In proteins, although proton chemical shift dispersion gives information about the protein folding, and despite the narrow linewidths obtained, site specific assignment, based uniquely on proton spectra, is often not possible to obtain due to overlapping proton resonances (Fig. 1.2A). Site specific resonance assignment is accomplished by acquiring multidimensional NMR experiments, 2D (Fig. 1.2B) or 3D. To this end, proteins have to be doubly ($^{13}\text{C},^{15}\text{N}$) or even triply ($^2\text{H}, ^{13}\text{C},^{15}\text{N}$) labeled in order to enhance sensitivity. In addition, the chemical shift assignment of the heteronuclei gives direct information about the protein secondary structure.

Using triple labeled ($^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$) samples for large molecular systems, with slow molecular tumbling, solution NMR application is limited to low molecular size systems of ~25-40 kDa. The size limitation comes from the relaxation processes occurring through the sample as discussed in more detail in section 2.1.1. The relaxation mechanisms reduce the coherence life time of the spin limiting resolution and the magnetization transfer efficiency between the nuclei. To overcome this limitation, it is possible to combine novel pulse sequences together with advance labelling schemes. By applying these methods, studies of large molecular systems with sizes up to hundreds of kilodaltons have become accessible to NMR (18).

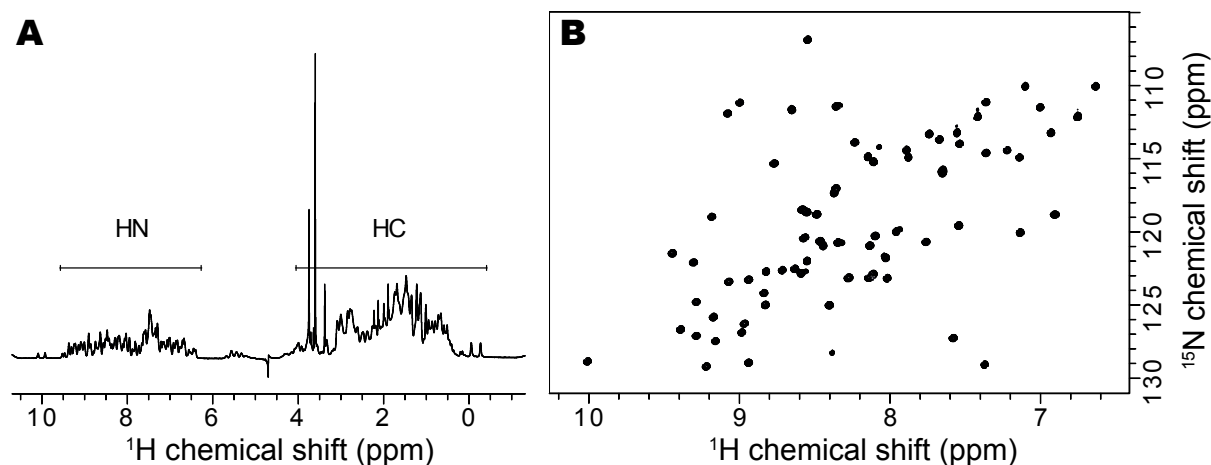


Figure 1.2. 1D and 2D NMR spectra of a 7 kDa protein in solution.

In A, the 1D ^1H spectrum is shown. In B, a two-dimensional (2D) NMR spectrum (^{15}N -HSQC) is shown. Although the 1D spectra inform on the protein secondary structure, higher dimensional spectra are needed in order to obtain reliable sequence specific resonance assignment.

Membrane proteins studies by solution NMR and X-ray crystallography often require the protein to be solubilized in detergents (Fig. 1.3A) that can result in important structural differences from the native protein fold, which sometimes are accompanied by alterations in protein function (19). Although preparation in nanodisc for NMR or lipid cubic phase in X-ray has become available, they are still difficulties to overcome regarding sample preparation and resolution (20, 21). Solid-state NMR (ssNMR) under magic angle spinning (MAS) conditions offers the possibility to carry out membrane protein studies in lipid bilayers (Fig. 1.3B). In chapter 4 and chapter 5 of this thesis, we applied ssNMR to characterize two membrane proteins in lipid environments, an inner mitochondrial membrane protein, Mic10, and a proton channel from influenza A, M2, respectively.

The importance of the environment while studying membrane proteins is shown for the case of the membrane protein M2 in figure 1.3. While small chemical shift changes are observed in the solution NMR spectra without rimantadine (rmt) (Fig. 1.3A, blue) or in the presence of rmt (Fig. 1.3A, orange), in the lipid bilayer large chemical shift differences are observed between apo (blue) and rmt pore bound (red) (Fig. 1.3B). In that particular case, the use of lipids was crucial to identify the functional binding site of the antiviral compound amantadine (22). Contrary to the ssNMR results where two binding sites, a pore and an external site, were observed in the lipid environment (22), solution NMR data in DHPC detergents suggested only an external binding site (23). Although limited by resolution and

sensitivity, so far, ssNMR has shed light on a variety of structures and biological processes such as transport and enzymatic mechanisms of membrane proteins in a more relevant context (24-26).

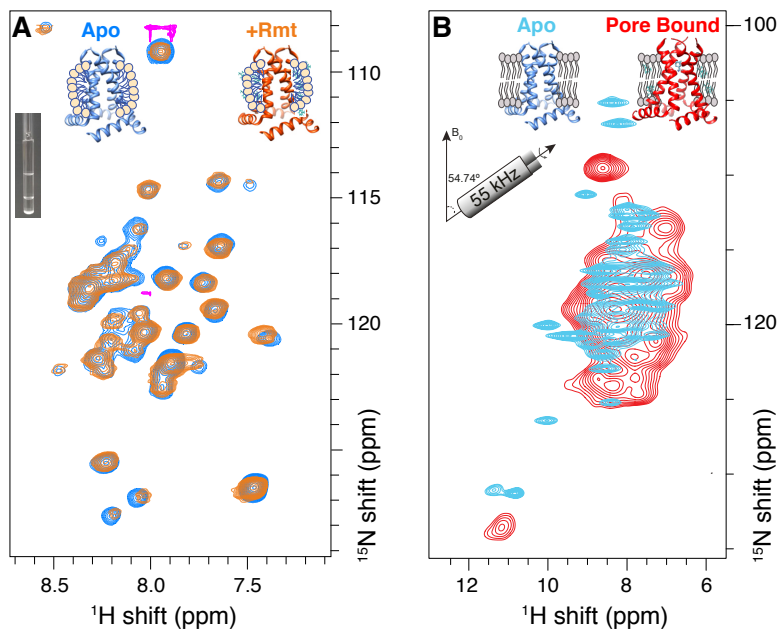


Figure 1.3. Binding study of the membrane protein M2 from Influenza A.

A and B panels show the 2D $^1\text{H}, ^{15}\text{N}$ -HSQC solution NMR spectra in 300 mM DPhPC micelle solution (A) and 2D $^1\text{H}, ^{15}\text{N}$ -(H)NH solid-state NMR spectra in DPhPC lipids (B) both in the apo state (dark blue in A and sky blue in B) and in the presence of the drug rimantadine (rmt) added (orange in A and red in B).

To perform, multidimensional NMR experiments efficient spin - spin magnetization transfer methods are required. Spin magnetization transfer can be realized either through covalent bonds via scalar couplings, e.g Inensitive nuclei enhancement by polarization transfer “INEPT” (Fig. 1.4A and B), or through space via cross-polarization (CP) (Fig. 1.4D), and cross-relaxation experiments such as nuclear Overhauser effect (NOE) (Fig. 1.4C). All these experiments are routinely used for assignment as well as for structure determination of proteins. Contrary to ssNMR, the long T_2 in solution samples allows efficient magnetization transfer via bonded spins (INEPT). These mechanisms will be encountered in the following chapters.

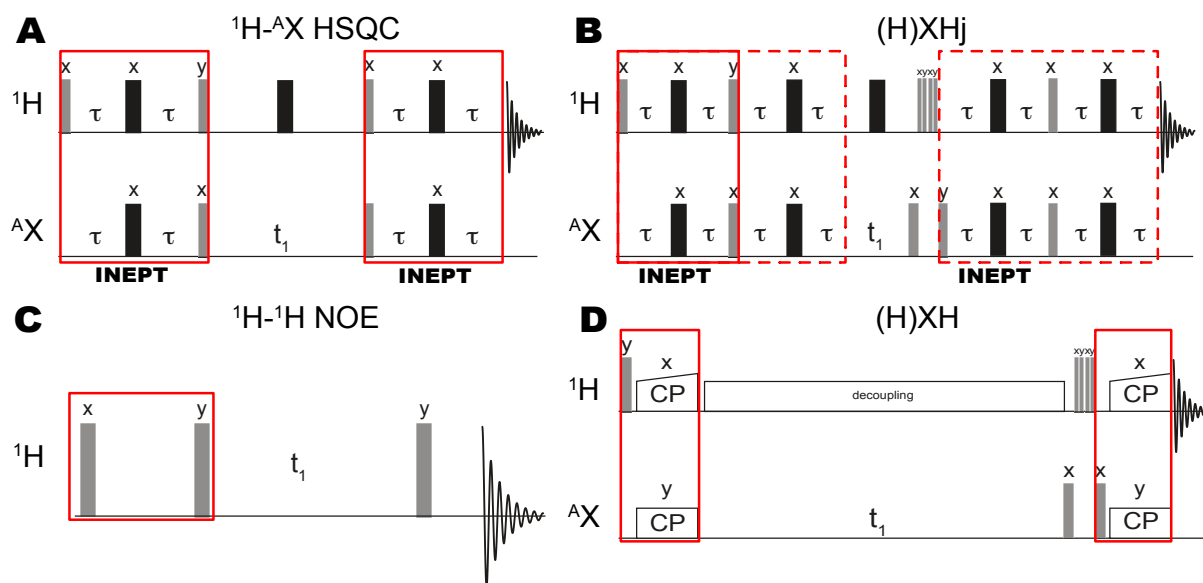


Figure 1.4. Different spin magnetization transfer mechanisms used in NMR.

A, shows an example of a solution HSQC pulse sequence using INEPT as spin magnetization transfer mechanism (solid red line). In B, similar to the INEPT used in solution, the J based spin magnetization transfer mechanism pulse sequence used in solid-state NMR $(\text{H})\text{XH}_j$ is shown. In the $(\text{H})\text{XH}_j$, the INEPT is extended (square dashed red line) to accomplish a full in-phase spin transfer from I_x to S_x . In C, a nuclear Overhauser (NOE) pulse sequence is shown. The ^1H - ^1H dipolar NOE transfer (red square) is accomplished by z-z mixing. In D, the cross-polarization (CP) mechanism used in solid-state NMR $(\text{H})\text{XH}$ sequence is shown. The black and gray rectangles indicate the π and $\pi/2$, respectively. The phase cycle is indicated by "x" or "y" on top of the pulse. And, the time period τ is set to $1/4J$, where J is the J-coupling between the two nuclei. The red square shows the CP mechanism which consists of applying a spin lock simultaneously to both nuclei to fulfill the Hartmann Hahn condition ($n\omega_r = \omega_I \pm \omega_S$) explained in section 1.3.2 "Solid-state NMR".

1.3.2 Solid-state NMR

Contrary to the case in solution, protein motion is restricted in the solid-state, hence the anisotropic interactions such as dipolar couplings won't be averaged out leading to faster T_2 relaxation. As introduced later in section 2.1.2, in Eq. 17 for dipolar couplings, in the high field approximation, the coupling strength of the anisotropic interactions is dependent on the orientation of the internuclear vector with respect to the magnetic field by the relation $(3\cos^2\theta - 1)$. Thus, there is a special angle where this term vanishes,

$$\begin{aligned}
 (3\cos^2\theta - 1) &= 0 \\
 \cos^2\theta &= \frac{1}{3} \\
 \theta &= \arccos \sqrt{\frac{1}{3}} \approx 54.74
 \end{aligned}$$

The 54.74° angle is known as the magic angle. The general idea of magic angle spinning (MAS) is to mechanically rotate the sample at the magic angle with respect to the applied field (B_0) which results in a time averaged anisotropic interaction with zero strength. However, to fully remove the effects of the anisotropic interaction, the spinning frequency has to be faster than the coupling itself. As can be seen from Eq. 16 (section 2.1.2 “Quantum mechanics treatments of spin interactions”), the dipolar coupling strength depends on the product of the gyromagnetic ratios of the coupled spins and the distance with r^{-3} . Table 1.2 summarizes the dipolar coupling strength for coupled single spins with the common distances found on a protein. In proteins, protons have both the strongest gyromagnetic ratio and the highest density. Hence, in solid samples such as membrane proteins, proton-proton dipolar coupling dominates T_2 relaxation leading to broad NMR lines.

In earlier times, high resolution ssNMR spectrum was achieved by taking advantage of the low gyromagnetic ratio nuclei (^{13}C or ^{15}N), which have a small dipolar coupling, with MAS of 10 to 20 kHz (achieved using large rotors with 3.2 mm in diameter) in combination with high power proton decoupling. Although high resolution was achieved for low gyromagnetic ratio nuclei, other drawbacks were limiting the application of ssNMR to biological systems: (i) the low sensitivity due to ^{13}C or ^{15}N detection, which requires large sample quantities; (ii) the use of long recycle delays needed due to the long ^{13}C and ^{15}N T_1 ($T_1 \sim$ seconds); (iii) the sample heating caused by the high proton power decoupling used during ^{13}C and ^{15}N acquisition.

Table 1.1. Dipolar coupling strength.

Atom 1	Atom 2	Distance (Å)	Dipolar coupling strength (kHz)
^1H	^1H	1.8	20.6
^1H	^{13}C	1.1	22.7
^1H	^{15}N	1	12.2
^{13}C	^{13}C	1.5	2.3
^{13}C	^{15}N	1.5	0.9

The calculation of the dipolar strength was performed using: https://www.nmrglue.com/jhelsing/dipole_dist_calc.html

As in solution NMR, sensitivity enhancement methods were introduced by using protons as starting nuclei and are now also used for detection (Fig. 1.5B). In contrast to

solution NMR, in ssNMR the fast T_2 relaxation does not allow INEPT transfer to be used efficiently, therefore the magnetization transfer is typically most efficient when based on the dipolar interaction and is performed using cross-polarization (CP) (Fig. 1.5B). CP utilizes the Hartmann-Hahn conditions to allow magnetization transfer through dipolar coupled spins.

$$\gamma_I B_I = \gamma_S B_S \quad (1)$$

Under magic angle spinning:

$$n\omega_r = \omega_I \pm \omega_S \quad (2)$$

Where ω_r is the MAS rotation frequency, and ω_I and ω_S are the nutation frequencies of the spin I and the spin S , respectively.

Although in solid samples solution line widths won't be obtained, sufficient proton resolution is achieved using 55 kHz MAS (Fig. 1.5) in perdeuterated, 100% back exchanged samples. Nonetheless, perdeuteration decreases the available information since only exchangeable sites are accessible by proton detected experiments. In Chapter 2, a novel labeling scheme is introduced to reintroduce backbone protons maintaining high resolution at 55 kHz MAS. This method offers the possibility to overcome exchange problems often encountered in membrane proteins. In the past decade, the increase in spinning rates above 100 kHz MAS has opened the door to address fully protonated samples. Nevertheless, often labeling methods are still required to achieve specific assignments for large proteins. (27). Proton detection in combination with 100 kHz MAS has been utilized in this thesis to investigate the proton channel M2 from influenza A in lipid bilayers. In particular, in chapter 5, we characterize the important residues, histidine 37 and tryptophan 41, involved in the proton conduction mechanism as well as the pore water bound to histidine 37.

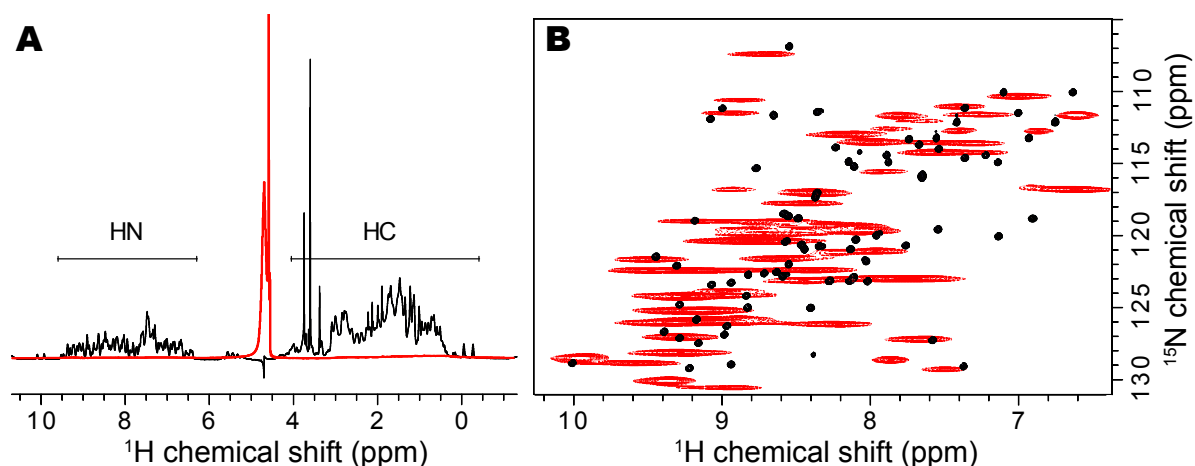


Figure 1.5. Comparison of 1D and 2D solution and solid-state NMR spectra of SH3.

In A, proton 1D spectra of a fully protonated SH3 sample in solution (black) and of a fully protonated microcrystalline SH3 sample in solids (red) are shown. In B, the 2D ^{15}N -HSQC spectrum in solution (black) and the 2D (H)NH spectrum in solids (red) are shown for the same protein. This shows the limited resolution in solid-state NMR spectra compared to highly homogeneous solution samples.

2. BRIEF THEORETICAL NMR INTRODUCTION

2.1 BASICS OF NUCLEAR MAGNETIC RESONANCE

This part of the thesis is meant to give to the reader a brief introduction on the nuclear magnetic resonance (NMR) phenomenon. NMR relies on the quantum mechanical properties of the nucleus. A nucleus is NMR active if it possess a non-zero spin angular momentum quantum number (I).

Naturally, proteins are mainly composed of ^1H , ^{12}C , ^{14}N and ^{16}O isotopes (Table 1.2). Of these nuclei, only ^1H and ^{14}N have a non zero spin, with $I_{^1\text{H}} = 1/2$ and $I_{^{14}\text{N}} = 1$. Nuclei with I greater than $1/2$ are quadrupolar nuclei. Such nuclei have a non zero electric quadrupole moment resulting in broad NMR lines. In general for NMR studies, nuclei with spin $1/2$ are preferred since they have enhanced coherence life time leading to narrow NMR lines and high sensitivity. Hence, for NMR studies proteins are isotopically enriched with ^{13}C and ^{15}N to enhance sensitivity (Table 1.2). This is a process referred to as isotopic labeling, which will be discussed in detail in section 2.2: “Labeling strategies in NMR”.

Table 1.2. Isotopes, natural abundance and spin number.

Isotope	Natural abundance (%)	I	Gyromagnetic ratio (rad.s ⁻¹ T ⁻¹) 10 ⁶
¹ H	99.9	1/2	267.5
¹² C	98.9	0	*
¹³ C	1.1	1/2	67.3
¹⁴ N	99.6	1	19.3
¹⁵ N	0.4	1/2	-27.1
³¹ P	99.9	1/2	108.4

The spin angular momentum quantum number, I defines the number of possible orientations of a spin in a magnetic field with $2I + 1$ possible orientations. For most common spins in proteins, $I = \frac{1}{2}$, in an external magnetic field (B_0), there are only two possible states: $+\frac{1}{2}$ (α state) and $-\frac{1}{2}$ (β state) (for a positive gyromagnetic ratio, γ). Here, the spins will align either parallel (α state) or antiparallel (β state) to B_0 . This phenomenon is called the Zeeman splitting (Fig. 1.6) and the energy of the states is given by:

$$E_m = -m\hbar\gamma B_0 \quad (3)$$

Where E_m defines the energy of the state, \hbar is the reduced Planck constant, γ is the gyromagnetic ratio of the spin and m the quantum number. The quantum number $m = [I, I + 1, \dots, I - 1, -1]$. The transition energy between the two states is:

$$\Delta E = \hbar\gamma B_0 \quad (4)$$

For an ensemble of spins, which is the case for any NMR sample, after a certain time (t), the spins will be distributed between the two energy states. The population distribution of the spins in the α or β states is described by the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{k_B T}} \quad (5)$$

where T denotes the temperature, k_B is the Boltzmann constant and ΔE the energy transition that is $\hbar\gamma B_0$ according to Eq. 4. The sum of all the individual magnetic moments, produces a macroscopic magnetic moment aligned along B_0 known as the bulk magnetization and noted M_0 . The small population difference is in part responsible for the low sensitivity in NMR experiments. The NMR sensitivity is measured by the signal to noise ratio (SNR):

$$SNR \propto N\gamma_{exc}\gamma_{det}^{\frac{3}{2}}B_0^{\frac{3}{2}}n^{\frac{1}{2}} \quad (6)$$

With N being the number of spins in the sample, γ_{det} is the gyromagnetic ratio of the detected nucleus, γ_{exc} is the gyromagnetic ratio of the starting nucleus (before magnetization transfer), B_0 is the magnetic field and n is the number of scans.

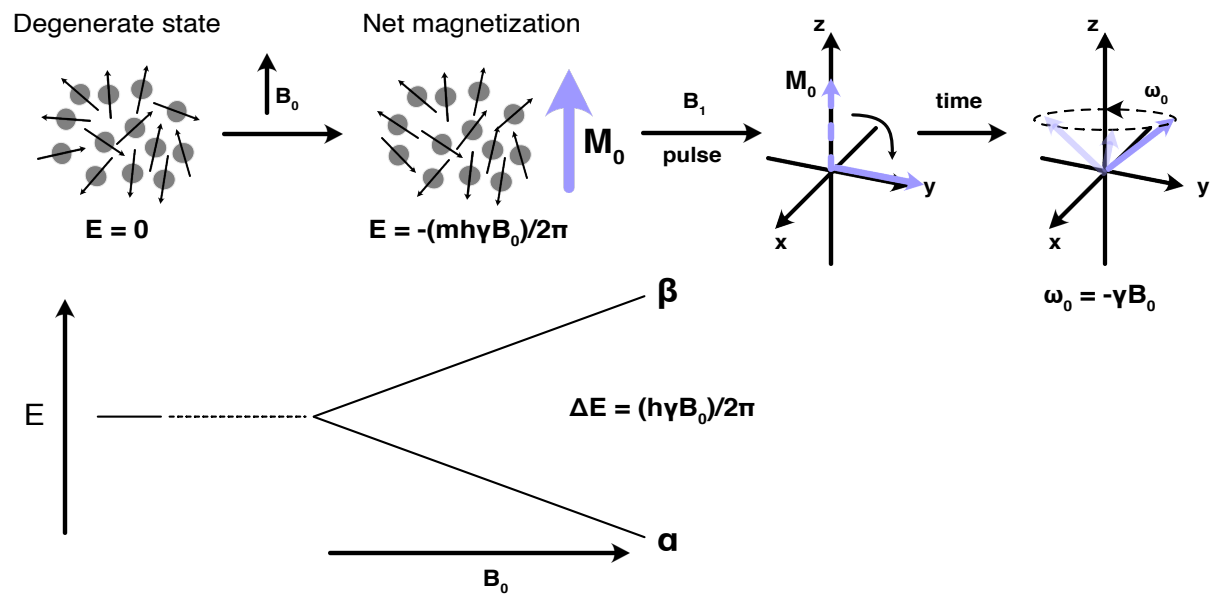


Figure 1.6: Representation of the spin states in a sample upon applying an external magnetic field (B_0) and a pulse (B_1).

The figure represents the degeneracy of the energy levels before the application of the magnetic field along the z-axis. With an applied B_0 field, the degeneracy is broken and the spins populate the α and β energy levels according to the Boltzmann distribution. Upon applying a radio frequency pulse (B_1) along the x-axis, the bulk magnetization is tilted from the z-axis to the xy-plane and start precessing along the z-axis once the pulse (B_1) is turned off.

In an NMR experiment, the bulk magnetization M_0 is tilted to the xy-plane using a radio-frequency pulse (rf) B_1 , which generates M_{xy} magnetization. B_1 is a short rf pulse with a limited duration. After the pulse the transverse magnetization will return to the initial state

(equilibrium) due to relaxation processes while precessing at the Larmor frequency (ω_0) about z.

$$\omega_0 = -\gamma B_0 \quad (7)$$

The M_{xy} precession induces an electrical signal that is detected by the NMR coil as a free induction decay (FID). The Fourier transform of the FID gives the NMR spectrum. The external magnetic field B_0 , induces an additional electro-magnetic field, which affects the local field perceived by the nucleus. Thus each nucleus in the sample experiences a slightly different local magnetic field that can be expressed using a local field, B_{local} by $B_{local} = B_0(1 - \sigma)$. Giving:

$$\omega = -\gamma B_{local} \quad (8)$$

Where ω is the chemical shift. A more detail explanation on the chemical shift will be given in a later section (2.1.2 “Quantum mechanics treatment of spin interactions”).

2.1.1 NMR relaxation processes

After manipulation of the spins by applying radio frequency pulses, the spins return to equilibrium due to relaxation processes. There are two mechanisms of relaxation observed in NMR, the spin-spin relaxation and the spin-lattice relaxation.

Firstly, the time for M_z magnetization to recover after tilting the magnetization to the xy-plane is known as **spin-lattice relaxation** or longitudinal relaxation given by $R_1 = \frac{1}{T_1}$.

Secondly, we have the time for the transverse magnetization M_{xy} to decay. This phenomenon is known as **spin-spin relaxation** or transverse relaxation and is denoted as T_2 , with the corresponding relaxation rate $R_2 = \frac{1}{T_2}$. T_2^* can be extracted from the NMR peak,

$T_2^* = \frac{1}{\pi FW}$. Where FW is the full width at half maximum of the peak. Note that T_2^* is

different from T_2 . Whereas T_2 accounts for the homogenous contributions, T_2^* contains both homogeneous and inhomogeneous contributions. There is inhomogeneity from the sample due to the spin orientation dependence with respect to the magnetic field. This phenomenon is very large in DNP experiments where by freezing the sample (100 K or lower) the wide conformational ensemble sampled by the protein is retained. The ensemble of orientations contributes to the inhomogeneous line width leading to broadening as we will see in chapter 5.6. There is also the magnetic field inhomogeneity (B_0) across the sample during the NMR experiment. B_0 inhomogeneities contribute to the line width (FW), while the homogenous contribution (T_2) is mainly coming from the properties of the sample occurring through the different spin-spin couplings mechanisms such as dipolar couplings or the presence of radicals in the sample commonly used in DNP (see section 2.1.2: “Quantum mechanics treatment of spin interactions” for more details on spin-spin interactions in NMR).

2.1.2 Quantum mechanics treatment of spin interactions

Spin in NMR can be treated either using classical or quantum mechanics. The classical mechanics model is limited to non coupled spins, but is easy to understand via a graphical picture using vectors. Although more abstract, the quantum approach is needed to explain coupled spin phenomena. In quantum mechanics a system is totally described by its wave function (ψ). By applying a special operator, the Hamiltonian operator \hat{H} , to the wave function, we can assess the energy of the system. The Hamiltonian is the operator which informs about the state of the system and is the solution of the Schrödinger equation regarding energy.

$$\hat{H} |\psi\rangle = E |\psi\rangle \quad (9)$$

Equation 9 is the time independent Schrödinger equation. However, in NMR during the manipulation of the spin by radio frequency pulses, the spin state is time dependent which implies the use of the time dependent Schrödinger equation, given by:

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}(t)\psi(t) \quad (10)$$

In NMR the total Hamiltonian is defined by:

$$\hat{H}_{total} = \hat{H}_{ext} + \hat{H}_{int} = \hat{H}_z + \hat{H}_{rf} + \hat{H}_{CS} + \hat{H}_Q + \hat{H}_D + \hat{H}_J \quad (11)$$

External Hamiltonian:

Zeeman interaction expressed as frequency:

$$\frac{\hat{H}_z}{\hbar} = \frac{-\hat{\mu} B_0}{\hbar} = -\gamma B_0 \hat{I}_z = \omega_0 \hat{I}_z \quad (12)$$

\hat{H}_z is the Zeeman Hamiltonian for a spin in a magnetic field B_0 . It describes the precession of the spin along B_0 at the Larmor frequency ω_0 . \hat{I}_z is the spin operator and γ is the gyromagnetic ratio of the involved spin.

Radio frequency:

\hat{H}_{rf} is the Hamiltonian corresponding to the B_1 field applied for flipping the spins during NMR experiments. The frequency of the applied rf has to match the energy difference between the states (Eq. 4) in order to allow the transition between the different states:

$$\Delta E = \hbar \gamma B_0 \quad (4)$$

The B_1 field determines the nutation frequency about the field direction B_1 :

$$\omega_1 = -\gamma B_1 \quad (13)$$

Internal Hamiltonian:

Chemical shift:

\hat{H}_{CS} is the chemical shift Hamiltonian defined by:

$$\frac{\hat{H}_{CS}}{\hbar} = \frac{-\hat{\mu} B_{local}}{\hbar} = -\gamma B_{local} \hat{I}_z = -\gamma(1 - \sigma) B_0 \hat{I}_z \quad (14)$$

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix}$$

\hat{H}_{CS} informs on the local environment of a nucleus (Eq. 8). With σ being the shielding tensor which is a rank two tensor informing about the orientation dependence of the spin with respect to the field B_0 . The \hat{H}_{CS} is typically separated into two parts, the \hat{H}_{iso} and \hat{H}_{CSA} , where the \hat{H}_{iso} is the isotropic chemical shift and the \hat{H}_{CSA} is the chemical shift anisotropy.

Quadrupolar interaction:

The quadrupolar Hamiltonian \hat{H}_Q is only present for spins with $I > \frac{1}{2}$.

$$\hat{H}_Q = \frac{eQ}{2I(2I-1)\hbar} \hat{I} \cdot V \cdot \hat{I} \quad (15)$$

Here Q is the electric quadrupole, I the spin quantum number, \hat{I} is the spin vector and V is the electric field gradient. In this work only spins with $I = 1/2$ will be considered.

Dipolar interaction:

Unlike these previous terms that are dependent on the intrinsic properties of a single spin, there are other terms involving spin - spin interactions. The dipolar Hamiltonian \hat{H}_D is the interaction between two dipoles in the sample and depends on the inverse of the distance between the interacting spins (dipoles) (r^{-3}) and the nuclei involved (γ_1, γ_2), by:

$$\hat{H}_{12}^D = \frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi} \left(\frac{\hat{I}_1 \cdot \hat{I}_2}{r^3} - \frac{3(\hat{I}_1 \cdot \vec{r})(\hat{I}_2 \cdot \vec{r})}{r^5} \right) = \frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi r^3} \left(\hat{I}_1 \cdot \hat{I}_2 - \frac{3(\hat{I}_1 \cdot \vec{r})(\hat{I}_2 \cdot \vec{r})}{r^2} \right) \quad (16)$$

In the presence of a high external magnetic field, only the spin operators along z (direction of the magnetic field) are retained. This is known as the secular approximation. Under this condition the dipolar coupling Hamiltonian reduces to:

$$\hat{H}_{12}^D \approx \frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi r^3} \left(\frac{(3\cos^2\theta - 1)}{2} \right) (3\hat{I}_{1z}\hat{I}_{2z} - \hat{I}_1 \cdot \hat{I}_2) \quad (17)$$

As seen in equation 17 the \hat{H}_{12}^D has an angular dependence of $(3\cos^2\theta - 1)$ where θ is the angle between the two interacting dipoles and the external magnetic field (B_0). Under isotropic conditions (fast molecular motion) as in liquids the term $(3\cos^2\theta - 1)$ averages to zero. Thus, under isotropic conditions $\hat{H}_{12}^D \approx 0$. On the contrary, in anisotropic systems such as solid samples, \hat{H}_{12}^D does not average to zero. This relation is essential for the concept of magic angle sample spinning used in ssNMR applications. As seen in section 1.3.2, in ssNMR, the sample is placed at the magic angle (MA) which corresponds to an angle of 54.74° with respect to the B_0 field and rotated mechanically. For all anisotropic interactions smaller than the rotation frequency, the anisotropic interactions will be reduced. Despite that proton detection is the standard goal for sensitivity due to its gyromagnetic ratio (γ) and structural information, proton-proton dipolar couplings are the major source of relaxation encountered in solids. Faster relaxation leads to broader NMR lines and, therefore, poor resolution. To overcome this limitation, in chapter 2, a novel labeling method is introduced to increase resolution maintaining almost full proton backbone information.

J-coupling:

The last term of the total Hamiltonian, \hat{H}_J , is the spin-spin coupling named J-coupling. The J-coupling is mediated by the electron coupling through the chemical bond between two nuclei. \hat{H}_J is defined by:

$$\hat{H}_{IS}^J = 2\pi \hat{I}_I \cdot J_{IS} \cdot \hat{I}_S \quad (18)$$

Where \hat{I}_I and \hat{I}_S are the spins I and S involved in the coupling and J_{IS} is the J coupling tensor. The J-coupling gives direct information on important features of the molecular systems such as the torsion angles (3J couplings) and hydrogen bonds (1J or 2J).

2.2. LABELING STRATEGIES IN NMR

Isotopic labeling methods are routinely used to enhance NMR sensitivity. These approaches are also used to reduce anisotropic interactions by diluting the spin content (e.g. ^1H to ^2H). Even in solution NMR, for large proteins with slow correlation time, typically encountered for membrane proteins, spin-spin interactions such as dipolar couplings and chemical shift anisotropy are reintroduced to a certain extent leading to faster T_2 relaxation, thus broadening the NMR lines. The removal of some problematic interactions can be achieved by applying tailored labeling methods (28).

The most utilized method to drastically reduce ^1H - ^1H dipolar interactions is perdeuteration (29). Introduced in the early 90s by LeMaster et al., the method uses minimal media containing D_2O and deuterated ^{13}C -glucose to carry out the protein expression. In these conditions, the protein is perdeuterated meaning that the protein is fully deuterated at the sidechains and backbone sites. In perdeuterated samples, only the amide protons will be exchanged in later steps using protonated buffers. Even using perdeuteration, at relatively low MAS < 30 kHz, the line broadening due to the proton density was such that it was common to vary the final buffer proton content to further reduce the proton dipolar network (30, 31). Nowadays, with the use of fast and ultra fast MAS (MAS > 55 kHz) 100% back-exchanged (100% protonated buffer) samples are routinely used (32).

Earlier work used different carbon sources such as ^{13}C -Pyruvate, ^{13}C -glucose and ^{13}C -glycerol containing different protonation patterns, diluting ^{13}C and ^1H content. This helps to obtain sparse carbon labeling removing strong one bond carbon-carbon and proton-carbon couplings as well as to reduce the protonation level at the protein backbone (28).

A remarkable advance in the labeling methods was achieved by using stereospecific labeling introduced by Kay and co-workers (33, 34). The authors used specifically labeled amino acid precursors in the amino acid metabolic pathway to selectively introduce labels at the desired positions maintaining high deuteration at other protein sites. Similarly, others introduced amino acid selective labeling (35) or amino acid reverse labeling methods, which consist in supplementing the media before induction with the amino acid containing the desired isotopically labeled atoms (28, 36).

The group of Bernd Reif introduced the Reduced Adjoining Protonation (RAP) labeling scheme. RAP reduces protein proton content in biological samples by varying the ratio of D₂O/H₂O used in the expression media (37). Since the labeling is random, several isotopomers (CD₂, CHD) co-exist in the sample, which reduces sensitivity and resolution. Yet it is the only inexpensive method for H α labeling while maintaining high deuteration (²H) levels. In this thesis, in Chapter 2 alpha-proton exchange via transamination (α -PET), a novel labeling scheme is introduced which consists of re-introducing backbone H α protons while maintaining high level of deuteration on the other proteins sites. This increases structural information while keeping high resolution and sensitivity at 55 kHz MAS.

3. AIMS AND THESIS STRUCTURE

Although a great deal of effort has been made to boost proton detected solid-state NMR to investigate fully protonated (^1H) biological systems on the atomic level, there is still a need for improvement regarding sensitivity and resolution in a cost-efficient manner. With the recent methodological advances including the availability of high magnetic fields (1.2 GHz, the first delivered mid-2020), which increases sensitivity, and faster spinning rates (> 110 kHz MAS) we are now one step further to address challenging systems. One of the persisting drawbacks is that despite increased proton resolution obtained in fully protonated systems, it often comes with reduced sensitivity compared to 1.3 mm rotors (~ 60 kHz MAS). To address similar limitations, the thesis focuses on both methodological development and applications of proton detected solid-state NMR under fast (~ 55 kHz) and ultra-fast MAS (~ 100 kHz) to investigate membrane proteins.

Chapter 2: In this chapter, we introduced a novel labeling strategy for protein NMR referred to as α -PET. α -PET involves using the inherent transamination activity of the expression system (*E. coli*) in order to selectively re-introduce protons at the $\text{H}\alpha$ position in the protein backbone. This chapter was published in the Journal of Biomolecular NMR (JBNMR) and has been adapted from the original publication with permission of the journal.

Chapter 3: Here, site-specific information regarding the environment of two membrane proteins in lipid bilayers was obtained using proton detected ssNMR. We have shown that the use of highly perdeuterated proteins in combination with 55 kHz MAS is sufficient to suppress spin-spin relaxation and thus achieve site-specific protein-lipid contacts by making use of NOE mixing. This chapter was published in the ChemPhysChem journal in 2019. This work was a collaboration between Eszter Éva Naubauer, who measured and interpreted NMR data of hVDAC, while I handled AlkL protein.

Chapter 4: This Chapter is part of a collaboration with Prof. Dr. Meinecke from the UMG Göttingen and Dr. Tavasenko from the group of Prof. Dr. Meinecke. The project consisted of carrying out a structural analysis of Mic10, a membrane protein of the MICOS complex. The MICOS complex is found in the inner mitochondrial membrane (IMM) and it is thought to be

the main protein responsible for the high degree of membrane curvature observed in the IMM leading to the formation of the cristae junctions. We, for the first time, assessed the secondary structure of this system in micelles by biophysical tools, namely CD, solution NMR and solid-state NMR. In addition, by using synthetic peptides labelled on specific positions, we found that the second transmembrane domain (TM2) of Mic10 shows a beta to alpha transition which seems to be regulated by temperature.

Chapter 5: In the last chapter, the work on M2 from influenza A is presented. M2 is a proton channel that arranges as a tetramer in membranes. We showed that M2 forms a dimer of dimers with a $^2J_{NN}$ inter imidazole hydrogen bond at residue H37. In addition, a *de novo* structure calculation of the protein was performed with a single 1H , ^{13}C , ^{15}N sample using ultra-fast MAS at 100 kHz with only 0.5 mg of sample. Furthermore, studies on the rimantidine binding were performed. We found that there is a thermodynamical limitation for pore binding in DHPC and that there is a high energy barrier of ~ 130 kcal/mol for pore binding to occur in DPhPC membranes consistent with the large structural rearrangement upon pore binding. Finally, a 1H chemical shift at ~ 11 ppm was identified and assigned to a pore water molecule hydrogen bonded to N δ 1 of the residue H37 by using a combination of ssNMR, DNP and DFT calculation.

Chapter 2

Alpha protons as NMR probes in deuterated proteins

This chapter is a publication in the journal of Biomolecular NMR with the doi: 10.1007/s10858-019-00230-y.

Authors: Kumar Tekwani Movellan, Eszter E. Najbauer, Supriya Pratihari, Michele Salvi, Karin Giller, Stefan Becker, Loren B. Andreas*

Contributions:

Eszter E. Najbauer contributed with the assignments of the human voltage dependent anionic channel (hVDAC). Supriya Pratihari performed the relaxation studies presented here with both perdeuterated and α -PET ubiquitin samples. Karin Giller and Dr. Stefan Becker carried out the expression and purification of Ubiquitin and hVDAC samples. Ph.D. Michele Salvi helped recording and assigning the NMR spectra. I contributed to this paper by preparing and designing the protocol for the deuterated α -ketoacids preparation. I measured the L-amino acid oxidase (LAAO) activity and quantified the incorporation of H α into the protein by NMR. I developed the protocol used for protein expression of hVDAC, Ubiquitin and SH3 using the deuterated α -ketoacids. I carried out the expression, purification and crystallization of SH3 samples. I also performed and analyzed solid-state NMR data.

1. Abstract

We describe a new labeling method that allows for full protonation at the backbone H α position, maintaining protein sidechains with a high level of deuteration. We refer to the method as alpha Proton Exchange by Transamination (α -PET) since it relies on transaminase activity demonstrated here using *E. coli* expression. We show that α -PET labeling is particularly useful in improving structural characterization of solid proteins by introduction of an additional proton reporter, while eliminating many strong dipolar couplings. The approach benefits from the high sensitivity associated with 1.3 mm samples, more abundant information including H α resonances, and the narrow proton linewidths encountered for highly deuterated proteins. The labeling strategy solves amide proton exchange problems

commonly encountered for membrane proteins when using perdeuteration and back exchange protocols, allowing access to alpha and all amide protons including those in exchange-protected regions. The incorporation of H α protons provides new insights, as the close H α - H α and H α - H^N contacts present in β -sheets become accessible, improving the chance to determine the protein structure as compared with H^N-H^N contacts alone. Protonation of the H α position higher than 90% is achieved for Ile, Leu, Phe, Tyr, Met, Val, Ala, Gln, Asn, Thr, Ser, Glu, Asp even though LAAO is only active at this degree for Ile, Leu, Phe, Tyr, Trp, Met. Additionally, the glycine methylene carbon is labeled preferentially with a single deuteron, allowing stereospecific assignment of glycine alpha protons. In solution, we show that the high deuteration level dramatically reduces R₂ relaxation rates, which is beneficial for the study of large proteins and protein dynamics. We demonstrate the method using 2 model systems, as well as a 32kDa membrane protein, hVDAC1, showing the applicability of the method to study membrane proteins.

2. Introduction

The study of proteins by Nuclear Magnetic Resonance (NMR) has been continuously evolving to improve sensitivity in order to resolve signals in multidimensional spectra, which serve as the basis for studies of structure and dynamics. For large proteins that tumble slowly in solution, as well as for proteins in the solid-state, a high level of deuteration with introduction of selective protons is used to improve proton relaxation, and therefore narrow lines, by elimination of strong proton-proton dipolar couplings.

Proton detected magic-angle spinning (MAS) NMR studies have employed different combinations of spinning frequency and deuteration to optimize sensitivity and resolution (30, 31, 38-42). Currently, many applications of proton detected MAS NMR are applied at about 60 kHz with 1.3 mm rotors, a spinning frequency that for fully protonated samples is not enough to average the strong network of ¹H-¹H dipolar couplings. This results in proton line broadening and about 200 to 300 Hz proton line widths (42). The advantage of this spinning frequency is that narrow lines are observed at high sensitivity when selected sites are labeled to 100 % incorporation of protons, while others are deuterated.

In the most straightforward approach, *E. coli* expression in D₂O is followed by exchange with H₂O/D₂O to produce a protein with a specific protonation level at amides, and perdeuteration at non-exchangeable sites (29, 31, 38). A protonation level of 100% at amide positions results in high resolution when using 40 - 60 kHz MAS in microcrystalline samples (39), enabling structure determination based on backbone resonances (30).

Accessing aliphatic protons is still an area of active development. Sidechain protons can be selectively introduced using metabolic precursors, which has the advantage that a single isotopomer is typically present and deuterium isotope shifts do not result in broadening, even at 100% protonation of the selected sites. In various ways, methyl groups of I,L,V,T,A can be incorporated (43-45). Exquisite control of labeling can be afforded using a synthetic approach known as SAIL labeling, with the only downside being the high cost, which has typically restricted applications to labeling only selected amino acid types in cell free expression systems (46). For fully protonated samples, high magnetic fields and very fast magic-angle spinning of ~100 kHz are required to sufficiently narrow proton resonances. At lower spinning frequencies, deuteration is still required, and approaches to resolve aliphatic protons involve using mixtures of H₂O and D₂O during protein expression, along with combinations of protonated or deuterated carbon sources such as glycerol or glucose (29, 47). While utilizing the residual protons in perdeuterated samples results in exquisite spectra (48), the low labeling level severely limits the ability to measure proton-proton distances. At higher proton concentrations, the carbon resonances are broadened due to deuterium isotope shifts (49). Although this is a small effect for H α , since only one proton is directly attached, it would still be desirable to limit the protons introduced in the sidechains, since such protons broaden the alpha resonance, and are a magnetization sink during proton-proton transfer.

We therefore sought a strategy that would allow labeling of alpha protons in *E. coli* at a cost that allows widespread adoption of the approach for structure determination and dynamics investigations. Previously, an approach for H α labeling was introduced, where the proton was chemically exchanged in a deuterated amino acid mixture, producing a D/L mixture of amino acids, the L portion of which can be utilized directly by bacteria (50). Although this previous method was successful, the alpha proton incorporation level was a problem for several amino acids, and serine and threonine were lost during the acetylation and deacetylation reaction. D-amino acids may also inhibit bacterial growth at high concentrations

(51, 52). It was also noted that during growth on deuterated amino acid media (53, 54), exchange of amide moieties occurs, but results in only 10-50 percent incorporation of alpha protons for hydrophobic residues (55).

We show an alternative approach that results in up to 100 percent H α incorporation by supplying keto acids. The keto acids are converted by *E. coli* transaminases to the respective amino acids, while adding a proton at the alpha position from the water pool. This avoids any problems due to racemic amino acid mixtures, since the correct L-amino acids are generated enzymatically. The major pathways of amino acid synthesis from glucose and glycerol carbon sources are depicted in figure 2.1A for *E. coli*. Keto acids are often the direct precursor to an amino acid, indicating that provided a source of keto acids, protons can be introduced via transaminase activity (Fig. 2.1B), a method hereafter referred to as ‘alpha proton exchange by transamination’ (α -PET). This method, as with any where the growth medium is based on H $_2$ O, results in protonation of the amide position during protein expression, such that both the alpha and amide positions of the protein are protonated.

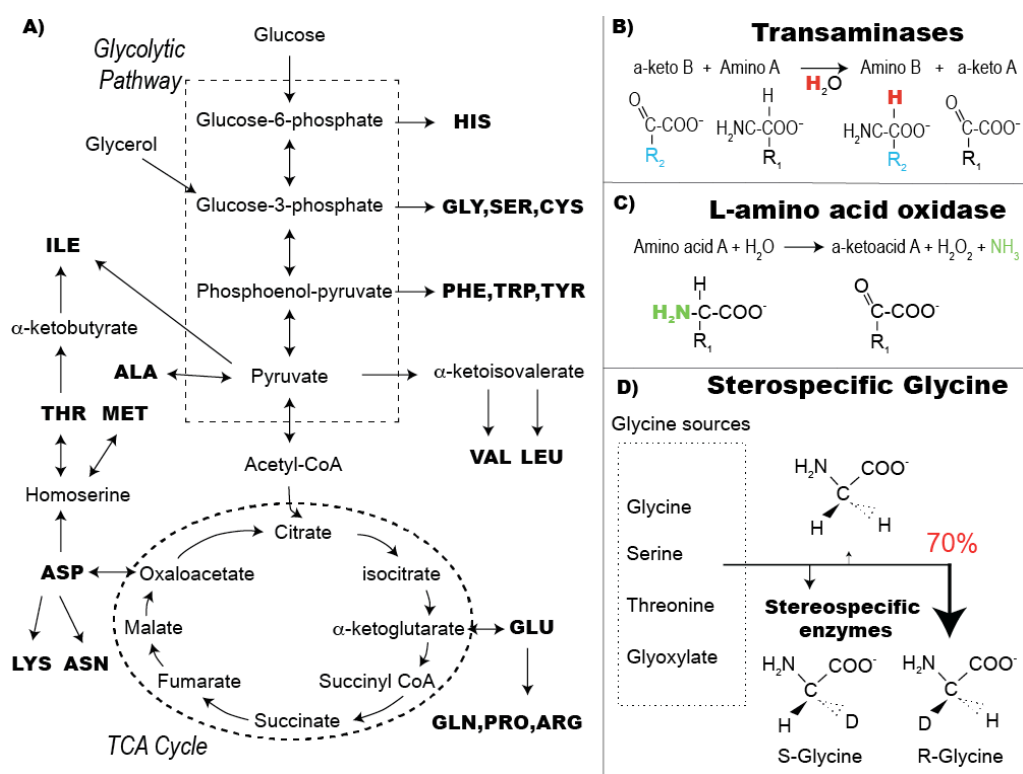


Figure 2.1. Amino acid metabolic pathways and the different enzymatic stages of the α -PET labeling method.

The metabolic pathways of the TCA cycle are shown in A. In B, the transamination reaction is shown, which is the main route for H α incorporation. In C, the generation of α -keto acids from amino acids by the enzyme LAAO is shown. D shows the main biosynthesis pathways of glycine with the observed stereospecific labeling.

Some transaminases are amino acid-specific, like the glutamate-pyruvate aminotransferase that transfers the NH_3 from glutamate to pyruvate forming alanine and α -ketoglutarate (56). Others are less specific for their substrate, such as branched-chain-amino-acid transaminase (BCAT) involved in leucine, isoleucine and valine anabolism (57). *E. coli* has a high diversity of such transaminase enzymes, resulting in effective labeling for the majority of residue types.

We generated keto acids by L-amino acid oxidase (LAAO) treatment of a commercial growth medium that is primarily comprised of ^2H , ^{13}C , ^{15}N -amino acids. LAAO enzymes are found in many organisms (58), with different specificity for the substrate amino acids (59, 60). We chose as the enzyme source a crude snake venom containing LAAO as the enzyme source, which can be applied directly to the commercial growth medium (Fig. 2.1C). Additional metabolic pathways might also be important, for example, we observed stereospecific labeling of glycine, which can occur through transaminase, but might also be influenced by conversion of serine and threonine (Fig. 2.1D).

Here we show successful introduction of alpha protons for 13 amino acids, with a high deuteration level that improves transverse relaxation rates in both solid and liquid samples.

3. Material and methods

3.1 L-amino acid oxidase stock

10 mg of L-amino acid oxidase (LAAO) powder (crude extract from the snake venom of *Crotalus adamanteus*, Sigma Aldrich) were dissolved in 1 ml of 100 mM sodium phosphate, 100 mM KCl at pH = 7.4. The solution can be kept at 4 °C for several weeks.

3.2 Preparation of keto acid mix

The amino acid mix (SILEX rich growth media as powder) from Silantes was used as starting material. To obtain keto acids, one gram of powder was dissolved in 150 ml of H_2O . To this mixture, 10 μl of bovine liver catalase solution was added at 0 and 12 hours (Sigma, 5-fold water dilution from crystalline suspension, 10,000-40,000 units/mg). In total, 3 – 4 mg of L-Amino Acid Oxidase (LAAO, Sigma) was used per gram powder media, added in equal

amounts at 0, 3, 6, 9 and 12 hours. The solution was kept shaking at 37 °C for 1 day, then lyophilized.

3.3 Protein expression

All proteins were expressed in *E. coli* BL21(DE3). Two NMR model proteins were used, ubiquitin in solution, and microcrystalline chicken alpha-spectrin SH3 (SH3). In addition, the 32 kDa voltage dependent anion channel (VDAC), a beta barrel membrane protein was prepared in lipid bilayers.

For α -PET ubiquitin, a change of medium was used prior to expression. *E. coli* cells were grown in 1 liter of M9 using 1 g/L of ^{15}N ammonium chloride and 4 g/L of ^{13}C glucose until the $\text{OD}_{600\text{nm}}$ reached 0.6-0.8. Then cells were spun down at 7000 g at 4°C for 20 minutes. The cells were re-suspended in 1L of M9 salts with 4 g of Silantes media either as received, or treated with L-amino acid oxidase (LAAO). The cells were adapted to the new media for 30 minutes before induction at $\text{OD}_{600\text{nm}} = 0.8$ with 1 mM of isopropyl β -D-1-thiogalactopyranoside (IPTG). Ubiquitin samples, including a $^{13}\text{C},^{15}\text{N}$ -ubiquitin reference sample were purified as previously described (61).

Using this media exchange protocol, 4 different samples of ubiquitin were produced, two using ^2H Silantes powder treated with LAAO or as received and two others using $^2\text{H},^{13}\text{C},^{15}\text{N}$ Silantes powder again LAAO treated or as received.

Two samples of α -PET SH3 were produced, a media-exchanged sample (as for ubiquitin), and a second α -PET SH3 grown in the presence of glucose. Specifically, the growth was started with a low concentration of 1.25 g/L ^{12}C -glucose in 800 ml of M9 media. Cells were grown until $\text{OD}_{600\text{nm}}$ reached 0.6-0.8. Then 4 g/L of treated Silantes media solubilized in 200 ml of H_2O were added. The culture was switched to 30 °C for about 30 min until $\text{OD}_{600\text{nm}} = 0.7-0.8$, and protein expression was induced using 1 mM IPTG. A reference sample ($^{13}\text{C},^{15}\text{N}$ -SH3) was expressed and all samples purified as previously described (62). In brief, the protein was purified by anion exchange chromatography (Q-TRAP, GE Healthcare) followed by gel filtration on a Superdex-75 column (GE Healthcare). The purified protein sample was extensively dialyzed against H_2O -HCl pH=3.5 for two days (exchanging the dialysis solution every 12h). The protein was then concentrated (Amicon, 3.5 kDa cut-off) to

20 mg/ml before lyophilization. The samples were resuspended in H₂O-HCl pH=3.5 or D₂O-HCl pH=3.5 at 15-20 mg/ml. Microcrystals were obtained using a pH shift protocol as previously described (63).

α -PET VDAC was expressed at 37 °C in dilute glucose media (as for SH3) and purified and reconstituted in 2D crystalline arrays as previously described (64, 65). The E73V, C127A, C232S variant of human VDAC was used.

3.4 NMR measurements

Solution NMR data were recorded in a 400 MHz Bruker spectrometer at 298 K. We recorded a set of spectra to characterize the labeling pattern: ¹⁵N-HSQC, ¹³C-HSQC in D₂O, ¹H-¹⁵N TOCSY-HSQC, and 1D proton spectra. Quantification of H α was done from a ¹³C-HSQC spectrum at 950 MHz at 310 K. Transverse relaxation rates (R₂) were measured at 277 K using a 600 MHz Bruker spectrometer equipped with a 5 mm cryoprobe.

The black spectrum of figure 2.6A was recorded at 105 kHz MAS on a 950 MHz Bruker spectrometer using a 0.7 mm HCND probe. All other solid-state NMR data were recorded on an 800 MHz Bruker spectrometer using a 1.3 mm narrow bore HCN probe and spinning at 55 kHz MAS. We recorded cross-polarization based (H)NH, (H)CH, and (H)CANH, (H)NCAHA for resonance assignment of VDAC and SH3. We measured contacts in H(H)CH, H(H)NH spectra (SH3) and (H)C(HH)CH (VDAC) using RFDR for the proton-proton mixing. The spectra were apodized with a squared cosine function (details in Table S2.3). The data analysis was performed using CcpNMR and Sparky.

4. Results and Discussion

4.1 CHARACTERIZATION OF THE LABELING PATTERN

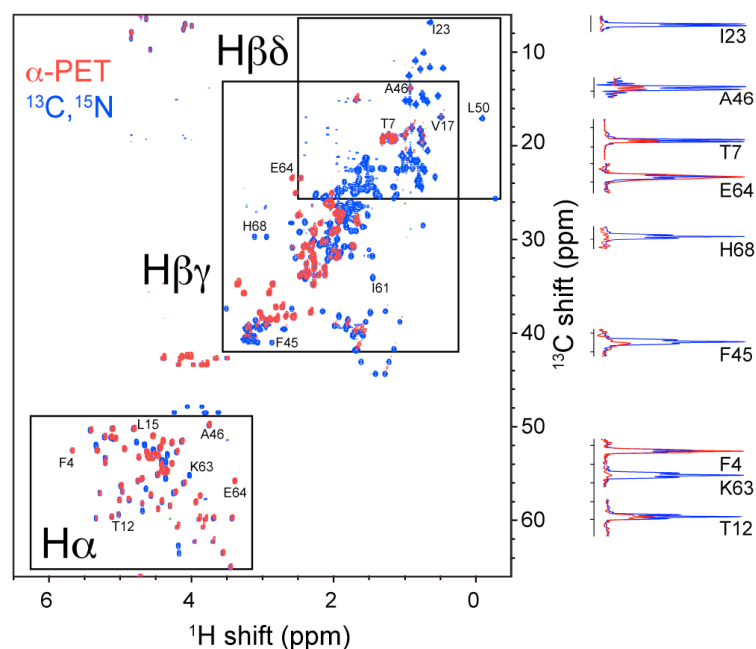


Figure 2.2. Effective incorporation of $H\alpha$ protons in a ubiquitin sample, while suppressing many side-chain signals.

The solution ^{13}C -HSQC of uniformly labelled ubiquitin (blue) is compared with α -PET ubiquitin (red). Selected slices show the intensity at backbone and sidechain sites. Intensities are not corrected for differences in T_2 .

To measure labeling patterns on an amino acid specific basis, we recorded a ^{13}C HSQC spectrum and integrated isolated peaks in the alpha region (Fig 2.2). The level of $H\alpha$ incorporation was determined assuming ideal incorporation of hydrophobic residues, based on complete reaction with LAAO. The uncorrected and T_2 corrected determinations are shown in Tables S2.2 and S2.3, respectively. A ^{15}N -TOCSY (Fig. 2.3) was recorded using a medium-range mixing time (75 ms) to assess suppression of sidechain protons. This spectrum cannot be used in a quantitative manner due to the potential for several isotopomers, differential relaxation, and relayed transfer. However, since the beta protons are relatively isolated from these effects, we could show effective suppression for most amino acid types. Figure 2.2 shows selected strips for each of the amino acid types of ubiquitin; the ^1H - ^{15}N TOCSY-HSQC of α -PET Ubiquitin (red) is compared to the ^{15}N , ^{13}C -labeled reference sample (black). The TOCSY was implemented with MLEV-17 mixing (66). The $H\alpha$ proton was detectable for 13 of the 16 (non-proline) amino acid types present in the ubiquitin sequence. Only lysine, arginine and histidine remained deuterated at $H\alpha$. This can be explained for lysine because *Cortalis adamanteus* LAAO is not able to use it as substrate (Fig. S2.1), and the deuterated amino acids are taken up in *E. coli*, while endogenous synthesis is suppressed (67). Although

LAAO showed some activity for arginine and histidine, these two amino acids are clearly relatively poor substrates of LAAO as reported in previous studies (68) and also herein (Fig. S2.1 and S2.4), and therefore it appears that the resulting keto acid could not be utilized by *E. coli*, while the remaining amino acid was effectively incorporated in the protein.

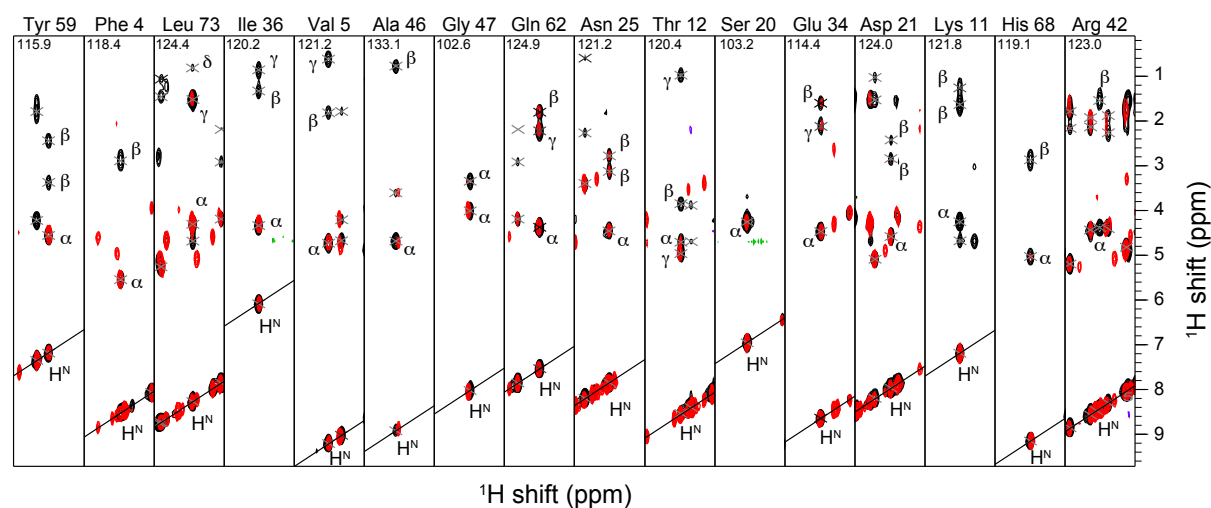


Figure 2.3. Residue specific characterization of labeling from ^1H - ^{15}N TOCSY-HSQC spectra of 1 mM ubiquitin using 75 ms MLEV-17 mixing. α -PET ubiquitin (red) is compared with ^{15}N , ^{13}C -ubiquitin (black).

Of the 13 successful amino acid types, tyrosine, phenylalanine, isoleucine, valine, alanine, threonine and aspartic acid residues show only $\text{H}\alpha$ signals in the ^1H - ^{15}N TOCSY-HSQC spectrum. The anabolic pathway of these residues ends with an aminotransferase reaction, with the exception of threonine, which explains the labeling. Effective aspartic acid labeling was unexpected since it enters and exits the TCA cycle, but is explained by the very high starting concentration.

The amino acid mix from Silantes (Table S2.1) is obtained from bacterial proteins by an HCl proteolysis and consequently glutamine, asparagine, tryptophan, and cysteine are not present in the media. Therefore, glutamine and asparagine require conversion from the respective acids, which explains protonation of beta and gamma protons for these residues (Fig. 2.1). Glutamic acid efficiently enters and exits the TCA cycle, which may explain the incomplete suppression of beta and gamma protons.

Leucine sidechain protons were not expected, but appear to some extent due to LAAO treatment (Fig. S2.2). If the LAAO treatment is not performed, this sidechain labeling is not

observed (Fig. S2.7), thus it is the crude snake venom extract that introduces leucine H γ protons. Details of this side reaction were not investigated further, however we did follow the reaction of LAAO to test efficiency in different buffer conditions for a variety of amino acids (Fig. S2.1-S2.5).

For most amino acids, the reaction proceeded as expected, and the snake venom LAAO was particularly efficient for hydrophobic amino acids such as phenylalanine and isoleucine (68). The degree of conversion to keto acids was also tested for all 20 amino acids directly in the Silantes medium. To distinguish the signal from the individual amino acid without significantly changing the composition, we used deuterated Silantes media, and added only 100 μ M of each protonated amino acid. In this way, we rule out potential issues such as competitive binding to the enzyme and determine the approximate starting concentration of all amino- and keto- acids in the medium (Table 2.1).

Quantification of the labeling for each residue type is tabulated in Tables S2.2 and S2.3 based on intensities extracted from 13 C-HSQC spectra. The intensities were corrected for the measured proton transverse relaxation rates (Fig. S2.11) and normalized based on the assumption of complete incorporation of isoleucine, phenylalanine, and leucine residues, which were cleaved completely and are known to effectively incorporate in *E. coli* (43).

Table 2.1. LAAO activity in deuterated Silantes media, as determined by solution NMR.

Keto acid conversion (%)	Residue	Measured H α incorporation (%)	Residue*
90-100	Ile, Leu, Phe, Tyr, Trp, Met	90-100	Ile, Leu, Phe, Tyr, Met, Val, Ala, Gln, Asn, Thr, Ser, Glu, Asp
10-50	Val, Arg, His	10-50	
0-10	Gly, Pro, Cys, Asn, Gln, Asp, Glu, Ser, Thr, Ala, Lys	0-10	Lys, Arg, His

Each amino acid was added in protonated form at a concentration of 100 μ M and LAAO was added exactly as described in the methods section for expression. The remaining alpha signal intensity was used to determine the degree of conversion to keto acid. *of 16 amino acids that could be quantified (see SI).

We also found that efficient transamination occurs when *E. coli* is grown primarily on amino acids. Some exchange still occurs at amide positions even without LAAO treatment

(Fig. S2.6-S2.7), consistent with a previous report showing significant H α labeling for TCA cycle amino acids, but only 10-50% H α labeling for hydrophobic residues (55).

4.2 GLYCINE IS LABELED STEREOSPECIFICALLY

The H α labeling of glycine attracts particular attention, since one of the two H α protons is labeled predominantly, resulting in stereospecific glycine labeling (Fig. 2.2 and Fig. 2.4). For glycine 28, the intensity ratio between the two alpha protons for microcrystalline $^{13}\text{C},^{15}\text{N}$ SH3 (Fig. 2.4, black) is 1 to 0.93 while the ratio is 1 to 0.30 for α -PET SH3 (Fig. 2.4, red). This effect was observed for glycine in all the samples tested, based on signal intensity in HSQC and CP-HSQC spectra. We also observed a considerable reduction in line width, by more than a factor of three.

Glycine can be produced from serine by hydroxymethyltransferase, from threonine by L-allo-threonine aldolase, or through serine-glyoxylate or alanine-glyoxylate transaminases. Information in *E. coli* is limited, but analysis of other organisms using tritiated water indicates that the stereo specificity depends on the pathways involved (69-71). If serine transhydroxymethylase acts in tritiated water, the resulting glycine will predominantly be the S configuration, but if liver transaminase acts then R will be the predominant configuration. Note that in our case, each enzyme will produce the reverse stereoisomer because the starting amino acid is deuterated, and the enzymatic reaction occurs in protonated water. By examination of NOE spectra of ubiquitin we observed a cross peak between the glycine 47H α and isoleucine 45H N , which according to the known structure, indicates that glycine was predominantly the R configuration. This is consistent with the stereospecific labelling approach reported previously using cell free extracts (72), but results in the opposite labeling, since we expressed in H $_2$ O rather than D $_2$ O. We can therefore rule out deuterated glycine from the medium as the main source of stereospecific glycine found in the expressed protein.

4.3 RESOLUTION AND STRUCTURAL DATA UNDER MAS CONDITIONS

To demonstrate that the α -PET labeling scheme results in improved resolution, we prepared a microcrystalline sample of α -spectrin SH3 according to established crystallization protocols (62). The H α line width is significantly reduced for α -PET SH3 and the effect is

particularly improved for certain residues, by a factor of 2 and above (Fig. 2.4). The proton resolution is also superior to labeling with deuterated glucose in otherwise protonated media (73) (Fig. S2.8). To characterize the narrowing of the homogeneous part of the lines, the bulk T_2' relaxation times at 55 kHz was measured for $H\alpha$, $C\alpha$, and CO from 1D (HCAN)H, (HCON)H and (HCA)HA spectra by integrating the full signal. The $H\alpha$ T_2' of 3 to 4 ms for α -PET SH3 is a dramatic improvement compared to 1 ms for the fully protonated sample (Fig. S2.9).

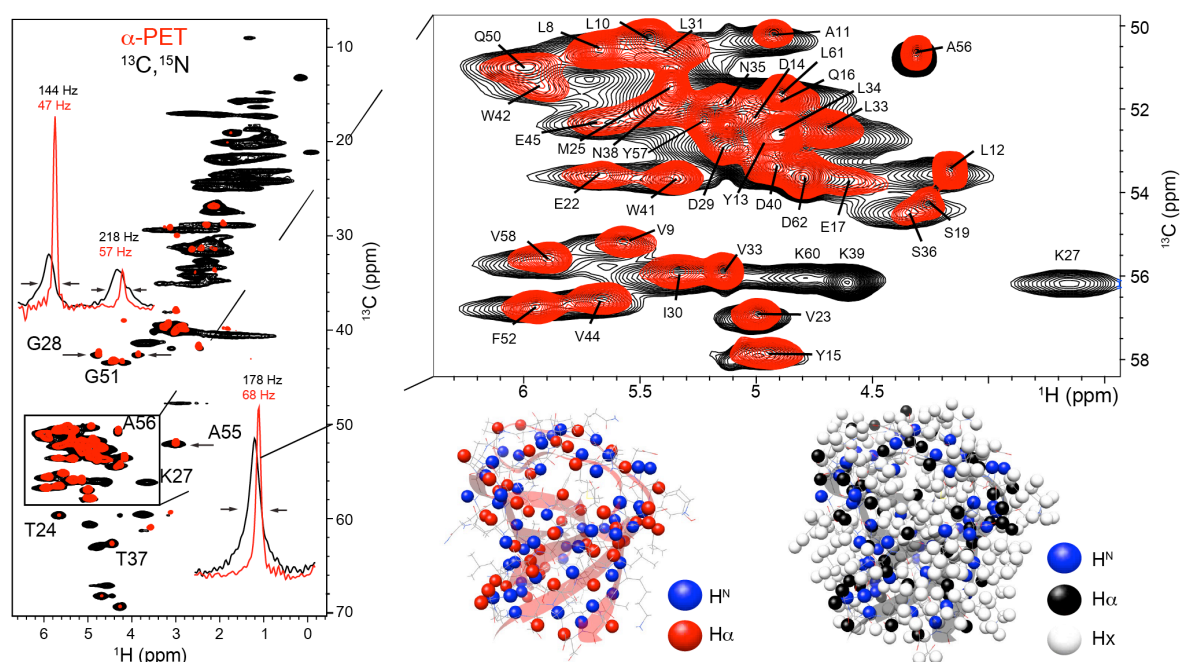


Figure 2.4. Cross-polarization based carbon-proton correlation spectra, (H)CH, of microcrystalline SH3 either uniformly α -PET labeled (red) and $^{13}\text{C},^{15}\text{N}$ -labeled (black) crystallized from a protonated buffer. Spectra were recorded at a magnetic field of 800 MHz and 30 °C, 55 kHz MAS. 1D slices from the spectrum indicate the improvement in linewidths for G28 (top left) and A55 (bottom right). The glycine peak intensities show stereospecific labeling with preference for R (α_3 protonated) over S (α_2) configuration. At the bottom right, the backbone and sidechain protons are indicated on the solution NMR structure (pdb: 1aey) for α -PET SH3 (red ribbon) and $^{13}\text{C},^{15}\text{N}$ SH3 (black ribbon).

The $H\alpha$ T_2' of α -PET SH3 crystallized in 100% D_2O buffer ranged from 7 to 15 ms, an improvement over the amide protonated sample large enough that we can directly observe an increase in resolution in the 1D spectrum (Fig. S2.9). The H^N signals were almost completely removed in the D_2O buffer.

Sequential resonance assignment in Fig. 2.4 were made using a (H)NCAHA spectrum, and are consistent with those previously reported (74). SH3 has 62 residues, of which, two are

prolines and the N-terminal 7 residues and residues 46-48 are flexible and are therefore not observed using cross-polarization based transfer experiments. Thus 50 H α peaks are expected for $^{13}\text{C},^{15}\text{N}$ SH3. For α -PET SH3 lysine, arginine and histidine are not expected. Thus only 41 H α peaks are expected and indeed 41 peaks were readily identified in (H)NCAHA spectra.

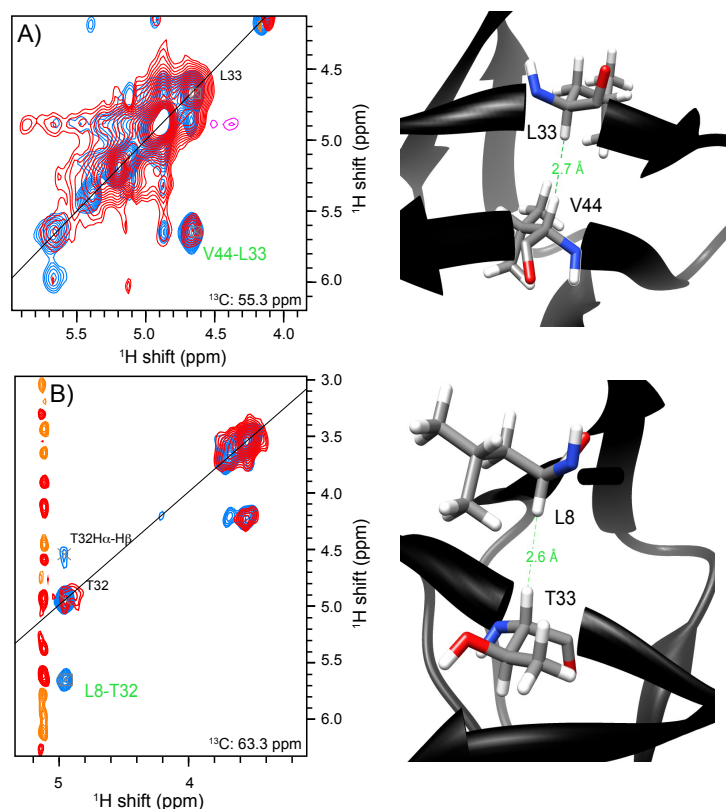


Figure 2.5. Long-range distance information is highlighted in a 3D H(H)CH spectrum of α -PET SH3 (pdb: 1aey) in D $_2$ O (blue) and in H $_2$ O (red).

A shows a contact between L33 H α and V44 H α . In B, the contact between T32 H α and L8 H α is readily observed in D $_2$ O (in blue) while it is much weaker in the presence of additional protons in H $_2$ O (in red). Recorded in a 800 MHz Bruker spectrometer at 30°C and 55 kHz MAS.

Figure 2.5 shows a comparison between α -PET SH3 in fully protonated buffer (red) and α -PET SH3 in fully deuterated buffer (blue) in which long-range structural restraints were measured. To characterize the benefit of the restraints present with α -PET labeling, we manually selected peaks in the H(H)CH and H(H)NH spectra, and used automated shift matching (0.05, 0.5 and 0.5 ppm tolerance, in ^1H , ^{13}C and ^{15}N , respectively) to identify contacts. Of 114 automatically assigned peaks from the 3D H(H)NH of α -PET SH3 in protonated buffer, 2 unambiguous contacts were identified, of which 1 is a long-range H $^{\text{N}}$ – H $^{\text{N}}$ contact. For the 3D H(H)CH, 132 peaks were selected, and 7 unambiguous contacts were identified, 5 of which are long-range restraints. However, for the H(H)CH spectrum in deuterated buffer we found 150 contacts, of which 8 are unambiguous restraints, 7 of which were long-range corresponding to either H $^{\text{N}}$ – H α or H α - H α . One of the additional contacts

identified in fully deuterated buffer is highlighted in fig. 2.5B. This method clearly improves the number of structural restraints available at 55 kHz MAS, and in particular, the unambiguous restraints, a metric that is crucial for the convergence of commonly used structure calculation methods. A concern with H α detection is the presence of water and other solvent signals in this spectral region. Therefore good water suppression is needed, but as demonstrated here for samples in both H₂O and D₂O, control of the water is possible even without gradient methods.

For resonance assignment, the α -PET labeling approach benefits from the implementation of new proton detected NMR pulse sequences focused on H α detection that were recently developed for >100 kHz MAS (75). So far, proton detected MAS NMR structures were mostly based on H^N detected experiments or more recently on fully protonated samples that are best investigated using 100 kHz MAS (76). New possibilities are opened with the α -PET approach, allowing effective structural measurements with the inherently more sensitive equipment for ~60 kHz spinning.

The method was also successful for a more challenging system, the human voltage dependent anion channel (VDAC). The lipid bilayer structure of this protein has been investigated through MAS NMR spectra of VDAC in liposomes (77) and in 2D crystalline arrays (78), and narrow proton resonances were reported for a perdeuterated sample (78, 79). With α -PET labeling, we also observed narrow amide proton linewidths of 150 Hz in H₂O, while ~100 Hz lines were observed using D₂O buffer, which is slightly better than the ~120 Hz linewidths observed for perdeuterated and H^N back-exchanged protein. This indicates that the non-exchangeable protons are slightly narrower, and that H α labeling does not significantly impact the spectral quality. In this D₂O-exchanged buffer, less improvement in H α T₂' (Fig. S2.9) was observed as compared with the SH3 domain, which is not unexpected, since approximately half the amide protons were protected from exchange (Fig. S2.9B).

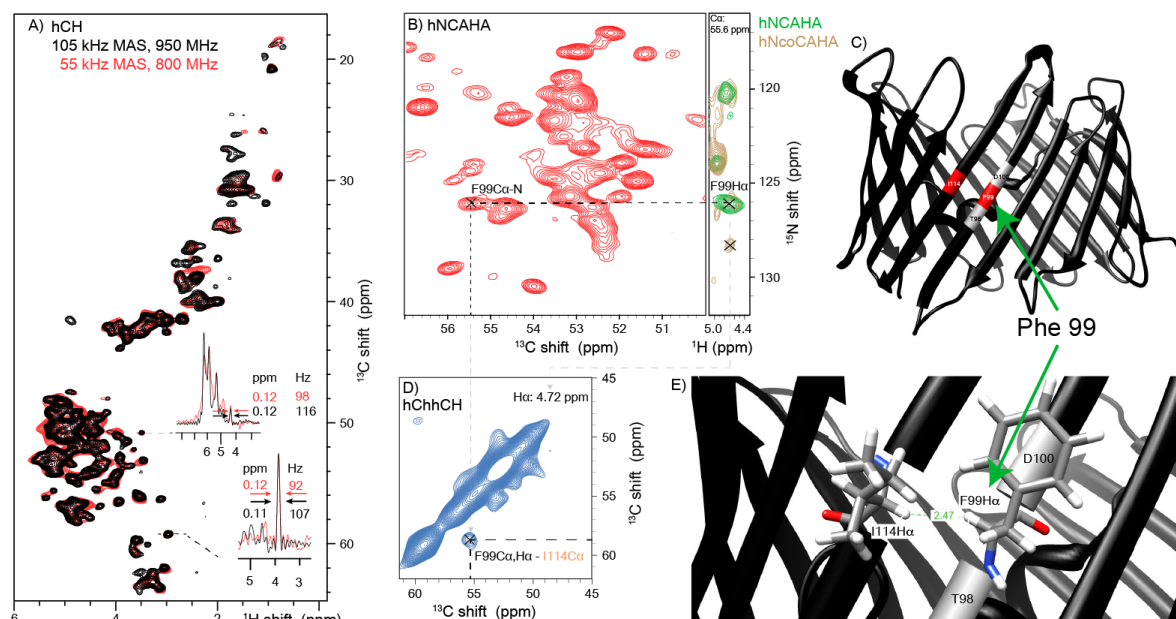


Figure 2.6. Identification of a cross beta strand contact (F99H α –I114H α) in the beta barrel membrane protein VDAC in lipid bilayers.

A shows, the comparison of the (H)CH spectrum at 105 kHz on a 950 MHz spectrometer (black) and at 55 kHz on an 800 MHz spectrometer (red). B shows a ^{13}C - ^{15}N projection of a (H)NCAHA spectrum. F99 H α is assigned from the strip comparing (H)NCAHA (green) and (H)N(CO)CAHA (brown). In C and E, the contact is shown on the X-ray structure of mouse VDAC (pdb: 2jk4). D shows the F99 – I114 cross-peak in the carbon-carbon 2D plane of the (H)C(HH)CH spectrum at the proton frequency of F99, 4.72 ppm.

To further characterize the potential spectral resolution, α -PET VDAC was measured at 105 kHz MAS at a 950 MHz spectrometer (black in Fig. 2.6A). Surprisingly, the same line width was obtained at 110 kHz MAS at 950 MHz (~110 Hz) and at 55 kHz MAS at 800 MHz (~95 Hz), showing that the inhomogeneous contributions are dominating the linewidth at 55 kHz. This shows that even for a highly homogeneous preparation of a membrane protein, α -PET labeling efficiently reduces the proton dipolar broadening at 55 kHz.

The protection from solvent exchange observed for VDAC highlights an issue with perdeuteration for proteins that lack refolding protocols. Perdeuteration of membrane proteins (73, 80) and large complexes (81) in *E. coli* often results in deuterated amides that cannot be exchanged with protons from water. Such exchange protected regions of the protein become inaccessible in the perdeuteration and back-exchange approach, limiting the analysis to solvent accessible regions (38, 42, 82-84), although such limited exchange phenomena can also be used to obtain functional information (80, 85, 86). Using α -PET labeling, we are now able to detect both exchangeable as well as non-exchangeable amide protons in highly deuterated samples as shown previously for amino acid based media (55).

Due to the size of the protein, unambiguous assignment of important cross-strand contacts was not possible in a 3D H(H)CH spectrum of VDAC. We therefore applied the better resolved 3D (H)C(HH)CH spectrum to measure cross-strand contacts (Fig. 2.6). VDAC assembles as a beta barrel, a topology that places cross-strand H α pairs in close proximity (~ 2.3 Å), and much closer than sequential H α spins, which are separated by about 4.5 Å. 28 H α – H α contacts were detected from this spectrum, of which we show the cross strand contact between residue phenylalanine 99 and isoleucine 114, which was assigned based on the existing ^{13}C and ^{15}N assignments of this protein (78, 79) and (H)NCAHA and (H)N(CO)CAHA spectra. The current published assignments (32% of 283 residues) of VDAC do not allow a characterization of all 28 peaks. However, resolving 28 peaks is significant, considering that only ~ 4 amide-amide or alpha-alpha contacts are available in each transmembrane beta sheet interface, of which VDAC has 19. Further analysis of the expected contacts in VDAC is shown in figure S2.10. This demonstrates a successful implementation of α -PET labelling for structure determination in a challenging 32 kDa membrane protein embedded in lipid bilayers, where structural restraints are particularly difficult to identify (78).

4.4 α -PET LABELING FOR SOLUTION NMR

The α -PET labeling approach is also beneficial for the study of proteins in solution, when deuteration is needed to reduce transverse relaxation rates (87, 88). Figure 2.7 shows the reduction in R₂ relaxation rates due to the high level of deuteration in α -PET labeled Ubiquitin. Such improvement in relaxation rates is important for the study of protein dynamics. For example in detection of H α relaxation dispersion, fractional deuteration was used to improve R₂ (89, 90). The current labeling incorporates the alpha positions at 100 percent for most residues, with a high overall deuteration level, which improves sensitivity as compared with random fractional deuteration.

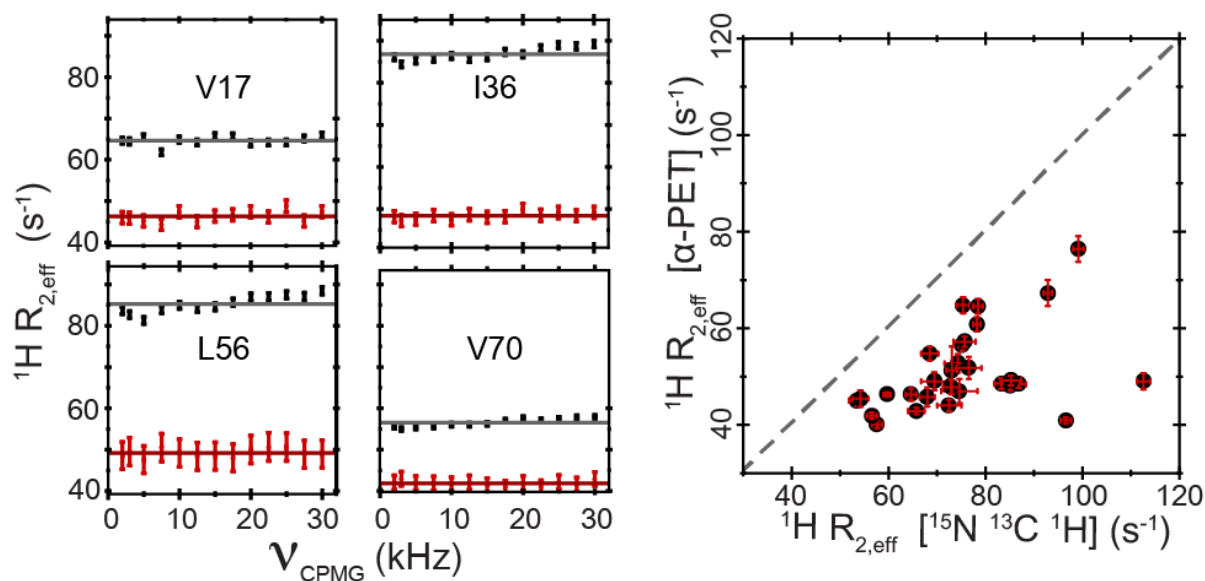


Figure 2.7. Selected residues showing the reduction in proton ($H\alpha$) R_2 relaxation rates with α -PET labeling (red) as compared with full protonation (black). The correlation plot (right) shows a reduction for all residues.

The data is from ubiquitin samples exchanged in 100% D_2O at 277 K and measured at a 600 MHz spectrometer.

5. Conclusion

Here, we introduced a new method to label $H\alpha$ protons in a protein without significant isotopic scrambling, and demonstrated how this new sensitive magnetic probe in the backbone of the protein adds new structural information even at below 60 kHz MAS. The α -PET labeling approach has several advantages, i) adaptation of the cells to D_2O is not required, ii) it gives similar yields as deuterated expression in M9 media iii) costs are similar to production of deuterated proteins. It is expected to be particularly useful for deuteration of proteins that lack refolding protocols, such as membrane proteins.

In this demonstration, we used a commercially available crude snake venom extract to generate keto acids. This approach results in the designed incorporation of alpha protons for Tyr, Phe, Leu, Ile, Gly, Gln, Asn, Asp, Glu and Met. In the future, further optimization of the method might entail other LAAOs with different substrate specificity, perhaps in combination with auxotrophic strains to limit unwanted reaction pathways. In addition, other amino acid mixtures or expression systems could be investigated. This might allow labeling of lysine, arginine, and histidine, which were currently left deuterated.

SUPPLEMENTARY

Quantification of labeling patterns, LAAO activity, measurement of relaxation times under 55 kHz MAS, $H\alpha$ R_2 in solution and spectral acquisition parameters.

Table S2.1. Amino acid composition of Silantes media as supplied by the manufacturer, with negligible concentrations of carbohydrates (less than 30 mg per liter).

Amino acid composition before treatment	
Amino acid	mole percent
Asp	20.6
Thr	3.7
Ser	4.1
Glu	11.1
Gly	11.1
Ala	14.8
Val	3.0
Met	1.8
Ile	2.1
Leu	5.3
Tyr	1.8
Phe	2.4
His	9.7
Lys	3.4
Arg	2.3
Pro	3.0

Table S2.2. Incorporation level at the H α position by amino acid type as estimated from ^{13}C -HSQC spectra without correcting for differences in T $_2$ relaxation.

Spectra were recorded on a 950 MHz Bruker spectrometer using 30 ms indirect evolution for both α -PET and U- ^{13}C , ^{15}N -labeled samples in D $_2\text{O}$. Water suppression by saturation or selective pulsing was not used, since the alpha protons overlap with water. The data was processed using a sine squared function in both dimensions and using 8k indirect points. Peak intensities were used to estimate labelling efficiency. Amino acids that do not scramble, and for which LAAO was effective (Ile, Phe, Leu) were used to normalize the spectra. The largest source of error is not due to the signal to noise ratio, and we therefore expect that it comes from differences in relaxation between the deuterated and protonated samples. We therefore corrected for T $_2$ in Table S3.

Residue	Number of residues	Reference sample	α PET sample	ratio	stdv	percentage
Tyr	1	2.18	3.16	1.451	*	145.1
Phe	2	2.24	2.34	1.046	0.04	104.6
Leu	8	3.57	3.63	1.015	0.11	101.5
Ile	7	2.79	2.64	0.944	0.08	94.4
Val	4	3.49	2.41	0.692	0.07	69.2
Ala	2	3.19	2.05	0.644	0.02	64.4
GlyHα3	3	2.53	3.24	1.280	0.07	128.0
GlyHα2	3	4.41	1.05	0.239	0.01	23.9
Gln	4	2.91	2.79	0.960	0.07	96.0
Asn	1	3.22	3.82	1.186	*	118.6
Thr	7	3.09	1.05	0.341	0.01	34.1
Ser	2	2.67	1.27	0.476	0.01	47.6
Glu	5	2.77	2.87	1.038	0.03	103.8
Asp	2	3.08	3.53	1.146	0.02	114.6
Lys	5	3.56	-	0.000	*	0.0
Arg	4	4.42	-	0.000	*	0.0
His	1	2.95	-	0.000	*	0.0
Met	1	4.76	4.52	0.950	*	95.0
	62					

Table S2.3. Incorporation level at the H α position by amino acid type as estimated from ¹³C-HSQC spectra corrected by the H α T₂.

The intensities from Table S2 were corrected by the T₂ measured at the H α position of each individual amino acid. Peak intensities were used to estimate labelling efficiency. Amino acids that do not scramble, and for which LAAO was effective (Ile, Phe, Leu) were again used to normalize the spectra. With T₂ correction, Val, Ala, Ser and Thr show higher incorporation, of about 100%. The correction of the T₂ could only be made reliable for isolated peaks that do not overlap with other peaks or with water (only 47 peaks have been used in this case).

Residue	Number of residues	Reference sample	αPET sample	ratio	stdv	percentage
Tyr	1	1.20	1.25	1.040	*	104.0
Phe	2	1.32	1.29	0.984	0.03	98.4
Leu	8	1.23	1.25	1.017	0.02	101.7
Ile	6	1.28	1.28	1.001	0.03	100.1
Val	3	1.22	1.26	1.039	0.04	103.9
Ala	2	1.21	1.27	1.046	0.01	104.6
GlyHaa	0	*	*	*	*	*
GlyHab	0	*	*	*	*	*
Gln	4	1.25	1.31	1.048	0.01	104.8
Asn	1	1.22	1.31	1.075	*	107.5
Thr	3	1.35	1.36	1.007	0.09	100.7
Ser	1	1.24	1.29	1.038	*	103.8
Glu	4	1.27	1.31	1.027	0.03	102.7
Asp	1	1.19	1.29	1.078	*	107.8
Lys	5	1.24	*	*	*	*
Arg	4	1.17	*	*	*	*
His	1	1.24	*	*	*	*
Met	1	1.20	1.26	1.044	*	104.4
	47					

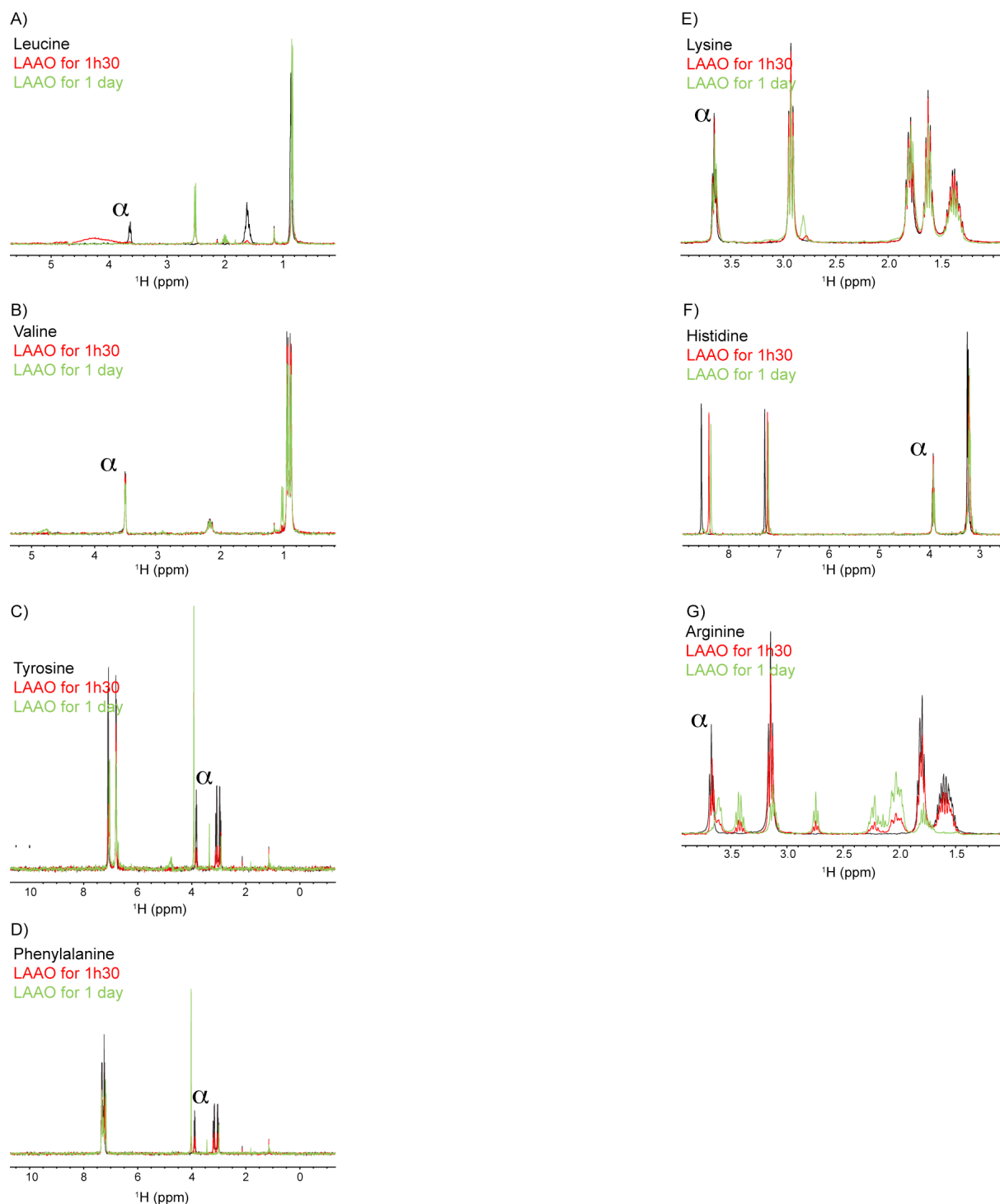


Figure S2.1. LAAO activity for different amino acids in solution.

Conversion was effective for leucine (A), tyrosine (C), phenylalanine (D) and arginine (G), and not effective for valine (B), lysine (E) and histidine (F). Some activity was observed for histidine at lower concentration (see Figure S2.4).

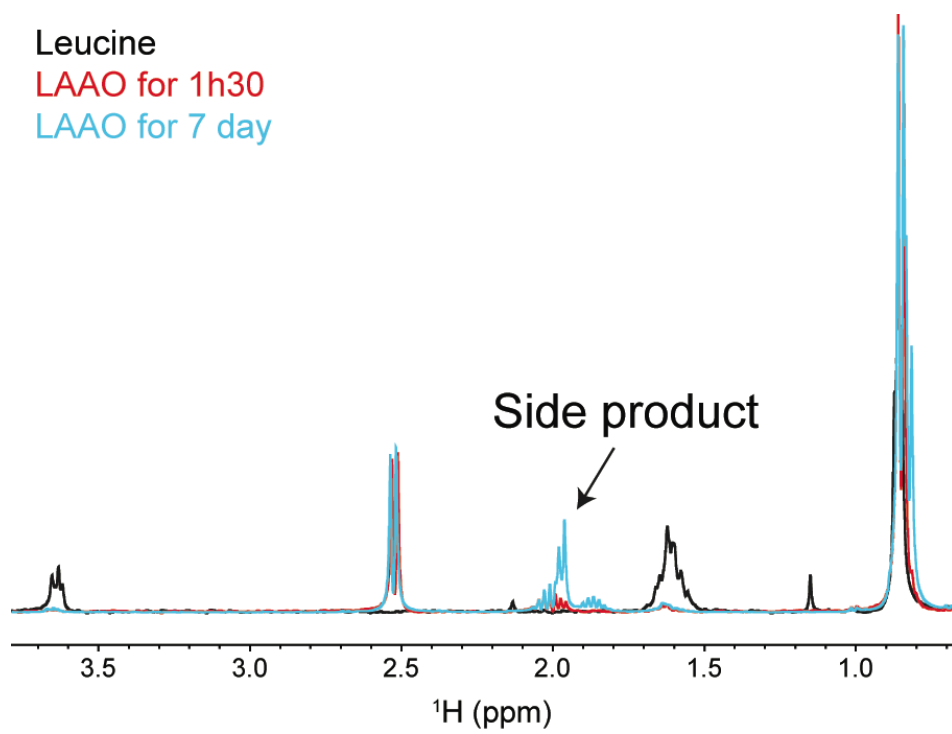


Figure S2.2. Side products from LAAO treatment on Leucine.

Other amino acids

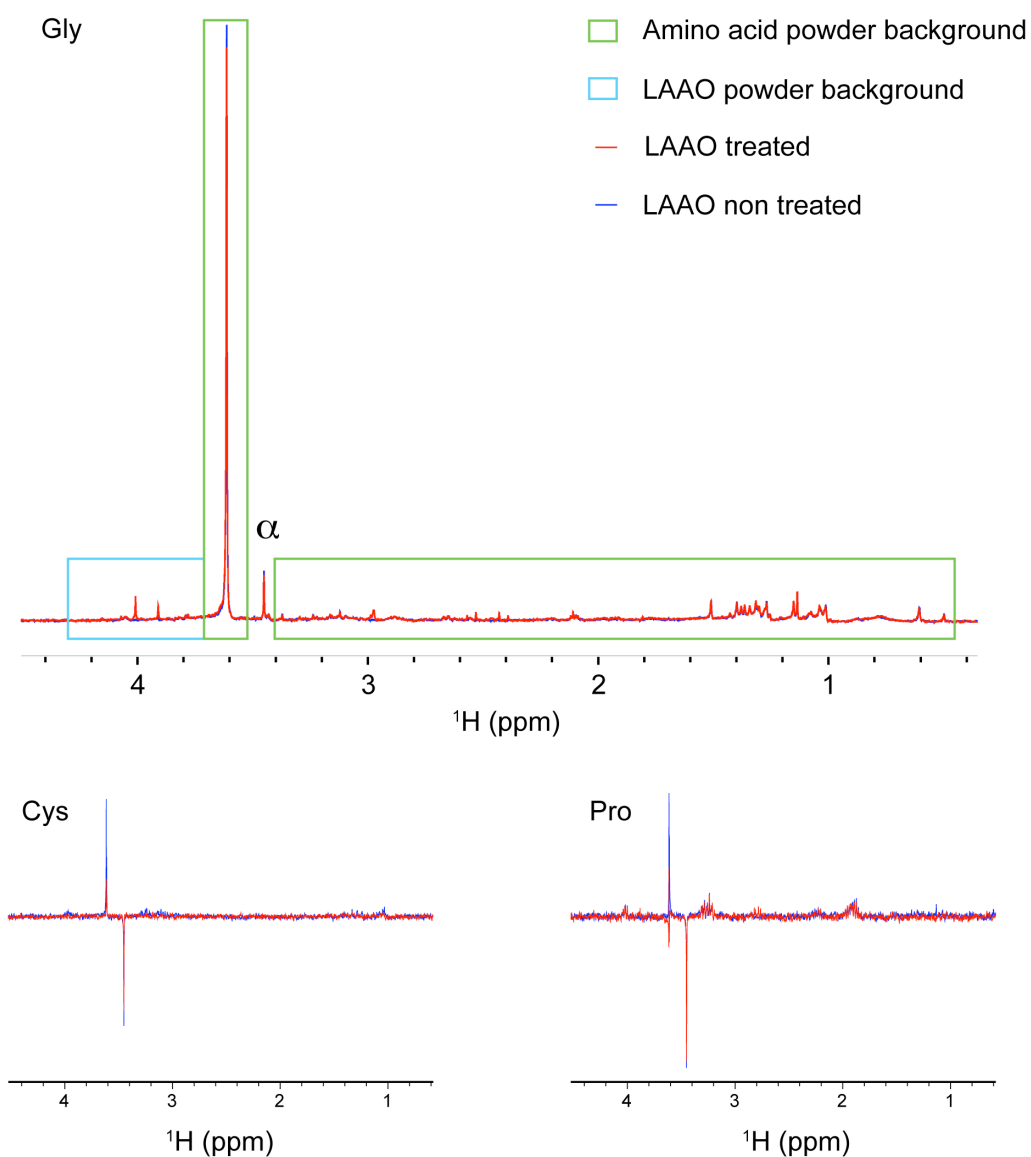


Figure S2.3. LAAO treatment efficiency on a deuterated amino acid powder for different amino acids.

Polar/Charged amino acids

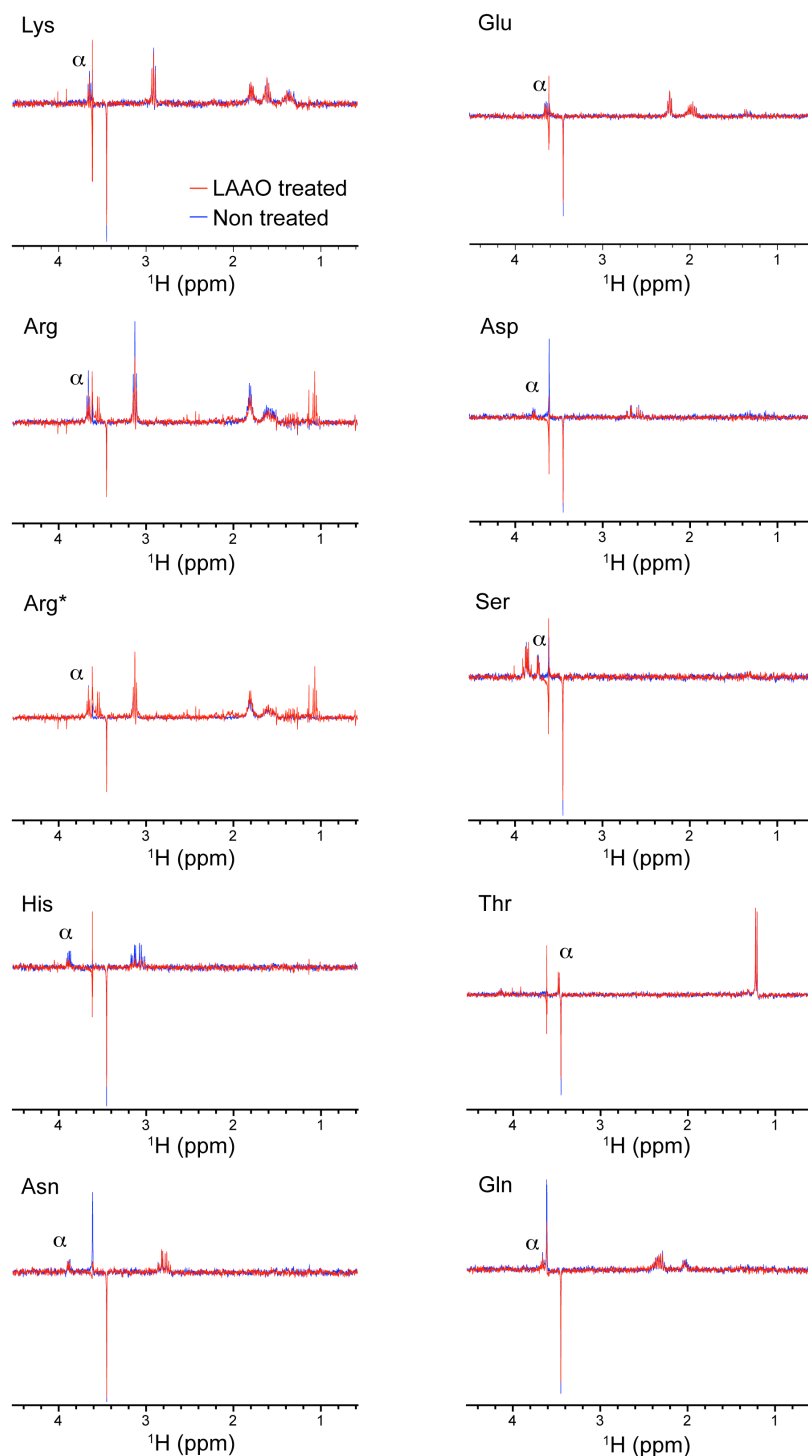


Figure S2.4. LAAO treatment efficiency on a deuterated amino acid powder for polar/charged amino acids.

50 μM of the respective amino acid were dissolved in deuterated powder. The Glycine spectrum was used to cancel the background by taking the difference (Glycine H_α appears as a negative peak). Arg* corresponds to the Arg spectrum scaled as a function of the triplet (CH_2), since the arginine concentration was different in the reference and LAAO treated samples. The LAAO enzyme is active on His and Arg, but results in only partial production of the keto acid. For all the other polar and charged amino acids Lys, Glu, Asp, Asn, Gln, Ser and Thr LAAO has no effect.

Hydrophobic amino acids

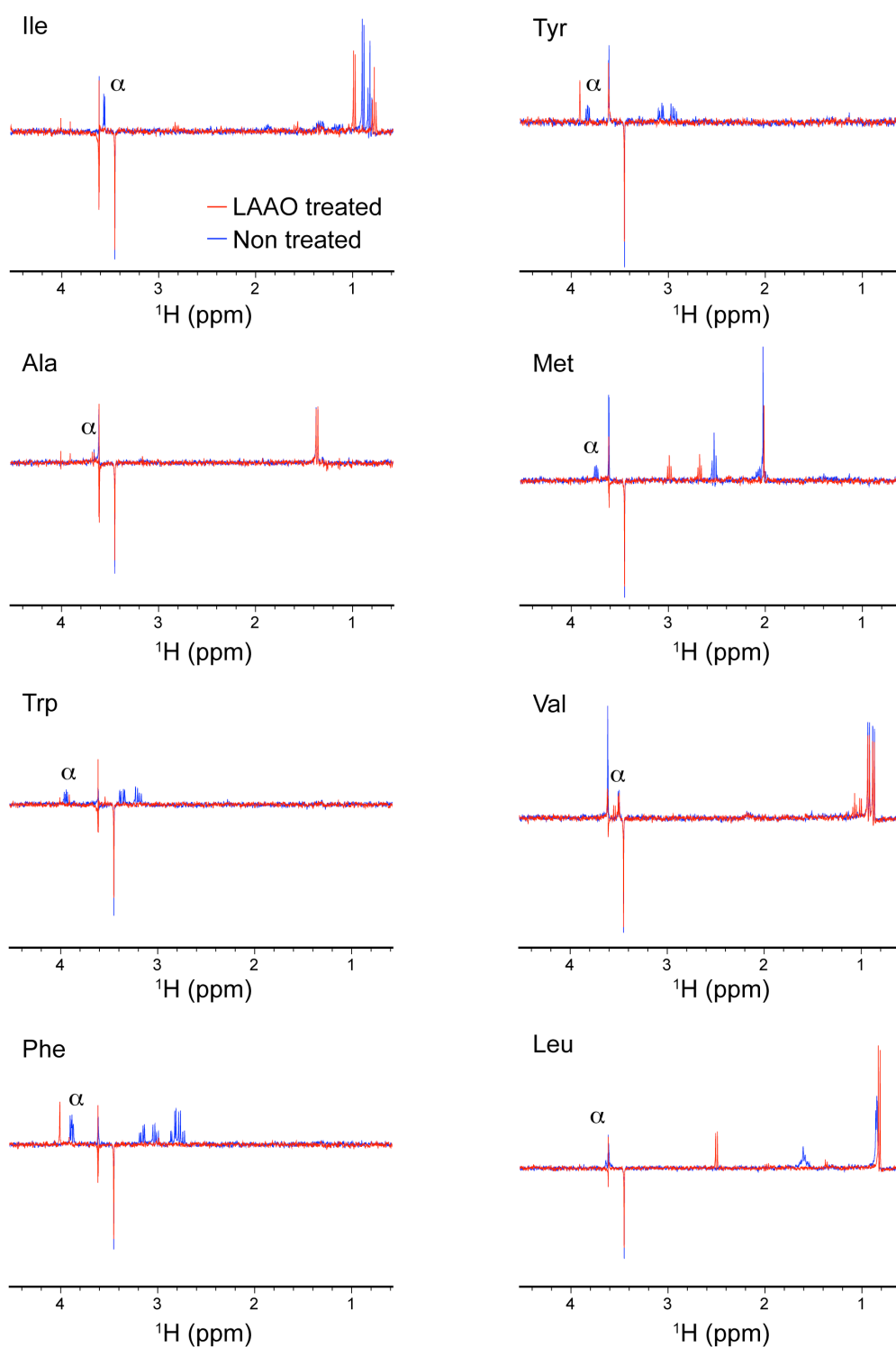


Figure S2.5. LAAO treatment efficiency on a deuterated amino acid powder for hydrophobic amino acids. 50 μM of the respective amino acid were dissolved in deuterated powder. The Glycine spectrum was used to remove the background signal by taking the difference spectrum ($\text{H}\alpha$ of Glycine appears as a negative peak). LAAO treatment is 100% effective for Ile, Leu, Phe, Trp, Tyr and Met. The LAAO treatment has some activity on Val, however, is not active for Ala in the deuterated amino acid media.

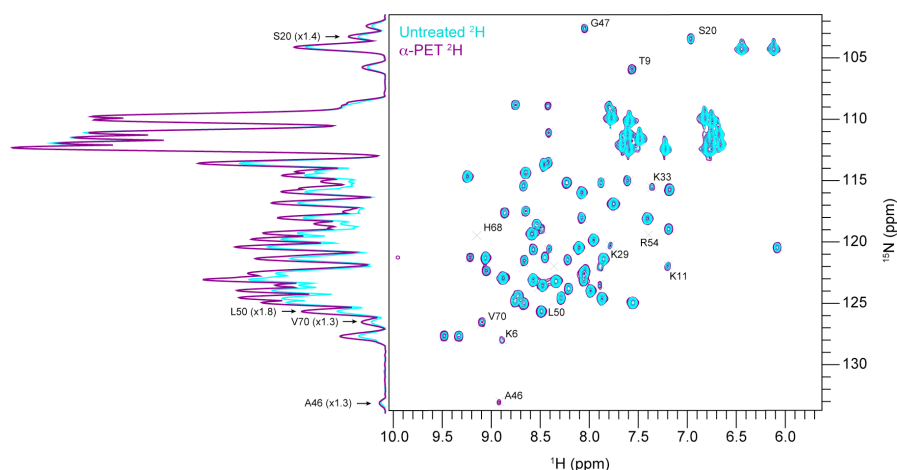


Figure S2.6. Transaminase and other enzyme activity in *E. coli* is sufficient to exchange the amide position of many residues, with and without LAAO treatment.

^2H Silantes (without ^{15}N labelling) was added to the culture after growth in ^{15}N M9 medium. Peaks that remain are labelled through amino acid synthesis pathways or through transaminase. Suppression of sidechain protons (Fig. S2.8-S2.9) indicate that synthesis pathways are not a significant contribution. In A, $^{13}\text{C},^{15}\text{N}$ -ubiquitin (black) and untreated ubiquitin (cyan) shows the inherent transaminase activity *E. coli*. In B, ^{15}N -HSQCs of LAAO treated (purple) and untreated (cyan) ubiquitin (starting with ^2H Silantes) shows that treatment helps to introduce a higher level of amide signal for many amino acids.

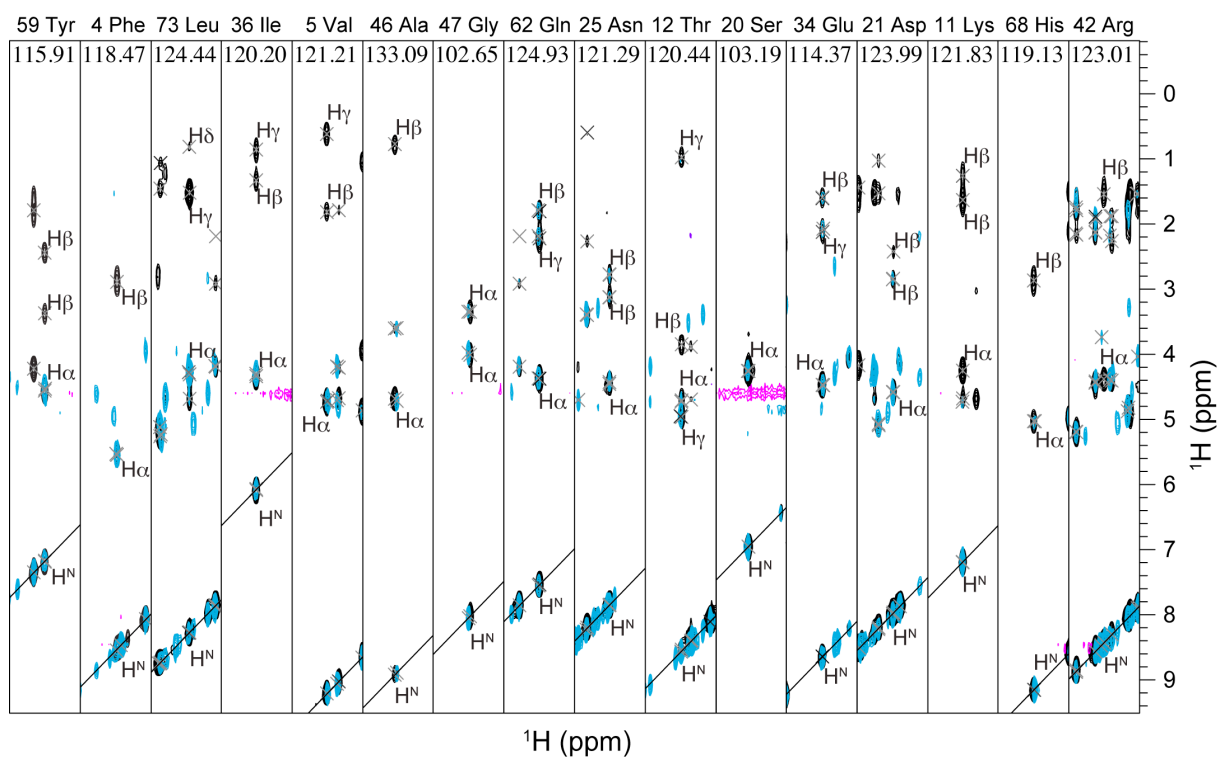


Figure S2.7. Amino acid labelling pattern of α -PET Ubiquitin without LAO treatment.
 The ^{15}N -TOCSY of ^{15}N , ^{13}C -Ubiquitin (black) is compared with untreated α -PET-Ubiquitin (cyan).

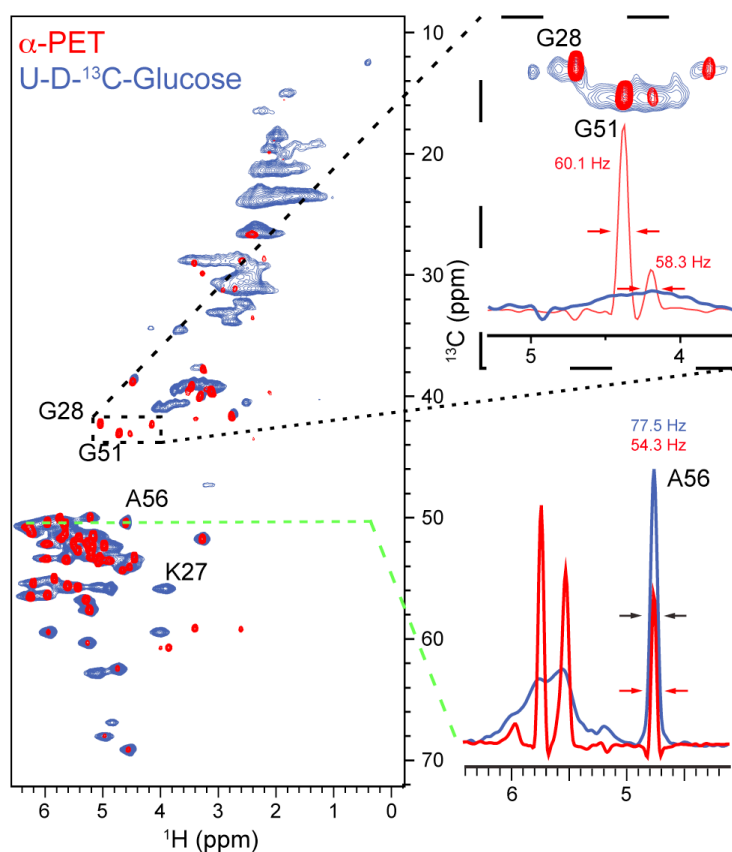
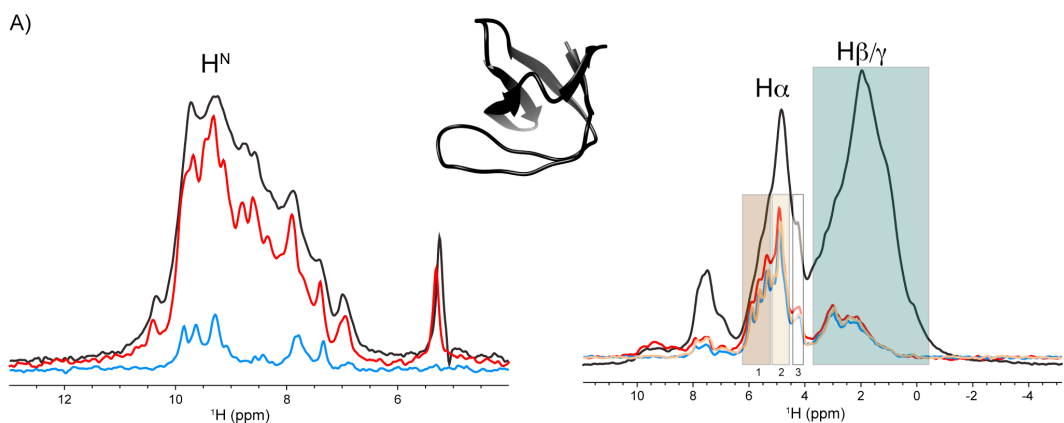
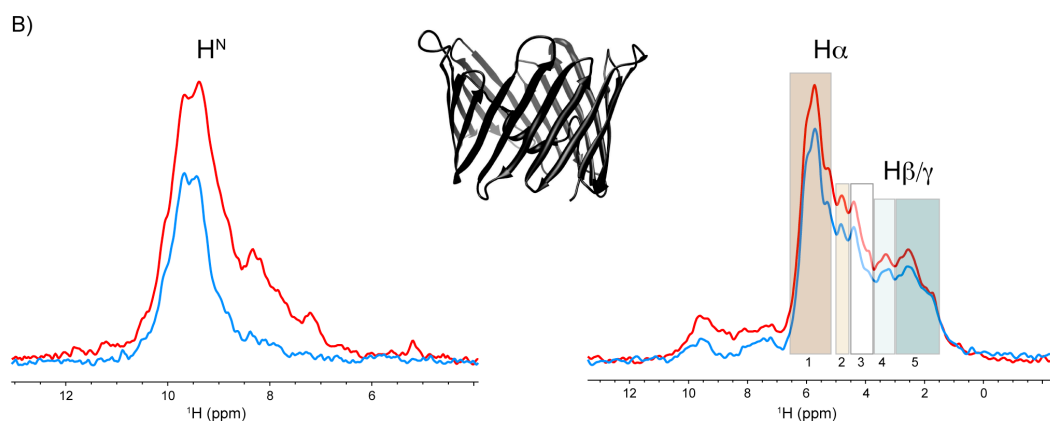


Figure S2.8. Comparison of spectral quality when labeling with deuterated glucose in otherwise protonated media.

Both spectra were recorded at a MAS spinning frequency of 55 kHz on a Bruker 800 MHz spectrometer. The improvement in resolution with α -PET labeling is compared for selected regions of the 2D spectrum. This labeling scheme has been dubbed inverse fractional deuteration (iFD) (Madeiros *et al.*).



T2' (ms)	H ^N	Hα ₁	Hα ₂	Hα ₃	Hβ/γ
Full protonation	1.245		1.39		0.878
α-PET in H ₂ O	3.58	2.79	3.1	5.64	2.36
α-PET in D ₂ O	*	7.04	8.59	15.22	4.47
α-PET in D ₂ O*	*	7.59	8.49	16.2	4.17



T2' (ms)	H ^N	Hα ₁	Hα ₂	Hα ₃	Hβ/γ ₄	Hβ/γ ₅
α-PET in H ₂ O	2.72	1.85	2.84	2.67	2.25	1.96
α-PET in D ₂ O	2.87	2.36	3.97	3.83	2.56	2.04

Figure S2.9. Transverse relaxation (T2') is compared for α-PET samples in 100% deuterated buffer and 100% protonated buffer.

Data were recorded at 800 MHz, with 55 kHz MAS and filtered through cross polarization based (H)NH or (H)CH spectra for H^N or H^C relaxation times, respectively. In A, H^N and H^C relaxation times for the microcrystalline α-spectrin SH3 domain labelled with α-PET in protonated buffer (red), α-PET in deuterated buffer (blue), α-PET using the media exchange protocol in deuterated buffer (brown), and ¹³C,¹⁵N in protonated buffer (black). The first point of the decay curves are shown above the table in the corresponding colors. In B, the bulk T₂' relaxation times of the α-PET labelled 32kDa membrane protein VDAC are tabulated. Again, the first point in the decay curves are shown in the corresponding colors

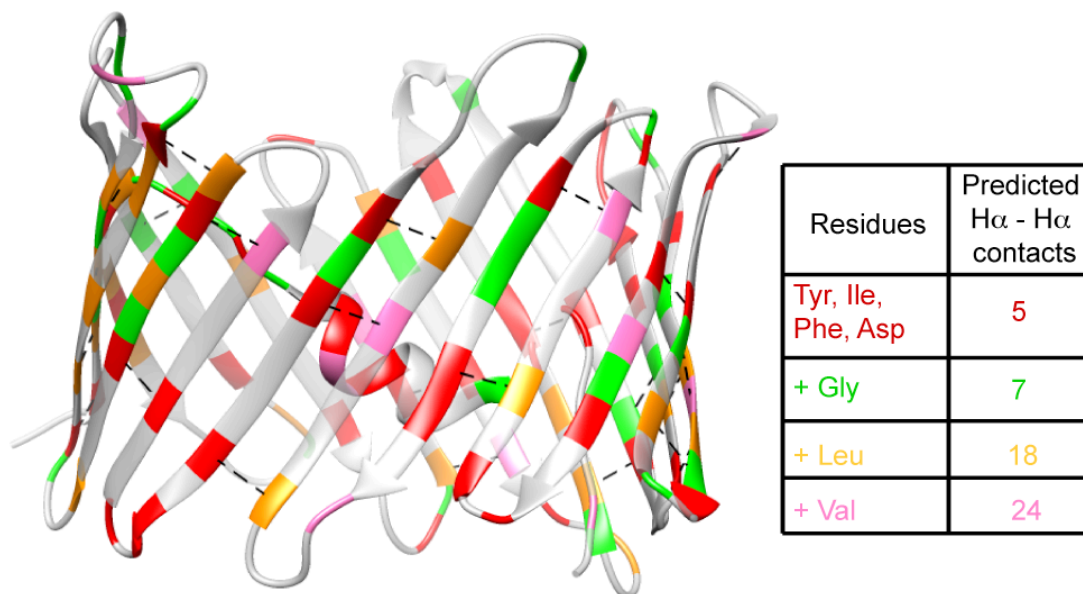


Figure S2.10. Predicted H α - H α contacts accordingly to α -PET labeling mapped into the VDAC NMR structure.

The residues showing 100% H α reincorporation are shown in red, green and yellow, 18 contacts could be predicted from those residues. Valine in pink add 6 more contacts. The predicted contacts are well distributed over the structure, showing the relevance of the H α - H α contribution for structure determination.

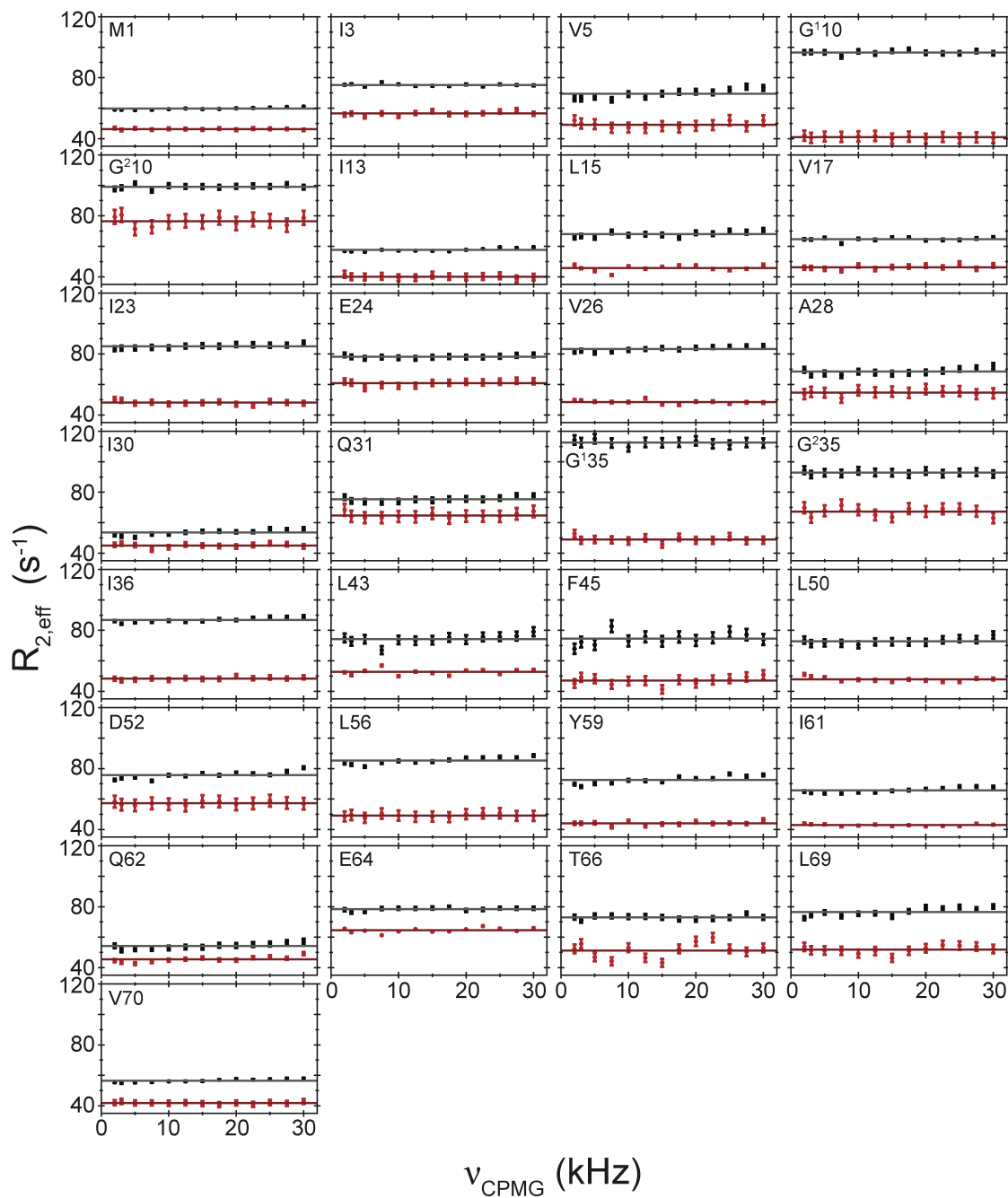


Figure S2.11. Transverse relaxation rates (R_2) using CPMG for fully protonated (black) and α -PET labeled ubiquitin (red).

The data is from isolated peaks in the $C\alpha$ - $H\alpha$ HSQC of ubiquitin samples exchanged in 100% D₂O at 277 K and measured at a 600 MHz spectrometer.

Table S2.4. Transverse relaxation rates (R2), obtained from CPMG measurements at 277 K for resolved residues in uniform ¹⁵N, ¹³C labeled and α -PET labeled ubiquitin measured at a 600 MHz spectrometer.

Residue	Transverse relaxation rate at 277 K (s ⁻¹)	
	¹⁵ N, ¹³ C ubiquitin	α -PET ubiquitin
M1	59.7 ± 0.7	46.3 ± 0.4
I3	75.2 ± 0.8	56.5 ± 1.2
V5	69.4 ± 2.8	49.0 ± 1.9
G¹10	96.5 ± 1.1	40.9 ± 0.7
G²10	99.1 ± 1.1	76.4 ± 2.7
I13	57.6 ± 1.1	40.1 ± 0.8
L15	68.0 ± 1.6	45.7 ± 2.0
V17	64.6 ± 1.1	46.3 ± 1.2
I23	85.1 ± 1.1	48.1 ± 1.0
E24	78.2 ± 0.8	60.8 ± 1.5
V26	83.3 ± 1.4	48.5 ± 1.1
A28	68.5 ± 1.8	54.7 ± 1.4
I30	53.6 ± 1.8	45.0 ± 1.1
Q31	75.4 ± 1.3	64.8 ± 1.7
G¹35	112.6 ± 1.4	49.0 ± 1.7
G²35	92.8 ± 0.9	67.3 ± 2.7
I36	86.8 ± 1.5	48.4 ± 0.8
L43	74.3 ± 2.7	52.7 ± 1.9
F45	74.6 ± 3.8	47.0 ± 2.5
L50	72.9 ± 1.7	47.8 ± 1.4
D52	75.7 ± 2.3	57.2 ± 1.1
L56	85.2 ± 2.1	49.2 ± 1.0
Y59	72.5 ± 2.5	44.0 ± 1.1
I61	65.7 ± 1.7	42.85 ± 0.7
Q62	54.2 ± 1.7	45.4 ± 1.7
E64	78.4 ± 1.0	64.6 ± 1.5
T66	73.0 ± 1.1	51.3 ± 5.0
L69	76.6 ± 2.6	51.8 ± 2.3
V70	56.5 ± 1.0	41.9 ± 0.6

Table S2.5. Transverse relaxation rate (R2), obtained from CPMG measurements at 308 K for resolved residues in uniform ¹⁵N, ¹³C labeled and α-PET labeled ubiquitin measured at a 600 MHz spectrometer.

Residue	Transverse relaxation rate at 308 K (s ⁻¹)	
	¹⁵ N, ¹³ C ubiquitin	α-PET ubiquitin
M1	21.1 ± 0.3	16.7 ± 0.5
Q2	27.3 ± 0.2	24.6 ± 0.6
F4	36.5 ± 0.5	23.3 ± 0.5
V5	24.7 ± 0.8	17.3 ± 0.3
T7	25.8 ± 0.2	20.5 ± 1.9
T9	18.4 ± 0.9	17.7 ± 1.6
G¹10	35.7 ± 0.6	31.9 ± 1.4
G²10	35.3 ± 0.3	34.0 ± 0.8
T12	32.6 ± 0.5	26.5 ± 1.1
I13	20.7 ± 0.3	15.5 ± 1.3
T14	34.0 ± 0.4	28.6 ± 1.1
L15	24.6 ± 0.9	16.8 ± 0.3
V17	24.0 ± 1.3	16.5 ± 5.0
E18	29.6 ± 0.6	18.7 ± 0.3
T22	34.1 ± 0.4	25.0 ± 1.2
I23	31.2 ± 0.6	19.5 ± 0.1
E24	28.2 ± 0.6	21.5 ± 0.3
V26	30.7 ± 0.4	19.4 ± 0.4
A28	25.4 ± 0.2	20.4 ± 0.8
I30	27.9 ± 0.2	18.9 ± 0.7
Q31	28.3 ± 0.5	23.9 ± 0.3
E34	19.9 ± 0.5	17.7 ± 0.1
G¹35	43.6 ± 0.4	35.0 ± 0.7
G²35	43.1 ± 0.3	38.7 ± 1.5
I36	32.6 ± 0.2	18.5 ± 0.6
D39	20.2 ± 0.4	19.5 ± 0.8
Q41	27.5 ± 0.1	21.5 ± 0.7
L43	27.3 ± 0.2	19.5 ± 0.5
I44	26.2 ± 0.2	19.3 ± 0.3
F45	25.5 ± 0.2	16.8 ± 0.3
A46	18.2 ± 0.5	14.8 ± 0.9
G47	43.0 ± 0.8	39.1 ± 2.5
Q49	24.7 ± 0.3	21.1 ± 1.2

L50	26.3 ± 0.3	18.1 ± 0.2
D52	28.0 ± 0.8	21.3 ± 0.5
T55	34.7 ± 0.2	25.5 ± 0.3
L56	27.7 ± 0.1	17.3 ± 0.6
S57	24.8 ± 0.4	19.7 ± 0.7
Y59	20.6 ± 2.0	15.8 ± 2.5
N60	22.6 ± 0.2	21.5 ± 0.4
I61	24.2 ± 0.4	18.0 ± 0.4
Q62	19.4 ± 0.5	16.5 ± 0.5
34.6	32.2 ± 0.4	27.2 ± 0.4
T66	34.6 ± 0.6	27.3 ± 1.4
L67	24.6 ± 0.2	17.1 ± 0.2
L69	27.1 ± 0.4	18.7 ± 0.4
V70	20.3 ± 1.6	16.2 ± 0.4
L71	21.6 ± 0.1	15.1 ± 0.2
L73	11.1 ± 0.1	87.7 ± 0.5
G75	9.2 ± 0.1	7.6 ± 0.5

Table S2.6. Spectrum acquisition parameters recorded for the different samples.

All were recorded at a magnetic field of 18.8 T (on an 800 MHz Bruker spectrometer) at 55kHz MAS using a 3 channel narrow bore HCN probe. The temperature was set at 250K for VDAC and 260K for SH3. 10kHz of waltz16 decoupling was used for ^{13}C and ^{15}N . 12 kHz swept TPPM decoupling was used during indirect evolution periods. Water was suppressed by saturation with at 13.75 kHz for 200ms (protonated buffer) and 50ms (deuterated buffer). The spectra were processed with a squared cosine function. The direct acquisition time was 10 ms and 7 ms for SH3 and VDAC, respectively.

Sample	Spectrum	TD F3	TD F2	TD F1	AQ F2 (ms)	AQ F1 (ms)	NS	Total time
SH3 ref	(H)CH	*	1024	402	*	9.9	4	1h30
α -PET SH3 in H ₂ O	(H)CH	*	1024	402	*	9.9	4	1h30
α -PET SH3 in D ₂ O	(H)CH	*	1024	402	*	9.9	4	1h30
α -PET SH3 in H ₂ O	H(H)NH	1024	102	80	5.3 (^1H)	14.9 (^{15}N)	4	24h
α -PET SH3 in H ₂ O	H(H)CH	1024	160	192	4.9 (^1H)	10 (^{13}C)	2	24h
α -PET SH3 in D ₂ O	H(H)CH	1024	160	192	4.9 (^1H)	10 (^{13}C)	2	24h
α -PET VDAC D ₂ O	(H)C(HH)C H	1024	294	170	4.6 (^{13}C)	4.2 (^{13}C)	2	8 d

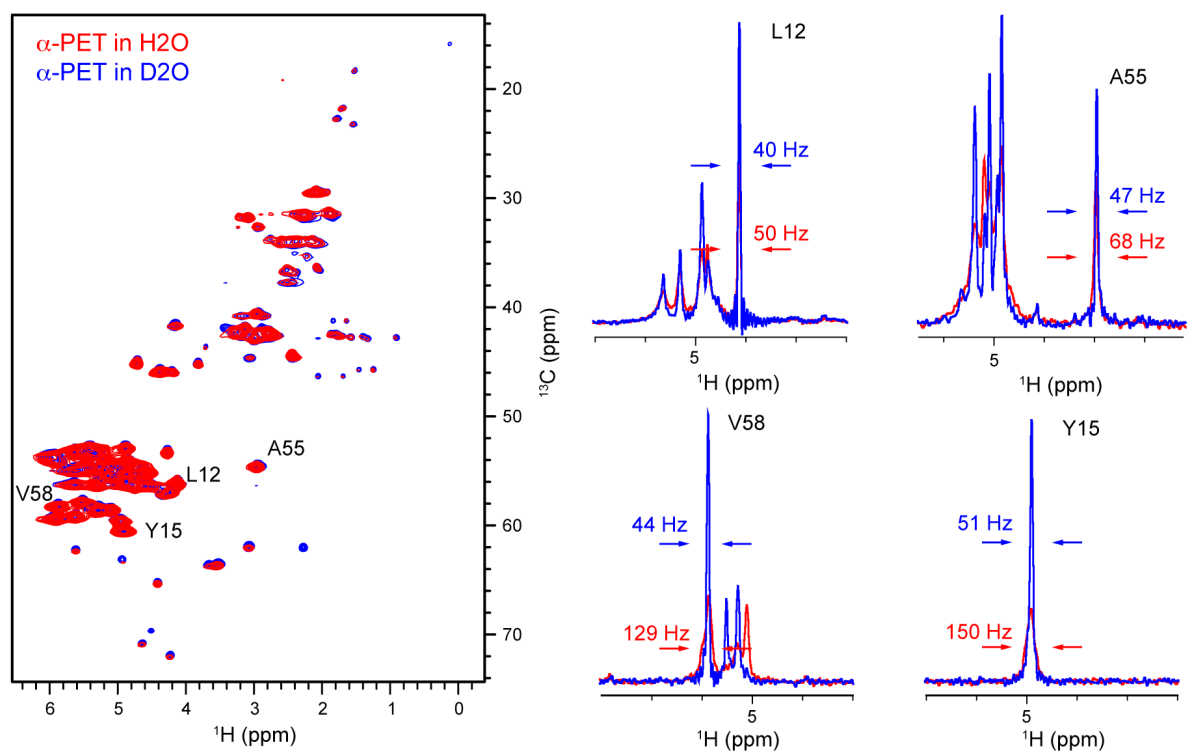


Figure S2.12. α -PET SH3 comparison using protonated buffer (red) or deuterated buffer (blue) for crystallization.

Most of the residues show an improvement in the line width when amide protons are exchanged using deuterated buffer for crystallization. When line widths are reported, no apodization was applied for the reported dimension.

ACKNOWLEDGMENT

We thank Pablo Trigo Mourino for helping to set up solution NMR measurements and Christian Griesinger, Tobias Schubeis and Guido Pintacuda for useful discussions. We also thank Ashok Kumar Rout for NOE data in ubiquitin. We acknowledge financial support from the MPI for Biophysical Chemistry, and from the Deutsche Forschungsgemeinschaft (Emmy Noether program grant AN1316/1-1, SFB803 grant INST 186/794-3).

Chapter 3

Probing membrane protein insertion into lipid bilayers by solid-state NMR

This chapter is part of a paper published in PhysChemPhys on the 19th of November 2019 with the DOI: 10.1002/cphc.201800793

Authors: Eszter E. Najbauer^{1†}, Kumar Tekwani Movellan^{1†}, Tobias Schubeis², Tom Schwarz³, Kathrin Castiglione⁴, Karin Giller¹, Guido Pintacuda², Stefan Becker¹, Loren B. Andreas¹

1 Department of NMR based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

2 Centre de RMN à Très Hauts Champs, Institut des Sciences Analytiques, UMR 5280/CNRS, ENS Lyon, UCB Lyon 1, Université de Lyon, Villeurbanne, France

3. Institute of Biochemical Engineering, Technical University of Munich, Boltzmannstraße 15, D-85748 Garching, Germany.

4 Institute of Bioprocess Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Paul-Gordan-Straße 3, 91052 Erlangen, Germany.

[†] These authors contributed equally to this work.

Contributions: Esther E. Najbauer and I measured and interpreted the data for hVDAC and AlkL, respectively. Additionally, Esther E. Najbauer fitted the intensity build ups. Tom Schwarz performed the expression and purification of the protein. Tobias Schubeis and Loren Andreas did the assignment of the AlkL protein. Karin Giller expressed and purify the hVDAC protein.

1. Abstract

Protein environment could play an important role on protein function, and can also be used to infer structural properties of the protein itself. Using proton-detected solid-state NMR, we show that reduced spin diffusion within the protein under conditions of fast magic-angle spinning, high magnetic field, and sample deuteration allows the efficient measurement of site-specific exposure to mobile water and lipids. We demonstrate this site specificity on two membrane proteins, the human voltage dependent anion channel, and the alkane transporter AlkL from *Pseudomonas putida*. Transfer from lipids is observed selectively in the membrane

spanning region, and an average lipid-protein transfer rate of 6 s^{-1} was determined for residues protected from exchange. Transfer within the protein, as tracked in the ^{15}N - ^1H 2D plane, was estimated from initial rates and found to be in a similar range of about 8 to 15 s^{-1} for several resolved residues, explaining the site specificity.

2. Introduction

Cellular membranes are essential for maintaining homeostasis in an organism by separating cellular processes and controlling transmembrane signaling and the flow of metabolites. These processes are regulated by various proteins, and invariably involve integral membrane proteins that span the lipid bilayer as beta barrels in mitochondria and prokaryotes, or as alpha helices in both eukaryotes and prokaryotes. Since the lipid bilayers are necessary to maintain membrane protein structure and function, it is crucial to develop methods to investigate membrane proteins in such near-native environments (91-93). A key advantage of magic-angle spinning (MAS) NMR in this context, is the ability to study membrane proteins prepared in planar lipid bilayers at atomic resolution for investigation of structure and dynamics (79, 94, 95). In addition, the proximity of mobile species such as lipids and water, that are difficult if not impossible to capture via cryo electron microscopy or X-ray crystallography, can be probed with MAS NMR. In this way, it is possible to investigate solvent accessibility (96-98) or membrane association (99).

The environment of the protein surface can be probed by magnetization transfer between the protein and components of its environment, which is most commonly water, but in the case of membrane proteins is also lipids. While there is a choice of NMR active nucleus for the lipids, the more universal approach, and indeed the approach most commonly implemented in literature is to transfer magnetization by longitudinal proton-proton mixing. At physiological temperatures, water rapidly diffuses, which results in long proton T_2 relaxation times. In the case of lipids, anisotropic diffusion results in proton T_2 times of tens of ms (100), which is still at least an order of magnitude longer than for the protein under moderate spinning of $\sim 10 \text{ kHz}$. This allows the water or lipid signals to be selected based on differential relaxation times (101-103) or to distinguish multiple water pools with distinct relaxation properties within the same sample (98, 104).

Different primary mechanisms for water-protein magnetization transfer are possible under this scheme, including the nuclear Overhauser effect (NOE), chemical exchange, and

dipolar mediated spin diffusion. Since the NOE mechanism relies on cross relaxation of dipolar coupled spins, and requires correlated motion of the two coupled spins, it is not usually considered in solid-state NMR. However, Zilm and coworkers showed that water-protein transfer was consistent with NOE relaxation in the solid-state for a deuterated protein using H^N detection (105). The conclusion was drawn by comparison of magnetization transfer rates with and without recoupling, and in the rotating frame. Under different experimental conditions, Böckmann and coworkers found chemical exchange to be an important factor governing the buildup, which in their case was detected through cross polarization to carbon resonances in protein microcrystals and fibrils (97, 106). The fact that these carbon resonances are closer to exchangeable NH- and OH-protons in the sidechains may in part explain the differing contributions of NOE and chemical exchange in the two studies. In addition, details of the sample condition may also play a role, since the rate of chemical exchange can be reduced by decreasing the sample temperature and changing the pH. For site specificity, the possibility for dipolar mediated spin diffusion complicates the picture as compared with membrane protein studies in solution (107). Spin diffusion in solids is typically mediated by dipolar couplings, although diffusion of magnetization also occurs through relayed NOE transfers at long mixing times. In general, magnetization transfer within the protein can occur through spin diffusion and NOE, while transfer between the protein and the surrounding mobile water or lipids can take place through NOE. For water chemical exchange is also possible. For lipids, the exchange pathway is not possible, as the lipid molecules do not have exchangeable protons, and dipolar spin diffusion is reduced by considerable anisotropic motion. A summary of the magnetization transfer pathways is shown in figure 3.1.

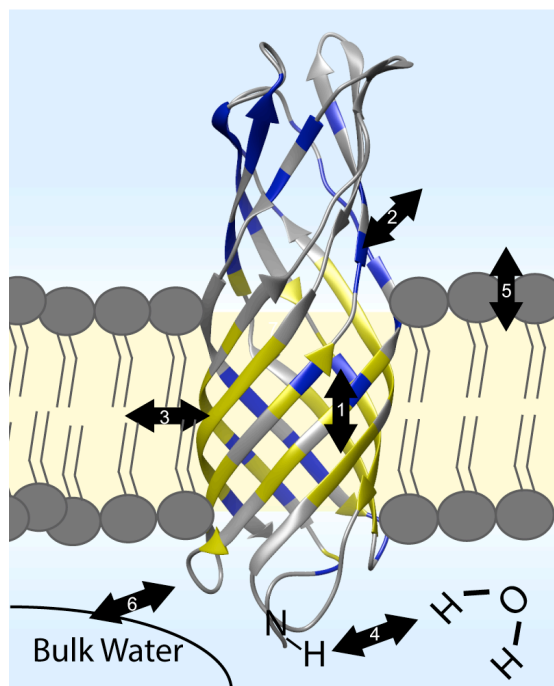


Figure 3.1. Schematic representation of longitudinal magnetization transfer in a lipid bilayer sample.

1-6 shows transfer between 1, protein and protein, 2, water and protein, 3, lipid and protein, 4, water and protein by chemical exchange, 5, lipid and water, and 6, bulk and associated water.

At low MAS rates and in fully protonated proteins, dipolar mediated spin diffusion is rapid, with efficient proton-proton transfer occurring in hundreds of microseconds (108). Typical NOE mixing times are orders of magnitude longer, since the NOE rates are lower. The result is that since the water-protein transfer is rate limiting, water magnetization is passed onto the protein primarily via fast exchanging groups and NOE, then spread quickly through the protein such that site specificity is obscured (101). Nevertheless, broader changes in the buildup of magnetization can be detected even at moderate MAS rates, for example, the reduction in exposed surface due to drug binding was evident in the influenza M2 protein (109). At high deuteration levels, high magnetic fields, and high spinning frequencies, the proton spectrum of the protein narrows as dipolar couplings are averaged at fast MAS (110), and the contribution of spin diffusion to magnetization transfer is reduced dramatically, allowing site resolved measurement of water proximity (97, 105, 106). We propose that this reduction in intra-protein transfer rates could be ideal for probing site specific lipid exposure, since the NOE transfer from the mobile lipid acyl chains could become similar to and potentially faster than the intra-protein transfer. Such site specificity would be consistent with previous measurement of specific water-protein transfer under similar conditions of deuteration and MAS rates, in which the mechanism was found to be mediated by chemical exchange (97, 106). In the case of transfer from the mobile aliphatic lipid chains, chemical

exchange is not possible, and we therefore are left with NOE as the dominant mechanism, although there may also be some residual transfer through spin diffusion.

Development of faster MAS frequencies, now applied in commercial probes at up to 111 kHz, have opened up new possibilities due to improved sensitivity, improved coherence lifetimes, and access to protons as the detection nucleus (30, 81, 111-116). Using ^1H detection, sensitivity is increased 8 and 32-fold on a per mole basis, with respect to ^{13}C and ^{15}N -detection. This assumes no changes in linewidth, which is reasonable with deuteration (116) and fast spinning. While fractional deuteration or smaller rotors capable of faster spinning reduces this gain somewhat, there is still an overall benefit in sensitivity with proton detection at 60 kHz MAS (42). Furthermore, the accessibility of a well-resolved proton dimension decreases degeneracy of chemical shifts and makes unambiguous assignments possible even on large proteins.

Early implementations for determining water accessibility rely on establishing polarization selectively on water, then transferring this polarization to the protein. Water can be selected using a T_2 filter (117, 118) or a selective excitation pulse (103). The T_2 filter uses a Hahn echo to select mobile molecules, as they can have much longer coherence times than rigid species such as protein. Subsequently, during a longitudinal mixing period, magnetization is transferred to the protein, and is then detected on ^{13}C or ^{15}N . In a bilayer, lipids are sufficiently mobile that the above mentioned strategies used for selection of water can also be applied to them (101). Alternatively, a spectral dimension can be used to distinguish mobile components from resonances of the protein (97, 99, 105, 106). This approach is effective for faster MAS rates, where there is less difference in protein and water or lipid T_2 relaxation times.

Here we apply fast MAS (55 kHz) in combination with perdeuteration and proton exchange for proton-detected measurement of water and lipids in contact with the protein. These conditions allow site-specificity within about 3 Å and determination of lipid insertion depth for membrane proteins, which is critical for evaluating the position of a membrane protein in its native environment, and thus for gaining a deeper understanding of its structural, functional, and evolutionary features. We demonstrate the applicability of the method on two integral membrane proteins, AlkL from *Pseudomonas putida*, and the human voltage-

dependent anion channel (hVDAC). Both proteins fold as membrane spanning beta barrels, with both lipid and water exposed surfaces.

3. Materials and methods

^2H , ^{13}C , ^{15}N -labeled AlkL was prepared as previously described (119, 120). The bilayer sample was reconstituted by addition of lipids to a 0.5 lipid to protein weight ratio and removal of detergent by dialysis. 2D crystalline ^2H , ^{13}C , ^{15}N -labeled hVDAC was prepared using established expression and refolding protocols (121-123). Preparation of 2D crystals was carried out using the protocol modified by Eddy et al. (65) from the original protocol published by Dolder et al. (64).

NMR experiments were performed on approximately 1 mg of sample packed into a 1.3 mm Bruker MAS rotor. All experiments were acquired on an 800 MHz Bruker Avance III spectrometer at a magnetic field of 19 T, and spinning at 55 kHz. The temperature of cooling gas was set to 250 K, which results in an estimated sample temperature of 20 °C. The HhNH and hNHH spectra were recorded by averaging 2 scans per point. Acquisition time in the direct dimension was 10 ms, in the indirect dimensions 22 ms on ^{15}N and 3.5 ms (HhNH) or 4.0 ms (hNHH) on ^1H . For ^1H to ^{15}N CP a contact time of 600 μs was used, with a linear ramp from 78 to 98 kHz on ^1H and constant irradiation of ^{15}N at 36.8 kHz. ^1H and ^{15}N RF field levels were determined from hard pulse calibrations. Cross-relaxation rates were determined from 3D spectra recorded with mixing times of 0, 5, 10, 15, 25, 50, 75, 100, 150, and 300 ms. A mixing time of 50 ms was chosen for the 4D spectra. For the HhnCANH spectra, 16 scans per point were averaged together and non-uniform sampling (NUS) was applied, selecting 1.98 % of the points. The sampling schedule was generated by Topspin 3.5, and the spectrum reconstructed using MDD in Topspin. The indirect acquisition times were 5.3 ms for ^1H , 10 ms for ^{13}C and 22 ms for ^{15}N . All experiments used waltz-16 heteronuclear decoupling at 10 kHz on both ^{13}C and ^{15}N . Protons were decoupled using 7 kHz TPPM decoupling during indirect acquisition times. The NOE mixing time was 50 ms. The recycle delay was 0.8 seconds. This resulted in a total experiment time of 6 to 7 hours for each mixing point of the HhNH, 7 to 8 hours for each mixing point of the hNHH, and 8 days for the 4D HhnCANH. Spectra were processed in Topspin 3.5 (Bruker) and analyzed in Sparky (124).

4. Results and discussion

Figure 3.2 shows the pulse sequences used to record 3- and 4-dimensional, z-mixing based proton detected spectra for determining site-resolved water and lipid proximities in integral membrane proteins. Figure 3.2A shows the 3-dimensional HhNH pulse sequence, which is only slightly modified from the hNHH sequence of Zilm and coworkers (105) by placing the z-mixing at the beginning rather than the end of the sequence. Proton excitation is followed by an indirect evolution period to encode the frequencies of water, lipids, and protein. Next, longitudinal mixing allows proton-proton transfer between mobile small molecules and the protein, via exchange and NOE. Magnetization is then transferred to ^{15}N , evolved, and transferred back to proton, where it is then detected. The out-and-back transfer to ^{15}N uses two short cross-polarization (CP) steps to ensure a selective transfer. The 3D experiment can be extended to 4 dimensions by including an additional transfer to CA. This HhNCANH sequence (Fig. 3.2B) proves useful when the HN spectrum is not sufficient to resolve signals from all residues. Similar in concept to the 3D, the out-and-back scheme in the 4D also directs transfer within one residue. The more sensitive HhCANH transfer scheme would allow longer range transfer during the ^1H - ^{13}C transfer step. Two different N-C-N out-and-back transfer schemes were considered, based on either CP (Fig. 3.2B) or transferred echo double resonance (TEDOR, Fig. 3.2C) (125). Similar to recent observations, the TEDOR transfer was found to be more efficient (126). Omitting ^1H - ^1H mixing, we found the CP based sequence to perform with 14% efficiency, as compared with 20% using TEDOR, both with respect to the hNH sequence.

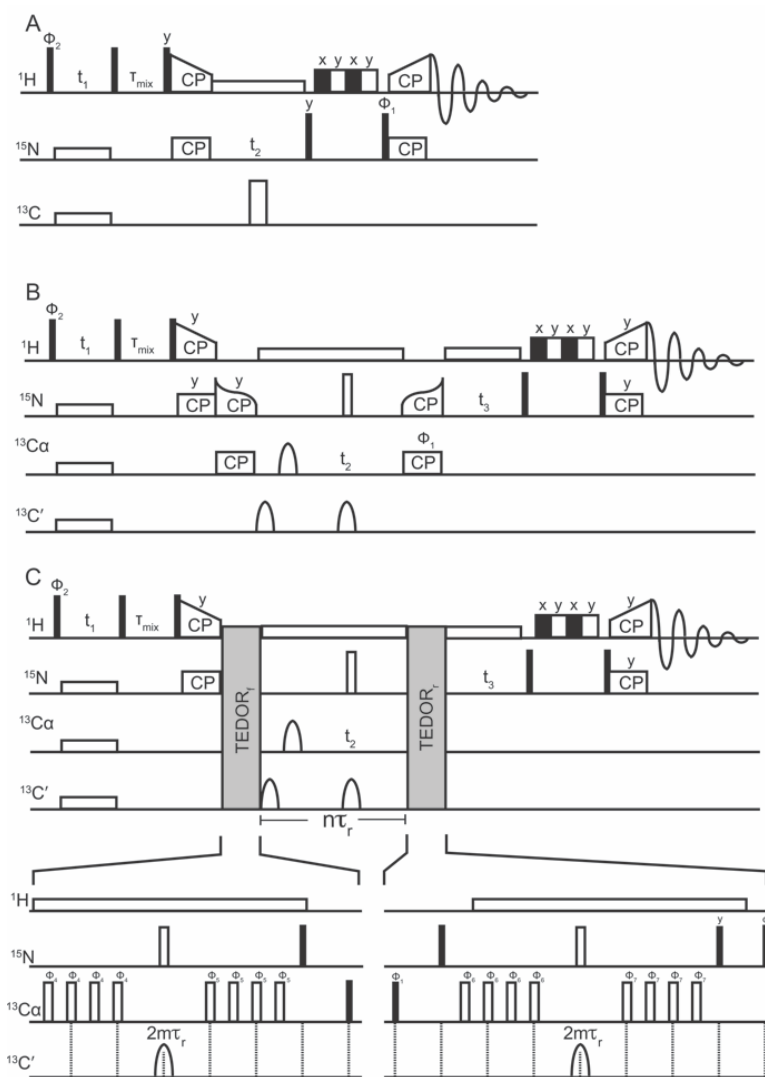


Figure 3.2. Pulse sequences used to record site resolved water and lipid in close proximity to membrane proteins.

In A, the 3D HhNH sequence. In B, the CP based HhNCANH sequence. In C, the TEDOR based HhNCANH sequence. Solid and empty rectangles represent 90 and 180 degree pulses, respectively. Empty half-ovals represent selective 180 degree pulses, Q3 pulses⁴³ in our implementation. Tangent ramped pulses were used on the ¹⁵N channel for ¹³C¹⁵N CP transfer. A selective pulse on the carbon channel during TEDOR prevents transfer to the carbonyl. 1.45 ms TEDOR transfer was used. Proton decoupling of 5 to 15 kHz WALTZ or swept TPPM decoupling was used during TEDOR, while 12 kHz swept TPPM was used during evolution periods. WALTZ decoupling on the carbon and nitrogen channels was set to 10 kHz. Rotor synchronization is indicated by vertical dashed lines. Phases were cycled, in A with $\phi_1 = y - y$, $\phi_2 = x x - x - x$, $\phi_{rec} = y - y - y y$, in B, $\phi_1 = x - x$, $\phi_2 = x x - x - x$, $\phi_3 = y y - y - y - y - y y$, $\phi_{rec} = y - y y - y - y - y - y y$, and in C, $\phi_1 = x - x$, $\phi_2 = x x - x - x$, $\phi_3 = x x x x - x - x - x - x$, $\phi_{rec} = x - x - x x - x x - x - x$. REDOR pulses on the carbon channel in C followed the xy-16 scheme. All other pulses were applied with phase x.

In all three pulse sequences, the ¹H-¹H mixing takes place at the beginning, preceded by evolution in an indirect proton dimension. This carries the advantage that the indirect sampling can be limited to reduce the overall data acquisition time. This results in lower resolution in the indirect proton dimension, which is sufficient as this dimension serves only to distinguish between lipids, water and protein resonances. In a perdeuterated protein, the

aliphatic proton resonances of the protein are suppressed, and lipid peaks of DMPC appear isolated at about 0.9, 1.3, and 3.3 ppm for acyl chain CH₃, acyl chain CH₂ and choline CH₃ moieties, respectively. The directly detected proton dimension, used for distinguishing between protein residues, can be sampled out as far as needed without additional cost in measurement time. Secondly, placing the NOE mixing directly before acquisition instead would allow water and lipid magnetization to build up during the mixing time, which results in problems with water suppression.

Figure 3.3 shows selected strips of the 3D spectra on AlkL and hVDAC at resolved amide sites, as well as the corresponding buildup of magnetization transferred from either lipid or water. Magnetization buildup increases on water and lipids until about 100 ms, however at the longer mixing times, magnetization may also be transferred to nearby protein residues, resulting in the loss of the experiment's specificity. Even without a detailed fitting of relaxation rates, it can be seen that the buildup of the inter-residue cross-peaks occurs with a similar rate as compared to lipid-protein and water-protein cross-peaks. This shows that spin diffusion within the protein is substantially suppressed by fast MAS and deuteration. Based on the relatively slow spin diffusion within the protein, a mixing time of 50 ms was chosen for both proteins and this value was used for the 4D spectra where more sites can be resolved. Site specificity is still imperfect under these conditions, since we do observe inter-residue transfer. However, as seen in figure 3.3A, we typically only observe transfer to the nearest proton, for example, in the cross-strand beta sheets of hVDAC. In this topology, the cross-strand distance is usually below 3.0 Å, while the next closest proton is above 4.3 Å away. For proton spin diffusion under fast MAS, or for proton driven spin diffusion, rate constants can be calculated from second order terms using perturbation or Floquet theory, and the transfer rate also becomes strongly chemical shift dependent (127, 128). In this case, since spin diffusion and NOE are second order effects involving two couplings, their rates scale down with the sixth power of distance, and we would therefore expect about an order of magnitude reduction in the rate of transfer to other protons. This explains why we typically observe one inter-residue cross-peak in a beta sheet topology, and the specificity is expected to be within about 3 Å.

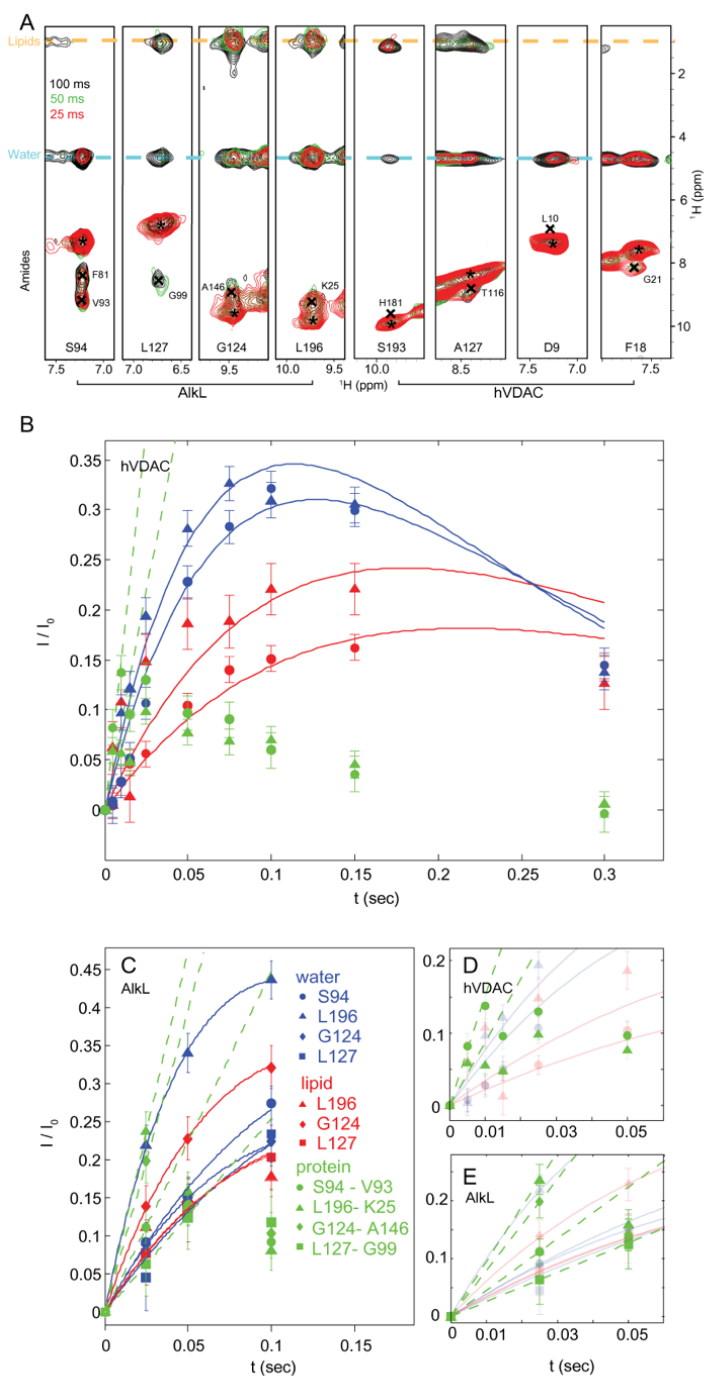


Figure 3.3. Evolution of magnetization in the HhNH spectrum of AlkL and hVDAC for selected residues.

In A, the proton-proton planes from isolated residues are shown for 25, 50, and 100 ms in red, green, and black, respectively. The evolution of cross-peak intensities in hVDAC in B shows the water-protein signal intensity for two residues shielded from lipids, D9 and F18, as well as the lipid-protein signal intensity for two lipid exposed residues, A127 and S193. Curves were fit with the analytical function described in the text. In C, the evolution of isolated peaks in AlkL are shown. In D and E, the initial buildup of protein-protein peaks is used to estimate the rate of spin diffusion within the protein. The intensities are scaled to the protein amide signal at zero mixing time, an approximation of the initial source polarization, $M_{0,m}$. The decay of these diagonal intensities are shown in figure S3.2. Error bars correspond to the root mean squared noise of the spectra.

An initial rate approximation shows transfer rates of 3-15 s^{-1} within the protein (green dashed lines in fig. 3.3). For hVDAC, these estimates are reasonably reliable, since the initial points are below the maximum in the curve. For AlkL, this is not the case, and the fits underestimate the true rates. Intra-protein rates of up to 15 s^{-1} represents a 2-3 orders of magnitude reduction in the deuterated sample at 55 kHz as compared with protonated samples at 4 - 5 kHz MAS (101, 109, 117). This comparison requires converting from the reported diffusion constants, D , to transfer rates, which can be done applying the relation $\Omega = D/a^2$,

where D is the diffusion coefficient and a is the spacing between spins in a lattice model (101, 117). A value of D of 0.8 nm²/ms was measured in rigid organic polymers (129). In biological polymers, a lower value of D of 0.3 to 0.4 nm²/ms has been applied under similar conditions of low spinning rates (101, 109, 117). At a lattice spacing of 2 to 2.4 Å, a spin diffusion rate constant of 0.3 nm²/ms, corresponds to a 5000-7500 s⁻¹ transfer rate for this lower spinning frequency in fully protonated samples. In deuterated proteins, protons are spaced by at least 3 Å. A very rough approximation for the expected proton-proton transfer rate at 5 kHz spinning can be calculated from a simple reduction by the 3rd power with distance, since the elementary transfer occurs via a first order process when the proton spectrum is broad and homogeneous due to ineffective averaging by MAS. This would result in an expected rate of about 2600-3800 s⁻¹. Clearly, both the fast spinning rate, as well as the deuteration, contribute to quenching of the intra-protein transfer.

While at the lower spinning frequencies, evolution of magnetization during longitudinal mixing can be modeled based on Fick's laws of diffusion (130), the rates of spin diffusion with deuteration and fast MAS are reduced and highly inhomogeneous. Under such conditions, acquisition of 2D proton-proton spectra and a matrix approach can be used to measure rates (100, 131). When spin diffusion is fast within the protein, a simplification based on an overall scaling by a single global longitudinal relaxation time (T_1) is sufficient to model the diffusive behavior and detect differences in overall exposure (103, 109). With fast MAS and deuteration, we found that a single T_1 did not fit the data and a more sophisticated model is needed. While a full rate matrix approach is clearly the gold standard for determination of cross-relaxation rates, protein resonances are woefully unresolved in 2D proton-proton spectra and improvement in resolution is afforded by filtering through a heteronucleus. This improvement in resolution in 3D spectra comes with the loss of the diagonal signals of the water and lipid species. We therefore chose to model the evolution of polarization, $M(t)$, with several assumptions about the relaxation matrix, starting from the assumption that the lipids and water may be treated separately. Then,

$$\frac{\partial M_p(t)}{\partial t} = (-R_{1,p} - R_{mp} - R_{pp})M_p + (R_{mp})M_m + (R_{pp})M_p \quad (3.1)$$

$$\frac{\partial M_m(t)}{\partial t} = \left(-R_{1,m} - \frac{1}{N} R_{mp} \right) M_m + \frac{1}{N} (R_{mp}) M_p \quad (3.2)$$

Where the subscript p is protein, and m is for the mobile species, either lipid or water. $R_{1,p}$ is the longitudinal relaxation rate of protein, and the cross relaxation between mobile species and protein is R_{mp} . R_{pp} refers to cross relaxation to all nearby protons of the protein. For simplicity, these other nearby protons are assumed to be in contact only with protein resonances. Lipids and water are assumed to equilibrate quickly among themselves, and N water or lipid protons are in contact with the amide resonance of the protein in fast equilibrium. Based on observed exponential decay of lipid signals, we can simplify to a single effective decay rate for lipid polarization, $R_{1,m_{eff}}$

$$\frac{\partial M_m(t)}{\partial t} = \left(-R_{1,m_{eff}} \right) M_m \quad (3.3)$$

We observed biexponential decay of water signal, and assume the protein is in contact primarily with the faster relaxing water population. This is consistent with observation of cross signals with only the broader water component in microcrystalline Crh samples under similar conditions (98). Then we can simplify the decay of water signal with a single rate, in the same way as for lipids.

We also chose to neglect magnetization transfer within the protein, since we observed moderate mixing of signals from R_{pp} , as indicated by the presence of relatively weak off-diagonal peaks in the amide region of the HhNH spectrum. We then find the solution,

$$M_p(t) = M_{0,m} \frac{R_{mp}}{R_{1,p} + R_{mp} + R_{1,m_{eff}}} \left(e^{-R_{1,m_{eff}}t} - e^{-(R_{1,p} + R_{mp})t} \right) \quad (3.4)$$

We applied this analytical solution to extract approximate cross relaxation rates in hVDAC. The scaling factor $M_{0,m}$ represents the initial polarization of the mobile species. The effective decay time for the lipids was set to the measured value of 5.3-5.5 s⁻¹. The effective decay time for water was allowed to vary between the two decay times observed in the biexponential decay, about 6 s⁻¹. The fit lipid- and water-protein transfer rates were found to be similar, with values that ranged from 3 to 12 s⁻¹ for the isolated peaks of hVDAC and AlkL shown in figure 3.3.

Since reliable fits require the correct scaling of the data, we took the initial protein amide peak intensity as an approximation for the initial lipid polarization. Due to differences in the recovery of lipids, water, and protein, this is estimated to introduce 10-20% error in the rates. To better focus on the lipids, we removed the influence of water by preparing a protein sample in D₂O. Many amides in the membrane-spanning portion of the protein do not exchange with buffer, and their bulk T₁ was found to be about 0.6 seconds, similar to the rate measured in H₂O. Since this measurement includes decay due to cross-relaxation to lipids (T₁~185 ms), we can safely assume that the protein T₁ is longer than the apparent decay of 0.6 seconds. Under this assumption, the long-time decay of magnetization tracks the decay of lipid polarization, which may further stabilize the fits.

As further verification of these rates, we recorded hNHH spectra, placing the z-mixing at the end of the sequence. In such spectra, cross-relaxation rates can be determined by measuring both the decay of starting signal on the protein, as well as the buildup on lipids (105). Tracking the decay of starting polarization allows the determination of the total initial intensity and thereby provides a more direct measure of the rates, as in the short mixing approximation or the full matrix approach based on proton-proton spectra (100, 131). Since imperfect water suppression severely affects hNHH spectra, we used the sample of hVDAC in D₂O and measured the cross-relaxation from protein to lipids. This has the further advantage that the measurement takes place in isolation from transfer to water, and from any chemical exchange process, for example from water or from exchangeable side-chain protons. The average protein-lipid cross-relaxation rate using this approach was 6 s⁻¹ (Fig S3.3). This rate is about an order of magnitude slower than the value of 90 s⁻¹ reported for the average water-amide cross-relaxation in microcrystalline ubiquitin (105).

Tracking the source polarization for rate determination is clearly preferred. However, this results in our case in detecting the water or lipid signal in the direct dimension, where there is loss of sensitivity due to imperfect solvent suppression and the resultant t_1 noise. There was still some t_1 noise evident near the lipid resonances in hNHH spectra of hVDAC in D₂O. We therefore still expect there will be utility in our approach using HhNH spectra, although one must be cautious about the assumptions made for the water, lipid, and amide proton T_1 relaxation times, as well as scaling the data by the initial protein signal which may introduce additional error if the initial polarization was not fully equilibrated. A more detailed investigation of cross-relaxation rates at different sites across the protein is left for the future.

Interestingly, analysis of average rates shows that cross relaxation to lipids or water can have a large influence on the measured longitudinal decay of amide proton magnetization. This can be appreciated by comparing the bulk decay of magnetization either after saturation of other protons, or starting with full signal on all protons. In hVDAC, the bulk ¹H T_1 of the amides as measured in these two ways is either 130 ms or 410 ms, respectively. The faster decay is due to cross relaxation to lipids and water and implies an average cross-relaxation rate similar to the 6 s⁻¹ determined for protein-lipid transfer. This observation has implications for the measurement of site-specific proton T_1 for dynamics or accessibility studies suggesting that a detailed analysis of the intrinsic protein T_1 must consider the effects of homonuclear magnetization transfer to water or lipids. On the other hand, solvent mediated PRE effects in samples with dissolved paramagnetic species still appear to highlight the exposed surface in accessibility studies; there was excellent agreement between the expected solvent exposed surface, and enhanced relaxation reported previously (132). The substantial cross-relaxation rates between protein and lipids or protein and water suggest that deuteration of water and lipids may be necessary to accurately measure site-specific proton T_1 even with protein deuteration and very fast MAS.

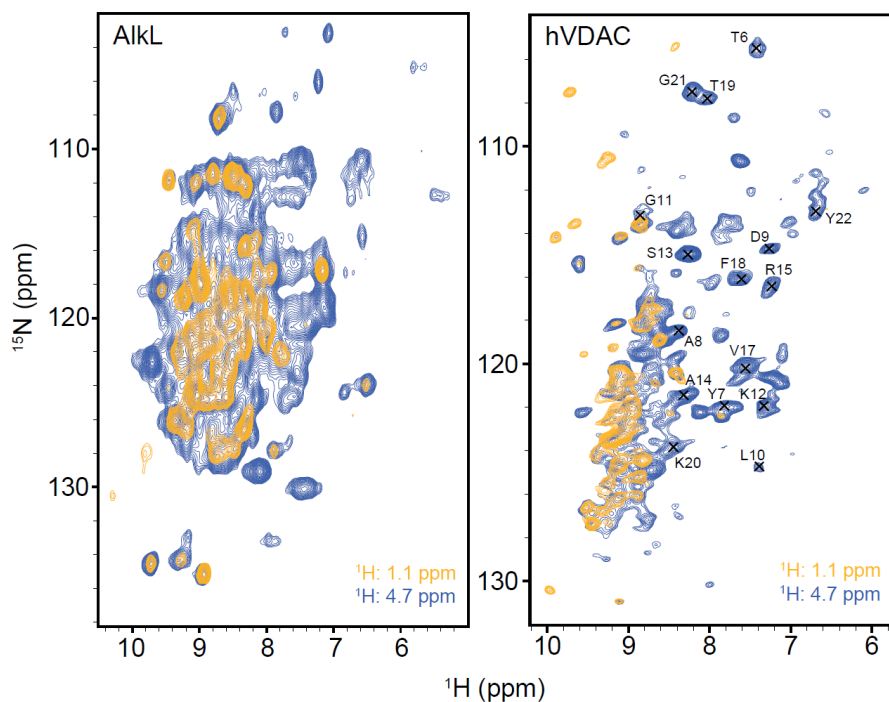


Figure 3.4. Lipid and water signal detected at the protein amides for AikL and hVDAC.

The 2D planes of the 3D HhNH spectrum are shown corresponding to the lipids (yellow, 1.1 ppm) and water (blue, 4.7 ppm). Assignments of the N-terminal helix are indicated for hVDAC. Proton-proton mixing proceeded for 50 ms.

Figure 3.4 shows 2D planes from the HhNH spectrum at the chemical shifts of both water (blue) and at the strongest signals of the lipid acyl chains (yellow). In hVDAC, although the amides are not fully resolved in an HN 2D plane, the helical residues of the N-terminus are well-separated from the more abundant beta sheet signals. Based on previously published assignments (79) these helical residues can be identified unambiguously, with the exception of D16. All of these show strong correlations with water, and no cross-peak with lipids. This is fully in accordance with the known 19-stranded beta barrel structure of the channel determined in detergent micelles (123), as well as the placement of the helix within the large water-filled pore in crystals and in micelles (122, 133, 134). In total, 53 correlations with water and 42 with lipids can be observed, however only approximately 19 water correlations and 17 lipid correlations can be assigned with certainty due to extensive signal overlap in the HN plane.

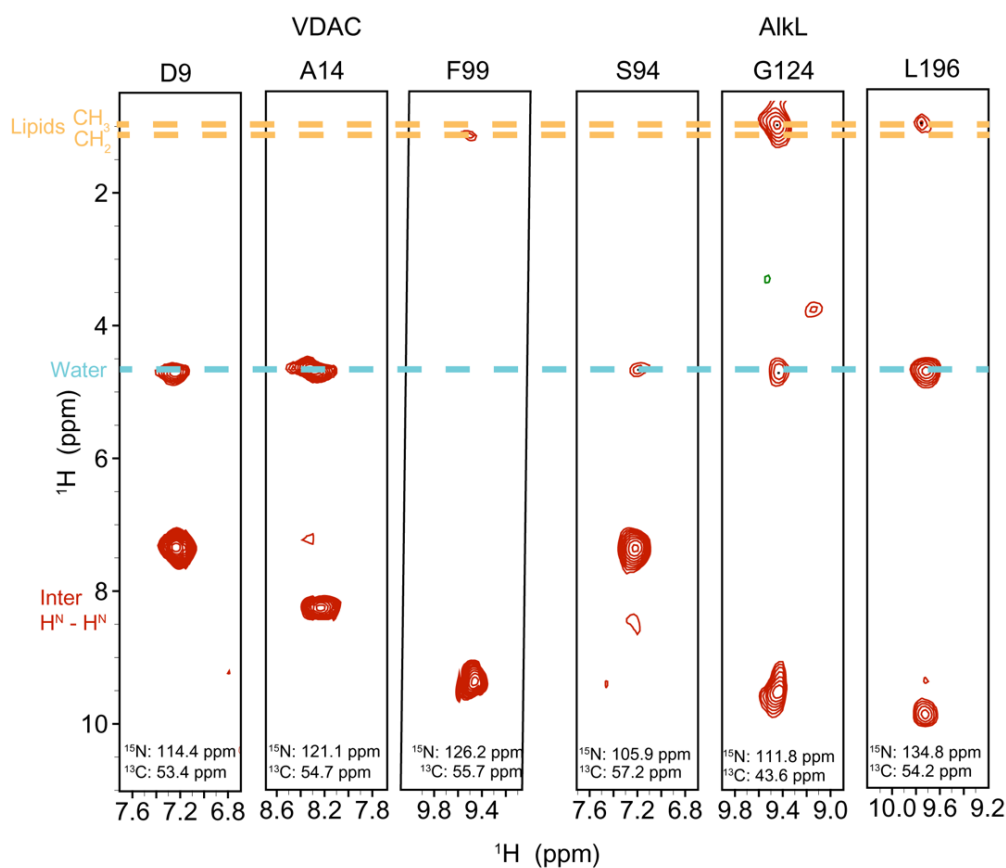


Figure 3.5. Selected strips from the 4D spectra of hVDAC and AlkL using 50 ms proton-proton mixing. The position of water and lipids are indicated by blue and yellow dashed lines, respectively. In most slices, only a single intense autocorrelation peak is observed in the amide region, indicating minimal magnetization transfer within the protein.

Site-resolved correlations between the protein and the surrounding small molecules can be obtained in the HhnCANH spectrum. Figure 3.5 displays strip plots from spectra acquired on both AlkL and hVDAC showing correlations between representative protein residues and lipids and/or water. In the case of lipid correlations, two separate shifts are observable, one at 1.3 ppm and one at 0.9 ppm, corresponding to the lipid CH_2 and CH_3 groups, respectively. For hVDAC we see 24 lipid correlations and 53 water correlations in total. Protein-lipid cross peaks tend to be weaker and therefore more difficult to observe. The fact that water-protein cross-peaks are stronger may be due to differences in mobility and structure, but are unlikely to arise from exchange pathways, considering that no significant correlation with exchangeable side-chains could be found. (Fig. S3.1) Despite the reduced sensitivity, the clear advantage of the 4D spectrum is that assignments are almost completely unambiguous, due to the resolution afforded by 3 spectral dimensions, ^{13}C , ^{15}N , and ^1H .

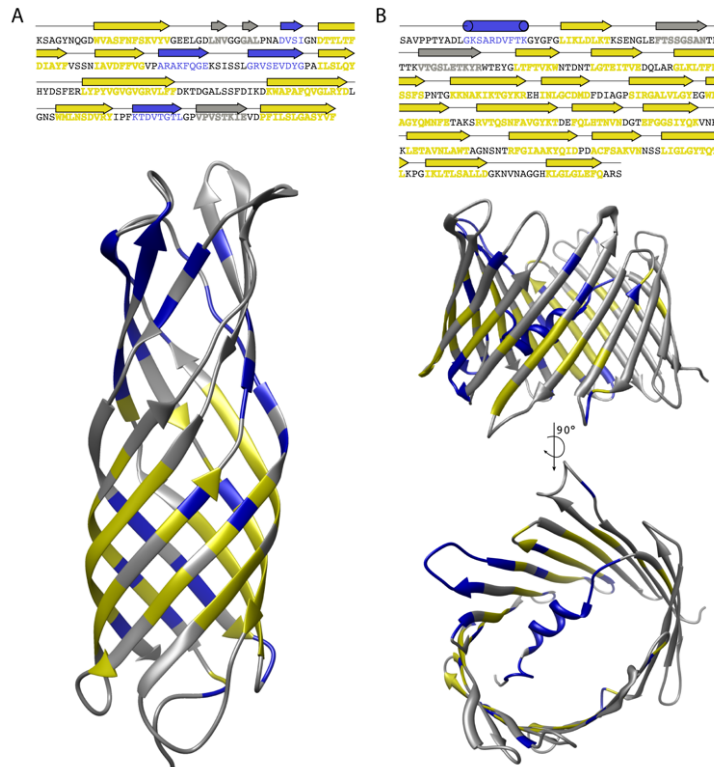


Figure 3.6. Water and lipid contacts shown on the homology model of AlkL using OmpW (PDB 2F1T) as a template (A) and the solution NMR structure (PDB 5JDP) of hVDAC (B).

All residues that show a lipid contact are colored yellow, while residues that show only water contact are colored blue, and residues for which no contact could be observed or assigned are colored in grey. The topology (by homology in a) and sequence are indicated above each structure.

Figure 3.6 shows water- and lipid-protein correlations mapped onto the OmpW homology model of AlkL, generated with swissmodel server (135) and the solution NMR structure of hVDAC (134). For the residue mapping we combined information obtained from the well-resolved HhnCANH and the high-sensitivity HhNH experiments and used contacts observed in both spectra. Homologs of AlkL (136-139) are composed of a transmembrane beta barrel, with long extracellular loops that are either flexible, or form a structured beta barrel extension, depending on the sample conditions, and the homolog. While the transmembrane beta barrel of homologous proteins is embedded in lipids, it is unclear from homology if the extracellular loops are in contact with lipids, are flexible, or are structured. At mixing times of 50 ms, we observed only water contacts from the extracellular loops, while the transmembrane beta barrel showed efficient contact with both the lipids and water. This is in agreement with the homology model based on OmpW, where the lower barrel is embedded in the membrane, but the extracellular loops form a barrel extension that is water-exposed. On the other hand, at longer mixing times of 100 ms, residues of the loops also show correlations with lipids. Further structural analysis of AlkL, as well as a more detailed investigation of

rates would be needed to understand the flexibility of loop residues, and their impact on lipid-protein transfer.

For hVDAC, combination of both 3D and 4D data (Fig. 3.6) provides a more complete picture of water and lipid accessibility. As already seen in the 3D data, in hVDAC, the helical residues are clearly water-exposed and do not show any correlations to lipids. This is in accordance with the known structure of the protein, as the N-terminal α -helix is positioned horizontally in a large, water-filled pore. Correlations to water can also be observed in the β -barrel, both in loop regions, and also with some residues in β -strands. Such loop residues with polar sidechains are energetically favored outside the membrane, but reside close to the lipids. That we do not observe lipid signal for some loop residues is therefore a testament to the high degree of selectivity possible under fast MAS, high magnetic field, and with deuteration. Protein-water proximities in β -strands are also not surprising, due to the barrel's large, water-filled pore. Lipid-protein contacts are only observed for residues in β -strands as well as in three loop residues, close to the end of β -strands (W64, Y146 and S240).

Insertion of membrane proteins in lipid bilayers has also been investigated with paramagnetic relaxation enhancement (PRE) measurements in solution (140). Whereas PRE has also been used for defining protein-protein (132) or water-protein (141) contacts in microcrystals, its potential for membrane insertion has not been exploited in solids. A drawback of the PRE technique is its indirect nature, as protein-lipid or protein-water contacts are not observed, but instead inferred due to an enhancement in relaxation where the solvent comes into contact with the protein. The method also may be less effective in larger, water accessible channels, such as hVDAC, since relaxation in the transmembrane part of the protein could be enhanced from the pore.

5. Conclusion

In conclusion, we have found that the combination of fast MAS, high magnetic field, and protein perdeuteration allows site specific detection of mobile species at the protein surface. We used the exchangeable amide proton as the detection nucleus, however, we anticipate that this methodology will be implemented in the future with different labeling schemes, for example, with methyl, or alpha proton labeling. The site specificity, as well as discrimination of water and lipid signals in a single spectrum, was used to probe the insertion of membrane proteins in lipid bilayers. We anticipate that the method presented here will

strengthen the role of solid-state NMR in structural biology. It can be extended in a straightforward manner to test the influence of lipid compositions, membrane curvature and protein aggregation, thus allowing an in-depth characterization of membrane proteins that cannot easily be studied by other techniques.

Acknowledgement

We acknowledge financial support from the MPI for Biophysical Chemistry, and from the Deutsche Forschungsgemeinschaft (Emmy Noether program grant AN1316/1-1, SFB803 grant INST 186/794-3).

SUPPLEMENTARY

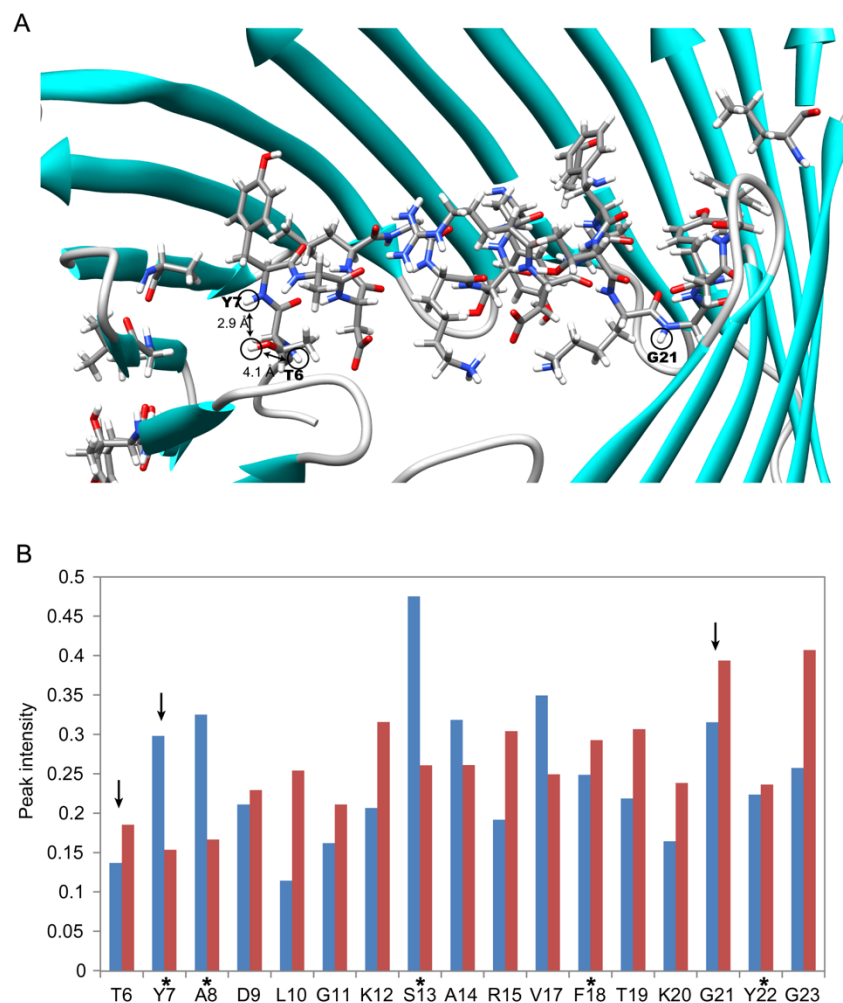


Figure S3.1. Proximity of exchanging residues has little or no effect on the observed buildup of the cross peak between a mobile species and an amide proton.

A) displays the water-exposed N-terminal α -helix of VDAC (pdb: 3emn). B) shows the intensity of water-amide cross peaks for the α -helix at 50 ms mixing time, either as raw intensity (blue), or scaled with the amide peaks at 0 ms mixing time to remove differences in HN CP efficiency (red). Residues marked with * are somewhat overlapped with another residue in the HN plane. Although T6 and Y7 have the exchangeable T6 sidechain in their close proximities (4.1 Å and 2.9 Å, respectively), the cross peak intensity to the water is low. G21 has no exchangeable OH- or NH- sidechains in its vicinity, however the residue shows a strong correlation to water.

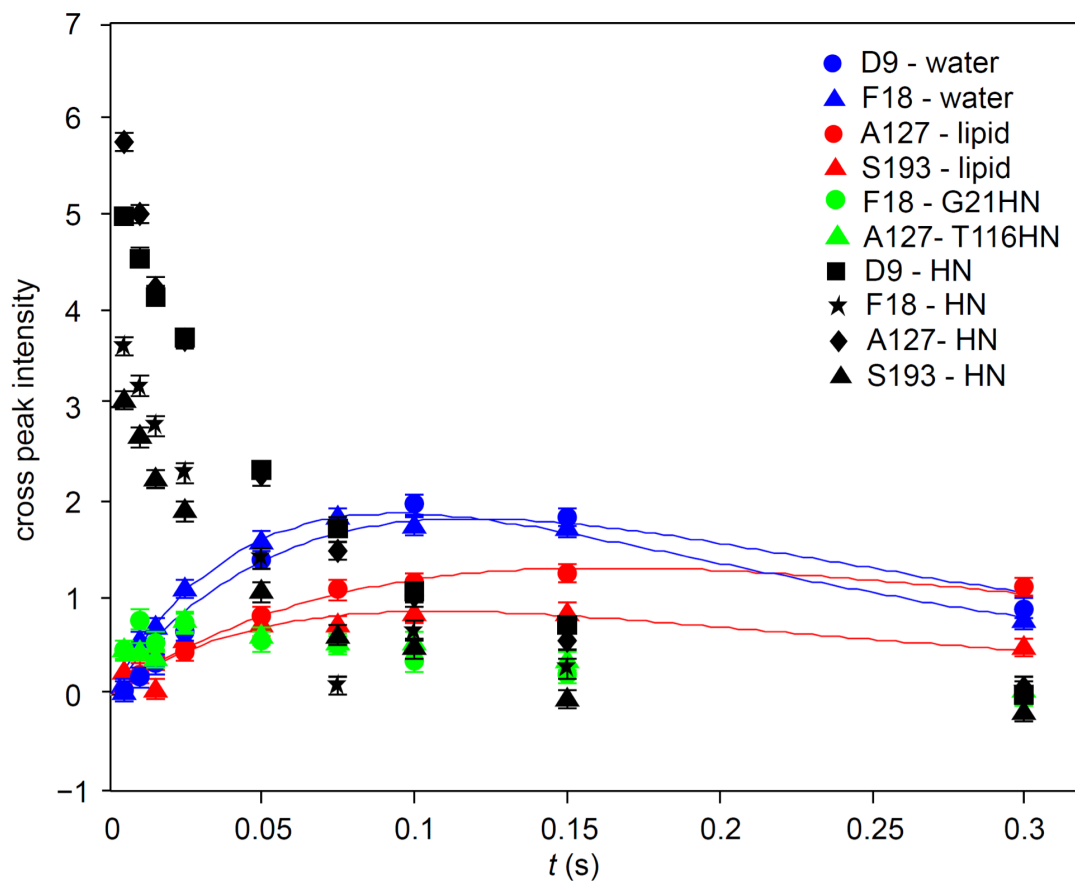


Figure S3.2. Cross-peak intensities for backbone amides, lipids, and water.

The starting polarization (black) decays as signal builds up on other nearby amides (green), lipids (blue) and water (red). The protein cross-peaks appear as a relatively small perturbation on the water-protein or lipid-protein exchange, justifying our choice to neglect them in the fits. The rate of transfer within the protein results in equilibration of the spin pair within about 100 ms. The higher intensity of lipids and water at long times represents the fact that there are many protons of water or lipids, as compared to the single amide proton of the protein.

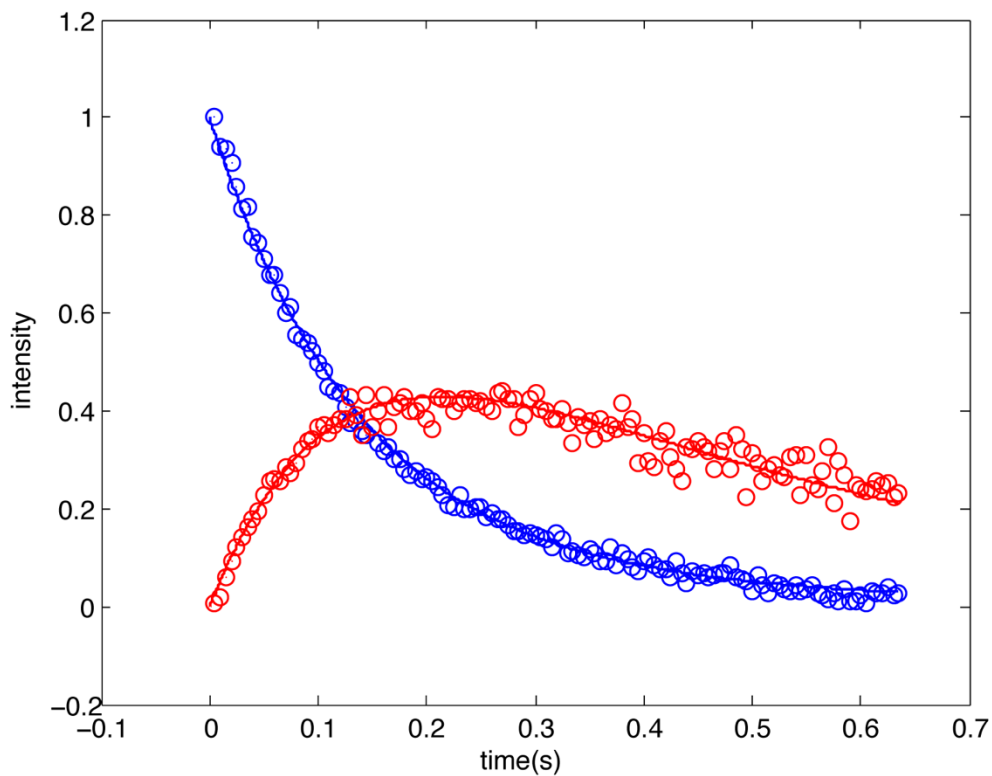


Figure S3.3. Measurement of the average protein-lipid cross-relaxation rate in hVDAC by tracking the initial protein signal decay (blue) and the buildup on lipid resonances (blue).

We used the hNHH sequence to generate initial polarization on amide protons only. The sample was exchanged in D₂O to eliminate the effects of water and exchangeable protons. The transfer is about 20 percent efficient on average at 50 ms.

Chapter 4

Structural and functional characterization of Mic10 from the MICOS complex

1. INTRODUCTION

1.1 CELLULAR ORGANIZATION

Evolution has led cells to evolve into a complex and optimized system. Like factories, cells have built a network system for inter- and intracellular communication and signal transfer. In this complex biological system, compartmentalization and signaling have become essential for performing and controlling complex biochemical processes in an optimal manner (142, 143). Cellular signaling, both inter- and intracellular, has become an indispensable property to fine-tune gene and protein expression in response to metabolic stresses (144).

Cellular compartments are usually separated by lipid bilayers and tightly regulated by an active and dynamic intercommunication of proteins, lipids and other factors between different cellular compartments (Fig. 4.1). Among these membrane enclosed ultrastructures, referred to as organelles, is the lysosome, which is well characterized by its low internal pH ($\text{pH} < 5$) and its role in degradation of macromolecules among others (145) and the endoplasmic reticulum (ER) that is involved in many roles including molecular trafficking as well as lipid and protein synthesis (146). There are also organelles devoid of lipid membranes referred to as “membraneless organelles” (147). Membraneless organelles are mostly constituted by protein condensates and are formed via liquid-liquid phase-separation (147).

Each cellular compartment has specialized cellular functions emphasizing the importance to unveil the biophysical properties of the proteins involved in their regulation. Cellular homeostasis, including spatiotemporal regulation of the different biochemical processes has to be maintained and fine-tuned to fit arising needs, which by consequence will demand a significant energetic consumption. In the cell aerobic respiration provides the energy that is mostly stored as ATP (148). The latter is generated by the F_1F_0 -ATP synthase of the oxidative phosphorylation complex (OXPHOS) in a process commonly known as “cellular respiration”. OXPHOS is localized in the mitochondria (149), more precisely, at the

cristae membranes (CM), which are an elongated tube-like structure observed in the inner membranes (150). Over the years, mitochondria have attracted the researcher's interest because of their central function in supplying the cell with ATP, and more recently through the discovery of additional functions, as detailed below.

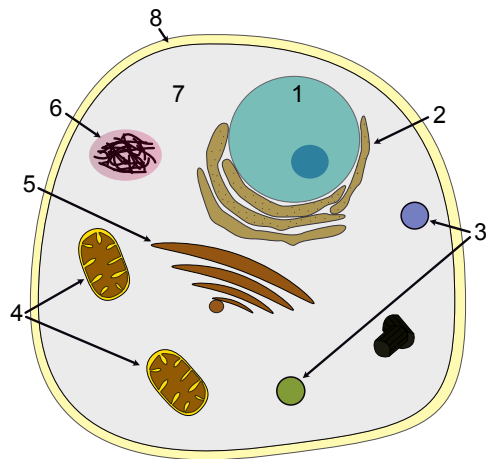


Figure 4.1. Schematic representation of an eukaryotic cell.

The numbers from 1 to 8 represent the different sub-structures found in a eukaryotic cell. 1, nucleus; 2, endoplasmic reticulum (ER); 3, lysosomes and endosomes (single membrane organelles); 4, mitochondria; 5, golgi apparatus; 6, droplet (liquid-liquid phase separation); 7, cytoplasm; 8, membrane bilayer.

1.2 MITOCHONDRIA

Mitochondria are highly dynamic organelles, which actively endure fusion and fission processes in a cellular metabolism-dependent manner (151). This organelle is often referred to as the “powerhouse of the cell”, as was first reported by Albert von Kolliker in 1857. In 1898, Carl Benda named these organelles, mitochondria, for the first time by reference to the observed shape (μίτος (mitos), thread and χονδρίον (chondrion), small grain). Despite the fact that mitochondria do not have an autonomous life-cycle and are instead highly dependent on their cellular hosts for survival, they contain their own DNA referred as mitochondrial DNA (mtDNA). MtDNA only encodes for 13% of the mitochondrial proteome including the proteins involved in the respiratory pathway (152). Mitochondrial phenotype can be highly dynamic and has been related to the metabolic state of the cell (153). Earlier studies have shown the involvement of mitochondria in energy production (154), by the so-called cellular respiration. However, now it is well established that mitochondrial functions are far more extensive than just energy generation. It has taken more than a century to reveal the complex role of mitochondria in cellular metabolism (155). Besides ATP synthesis, mitochondria have been found to be involved in metabolic pathways, such as amino acid metabolism, lipid synthesis (156), differentiation and even programmed cell death, known as apoptosis (157). Due to their wide range of functions in cell metabolism, it is not surprising that aberrant

mitochondria function has been related to severe human diseases, such as neurodegenerative or myopathy disorders (158-162).

Phenotypically, the mitochondrion is a unique organelle in the eukaryotic cell with a double lipid membrane: the outer mitochondrial membrane (OMM) and the inner mitochondrial membrane (IMM). Both membranes have their own structural characteristics, as well as functionally relevant differences (Fig. 4.2). The OMM provides the first barrier between the cytoplasm and the mitochondrial interior. Although it serves as a barrier, ions and small molecules can freely diffuse through the porins encountered. While, the IMM is a sealed membrane with a voltage across it, which has a particular architecture forming elongated protuberances called cristae. This gives to the IMM a larger membrane surface area and allows for the formation of sub-compartments that are used for specific functions.

1.2.1 Mitochondrial membrane

In order to understand the morphology and formation of cristae, a brief overview of mitochondrial membrane composition and function is needed. Both IMM and OMM have an asymmetric lipid composition. Lipids play important roles in protein activation, folding and signaling (163, 164). Both membranes are essentially composed of phosphatidylcholine and phosphatidylethanolamine lipids (>50%) (165). Interestingly, IMM is enriched in cardiolipin (CL), which has been recently shown to actively participate in the regulation of different metabolic pathways, such as lipid synthesis or in the mitochondrial cristae organization system (MICOS) by acting as a modulator of protein interactions (166, 167).

Just like the lipid composition, the protein composition differs between the OMM and IMM as well. The OMM provides a “sealed“ environment for mitochondria, ensuring the protection for both the cell and the mitochondria concomitantly. While the OMM is largely populated by beta barrel proteins referred as outer membrane proteins (OMPs, figure 4.2) involved in transport and protein folding, the IMM presents a large majority of helical proteins. Among the beta barrel OMM proteins, there is the translocase of the outer membrane (TOM), which has a role in protein import from the cytosol (168), and the voltage dependent anionic channel (VDAC), which has been shown to have a variety of functions including ion conduction and apoptosis regulation (169).

Proteins of the IMM, referred as inner membrane proteins (IMPs, figure 4.2), play important roles for a wide range of cellular functions, such as aerobic respiration, lipid metabolism or even programmed cell death by regulating the release of cytochrome *c* from the cristae lumen (170). Interestingly, many seminal works have shown striking evidence for the importance of the IMM shape in functional compartmentalization efficiency. Indeed, the OXPHOS is preferentially localized at the rims of the CM regions. OXPHOS makes use of an electrochemical and pH gradient across the IMM in order to activate the ATP synthase, which generates ATP from ADP (171-174). In the early 2000s, Strauss et al. showed that dimerization of the F_1F_0 -ATP synthase promotes the curvature at the CM rims (ATPase, figure 4.2), suggesting that the curvature is a key feature to generate the pH gradient across the CM (174). An electric gradient simulation indicated that the membrane curvature increases the charge density at the apex and that consequently increases the concentration of protons in this region. As a result, it promotes efficient ATPase activity at the CM rims (174).

Historically, in addition to its peculiar phenotype, the presence of circular DNA and the membrane composition of mitochondria have led scientists to speculate on a potential earlier event of bacterial-cell endosymbiosis (175, 176). Recent *in silico* data suggest that the mitochondria might have evolved from alpha-proteobacteria, which were engulfed by a host cell through an earlier endosymbiosis event (177). The difference in membrane architecture, lipid and protein composition, between the two membranes, along with the specialized and localized functions of the IMM and OMM, reflects the high degree of regulation demanded to maintain a well-functioning mitochondria-cell symbiosis.

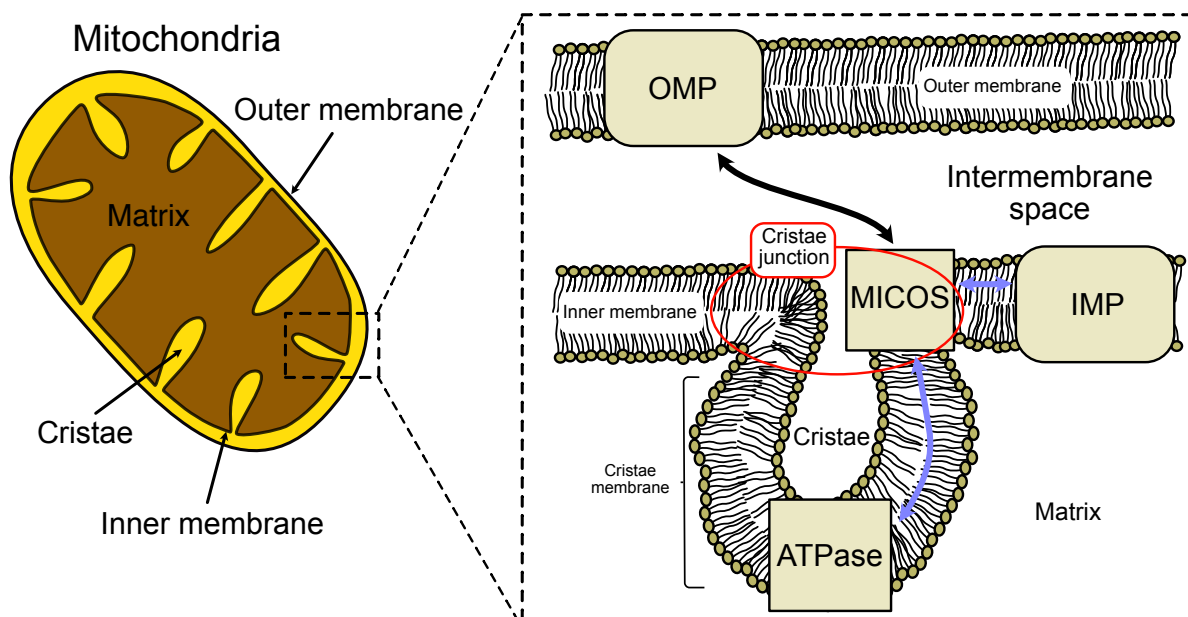


Figure 4.2. Mitochondrial membrane architecture and composition.

The double membrane of mitochondria consists of two lipid bilayers. Within the membranes, different proteins can be found: Outer membrane proteins (OMP) in the outer membrane and inner membrane proteins (IMP) in the inner membrane. Specific sub-structure located in the inner membrane are inner boundary membrane (IBM), cristae and cristae junctions (CJ). At the CJ, we find the MICOS complex. The arrows represent the network of MICOS to protein contacts. The respective roles of this membrane sites are explained in the main text.

1.2.2 The inner mitochondrial membrane (IMM) and cristae membrane (CM)

Cristae membranes (CMs) are part of a larger architectural arrangement of the IMM that has three distinct membrane sub-structures with specialized functions: the inner boundary membrane (IBM) is a mostly flat surface that lies parallel to the OMM (178, 179); the CM are tubular protuberances or invaginations and the cristae junctions (CJs) are tubule-shaped membranes forming the junction between IBM and CM (red circle, figure 4.2).

Interestingly, mitochondrial membrane sub-structures are highly dependent on the metabolic state of the cell, which changes the protein distribution across the membranes. *In vivo* fluorescence data using immunoelectron-microscopy (180) and stimulated emission depletion microscopy have been applied to monitor the localization of mitochondrial proteins depending on the cellular activity and showed that the respiratory complex is found in the CM (181). Similarly, Korsmeyer S.J. and co-workers showed that there is a concomitant redistribution of cytochrome *c* with an opening of the CM and a re-structuring of IMM (182, 183). Other proteins like the translocases of the inner mitochondrial membrane (TIM) and the sorting and assembly machinery (SAM) co-localized at the IBM, consistent with the close

contact observed between IMM and OMM at these regions. Further evidence of this close contact between IBM and OMM are provided by the complex formed by TIM and TOM involved in IMM protein import (184). Similarly, the interaction between SAM and Mic60, an IMM protein part of the MICOS complex, has been shown to be crucial in mitochondrial calcium and lipid uptake (185). Indeed, the spatial rearrangement of these proteins depending on the cellular metabolic state and their role provides strong evidence for the importance of compartmentalization in the regulation of biochemical processes.

CMs have been proposed to play a role in compartmentalization and functional specialization. Fuhrmann et al. and others have shown that the phenotype of CMs depend on the cellular energetic demands and oxygen availability (186, 187). The CM is maintained by CJs, which are tight tubular structures of about 12 to 40 nm of diameter (173), regulating proteins and metabolites trafficking between the inter-membrane space and the cristae lumenal compartments. However, mitochondria are dynamic organelles that undergo many dynamical processes, such as fission and fusion in response to external factors sensed by the cell (186). These processes are regulated by a set of proteins of the dynamic GTPase family (151). The mitochondria restructuring has to occur concomitantly with a complete remodeling of the IMM. Since mitochondria have an essential role as the cellular energy generator, such mechanisms have to be highly regulated to maintain an optimal, efficient and functional structure. How do mitochondria regulate these processes? And how is the membrane shaping controlled and regulated? These are processes that are not completely understood yet.

The formation of CM requires a high degree of membrane curvature, which can only occur under energetically favorable conditions. This has been attributed to certain membrane proteins found in the IMM like F_1F_0 -ATP synthase and Opa1/Mgm1 (188). Membrane curvature by F_1F_0 -ATP synthase is generated by the dimerization of the subunit e. Knockout (172) or mutation of G to L in the GxxxG motif of subunit e lead to a large disruption of IMM morphology forming so called “onion-like” structures (189) suggesting that this motif is responsible for the dimerization of F_1F_0 -ATP. More than a decade ago, several studies identified an important complex affecting the CJ formation. The complex was localized in the CJs and initially termed MITOS, then MINOS and finally, in 2014, it was termed, with a universal nomenclature, as mitochondrial contact site and cristae organizing system (MICOS) (178, 179, 190, 191) (Fig. 4.3).

1.3 MITOCHONDRIAL CONTACT SITE AND CRISTAE ORGANIZING SYSTEM (MICOS)

MICOS is a multi protein complex where certain proteins have been evolutionarily conserved through different species and kingdoms, supporting the endosymbiotic origin of mitochondria (177). In yeast, two main sub-complexes of MICOS have been identified so far, Mic10 and Mic60. The Mic10 sub-complex has four sub-units namely Mic10, Mic12, Mic27 and Mic26, where the number is an indication of the approximate molecular weight. The second sub-complex, Mic60, has only two subunits: Mic60 and Mic19 (Fig. 4.3) (192, 193). In order to understand the role of MICOS in CM formation many biochemical studies have been performed. In yeast, depletion of either of the core components, Mic10 or Mic60, strongly impacts other sub-units and showed severe effects on CM phenotype (193, 194).

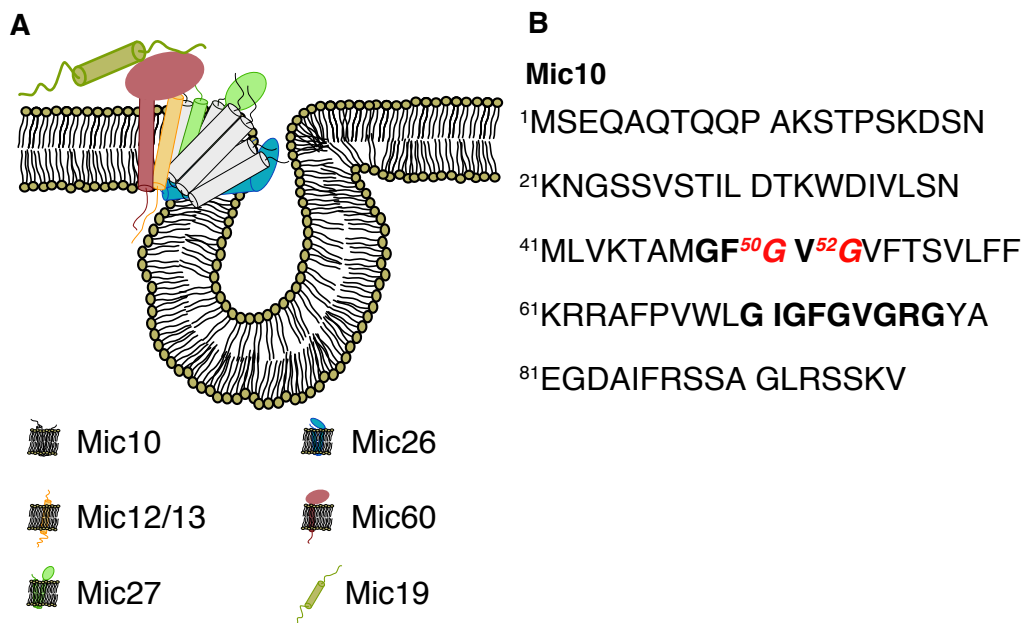


Figure 4.3. Schematic representation of the sub-complexes of the MICOS found at the cristae junction and Mic10 amino acid sequence.

Panel A shows the MICOS complex, which is composed of two main sub-complexes, Mic10 and Mic60. In yeast, the Mic10 sub-complex is subdivided in four sub-units Mic27, Mic26, Mic10 and Mic12. And, the Mic60 sub-complex composed of two sub-units only: Mic60 and Mic19 (the only soluble proteins found in the MICOS complex). The number indicates the approximate molecular weight of the protein. The subunits shown in the graph are those found in yeast. Panel B shows the amino acid sequence of Mic10 from *Saccharomyces cerevisiae*.

1.3.1 Mic60

Mic60 homologs have been found in the membranes of oldest known organisms presenting a cristae-like membrane shape in α -proteobacteria (195). Knockdown of this protein leads to a disruption of CJ structures and detachments of the IMM from the OMM,

inducing sheet-like-structures or stacks of IMM within the mitochondria (192, 193, 196, 197). Mic60 is a 60 kDa protein containing a pre-sequence at the N-terminus which targets the protein to the IMM (198). Interestingly, recent biochemical data showed that the soluble domain of Mic60 retains the ability to induce membrane curvature (198). The IMM phenotype has partly been related to Mic19 as well, due to its role in linking Mic60 to SAM, which is in part responsible of the formation and stabilization of IBM (199). Thus, losing Mic60, which interacts with OMM proteins such as TOM and VDAC, is accompanied by a detachment of IMM from the OMM and leads to the formation of IMM stacks. In fact, all the subunits found in the MICOS complex have at least one helical transmembrane domain, with the one exception being Mic19, as it is found soluble in the inter mitochondrial space (200, 201)(Fig 4.3). It has been recently shown that Mic19 interacts with an amphipathic helix of Mic60 and concomitantly interacts with the outer membrane protein Sam (199). In addition, Zhiyin Song and co-workers proposed Mic60 sub-complex as the hub and master regulator of MICOS location by showing that Sam-Mic19-Mic60 interaction can be regulated by OMA1, a mitochondrial protease (202). The cleavage of Mic19 by OMA1 results on a disruption of the Sam-Mic19-Mic60 complex leading to an abnormal mitochondrial architecture. Despite its key role on mitochondrial membrane architecture, there is no high resolution structural data available yet.

1.3.2 Mic10

Similarly, Mic10 is the smallest subunit of MICOS with a molecular size of ~10 kDa that contains two transmembrane helices (194). Proteomic studies as well as *in cell* and *in vitro* studies have shown the presence of multiple copies of Mic10 in the MICOS complex. Independent studies showed that Mic10 has a high propensity to homo-oligomerize (194, 203). Mic10 knockout studies in yeast showed an alteration of the inner mitochondrial membrane architecture (194, 203). Complementary studies on Mic10 sub-complexes have shown the interdependence of the different components on the maintenance and stability of MICOS. Furthermore, elegant studies using co-immunoprecipitation in depleted Mic27 or Mic26 yeast cells have shown that these proteins can either promote or inhibit Mic10 oligomerization (204). Whereas Mic27 depletion results in Mic10 oligomer destabilization, Mic26 knockout has the opposite effect, enhancing the size of Mic10 oligomers. In the same study, Nikolaus and co-workers showed the stabilizing effects of cardiolipin on Mic10

oligomers and that mutations in Mic26 and Mic27 have a cooperative effect on Mic10 oligomer destabilization (204). However, depletion of any other MICOS sub-unit, except for the two core components Mic10 or Mic60, have not been related to a severe impact on the IMM phenotype.

A closer look into the amino acid sequence of Mic10 (Fig. 4.3B) shows the presence of several GxxxG motifs in both transmembrane domains. GxxxG motifs are well known to promote helix-helix interactions in transmembrane domains (205). A well-known example is the dimerization of F₁F₀-ATPase, which is promoted by the GxxxG motif found in the transmembrane domains of sub-unit e in yeast (172, 206, 207). Martin van der Lann and co-workers used point mutation studies to show that GxxxG motifs are indispensable for Mic10 homo-oligomerization but do not participate in the interaction between Mic10 and the other MICOS subunits (203). Combining single and double mutation of glycine to alanine, Meinecke and co-workers were able to identify two glycine residues, which are part of two independent GxxxG motifs, at position 50 and 52, responsible for Mic10 homo-oligomerization (194). In the same work, they showed that Mic10 alone is able to induce membrane curvature in liposome preparations but that the double mutant loses the ability to induce membrane curvature (194). Interestingly, G50 and G52 are located in the same transmembrane helix 1 and face in opposite directions regarding the helix surfaces indicating a helix-helix interaction involving a more complex GxGxGxG domain which could explain the high degree of oligomerization observed for Mic10 in these studies.

Despite all these pieces of evidence suggesting the role of MICOS in the IMM architecture and GxxxG motifs as a driving force for protein-protein interaction, little is known about the structure of the individual components and its interactome on an atomic scale. Insights on an atomic level are necessary to shed light on the regulation and function of this important biomolecular complex. In this chapter of the thesis, a combination of *in silico* and *in vitro* approaches have been applied to understand the structure-function relationship of Mic10. On the one hand, tools from the Swiss Institute of Bioinformatics (SIB) and other sources have been used for predicting Mic10 structure and TM-TM interactions. On the other hand, solution and solid-state NMR spectroscopy have been used to identify key structural features of Mic10 in detergent and lipid environments.

2. MATERIALS AND METHODS

2.1 PROTEIN DESIGN

2.1.1 Constructs used for this work

During the workflow of this thesis project, seven distinct constructs were used. Tables 4.1 includes the name, plasmid, purification/solubilization tag(s) and size of the different constructs of Mic10, Mic δ s, Mic12 and Mic13. The specification regarding DNA and protein sequences as well as the analysis of the protein sequence using ProtParam from Swiss Institute of Bioinformatics (SIB) server is provided in the appendix 4.I. The Mic10His construct was kindly provided by Professor Dr. Meinecke at the University of Göttingen. Mic10strep, MBPMic12 and MBPMic13 were purchased from Genscript (genomic sequences encoding for the different proteins can be found in appendix 4.I). The truncated versions of Mic10 (Mic10 δ) were made in-house by PCR using Mic10strep as a template.

Table 4.1. Mic10, Mic12, and Mic13 constructs used.

Name	Plasmid	Tag	Selection	Restriction enzyme	Construct size (bp)	
Mic10-His	pPROEXHTc	6xHis	Amp			Dr. Meinecke lab
Mic10strep	pET-28b(+)	Strep	Kan	NcoI/NotI	344	This work
Mic10 δ 10	pET-28b(+)	Strep	Kan	NcoI/NotI	314	This work
Mic10 δ 20	pET-28b(+)	Strep	Kan	NcoI/NotI	284	This work
Mic10 δ 30	pET-28b(+)	Strep	Kan	NcoI/NotI	254	This work
Mic10 δ 40	pET-28b(+)	Strep	Kan	NcoI/NotI	224	This work
MBPMi12	pMAL-c4x	MBP and 6xHis	Amp	EcoRI/XbaI	375	This work
MBPMic13	pMAL-c4x	MBP and 6xHis	Amp	EcoRI/XbaI	375	This work

Two distinct approaches were followed to gain structural insights of Mic10. On one hand, the full-length protein Mic10 was purified, but due to its aggregation propensity it was not suitable for the following NMR experiments. Thus, two shorter constructs containing only the transmembrane regions separately were designed and purified.

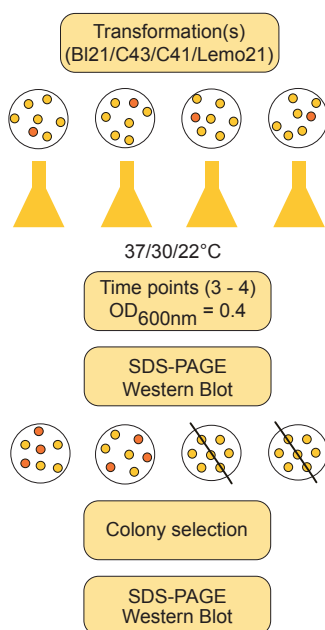
2.1.2 Constructs production

Four N-terminally truncated versions of Mic10strep (Mic10 δ 10, Mic10 δ 20, Mic10 δ 30 and Mic10 δ 40 constructs) were synthesized, using a standard PCR protocol (Table 4.2) (208).

Table 4.2. PCR buffer composition used for *Micδs* preparation.

Products		Purchase
Phusion Polymerase buffer	12.5 μ l	New England Biotechnologies
dNTPs (at 10 mM)	0.4 μ l	New England Biotechnologies
Forward primer (at 10 μ M)	1.5 μ l	Sigma
Reverse primer (at 10 μ M)	1.5 μ l	Sigma
DNA template	x (~50 - 100 ng)	Genscript
DMSO	0.6 μ l	
H ₂ O miliQ	x (up to 25 μ l)	
Phusion DNA Polymerase	0.2 μ l	New England Biotechnologies

Detailed information regarding the primers and the PCR run conditions is provided in appendix 4.III. Appendix 4.III, table S4.2 shows the specifics (sequence and melting temperature) of the primers. The PCR was evaluated by a 1% agarose gels. DNA fragments of interest were purified from the agarose gel using a Macherey Nagel purification kit Plasmid DNA. The PCR product (insert) and plasmid (pET28(+)) were treated with NotI-HF and NcoI-HF restriction enzymes (from NEB) for 15 min at 37°C, followed by heat shock deactivation at 65°C for 20 min. The linearized plasmid was purified using an agarose gel purification kit from Macherey Nagel.



Scheme 4.1. Test expression for optimizing protein yields.

The ligation was performed following the protocol table 4.3. The ligation reactions were incubated over-night (16 hours) at 4°C. Following the ligation, 5 to 10 µl of the ligation reaction were used to transform DH5α cells and plated on agar plates containing 50 µg/ml of kanamycin and incubated over-night (~16 hours) at 37°C. To identify positive colonies (having the plasmid with the full insert) we performed a colony PCR (209). For this purpose, six colonies from each of the agar plates were selected (each construct) and used for the colony PCR following the protocol described in table 4.4. The results were checked by a 1% agarose gel (210). Positive colonies, meaning presenting the plasmid with the inserted fragment, were selected for sequencing.

Table 4.3. Ligation protocol used.

	Volume for 25 µl reaction (µl)	Purchase
T4 DNA ligase	1	New England Biotechnologies
Insert:Vector (in mass)	1:1 / 1:3 / 1:5	Vector from genscript
T4 DNA ligase buffer	x (up to 25 µl)	New England Biotechnologies

Table 4.4. PCR protocol used.

Step	Temperature (°C)	Time (min)	Cycles
Denaturation	95	5	25
Denaturation	95	1	
Annealing	55	1	
Extension	72	5	
-	4	-	

2.1.3 Test Expression of the different constructs in minimal media

To carry out biophysical studies by nuclear magnetic resonance (NMR), proteins have to be isotopically enriched with ¹³C and ¹⁵N, to increase NMR sensitivity. ¹³C and ¹⁵N isotopically enriched protein was obtained by expression in *Escherichia coli* (*E. coli*) grown in M9 minimal media (Appendix 4.II) supplemented with 4 g ¹³C-glucose and 1 g ¹⁵NH₄Cl. The M9 media was derived from the media used in Marley et al. (211) which allows full control on the carbon (¹³C-glucose) and nitrogen (¹⁵N-NH₄Cl) sources metabolized by the expression system chosen. A systematic search for the optimal protein expression conditions (Scheme 4.1) of Mic10His, MBPMic12 and MBPMic13, was performed in small scale expression (5 to 20 ml culture) by varying temperature (room, 30 and 37°C), bacterial strains

(BL21(DE3), Lemo21(DE3), C41(DE3)pLysS and C43(DE3)pLysS, induction time (2h, 4h, 6h, 8h, and 16h) and Isopropyl β -D-1-thiogalactopyranoside (IPTG at 0.5 and 1 mM). Additionally, T7 lysozyme expression was tuned in Lemo21(DE3) by varying the rhamnose concentrations (0, 250, 500 and 1000 μ M). Prior to carrying out large protein expressions a colony selection step was added to improve final protein yields. Protein yields during expression were followed by sodium dodecyl-sulfate gels (SDS-PAGE gels)(212) and Western blots (WB)(213) (Appendix 4.V).

2.2 PEPTIDE SYNTHESIS

Two peptides (TM1 and TM2) each containing one of the transmembrane domains of Mic10 were synthesized for NMR studies. TMpred from SIB was used to predict the two transmembrane (TM) helices in Mic10 TM domain 1 (TM1) from 42 to 62 and TM domain 2 (TM2) from 64 to 82 (Fig. 4.5). Both peptides, TM1 and TM2 were synthesized in-house using standard solid-phase peptide synthesis by Kerstin Overkamp from the NMR-based structural biology department at the Max Planck Institute for biophysical chemistry Göttingen. Four distinct peptides were synthesized, two non-labelled named TM1 and TM2 and, two containing ^{13}C and ^{15}N labelled amino acids at specific positions named TM2-GA and TM2-IGG (Table 4.5). The peptide synthesis was evaluated by HPLC and mass spectrometry (Appendix 4.IX).

Table 4.5. Transmembrane domain constructs.

Name	Amino acid sequence	Construct length (aa)
TM1	Acyl- ^{33}K WDIVLSN <i>MLVKTAMGFG VGVFTSVLF</i> F^{61}K -NH ₂	29
TM2	Acyl- ^{61}K RR <i>AFPVWLG IGFVGRGY</i> EGDAIFRSSA GLRSSKV ⁹⁷ -OH	37
TM2-GA	Acyl- ^{61}K RR <i>AFPVWLG IGF^{13/15}VGRGY^{13/15}A</i> EGDAIFRSSA GLRSSKV ⁹⁷ -OH	37
TM2-IGG	Acyl- ^{61}K RR <i>AFPVWLG ^{13/15}I^{13/15}VGRGY^{13/15}A</i> EGDAIFRSSA GLRSSKV ⁹⁷ -OH	37

The ^{15}N , ^{13}C labeling positions are indicated by “ $^{13/15}$ ”. The amino acids in bolded and italic represents the predicted transmembrane region of Mic10.

Acyl = HCO group at the N-ter of TM2 group. The group is used at the N-terminal of these peptides to mimic the N-terminus of a peptide bond. In the same context, TM1 has the acyl group at the N-terminus and an amine group at the C-terminus.

2.3 PROTEIN PURIFICATION, REFOLDING AND RECONSTITUTION

2.3.1 Protein purification

Mic10His, MBPMic12 and MBPMic13 were expressed in inclusion bodies using Lemo21(DE3). The proteins were purified using a standard inclusion body purification protocol. Briefly, after induction the cells were harvested at 7 000 g at 4°C for 15 min. The pellet was resuspended and homogenized in 20 mM HEPES, 150 mM NaCl, 5% glycerol, 2 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride (PMSF) at pH 7.4. 1 mg Lysozyme and 1 mg DNase per 15 ml of resuspended solution were added in addition to 1 mM of MgCl₂. The sample was stirred for one hour at 4°C. After, the cells were submitted to three cycles of sonication using 30 – 40% power in intervals of 20 seconds pulsing with one minute off (total time 1 min 20 sec per cycle) at 4°C. Finally, the cells were ruptured with an EmulisFlex C3 (Avestin) using 1 000 psi at 4°C (3 to 4 times). Cell debris were pelleted by 7 000 g centrifugation for 20 min at 4°C, the supernatant collected and recentrifuged at 10 000 g for 15 min at 4°C. The inclusion body were pelleted from the supernatant by centrifugation at 20 000 g for 60 min at 4°C. The pellet, inclusion body, was washed in three steps each including resuspension, homogenization, sonication and centrifugation before loading the sample on the affinity column (TALON). In the first wash, the sample was resuspended in 20 mM HEPES, 150 mM NaCl, 5-10% glycerol, 2 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride (PMSF) at pH 7.4 with 5 mM EDTA and 2% triton X100. The sample was sonicated using the same conditions as previously used. After sonication, the sample was centrifuged at 20 000 g for 45 min. For the second wash, the pellet was resuspended in 20 mM HEPES, 150 mM NaCl, 5-10% glycerol, 2 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride (PMSF) at pH 7.4 with only 1% triton X100 and the same procedure was performed. For the last wash, the pellet was resuspended in buffer A (20 mM HEPES, 150 mM NaCl, 5-10 % glycerol, 2 mM β-mercaptoethanol at pH 7.4 and 6 M guanidinium hydrochloride (GdHCl) containing one tablet for 100 ml of buffer of cOmplete EDTA-free protease inhibitor cocktail (Roche)). Then, the sample was centrifuged at room temperature at 20 000 g for 45 min. NaCl up to 300 mM and Imidazole up to 5 mM final concentrations were added to the supernatant that was incubated for 15 min at room temperature with half of the affinity column beads previously washed with water and equilibrated with binding buffer B (20 mM HEPES, 300 mM NaCl, 5% glycerol, 6 M GdHCl, 5 mM imidazole at pH 7.4). After incubation, the sample was

loaded onto the column (the loading was repeated twice, meaning that the flow-through from the first loading was reloaded on the column). The protein was eluted stepwise with buffer B containing 10, 150 and 500 mM imidazole pH 7.4. The protein was further purified by gel filtration in buffer C (20 mM HEPES, 150 mM NaCl, 5-10% glycerol at pH 7.4 and 6 M GdHCl with one protein inhibitor EDTA-free tablet for 100 ml of buffer). The purification was followed by 15% SDS-PAGE gels using Coomassie staining (note: prior to SDS-PAGE, the samples were treated by an ethanol precipitation protocol to precipitate proteins and wash out the GdHCl that causes SDS precipitation)(214).

For Mic10strep and Mic10 δ s the protocol used was a slightly modified protocol from the IBA-life science protocol provided by the supplier (strep-XT data sheet). Briefly, the cells were resuspended in 100 mM Tris, 150 mM NaCl, 1 mM EDTA, 8 mM Urea at pH 8 solution. The solution was stirred at 4°C for one hour, then sonicated as previously described (three runs of sonication using 30 - 40% power, with 20 sec on and 1 min and 20 sec off). The solution was centrifuged at 10 000 g for 20 min at room temperature. The supernatant was diluted with 6 M Urea in 100 mM Tris, 150 mM NaCl, 1 mM EDTA at pH 8 before starting loading the strep-XT column (note: no further optimization on the purification protocol was attempted to maximize protein yields).

Table 4.6. Optimal expression and purification obtained conditions for the different constructs.

Name	Plasmid	Expression	Buffer	Affinity column	Gel filtration
Mic10His	pPROEXHTc	Lemo21(DE3) In M9 containing 500 µM of rhamnose and ampicillin and chloramphenicol	20 mM HEPES, 150 mM NaCl, 6 M GdHCl, 5% Glycerol, 2 mM β-mercaptoethanol at pH 7.4	(Talon) 20 mM HEPES, 150 mM NaCl, 6 M GdHCl, 5% Glycerol, 2 mM β-mercaptoethanol at pH 7.4 Elution 50, 250 and 500 mM	20 mM HEPES, 150 mM NaCl, 6 M GdHCl at pH 7.4
Mic10strep	pET-28b(+)	Lemo21(DE3) In M9 with Kan /Chl	100 mM Tris, 150 mM NaCl, 1 mM EDTA, 8 M Urea, pH 8	(Strep XT) 100 mM Tris, 150 mM NaCl, 1 mM EDTA, 6 M Urea, pH 8 with 50 mM biotin Elution in linear gradient mode	/
Mic10c30	pET-28b(+)	Lemo21(DE3) In M9 with 500 µM of rhamnose, Chi and Kan	100 mM Tris, 150 mM NaCl, 1 mM EDTA, 8 M Urea, pH 8	(Strep XT) 100 mM Tris, 150 mM NaCl, 1 mM EDTA, 6 M Urea, pH 8 with 50 mM biotin Elution in linear gradient mode	/
MBPMic12	pMAL-c4x	Lemo21(DE3) In M9 with Amp/Chl Over night at 37°C	20 mM HEPES, 150 mM NaCl, 6 M GdHCl, 5% Glycerol, 2 mM β-mercaptoethanol at pH 7	(Ni-NTA) 20 mM HEPES, 150 mM NaCl, 6 M GdHCl, 5% Glycerol, 2 mM β-mercaptoethanol at pH 7 Elution 50, 250 and 500 mM	20 mM HEPES, 150 mM NaCl, 6 M GdHCl at pH 7.4
MBPMic13	pMAL-c4x	Lemo21(DE3) In M9 with Amp/Chl Overnight at 37°C	20 mM HEPES, 150 mM NaCl, 6 M GdHCl, 5% Glycerol, 2 mM β-mercaptoethanol at pH 7	(Ni-NTA) 20 mM HEPES, 150 mM NaCl, 6 M GdHCl, 5% Glycerol, 2 mM β-mercaptoethanol at pH 7 Elution 50, 250 and 500 mM	

Note: All columns were used at room temperature except the strep-XT column that was used at 4°C. Before running the columns, the system was equilibrated with their respective buffers. A particular attention was made on the TALON column due to the presence of β-mercaptoethanol that decreases the efficiency of the column. Thus, only ~50% of the bead were incubated for 10 to 15 min in a rocky shaker. After incubation, the protein was poured slowly into the column and passed through the rest of the resin, which was first loaded into the gravity column. The samples were passed 2 - 3 times over the column and then washed and eluted with the buffers indicated above in the table.

2.3.2 Refolding and reconstitution protocol

Different conditions were tested for Mic10His refolding in terms of detergents and refolding methods (step-wise, column or dilution). Briefly, the denatured protein in 6 M GdHCl was transferred to a dialysis cassette (3.5 kDa cut off) together with the desired detergent in a concentration of five times the critical micellar concentration (CMC). The GdHCl was dialyzed out stepwise by reducing the GdHCl concentration from 6 to 3 M, 1.5 M and 0 M. In each step the dialysis was allowed for at least three to four hours. After removal of GdHCl, two dialysis steps were performed for 16 hours each using 20 mM HEPES, 150 mM NaCl and 5% glycerol at pH 7.4 with detergent concentration at the CMC.

After refolding, a screening of different lipids was performed to obtain suitable sample conditions which enables high resolution solid-state NMR studies. Many reconstitution conditions were tested by varying the initial protein conditions from a prior refolded protein in detergents to a single step refolding/reconstitution or even changing the reconstitution method using preformed large unilamellar vesicles (LUVs) or by a mix of detergent and lipid in solution. Excess detergent was removed with biobeads (215) or cyclodextrin (216). Table 4.7 shows the different detergents and lipids used during this work. A more thorough explanation of the reconstitution is discussed and explicitly developed in the results section.

Table 4.7. Detergents and Lipids used for Mic10 and TM studies.

Detergent and lipids	Molecular weight (g/mol)	CMC	Purchase from
Sodium Dodecyl sulfate (SDS)	288.4	0.23% (6-8 mM)	
n-Dodecyl β -D-maltoside (DDM)	510.6	0.0087% (0.17 mM)	Antrace
n-Octyl β -D-glucopyranoside (OG)	292.4	0.53% (18-20 mM)	
Dodecyltrimethylammonium bromide (DTAB)	229.4	0.023% (1-2 mM)	Antrace
n-Dodecylphosphocholine (DPC)	351.5	0.047% (1.5 mM)	Antrace/Avanti
1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC)	846.7	/	Avanti
1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC)	677.5	/	Avanti
1,2-dioleoyl-sn-glycero-3-phospho-L-serine (DOPS)	809.5	/	Avanti
1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC)	785.5	/	Avanti

2.4 STRUCTURAL AND FUNCTIONAL DATA ACQUISITION

2.4.1 Circular dichroism (CD)

Information regarding the secondary structure of proteins can be obtained by a variety of biophysical methods. We used circular dichroism (CD) as an indicator for the overall secondary structure of Mic10His when refolded in detergents. CD allows a rapid distinction between the different protein conformations using low protein concentrations. Whereas the CD spectrum of alpha-helical structures show two negative peaks at 209 nm and 222 nm and a positive peak at 190 nm, a beta-sheet conformation shows a negative peak at 218 nm and a positive peak at 195 nm. For random coil structures, they are characterized in CD by a negative peak at 195 nm (217).

The influence of the buffer and detergents on the secondary structure of Mic10His, TM1 and TM2 were analyzed by CD (Table 4.8). All the CD spectra were recorded using a JASCO instrument. Spectra were recorded from 260 to 180 nm using a cuvette with 1 mm length. The scanning speed was 1 min / 20 nm at room temperature (~25°C). Triplicates of each spectra were recorded and then the average was taken. All spectra were processed with the JASCO software provided; first the spectrum was cut after the conductivity HT raised more than 600 - 700 mV (the data lose reliability). Then, the data points were smoothed using the lowest smoothing function for two iterative times with means-movement function. For all spectra, the corresponding buffer spectrum was subtracted from the protein spectrum before processing.

Table 4.8. Constructs and conditions used for the acquisition of the circular dichroism data.

CONSTRUCT	RANGE OF WAVELENGTH (NM)	SCANS	TEMPERATURE (DEGREE CELSIUS)	BUFFER	DETERGENT OR LIPIDS	Concentration (MM)
Mic10His	260 to 180	3	20	20 mM HEPES, 100 mM NaF at pH 7.4	DPC at 0.1%	~100
Mic10His	260 to 180	3	20	20 mM HEPES, 100 mM NaF at pH 7.4	SDS at 0.2%	~100
Mic10His	260 to 180	3	20	20 mM Phosphate, 150 mM NaCl at pH 7.4	1 to 1 lipid protein ratio (DPhPC)	From solid state roto
TM2	260 to 180	3	20	5 mM HEPES, 25 mM NaCl at pH 6.8		100
TM2	260 to 180	3	20	5 mM HEPES, 25 mM NaCl at pH 6.8	SDS at 1 %	100
TM2	260 to 180	3	20	5 mM HEPES, 25 mM NaCl at pH 6.8	DPC at 1 %	100
TM2	260 to 180	3	20	6.6 mM Sodium Phosphate at pH 6.8		100
TM2	260 to 180	3	20	6.6 mM Sodium Phosphate at pH 6.8	SDS at 1 %	100
TM2	260 to 180	3	20	6.6 mM Sodium Phosphate at pH 6.8	DPC at 1 %	100
TM1	260 to 180	3	20	5 mM HEPES, 25 mM NaCl at pH 6.8		180
TM1	260 to 180	3	20	5 mM HEPES, 25 mM NaCl at pH 6.8	SDS at 1 %	180
TM1	260 to 180	3	20	5 mM HEPES, 25 mM NaCl at pH 6.8	DPC at 1 %	180
TM1	260 to 180	3	20	6.6 mM Sodium Phosphate at pH 6.8		180
TM1	260 to 180	3	20	6.6 mM Sodium Phosphate at pH 6.8	SDS at 1 %	180
TM1	260 to 180	3	20	6.6 mM Sodium Phosphate at pH 6.8	DPC at 1 %	180
TM1	260 to 180	3	20	6.6 mM Sodium Phosphate at pH 6.8		180

2.4.2 Nuclear magnetic resonance (NMR)

Similar to the CD, a variety of conditions regarding detergents, pH, buffers and lipids were tested (Table 4.9, 4.10 and 4.11 for Mic10His solution NMR, TMs solution NMR and Mic10His solids-state NMR data respectively). Note that these tables indicate the major experiments recorded and acquisition parameters, however more specific details as to the processing parameters used will be described in the respective figure captions.

Table 4.9. Mic10His sample conditions used for solution NMR.

BUFFER	TEMPERATURE (K)	EXPERIMENT	SCANS	TD1 H / ms (SPECTRAL WIDTH/PPM)	TD2 N / ms (SPECTRAL WIDTH/PPM)	TD3 C / ms (SPECTRAL WIDTH/PPM)
20 mM Phosphate 150 mM NaCl at pH 7.4 and 1% SDS	310	trosetf3gpsi	8	14	35	x
20 mM Phosphate 150 mM NaCl at pH 7.4 and 1% SDS	310	trhnaetgp3d	20	121 (12)	16.4 (30)	10.2 (40)

Three distinct buffer conditions for the TM2 NMR measurements were tested, HEPES, sodium phosphate and sodium carbonate. Since, no major differences in the NMR spectrum were observed, the NMR studies were performed using 20 mM sodium phosphate buffer, 100 mM NaCl at pH 6.5 with 2% DPC. In addition, three distinct pH values (7.4, 7.0 and 6.8) were tested in a 400 MHz Bruker spectrometer using the “zgpgw5” 1D Bruker standard NMR pulse sequence. For assignment and distance restraints measurements a proton-proton mixing of 80 ms for DIPSI and 120 ms or 200 ms for NOESY spectra were used respectively. Prior to measurement, 10% D₂O was added to the NMR buffer for the solution samples.

Table 4.10. Sample conditions and NMR experiments used for the TMs constructs.

CONSTRUCT	BUFFER	TEMPERATURE (K)	EXPERIMENT	SCANS	TD1 H / ms (SPECTRAL WIDTH/ PPM)	TD2 H / ms (SPECTRAL WIDTH/ PPM)
TM2	20 mM HEPES 100 mM NaCl at pH 6.5	310	zgpgw5	256	77 (14)	
TM2	20 mM HEPES 100 mM NaCl, 20 mM Arg, 20 mM Glu at pH 6.5	310	zgpgw5	256	77 (14)	
TM2	20 mM sodium phosphate 100 mM NaCl at pH 6.5	310	Dipsi2gpph19	32 - 48	77 (14)	29 (14)
TM2	20 mM sodium phosphate 100 mM NaCl at pH 6.5	310	noesygpph19	8 - 16	77 (14)	29 (14)
TMI	20 mM sodium phosphate 100 mM NaCl at pH 6.5	310	Dipsi2gpph19	48	77 (14)	29 (14)
TMI	20 mM sodium phosphate 100 mM NaCl at pH 6.5	310	noesygpph19	8	77 (14)	29 (14)
TMI-TM2	20 mM sodium phosphate 100 mM NaCl at pH 6.5	310	Dipsi2gpph19	40	77 (14)	29 (14)
TMI-TM2	20 mM sodium phosphate 100 mM NaCl at pH 6.5	310	noesygpph19	32	77 (14)	29 (14)

Table 4.11. Mic10His sample conditions used for solid-state NMR.

CONSTRUCT	BUFFER	Lipids	TEMPERATURE (K)	EXPERIMENT	SCANS	TD1 H / ms (SPECTRAL WIDTH/ PPM)	TD2 N / ms (SPECTRAL WIDTH/ PPM)
Mic10His	20 mM potassium phosphate, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	DPhPC	250	(H)NH	96	55 (30)	16 (200)
Mic10His CLEAVED	20 mM potassium phosphate, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	DMPC	250	(H)NH	4	55 (30)	15.7 (200)
Mic10His	20 mM potassium phosphate, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	Micelles	250	(H)NH	32	55 (30)	15.7 (200)
Mic10His	20 mM potassium phosphate, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	E. coli lipids extract	250	(H)NH	32 - 48	77 (14)	29 (14)
Mic10His	20 mM HEPES 100 mM NaCl, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	DOPS/DPhPC	250	(H)NH	128	21 (30)	25 (120)
Mic10His	20 mM HEPES 100 mM NaCl, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	DMPS/DPhPC	250	(H)NH	128	42.5 (30)	26 (120)
Mic10His	20 mM HEPES 100 mM NaCl, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	DOPS	250	(H)NH	128	42.5 (30)	31.6 (200)
Mic10His	20 mM HEPES 100 mM NaCl, 20 mM Arg, 20 mM Glu, 0.02 % NaN3 at pH 7.4	DOPC	250	(H)NH	8	55 (30)	15.7 (200)

Solid-state NMR (ssNMR) data was acquired at both 600 and 800 MHz Bruker spectrometers using 1.3 mm rotors and 55 kHz magic angle spinning (MAS) in a HCN channel probe. (H)NH spectra were recorded using ^1H - ^{15}N cross-polarization (CP) with a square pulse on nitrogen and ramped pulses in the proton of 100 to 80 percent (proton to nitrogen CP) and 80 to 100 percent (nitrogen to proton CP) with a proton RF power of ~ 120 kHz for the proton channel and 60 kHz for the nitrogen channel (218). All NMR data was processed using Topspin 3.6 or 4.0. The spectra were analyzed in CcpNmr (219) and/or Sparky (220).

2.4.3 Enzymatic treatment

To assess the membrane embedded region of Mic10, the soluble domains of Mic10His were targeted by an endoproteases enzymatic treatment. For this purpose, Mic10His was treated with two different endoproteases, α -Chymotrypsin and trypsin. Two separated batches of Mic10His were used, one containing 10 mM of arginine and glutamic acid and another without any additive. Both samples were refolded using DPC at five times CMC in a stepwise manner (as previously described), from a buffer containing 20 mM HEPES, 15 mM NaCl at pH 7.2 and 6 M GdHCl, to a buffer containing 20 mM HEPES, 15 mM NaCl at pH 7.2. The denatured samples were dialyzed overnight against 50 mM Tris, 150 mM NaCl, 2 mM CaCl_2 at pH 7.2 with DPC at $\sim 0.1\%$, buffer conditions were both enzymes are active, supplemented either with arginine and glutamine at 10 mM each or without both amino acids. Stocks of trypsin and α -chymotrypsin were prepared by resuspending 1 mg of the enzyme in 1 ml of H_2O containing 1 mM HCl and 1 mM CaCl_2 . The cleavage was performed using 0.75 U/mg of protease and follow for 48 hours (h) with digestion samples taken at 1, 2, 3, 12, 24, 36 and 48 h. 5 μl of the samples were directly mixed with the SDS loading buffer (62.5 mM Tris-HCl, 2.5% SDS, 0.002% bromophenol blue, 0.7 M β -mercaptoethanol and 10% glycerol) and heated to 95°C for 10 min before loading. A part of the cleaved sample was reconstituted in lipids by using a dialysis method as described in the previous section (2.3.2 Refolding and reconstitution protocol).

3. RESULTS

3.1 INCREASING MIC10HIS YIELDS IN MINIMAL MEDIA FOR STRUCTURAL STUDIES

Obtaining at least 1 mg of ^{13}C and ^{15}N enriched Mic10His per liter of minimal media is a prerequisite for NMR structural characterization studies. The full-length Mic10His construct from *Sacharomyces cerevisiae* as N-terminal hexahistidine fusion used in this work was kindly provided by Professor Dr. Meinecke at the University of Göttingen. The previously reported protocol from Barbot et al. was slightly modified (194) using the optimization protocol described in the materials and method section (see section 4.3.1 and 4.3.2) to improve the expression yields in minimal media. A three to five times higher expression yields was achieved by changing the bacterial strain and induction times. The western blot and properties of the bacteria used for the small scale test expression are shown in appendix 4.V, The best expression conditions were found to be 6 to 7 hours of expression in Lemo21(DE3) and overnight (16 hours) expression in BL21(DE3) at 37°C using 1 M IPTG. Both conditions were taken and used for large scale expression.

To fine tune the expression protocol the two best expression conditions were used in a large scale expression volume (500 ml). Figure 4.4 shows the amino acid sequence of the Mic10His construct used (Fig. 4.4A) and the western blot (Fig. 4.4C; see appendix 4.V). Figure 4.4B shows the bacterial growth over time during protein expression. Both bacterial strains seem to grow up to an expected $\text{OD}_{600\text{nm}}$ of three to four consistent with that reported using this minimal media protocol (4 grams of glucose, by shaking at 120 to 150 rpm). Here, protein concentration was determined by UV-vis at 280 nm after affinity column (Talon). The purification process was monitored by SDS-PAGE (Fig. 4.4C). The expression tested gave final yields of ~2 – 3 mg and ~1 – 2 mg of pure Mic10His per liter of M9 culture with Lemo21(DE3) and BL21(DE3), respectively. Since protein yields were 15% to 20% higher for Lemo21(DE3) than for BL21(DE3), the Lemo21(DE3) was selected as the optimal expression system.

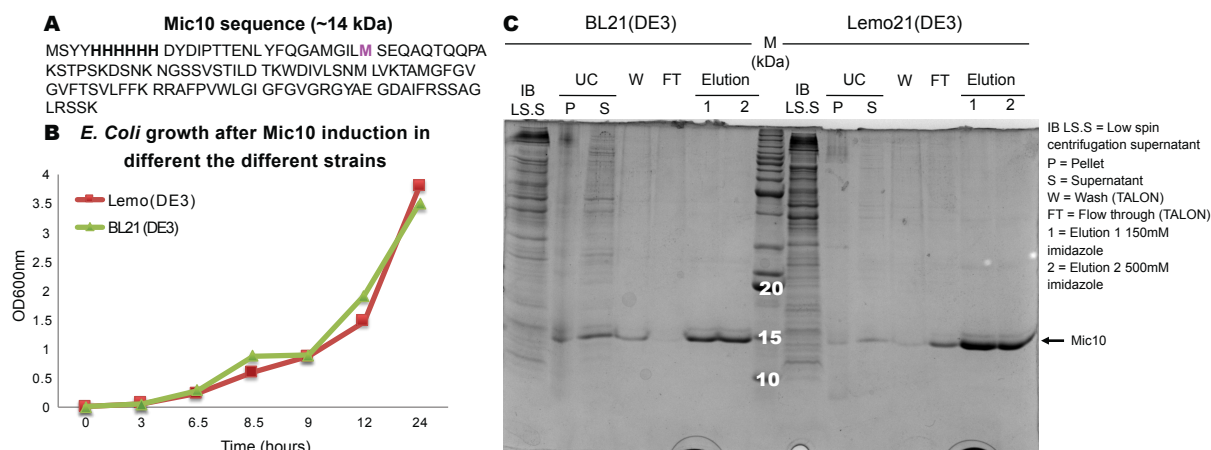


Figure 4.4. Comparison of Mic10His expression and growth in both BL21(DE3) and Lemo21(DE3). The highest yields were obtained using the Lemo21(DE3) strain. In A, the amino acid sequence of Mic10His is shown with the bold purple residue indicating the starting residue of Mic10. B, shows the bacterial growth by following the OD_{600nm} obtained in minimal media for Lemo21(DE3) (red) and BL21(DE3) (green) and C shows the SDS-PAGE of the first part of the purification including elution from the affinity column (Talon).

From the trials, expression in Lemo21(DE3) using 500 μ M of rhamnose and inducing at OD_{600nm} of \sim 0.5 with 1 mM IPTG at 37°C for about seven hours gave \sim 2 mg/L of Mic10His, sufficient to initiate ssNMR tests.

3.2 BIOPHYSICAL EVIDENCE FOR MIC10 GXXXG MOTIFS ADOPTING A BETA SHEET AND/OR ALPHA-HELICAL CONFORMATION IN DETERGENT ENVIRONMENT

3.2.1 Bioinformatic studies of Mic10 - Predicted domains

To quickly assess the expected secondary structure of the different Mic10 constructs used in here, we made use of available bioinformatic tools TMpred (221), SWISS-MODEL (135) and GalaxyTBM (222). These algorithms are based on statistical analysis of a transmembrane protein database and provide quick information regarding the propensity of certain protein regions to be either soluble or membrane associated, and predict the secondary structure of the different protein regions.

Figure 4.5 shows the predicted membrane topologies with their corresponding scores obtained using TMpred (221). As expected, since the difference between the constructs (Mic wild type, Mic10His and Mic10Strep) is basically due to the tags, the score regarding the predicted transmembrane domains (TM) and their respective length is identical. Surprisingly, the highest score is obtained for a structure with an equal length for both TM1 and TM2 domains of 19 residues, from residues L42 to F60 for TM1 domain and A64 to G82 for TM2 domain (Fig. 4.5A, bold). This opposes to the previous studies suggesting that the unequal length of Mic10's TM domains is the driving force for inducing membrane curvature (194).

Although, TMpred also predicts a less likely but possible situation where the TM1 domain is extended from 19 to 25 residues from residue 36I to 60F (Fig. 4.5A, red bold). This situation, with a longer TM1 domain, would be in agreement with the previously proposed mechanism of Mic10-induced membrane bending, where the asymmetric length in between TM1 and TM2 results in a larger surface area occupied by Mic10 towards the inter-mitochondrial space, which drives the IMM curvature (194). However, there is no direct biophysical evidence of the asymmetry or even of the helical conformation of both TM domains in lipid membranes.



Figure 4.5. Predicted transmembrane domains of Mic10 by TMpred.

In A, the amino acid sequence of Mic10 is shown with in bold the amino acids predicted to be in the TM domains and in red bold the extended TM1 from the alternative model propose by TMpred. B shows the prediction table from TMpred, showing the length of the TMs and the score for the two predicted models.

To build the potential structural model for the full-length Mic10, its primary amino acid sequence was submitted to two secondary structure modeling servers GalaxyTBM (222) and SWISS-MODEL (135). Both servers use homology-based algorithms. SWISS-MODEL predicted Mic10 secondary structure based preferentially on the Human membrane protein TMEM141 (pdb: 2lor); the model shows a hairpin-like structure with two TM domains (Fig. 4.6 Mic10, gray). A first TM domain spans residues 35D to 58L (TM1 domain) and a second spans 64A to 80A (TM2 domain). GalaxyTBM predicts a more elongated structure with several helical and random-coiled regions based on the Human potassium channel protein, TRAAK (pdb: 4WFE_B) as the top hit. Interestingly, all five models proposed by galaxyTBM (Fig. 4.6, Mic10_1 to Mic10_5) predicted a long helical domain for the N-terminal residues 31L - 32D to 60F that could be the entirety, leaving a relatively long TM1 domain with a length of 28 to 29 residues, or part of the TM1 domain and an unstructured C-terminal domain with stretches of helical regions. Nevertheless, in all five models, the helical TM1 domain is broken between residues 40N to 43V producing a kink in the membrane segment.

This leads to a hypothetical TM1 domain length of ~16 residues, from 43V to 58L. A TM domain composed of 16 residues is atypically short for helical domains with ~21 residues being the average (223) or 24 residues proposed for Mic10 in previous studies (194). Overall, the predicted secondary structure of the different models and the location of TM region are consistent regarding the TM1 domain.

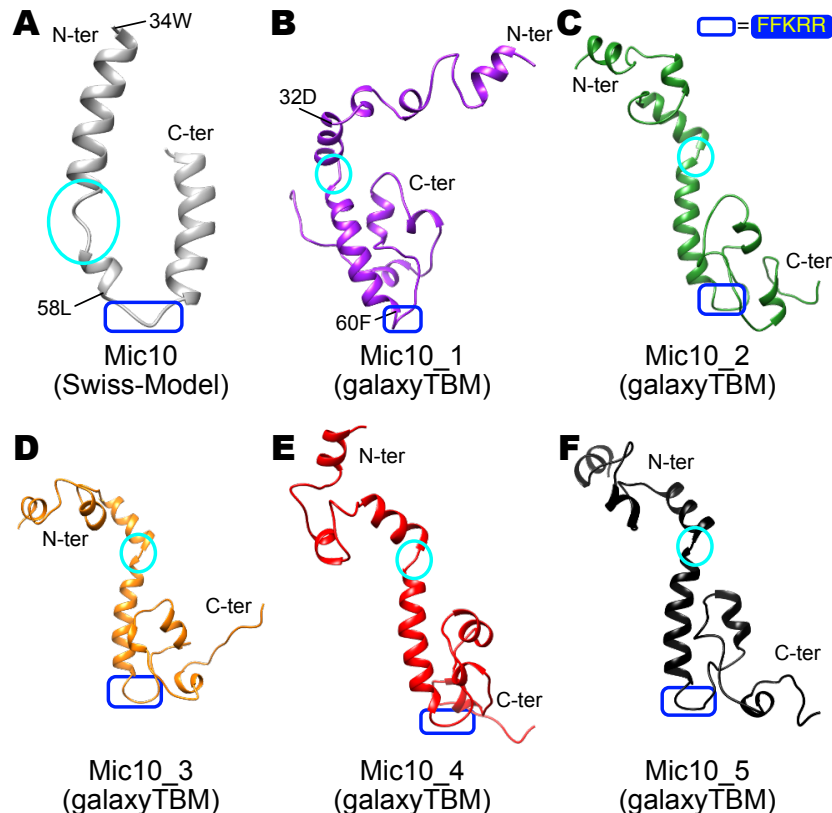


Figure 4.6. Mic10 structural prediction based on homology models from two structure prediction servers: GalaxyTBM and SWISS-MODEL.

In A, prediction from the Swiss-model software from SIB. B to F, the 5 predicted models from GalaxyTBM. The two servers found different hits in their databases for structure prediction. Both are able to predict the helical structure for TM1 domain and some helical contain in the TM2 domain region. All the models show a loop located in between the TM1 and TM2, the region FFRKK, which is consistent with biochemical data and found facing the mitochondrial matrix compartment.

Additionally, all models showed that residues 59-60F to 63R form a small loop which is in perfect agreement with the findings of Mariam Barbot et al., where the FFKRR region (Fig. 4.6, blue box) is reported to be exposed towards the mitochondrial matrix (194). In fact, this positively charged region domain “FFKRR” is responsible for targeting Mic10 to the IMM (203). Despite the consistency between predictions and biochemical data regarding the TM1 domain, the TM2 domain is less well defined using the available bioinformatics tools.

As shown by the prediction, in figure 4.6B-F, TM2 does not show an entire single helical domain. Moreover, for both TM1 and TM2 domains, the structural data is still missing. However, structural modeling has to be taken with caution, since the algorithms used for this purpose only rely on sequence alignment of previously reported protein structures and could be unreliable since the protein may exhibit different structural features. Another source of error is that the database is not selective for membrane bound or soluble proteins. Thus the change in the chemical environment of the protein is also ignored in the structural modeling. Therefore, in more hydrophobic environments such as in lipid membranes, the fragmented helical/unstructured TM2 domain predicted by GalaxyTBM could be different.

Overall, these results strongly support the necessity to carry out the *de novo* structural determination in a membrane environment to understand the relationship between structure and function of Mic10 in mitochondrial membranes.

3.2.2 Biophysical evidence for the secondary helical conformation adopted by Mic10

Following the optimization of the expression protocol for Mic10His (see section 2.1 of the Materials and Method section for more information), a screening of detergents and refolding methods (step wise dialysis (SW), fast dilution and on-column) gave similar results, thus we opted for using the dialysis method since there is less loss of material (Fig. 4.7A, DPC lines). Figure 4.7A shows the SDS-PAGE gel of the detergent refolding screening in sodium dodecyl sulfate (SDS), n-dodecyl phosphocholine (DPC), n-octyl- β -D-glucopyranoside (OG) and n-dodecyl- β -D-maltopyranoside (DDM from Avanti). The SDS-PAGE gel of figure 4.7A shows a major band at ~15 kDa corresponding to Mic10His, however, other higher molecular weight bands are present in the gel. Due to the well known propensity of Mic10 to homo-oligomerize, we wondered if the higher molecular weight bands observed might correspond to Mic10 homo-oligomers. Thus, with the help of our collaborators (Dr. Daryna Tavarsenko University of Göttingen) the presence of Mic10 homo-oligomers was confirmed by western blot using a polyclonal anti Mic10 and anti-His antibody (Appendix 4.V). This finding is consistent with previously reported data, that support a strong Mic10 protein-protein interaction and a strong stability of the oligomeric state (194, 203).

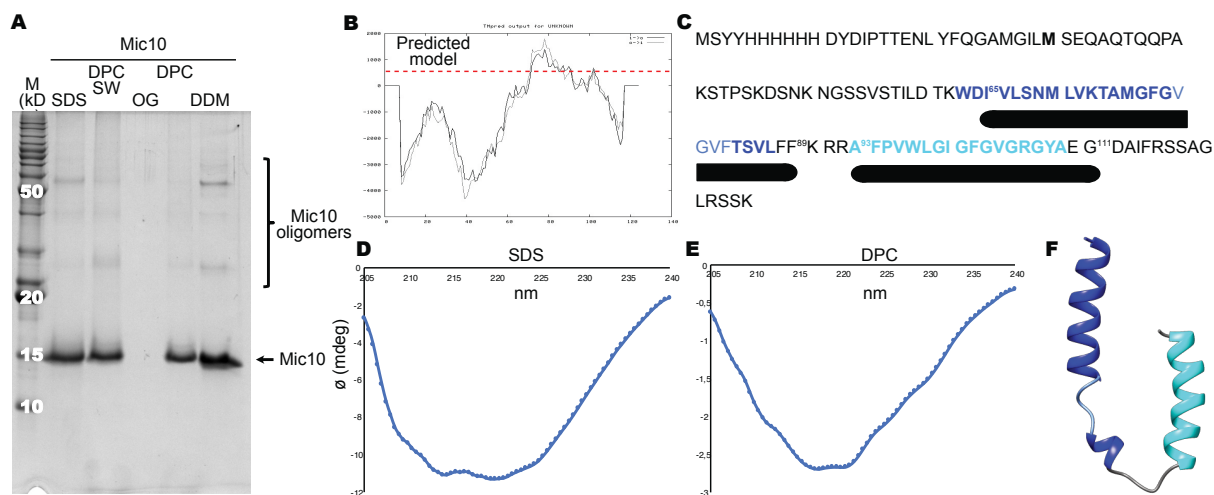


Figure 4.7. Heterogeneous contribution from alpha and beta sheet secondary structure in refolded Mic10. In A, Mic10 refolding in different detergents using 15 % SDS-PAGE and coomassie staining. In B, predicted model of Mic10His by the SWISS protein servers and C shows the helix position in the amino acid sequence from the predicted structure F. D and E are the CD spectra recorded for Mic10-His in different detergents SDS and DPC, respectively.

The secondary structure of the refolded Mic10His was inspected by CD. The CD spectra recorded from refolded Mic10His in both SDS and DPC show a negative absorption from 240 to 200 nm (Fig. 4.7D and 4.7E). Contrary to what we expected, a helical secondary structure, the CD spectra measured show a large heterogeneity regarding the secondary structures present in the samples. The CD spectrum in DPC micelles shows a higher contribution of beta conformation with a single minimum at 217 nm (Fig. 4.7E), while in the CD spectrum in SDS micelles, shows two minima at about 213 nm and 222 nm are observed (Fig. 4.7D) indicating a larger helical contribution in this environment. In both cases, DPC and SDS, the helical content predicted by the JASCO software is below 50%. Thus, these data suggest that Mic10His in detergent environments adopts secondary structures with a different contribution of beta and alpha-helix secondary structure that depends on the detergent used. To obtain more detailed information of the secondary structure of Mic10His, the system was investigated using solution NMR spectroscopy. Table 4.12 summarizes the different detergents tested.

Table 4.12. Detergent properties.

Detergent	Critical micellar concentration (mM)	Molecular weight (Da)	Purchase
SDS	7 - 10	288	Merck
DPC	1.5	351	Avanti
OG	20	292	Sigma
DM	0.2	510	Avanti
LDAO	1.7	229	Sigma

Figure 4.8 shows the ^{15}N -HSQC spectrum recorded in the different detergent conditions. The sample temperature was increased from 30 to 37°C to decrease the correlation time (τ_c) of the particle in solution. Out of the four detergents tested, SDS gives the best sensitivity and spectral quality (Fig. 4.8A). DM (Fig. 4.8B) and DPC (Fig. 4.8D) have almost no detectable signal in the ^{15}N -HSQC. The lack of sensitivity could be caused by the slow motion of the protein in the presence of detergent, fast exchange with the solvent, which will increase with temperature at 37°C, and/or the coupling value of the covalent bond used during the magnetization transfer ($^1J_{\text{HN}} \sim 95$ Hz). ^{13}C -HSQC spectra of SDS and DM show sufficient spectral quality (Appendix 4.VI) suggesting that the lack of sensitivity in the ^{15}N -HSQC is most likely due to exchange with the solvent. Although, this does not totally exclude the change of the HN coupling ($^1J_{\text{HN}}$) which will considerably decrease the magnetization transfer efficiency.

The 2D ^{15}N -HSQC NMR spectra recorded for Mic10His refolded in SDS (Fig. 4.8A) and LDAO (Fig. 4.8C) showed narrow amide proton (H^{N}) chemical dispersion of about 1 ppm, from 7.5 to 8.5 ppm, suggesting that Mic10His adopts a helical conformation. This is consistent with the bioinformatics predictions and the CD data. Detergents such as SDS have been characterized as being harsh and promoting helical formation (224). In order to test the effect of SDS itself on the protein secondary structure, ^{13}C -HSQC spectra in DM and SDS were acquired (Appendix 4.VI). The chemical shift changes in the spectra are due to the difference in the sample temperatures of 37°C and 30°C for SDS and DM, respectively. No other significant chemical shift differences were observed. Thus, large structural influence of SDS in Mic10 can be ruled out.

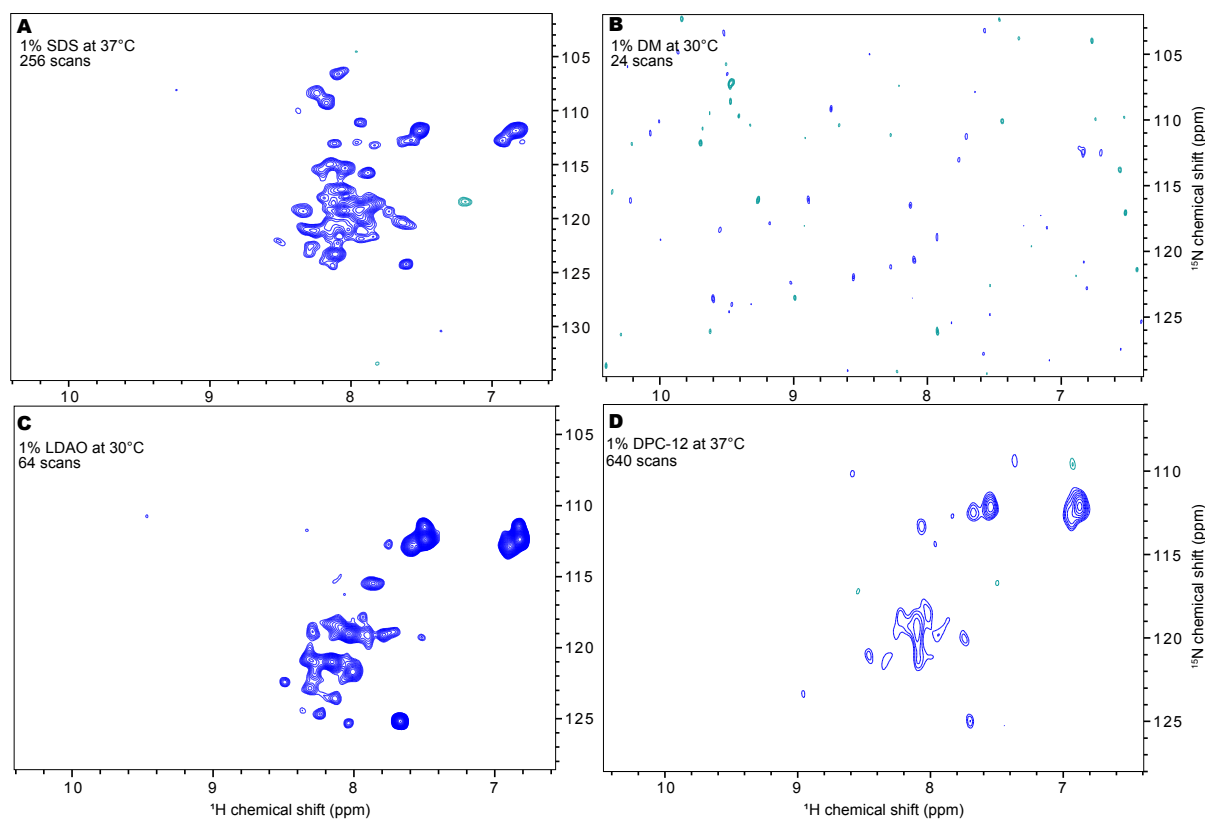


Figure 4.8. Detergent screening by NMR.

Four different detergent conditions were tested for Mic10-His, SDS (A), DM (B), LDAO (C) and DPC (D). SDS turns out to be the optimal detergent in these conditions $\sim 80 \mu\text{M}$ (note concentration before refolding) of Mic10 refolded in 20 mM Sodium phosphate buffer 150 mM NaCl at pH 7.4 using 1%. ^{15}N -HSQC spectra were recorded on a 400 MHz Bruker spectrometer with a 3 channel TCI probe using 106 ms in the direct dimension (^1H) and 46 ms in the indirect dimension (^{15}N) with spectral width of 12 and 24 ppm, respectively.

SDS was used to further investigate Mic10 structure since it provides both the best CD and NMR spectral quality. A $^{13}\text{C},^{15}\text{N}$ -Mic10His enriched sample was prepared and refolded in a 1% SDS solution. The sample was used for acquiring assignment data in a 700 MHz spectrometer equipped with a 5 mm cryoprobe (Fig. 4.9). TROSY based experiments that give higher sensitivity and resolution for the slowly tumbling membrane proteins were used (225). Using the automatic peak picking tools from CcpNmr (219), only about half of the expected HN peaks (65 H^{N} peaks) were picked from the ^{15}N -HSQC out of 120 H^{N} peaks expected (125 – 5 prolines – 1 N-terminus = 120 peaks). Two peaks at about 10 ppm in proton and 130 ppm in nitrogen are observed corresponding to $^{15}\text{HN}\epsilon$ from the two tryptophan sidechains. The assignment of those peaks belonging to the tryptophan sidechain was confirmed by acquiring 3D HNCA. The $\text{N}\epsilon\text{H}\epsilon$ peaks are absent from the HNCA spectrum because there is no NCA correlation in the tryptophan sidechain. During the course of acquisition, 72 hours, several peaks on the spectrum cease to be visible (Fig. 4.9, red spectra, peaks represented by blue arrows), suggesting that over time the Mic10His protein aggregates or/and degrades. Spectral

changes over such a short period of time limit the NMR applicability for assignment purposes since the acquisition of the complete set of spectra required for the assignment and the structure calculation need several weeks.

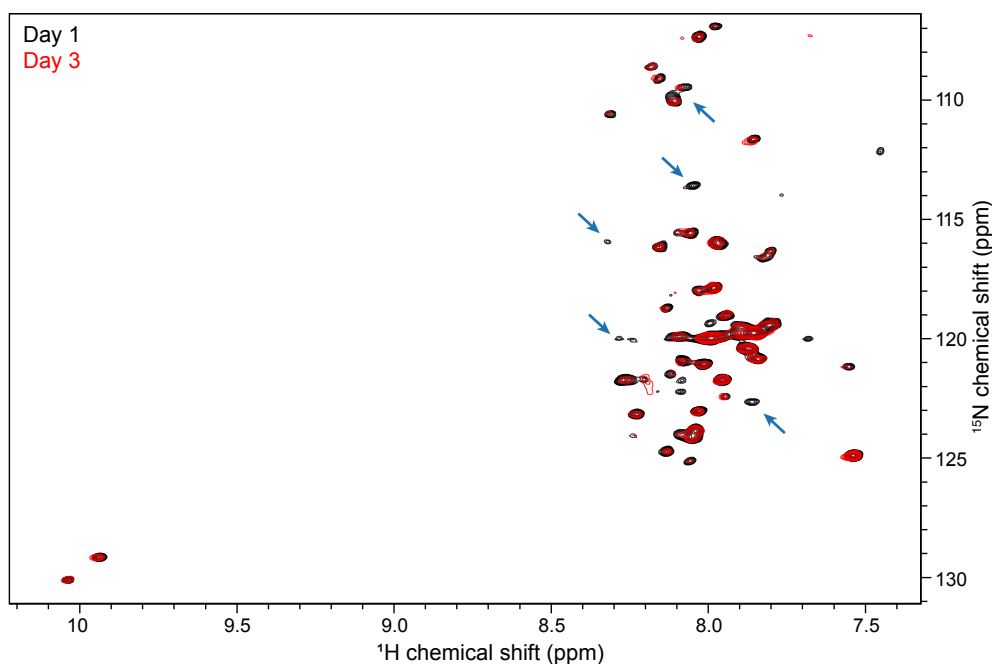


Figure 4.9. Poor stability of Mic10His in SDS detergent micelles conditions.

TROSY based ^{15}N -HSQC (trosetf3gpsi) using $\sim 300 \mu\text{M}$ of Mic10His refolded in 20 mM Sodium phosphate buffer 150 mM NaCl at pH 7.4 using 1% of SDS. spectra were recorded on a 700 MHz Bruker spectrometer with a three channel TXI probe using 121 ms (2048 points) in the direct dimension (^1H) and 60 ms (256 points) in the indirect dimension (^{15}N) with spectral width of 14 and 35 ppm, respectively. Processing was performed applying a square sine function in both dimensions with only 60 ms (1024) points in the direct dimension (^1H).

Although, we could not obtain all the data for this purpose, we attempted to obtain information on the conformation adopted by the glycine residues which might be involved in GxxxG motifs. NMR chemical shift provides qualitative information regarding the protein secondary structure in a residue-specific manner (226). For this purpose, a 3D HNCA TROSY spectrum was recorded to access $\text{C}\alpha$ chemical shifts. Glycine residues have particular ^{15}N and ^{13}C chemical shifts: a high field carbon alpha chemical shift between 40 and 50 ppm and a nitrogen chemical shift between 100 and 115 ppm. Figure 4.10A, shows the CN projection of the 3D HNCA spectrum recorded on a 700 MHz Bruker spectrometer. Nine out of the thirteen expected glycine peaks are observed in the HNCA spectra (Fig. 4.10D, bold). Based on the models and prediction, the number of glycine residues excluding the TM regions can be estimated to be four (Fig. 4.5 and 4.6). Therefore, at least five out of these nine $\text{C}\alpha$ glycine

peaks identified are expected to be part of the transmembrane domains. Figure 4.10B indicates the chemical shift expected from a beta-sheet (red), random coil (white) or alpha-helical (blue) structure for glycine. Three $C\alpha$ glycine residues have their chemical shift around the values expected for a random coil region with an average chemical shift of ~ 45.4 ppm and the others show chemical shifts values above 46 ppm which suggest an alpha helical conformation. Furthermore, the HNCA spectrum shows $C\alpha$ peaks shifted to lower field in the carbon dimension with chemical shifts above 60 ppm. These chemical shift values are a signature of $C\alpha$ of Ile, Val, Phe, Pro, Ser and Tyr in a helical conformation which is consistent with the structural model of Mic10 adopting a helical conformation.

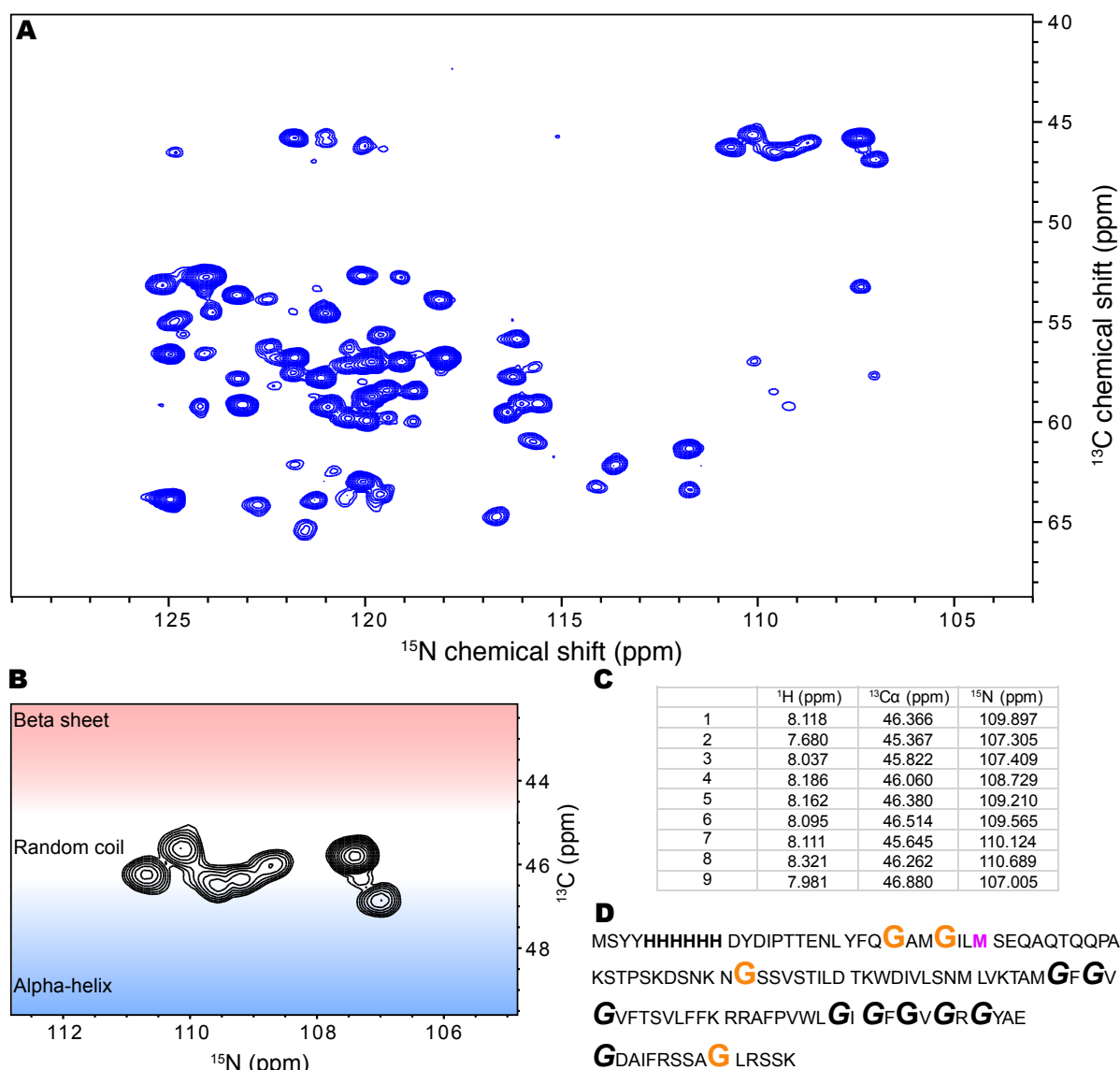


Figure 4.10. Part of the glycine content in Mic10His adopts a helical conformation in SDS micelles.

A, shows the 2D plane of a TROSY based HNCA spectrum acquired on a 700 MHz Burker spectrometer at 37°C. B, shows an expansion of the glycine region of the 2D plane, with a color code indicating the expected

glycine $C\alpha$ chemical shift from helical (in blue), random coil (in white) or beta sheet (in red). In C, a table representing the chemical shift values of the nine glycine residues that have been observed (no assignment available so far). In D, the amino acid sequence of Mic10His with the glycine residues (in bold), the glycines that are expected to be outside the membrane region (orange) and the glycines that are expected to be in the TM regions (black and italic).

Altogether, the biophysical data recorded are consistent and support the model that Mic10 adopts a random-coil/alpha helical structure in detergent environments. With the available data, we could not obtain residue-specific assignment of the glycine peaks observed in the 3D NMR spectra, which shows the mostly alpha helical conformation based on their chemical shift. However, it is worth noting that all the data were obtained in detergent micelles which differ from the native membrane environment. This could lead to a looser protein fold and therefore a different protein folding from that in *in vivo* conditions. Even though all detergents tested here have been successfully used in other NMR studies with alpha-helical membrane proteins, one should be particularly cautious with the potential artefacts due to the detergent used. Therefore, we further investigated Mic10's structure in a more a native-like environment, lipid bilayers, using solid-state NMR.

3.3 INHOMOGENEOUS NMR LINE SHAPE FROM MIC10 RECONSTITUTED IN NATIVE-LIKE LIPID ENVIRONMENT

So far, our data indicate that Mic10 forms a helical structure in micelles, however, we wanted to obtain biophysical evidence of Mic10's secondary structure in membrane-like environments and additionally obtaining insights into the processes driving membrane curvature. We performed small-scale lipid screening in order to find the best lipid candidate for ssNMR studies. Figure 4.11 shows (H)NH spectra of the different lipids tested using 1 to 1 lipid to protein ratio by weight. None of the tested lipids gave sufficient spectral quality to obtain sequence specific assignments in uniformly labeled ^{13}C and ^{15}N Mic10His in lipid bilayers by ssNMR. To overcome the spectral overlap, different reconstitution methods using DPhPC, which gives the most resolved spectrum (Fig. 4.12, black), were tested. The reconstitution protocols tested were i) **direct reconstitution** from denaturing conditions (6 M GdHCl solution) in a detergent/liposomes mix together with a buffer containing β -cyclodextrin during dialysis to remove detergent and GdHCl (Fig. 4.12, green), ii) reconstitution in **large unilamellar lipid (LUV)** by performing the LUV using a standard extrusion protocol and mixing it with refolded protein in DPC, in the later the detergent was

removed by bio beads (Fig. 4.12, red), and iii) a reconstitution in **cyclodextrin detergent removal** (CDR): after mixing the lipids with DPC detergent, the mix was added to the protein previously refolded in DPC and finally the detergent was removed by adding β -cyclodextrin in the dialysis buffer (Fig. 4.12, black). Figure 4.12 shows the different (H)NH spectra recorded for the different reconstitution conditions. The LUV and CDR samples are the best methods to reconstitute Mic10His in lipids as shown by the better resolved glycine region of the (H)NH spectrum (105 to 110 ppm) obtained in comparison to the GdHCl sample

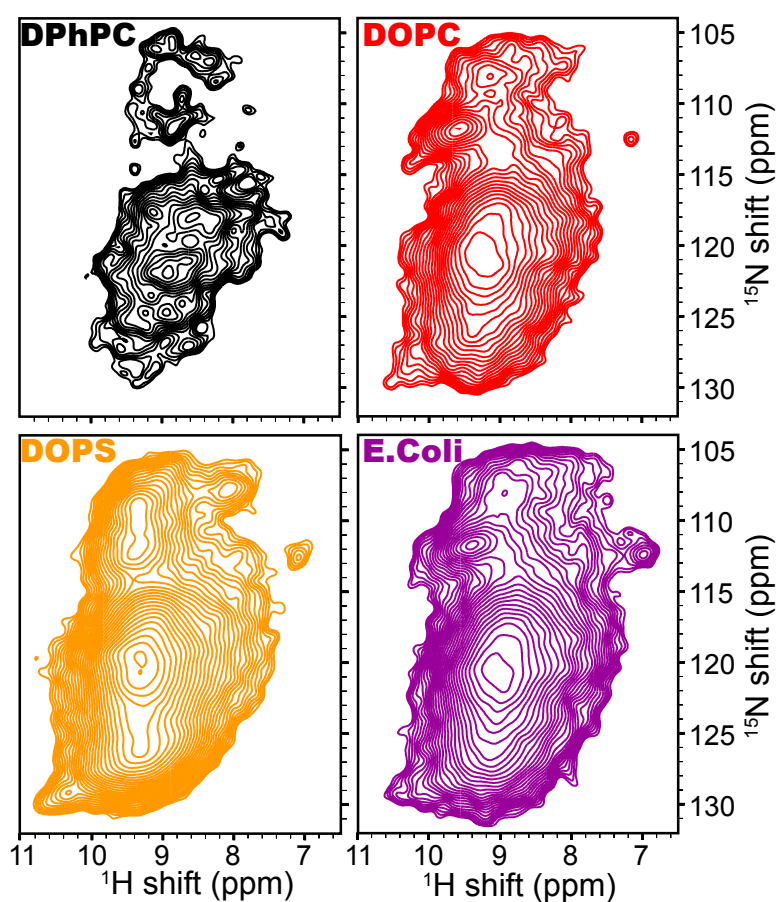


Figure 4.11. (H)NH spectra of Mic10-His reconstituted in different lipids.

The (H)NH spectra of Mic10His reconstituted in DPhPC (black), DOPC (red), DOPS (orange) and *E. Coli* (purple) lipids are shown. All the samples were reconstituted using a lipid protein ratio of 1 to 1 by weight. The (H)NH spectra were processed using a squared sine function with 10 ms in the direct dimension (^1H) and 15 ms on the indirect dimension (^{15}N) in a 800 MHz Bruker spectrometer using a 1.3 mm HCN probe at 55 kHz at 250 K.

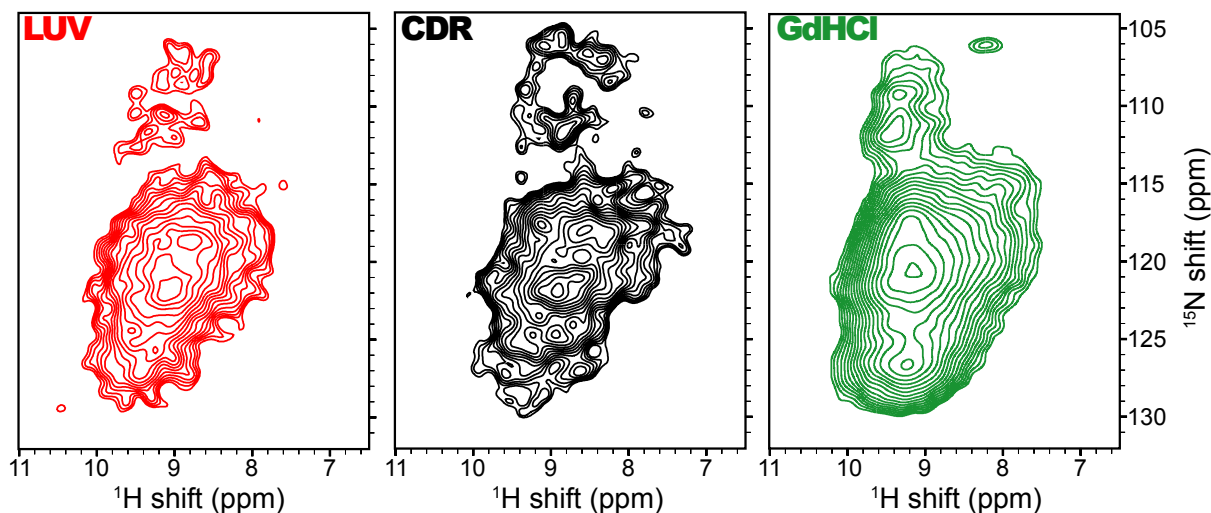


Figure 4.12. (H)NH spectra of Mic10-His using different reconstituted protocols.

The (H)NH spectra show the spectral quality obtained for Mic10His reconstituted in DPhPC using different reconstituted protocols: LUV (red), CDR (red), GdHCl (green). All the samples were reconstituted using a lipid protein ratio of 1 to 1 by weight. The (H)NH spectra were processed using a squared sine function with 10 ms in the direct dimension (^1H) and 15 ms on the indirect dimension (^{15}N) in a 800 MHz Bruker spectrometer using a 1.3 mm HCN probe at 55 kHz at 250 K.

After screening different lipids for spectral quality, the DPhPC containing membranes were chosen for further investigation. We decided to explore the influence of the lipid environment by mixing DPhPC with charged lipids. The choice of the lipid was made to take into account the composition of the mitochondrial membrane, which is composed of a variety of lipids including negatively charged lipids. The mitochondrial yeast membranes, from which Mic10His construct is derived (*S. cerevisiae*), contain mostly phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI), cardiolipin (CL) and phosphatidylserine (PS) at 38.4, 24.0, 16.2, 16.1 and 3.8 mol%, respectively (227). PE and CL lipid were already included in our precedent tests since it is one of the components of the *E. coli* lipid extracts (PE is the major component found in the *E. coli* lipids extract from Avanti; Fig. 4.11, purple). Therefore, PS-containing lipids were our next choice (Fig. 4.13). Figure 4.13 shows the comparison of the (H)NH spectrum of DPhPC/DMPS and DPhPC/DOPS using a LPR of 1.5 to 1 by weight, such as the additional lipid (DMPS or DOPS) is added extra to the DPhPC/Mic10His mixed at a LPR of 1 to 1 by weight. Neither DMPS (light green) nor DOPS (brown) give a substantial improvement in the spectral quality (Fig. 4.13).

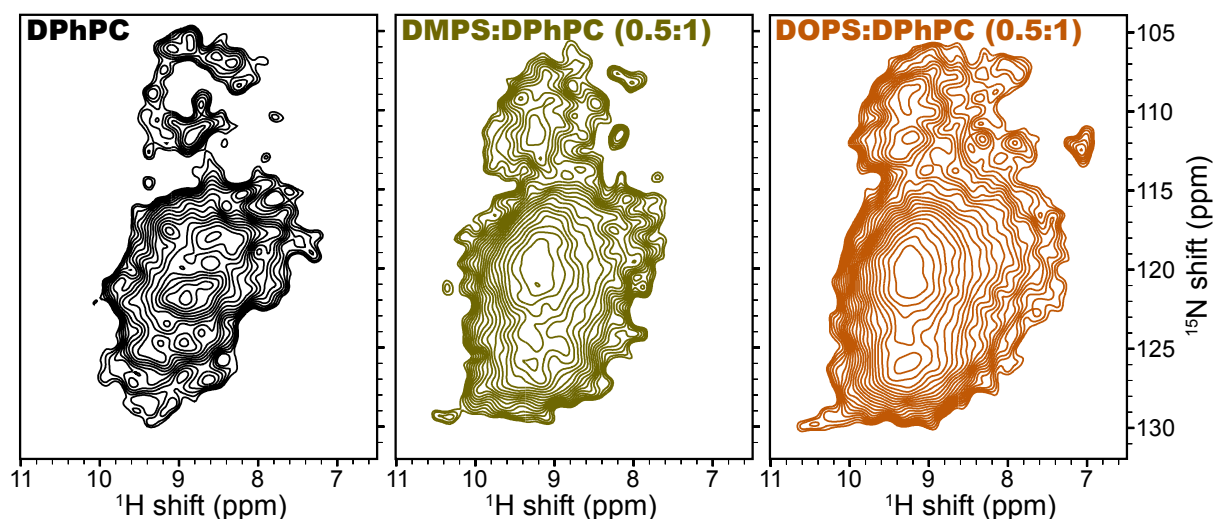


Figure 4.13. Negatively charged lipids do not improve Mic10His spectral quality in ssNMR.

The best previous reconstitution condition, DPhPC, in Black and two mixtures of lipid membranes to add negative charges on the membrane surface DMPS:DPhPC (light green) and DOPS:DPhPC (brown). The (H)NH spectra were process using a square sine function with 10 ms in the direct dimension (^1H) and 15 ms on the indirect dimension (^{15}N) in a 800 MHz Bruker spectrometer using a 1.3 mm HCN probe at 55 kHz at 250 K.

The ssNMR data did not provide enough information to gain insights into either the conformation of Mic10 or the GxxxG motif functionality in lipid bilayers. The broad peaks obtained in the ssNMR spectra suggest that there is a large sample inhomogeneity, which suggests that the sample preparation requires further optimization or that inhomogeneity is characteristic of the protein in the absence of the others MICOS complex subunits. The inhomogeneity may also arise from multiple modes of homo-oligomerization that can be imagined for the overlapping GxxxG motifs of Mic10. Indeed, different homo-oligomeric states of Mic10 have been observed in the SDS-PAGE and native polyacrylamide gels (194, 203).

3.4 THE GLYCINE RICH MOTIFS IN MIC10'S MIGHT NOT BE THE ONLY SITE DRIVING THE OLIGOMERIZATION

3.4.1 TM2 domain homo-oligomerization is highly likely by bioinformatics predictions

Since both Mic10's TM domains have a high content of glycine rich motifs (GxxxG), we wonder about the propensity of these TM domain to interact individually. These motifs are well known to promote helix-helix interaction (228-230). Anderson et al. observed high propensity of the helix-helix interaction in peptides containing ZxxxZ motif (Z = small amino acid residues like G, S or A). Interestingly, including this definition into Mic10 an extended packing-packing region is observed in both TM domains (Fig. 4.14B). On the one hand, TM1

shows a AxGxGxGxxxS motif, thus three ZxxxZ motifs (Fig. 4.14B) or one GxxxGxxxZ motif (Fig. 4.14B, red). On the other hand, TM2 has a GxGxGxGxGxG motif depicted in four ZxxxZ motifs (Fig. 4.14B) or two ZxxxZxxxZ motifs (Fig. 4.14B, red for GxxxGxxxZ and black bold for GxxxGxxxG). To investigate the strength of the interaction between the TM domains, both TM domains were submitted to preddimer a bioinformatic tool, which predicts the probability of two transmembrane α -helical domains to dimerize (231). The three combinations TM1-TM1, TM2-TM2 and TM1-TM2 were submitted. Interestingly, different dimer structures were predicted by preddimer: whereas for TM1 a higher dimerization score is obtained for a parallel TM1-TM1 homo-dimer, referred to “parallel-dimer” (score of 4.8, fig. 4.14A, TM1); for TM2 the highest score is obtained for two helices interacting with each other at a specific location of the TM2 domain, the RGY region, referred to “cross-dimer”(score of 6.8, fig. 4.14B, TM2).. In the cross-dimer situation, where the score obtained is the highest, the dimer seems to be stabilized by a cation- π interaction between the arginine and tyrosine side-chains. Interestingly, the score obtained for dimerization between the two TM domains (TM1-TM2) drops to 3.0 suggesting that Mic10 homo-oligomerization is more likely to occur between the interfaces of the “parallel-dimer” (TM1-TM1) and “cross-dimer” (TM2-TM2) rather than the interface of the hetero-dimer (TM1 – TM2). Judging from the number of the ZxxxZ motifs, it is not surprising that the “cross-dimer” (TM2-TM2) interaction obtains the higher score since it contains more GxxxG motifs. Additionally, the presence of L, V and I residues neighboring the glycines of the GxxxG motifs in both domains, TM1 and TM2, could further stabilize the helix-helix interaction. Such residues have been reported to increase the stability for helix-helix packing (230).

Overall, the bioinformatics predictions are consistent with the available biochemical studies, suggesting that oligomerization is in fact a prominent behavior for Mic10. However, the predictions that TM2 domain has a higher propensity to dimerize opposes to the biochemical data. Previous biochemical studies showed that TM1 domain is responsible for the oligomerization of Mic10. Mutations at residues G50 and G52, found facing in opposite direction of the TM1, to alanine induce the complete loss of Mic10 homo-oligomerization (194). However, so far no atomic data reported have determined the helix-helix packing interface or explained the relationship between the ZxxxZ motifs present in Mic10 regarding oligomerization and membrane curvature. Additional biochemical data targeting the RGY

region to disrupt the predicted cation- π interaction, which seems to be involved in stabilizing the “cross-dimer”, would be necessary to further shed light into the possible role of the TM2 domain in the Mic10 oligomerization process. Considering that the predicted length of the TM domains are of 19/25 residues for the TM1 domain and 19 residues for the TM2 domain, and that the total number of residues is of 96, Mic10 has more than half of the protein exposed to the mitochondrial interspace. This suggests the possibility that the soluble domains are involved in the oligomer formation as well.

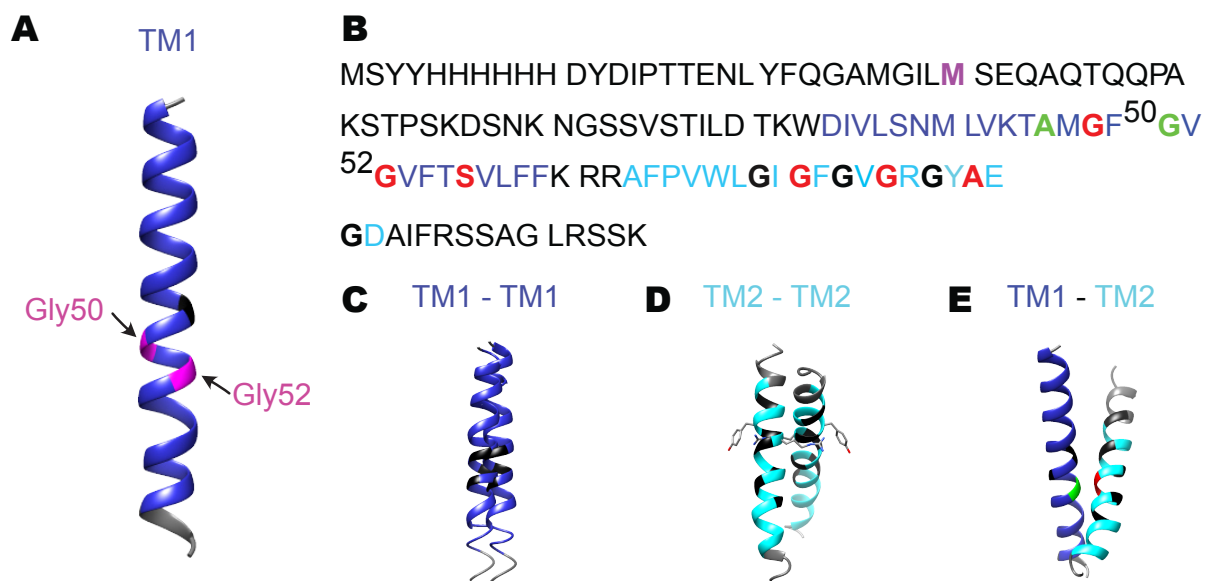


Figure 4.14. Predicted helical-helical interaction from the different Mic10 transmembrane domains.

A, shows the TM1 helical structure prediction including the position of the two glycine (Gly50 and Gly52, in pink) responsible for homo-oligomerization. B, shows the amino acid sequence of Mic10 with the (Z)xxx(G) motifs in green, (G)xxx(G)xxx(Z) motifs in red and the (G)xxx(G)xxx(G) motifs in black bold. From C to E, the predicted parallel-dimer (TM1-TM1, in C), cross-dimer (TM2-TM2, in D) and the hetero-dimer (TM1-TM2, in E) by preddimer from SIB tools are shown. The TMs are color coded on the amino acid sequence with TM1 in purple and TM2 in cyan

3.4.2 Soluble domains might have a role in the stabilization of oligomers

In order to disrupt the homo-oligomerization of Mic10 and thus decrease spectral complexity for solid-state NMR studies, a double mutant G50A and G52A which was kindly provided by our collaborators, Professor Dr. Meinecke at the University of Göttingen, was expressed. However, yield of the expression was too low to pursue NMR studies. Moreover, after the affinity chromatography column, the eluted double mutant Mic10 protein, when analyzed by SDS-PAGE exhibited high molecular bands as observed for Mic10His.

To alleviate the NMR spectral complexity, we opted for reducing the length of the protein. Since our focus is in the TM domains (residue ~33 to 82) of Mic10, we wonder if by

removing the soluble domains one could increase spectral resolution. To evaluate this possibility, an enzymatic treatment was performed in the refolded Mic10His to cleave out the accessible regions of the protein. Using this approach, we expected to cleave out the extracellular domains of Mic10His while keeping the TM domains since they should be protected from the proteases by the membrane. We tested two distinct endoproteases, trypsin and chymotrypsin.

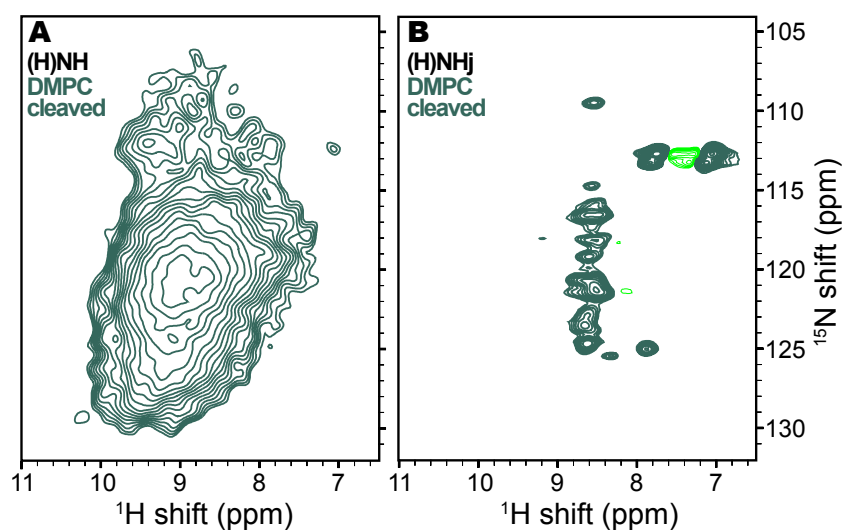


Figure 4.15. ssNMR spectra of trypsin treated Mic10 reconstituted in DMPC.

A, shows the cross polarization (H)NH spectra obtained with Mic10His after trypsin treatment reconstituted in DMPC (DMPC cleaved). Panel B, shows the (H)NHj spectra of the DMPC cleaved sample recorded using INEPT. The bright green is the negative phase of the spectrum due to baseline distortions, which in here it indicates truncation of the FID.

Figure 4.15 shows the (H)NH and (H)NHj spectra of Mic10His recorded in DMPC lipids after trypsin treatment. The trypsin treatment does not improve spectral quality (Fig. 4.15A, DMPC cleaved). The (H)NHj shows that there is still flexible regions on the sample (Fig. 4.15B). Although the TM domains were not accessible for assignment purposes, we wonder if there have been any changes on the flexible domains compare to the micelles prior to trypsin treatment. Slightly differences on the glycine region from LDAO to DMPC cleaved sample (above 110 ppm nitrogen peak) can be observed (Appendix 4.VII). Other peaks seemed to be affected as well, however it is still to be seen if the (H)NHj spectra of a non cleaved Mic10His in DMPC shows the same differences.

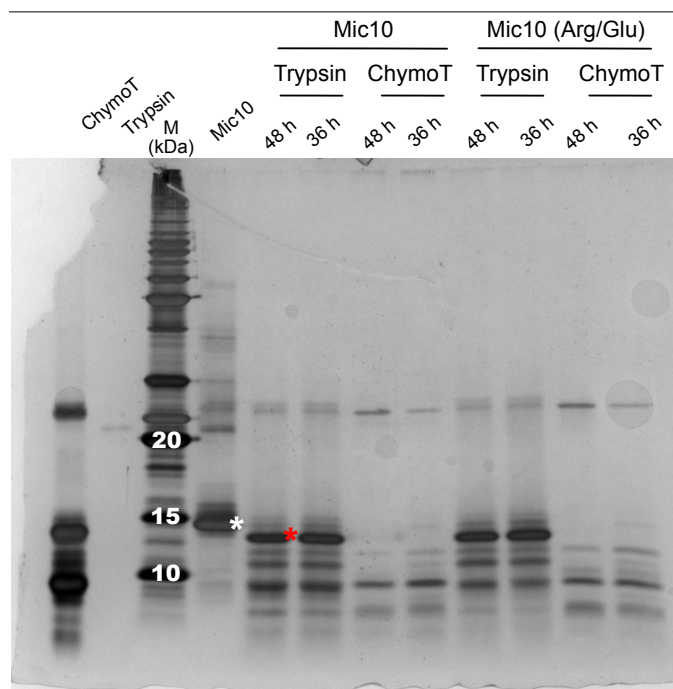


Figure 4.16. Enzymatic treatments disrupt Mic10His oligomers.

White and red stars show the lines where Mic10His was found by mass spectrometry analysis (Appendix 4.VIII). From left to right of the gel is Chymotrypsin (1), Trypsin (2), leader marker (3), Mic10-His (4) as control for the next lines where we treated Mic10His with Trypsin for different times of 48 and 36 hours (5 and 6) at room temperature. The same treatment was performed with a buffer containing 10 mM arginine and 10 mM glutamic acid (9 and 10). The same treatment conditions were used for Chymotrypsin treatment without additives (lines 7 and 8) and with additives (lines 11 and 12). The cleavage sites are represented over Mic10His amino acid sequence. A total of 15 and 11 cleavage sites are found for Chymotrypsin and Trypsin over the full Mic10His sequence, respectively.

MSY|Y|HHHHHH DY|DIPTTENL Y|F|QGAMGILM SEQAQTQQPA
 K|STPSK|DSNK| NGSSVSTILD TK|W|DIVLSNM LVK|TAMGF|GV
 GVF|TSVLF|F|K| R|R|AF|PVW|LGI G|FGVGR|G|YAE
 GDAIF|R|SSAG LR|SSK

Theoretical cleavage sites:
 | : Chymotrypsin (15 sites)
 | : Trypsin (11 sites)
 ChymoT = α - Chymotrpsyn
 M = Marker
 h = incubation hours

The efficiency of the cleavage and the accessible regions of the enzyme tested was followed by SDS-PAGE (Fig. 4.16). Figure 4.16 shows the SDS-PAGE gels stained using silver nitrate (232). In both enzymatic treatments, enzyme was added at 0.75 U/mg and the reaction was followed by SDS-PAGE at different time points. Previous data reported the role of arginine as stabilizer for membrane proteins (233), thus Mic10 was refolded in two conditions: one with and one without the addition of 20 mM arginine in 0.5% DPC. The SDS-PAGE gel shows that the oligomers of Mic10His were largely digested by the treatment of both enzymes (Fig. 4.16). Surprisingly, in contrast to the sample treated with chymotrypsin, the trypsin treated sample showed a major band above ~12 kDa. To confirm the presence of Mic10His in this band, the bands present on the gel were cut out and analyzed by mass spectrometry. Even though the sequence coverage was low (~30%, see appendix VIII), Mic10His sequence was identified from residue W34 to R87, which is comprised in the TM domains (residue 34 to 62) including residues G50 and G52 shown to be important for Mic10 oligomerization (Barbot et al.). None of the mass spectrometry data showed the presence of

the N-terminal including the His tag or the C-terminal of Mic10His (Appendix 4.VIII). This indicates that the flexible regions are likely to be cleaved by the enzymatic treatment. Although further investigation is needed, these results suggest that Trypsin cleavage can be used in the presence of DPC to reduce the presence of oligomers and obtain a ~12 kDa domain.

To conclude, the combination of bioinformatics, ssNMR in more native environments and enzymatic treatments suggest a strong protein-protein interaction behind the stable non homogenous homo-oligomers formed by the full-length protein. The enzymatic treatment raises additional questions regarding the oligomerization of Mic10. Is the oligomerization in part modulated by the soluble domains of Mic10His as well? Our data suggest that it might be a more complex mechanism behind the formation of Mic10 homo-oligomers and that including other proteins of the MICOS complex may be important for the further progress in structurally defining the Mic10 protein.

3.5 BOTH TRANSMEMBRANE DOMAINS OF MIC10 SHOW HELICAL PROPENSITY IN DETERGENT

3.5.1 The presence of detergent is required for driving the TM2 towards a helical conformation

As previously discussed in section three of the "Results", the inhomogeneity of the complex form by the oligomerization of Mic10His full-length protein hindered the structural studies in a lipid bilayer. Several studies (194, 203) including our own bioinformatic analyses showed that Mic10 contains two helical transmembrane domains. To further characterize Mic10's oligomerization, we decided to assess the secondary structure of the transmembrane domains TM1 and TM2 of Mic10 individually. To this end, the two peptides TM1 of 29 residues and TM2 of 37 residues both containing the transmembrane domain 1 and 2 regions of Mic10, respectively were synthesized. The synthesis was carried out by standard Fmoc solid phase peptide synthesis (Appendix 4.IX for HPLC chromatogram and mass spectrometry analysis of the peptides).

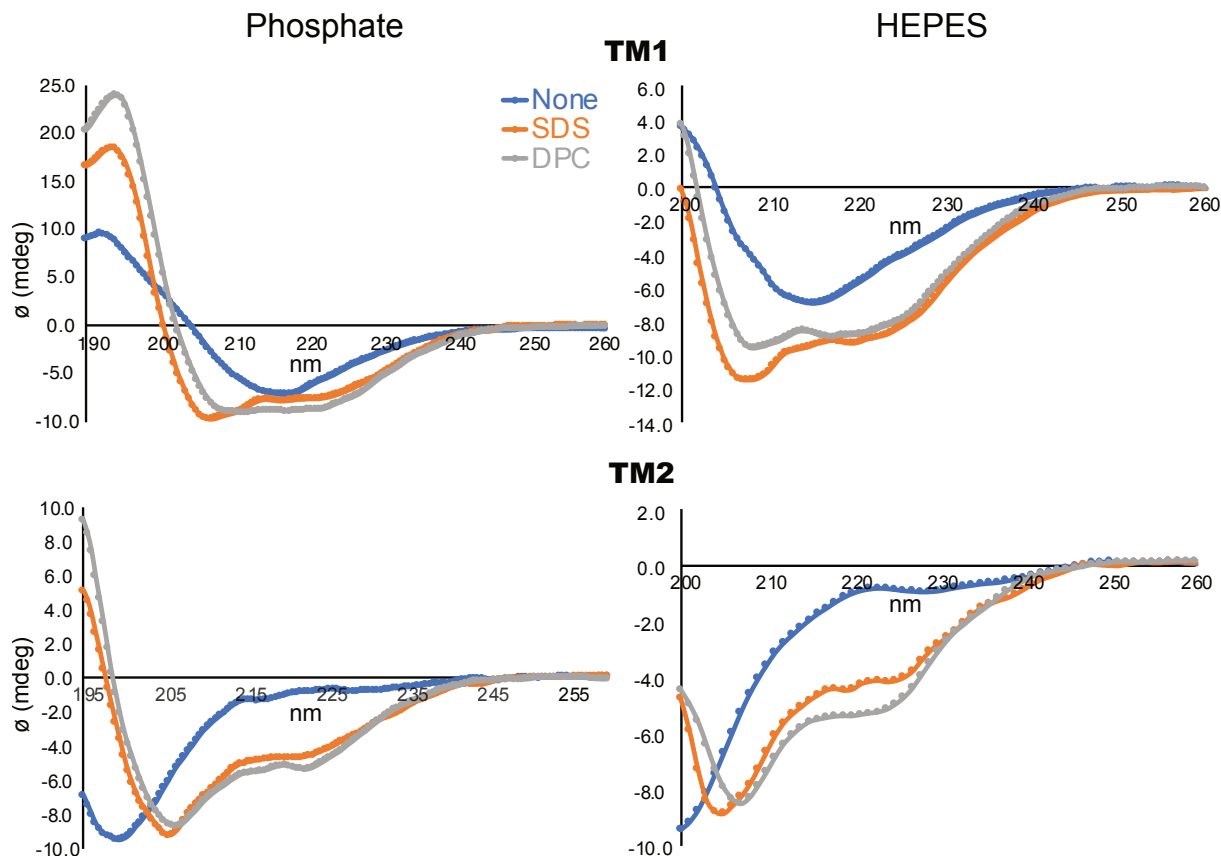


Figure 4.17. TM1 and TM2 adopt a helical conformation in the presence of detergents.

TM1 (top panels) and TM2 (bottom panels) in 20 mM sodium phosphate, 100 mM NaCl at pH 6.5 (left) and in 20 mM HEPES, 100 mM NaCl at pH 6.8 (right). The CD spectra were acquired at room temperature $\sim 25^\circ\text{C}$ in different conditions without detergents (blue), in 2% SDS (orange) and in 2% DPC (grey) using a JASCO CD spectrophotometer. The data shown was processed by subtracting the reference spectrum corresponding to the buffer without protein and then smooth function was applied.

The peptides were analyzed both separately, as well as in equimolar mixtures by CD and NMR. Both CD (Fig 4.17) and 1D proton NMR spectrum (Appendix 4.X) of the TM2 peptide in different conditions showed that TM2 domain adopts a helical structure in the presence of both SDS and DPC. Likewise TM2 peptide, the TM1 peptide is mainly in a secondary helical structure in the presence of detergent (Fig. 4.17). Furthermore, the influence of the buffer on the structure was tested for the two TM peptides. None of them, neither the TM1 nor the TM2 peptide, showed changes in their secondary structure regardless of the conditions of the buffer or the detergents tested (HEPES and phosphate) (Fig. 4.17).

To conclude, both CD and NMR data showed that the presence of detergents is necessary for the TM domains to adopt a stable helical conformation. These results, underline

the importance of studying membrane proteins in the condition closest to the native environment as possible.

3.5.2 TM2 domain forms a helical domain of at least 19 residues

By using the peptide synthesis which confers complete control of residue position specific labeling, we attempted to obtain the length of helical domain of the TM2 peptide as well as information regarding the GxxxG motifs. To these aims, three distinct TM2 peptides were synthesized: one completely unlabeled, and two others labeled at specific positions, one ^{13}C and ^{15}N labeled at position Ile11, Gly12 and Gly18 (residues Ile71, Gly72 and Gly78 in Mic10) referred as TM2-IGG. And, a second peptide with ^{13}C and ^{15}N at Gly14 and Ala20 (Gly74 and Ala80 in Mic10) referred as TM2-GA, see table 4.5. Figure 4.18 and 4.19 show the assignment of TM2 peptide using HH-TOCSY and HH-NOESY proton spectra in deuterated DPC micelles, respectively. From the HH-TOCSY spectrum, we expected 36 alpha protons in the amide region (10 to 7 ppm), however only ~28 peaks have been identified. Missing peaks are expected from overlapping peaks due to the low dispersion of the amino region and other peaks from dynamic regions that prevents NMR detection. The labeling peptides was used for confirmation of the assignment, since the labeled residues, Gly18, Gly12, Ala20 and Ile11 showed the peak splitting from $^1J_{\text{HC}}$ coupling of ~140 Hz and the $^1J_{\text{HN}}$ coupling of ~95 Hz (Fig. 4.18).

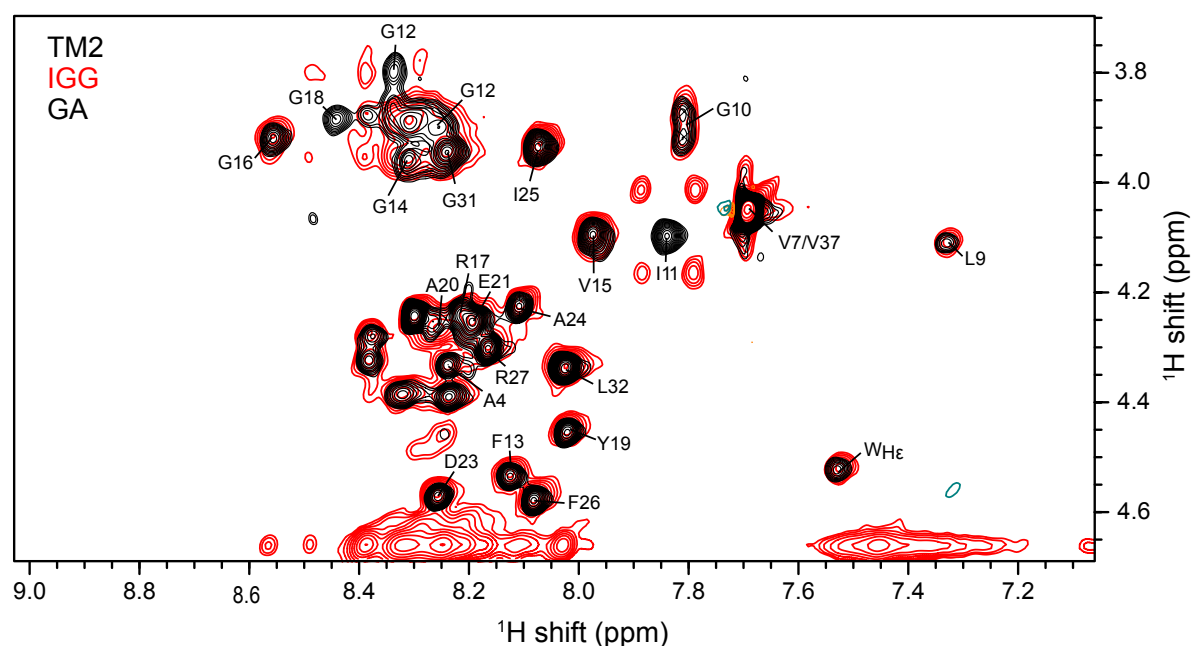


Figure 4.18. HH-TOCSY assignments of TM2.

In red, HH-TOCSY from ~800 μM of TM2-IGG labeled at position I11, G12 and G18. In black, HH-TOCSY from ~500 μM TM2-GA peptide labeled in position G14 and A20. HH-TOCSY was recorded in a 950 MHz

Bruker spectrometer using a 5 mm cryo-probe at 310 K. A spectral width of 14 ppm was used and the acquisition time was set to 77 ms in the direct dimension and 38 ms on the indirect proton dimension. Water suppression was implemented with WATERGATE. Both samples were resuspended in 20 mM sodium phosphate, 150 mM NaCl at pH = 6.5 with 2% DPC.

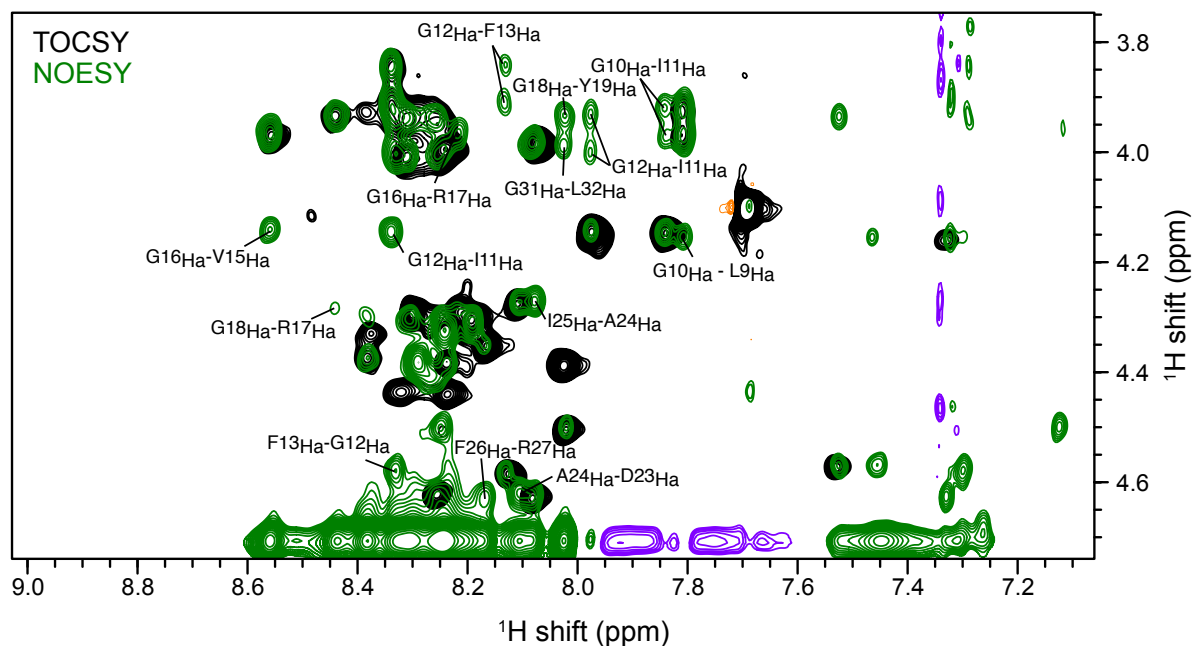


Figure 4.19. HH-NOESY assignments of the TM2-GA.

Overlay of HH-TOCSY used for assignment purposes (black) and HH-NOESY for structure determination (green) using 80 and 120 ms proton mixing, respectively. A sample of ~500 μ M of TM2-GA peptide labeled in position G14 and A20 in 20 mM sodium phosphate, 150 mM NaCl at pH = 6.5 with 2% DPC was used. NOESY and TOCSY spectra were recorded in a 950 MHz Bruker spectrometer using a 5 mm cryo-probe at 310 K. A spectral width of 14 ppm was used and the acquisition time was set to 77 ms in the direct dimension and 38 ms on the indirect proton dimension.

Further spectra, ^{13}C -HSQC and ^{15}N -HSQC, were recorded for assignment confirmation and determination of ^{13}C and ^{15}N chemical shifts. The ^1H , ^{13}C and ^{15}N backbone chemical shift are direct probes of the secondary structure in a site-specific manner. The chemical shift is directly related to the Psi, Phi and Chi torsion angles of the protein (226), which are related to the secondary conformation adopted by the protein. The most direct method is to calculate the chemical shift index (CSI) by comparing the measured chemical shifts to the random coil chemical shifts. We predicted the random coil chemical shift from the TM2 peptide amino acid sequence by using the Neighbour Corrected Intrinsically Disorder Protein Library software (NCIDP) (234). The assignment of the backbone available for most of the amino acids, including alpha protons, from residue Leu9 to Leu32 (Leu69 to Leu92 in Mic10), does not give a clear indication of TM2 peptide adopting a helical structure. All alpha proton CSI values are below the threshold of ± 0.19 (Appendix 4.XI, left), except for Ala4 (residue Ala64 in Mic10) with a positive alpha proton CSI value, indicative of a beta

sheet. Similar behavior is observed for the amide protons CSI where the threshold is only reached by Gly10 (Appendix 4.XI, right). In contrast to Ala4, Gly10 (Gly70 in Mic10) shows a helical secondary structure. Interestingly, neighboring residues Ile11 and Gly12 (Ile71 and Gly72 in Mic10), follow the same CSI trend suggesting that TM2 peptide adopts a helical structure from at least residues Gly10 to Gly12 (residue Gly70 to Gly72 in Mic10). Since in our case proton chemical shifts are not sensitive enough to report precisely the secondary structure, we measured CSI from the ^{13}C and ^{15}N chemical shifts as well. Figure 20 shows positive values of the $\delta\text{C}\alpha - \delta\text{C}\beta$ CSI for residues Ile11, Val15, Tyr19, Ala20, Ala24 and Ile25 and negative values of the measured ^{15}N CSI for residues Ile11, Gly12, Gly14, Gly18 and Ala20 indicating that these residues are in a helical structure.

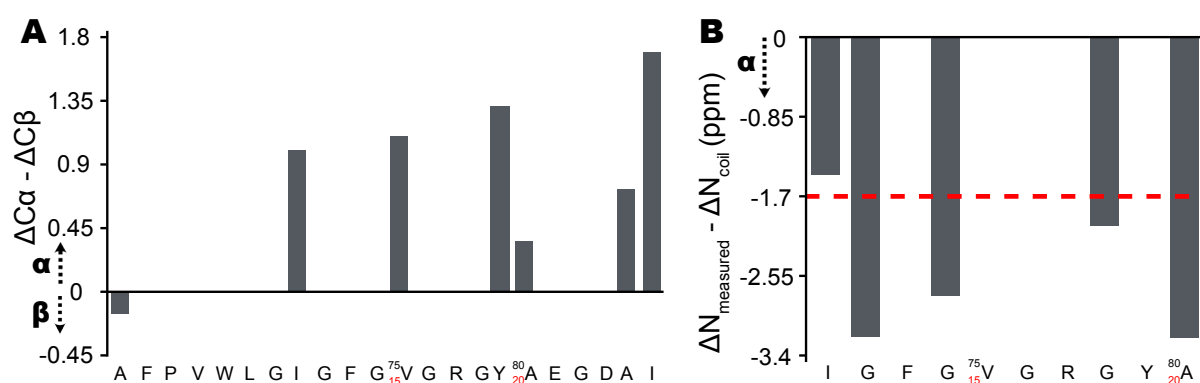


Figure 4.20. TM2 adopts an alpha helical conformation based on the chemical shift index (CSI).

The amino acid sequence is shown with the amino acid number corresponding to the Mic10 sequence in black and the peptide sequence in red. A, shows the difference between the experimental $\text{C}\alpha$ or $\text{C}\beta$ chemical shift and that of the calculated random coil carbon chemical shift. B shows the difference between observed nitrogen chemical shift and random coil nitrogen chemical shift. The red dashed line indicates the threshold for clear determination of helical structure as suggested by CSIC (Consejo Superior de Investigaciones Científicas) web page (<http://triton.iqfr.csic.es/guide/eNMR/proteins/chemshift.html>). The random coil values were calculated using the program NCIDP.

Taken together, our data suggest that the helical domain of TM2 peptide is extended at least from residue Gly10 (Gly70 in Mic10) to Ile25 (Ile85 in Mic10). Surprisingly, amide proton CSI shows a negative tendency of residues Phe26 to Leu32 (Phe86 to Leu92 in Mic10) indicating a helical conformation, which suggests an extended TM2 domain of 24 residues. Although a more extended TM2 domain is shown by our NMR data, overall, the data is consistent with prediction (see section 2.1, "Results"). The helical domain of TM2 is predicted to have a length of 19 residues that correspond to residue Ala4 to Gly22 (Ala64 to Gly82 in Mic10). It is important to recall that the TM2 peptide was measured in DPC micelles by solution NMR. Therefore, the data does not provide specific information about the length of

the TM2 domain spanning the lipid bilayer. Alternatively, the extended helical structure of the TM2 peptide shown only indicates that this region adopts a helical structure in a DPC micelles solution. This is further supported by the predicted structures which showed helical structures all along the C-terminus of Mic10. A more accurate measurement including labeled positions towards the C-terminus of the TM2 peptide would be ideal to explore this more thoroughly in micelles and lipid environments in order to fully define the length of the helical domain of Mic10's TM2 domain.

3.5.3 Structure of TM2

Finally, the NMR data obtained from the isotopically labelled TM2 peptides was used as an attempt to calculate the structure of TM2 peptide using CYANA (235). Figure 4.21 shows the ten best structures of the TM2 calculated using NOE data in combination with manual TOCSY assignments. The NOE data comprises mostly the automatically picked peaks, except for labeled residues showing splitting (these were then removed in the NOESY and TOCSY spectrum).

The final average backbone RMSD is below 1 Å and the target function, which informs about the agreement between calculation and measured data, is below four for all the cycles of the CYANA calculation run showing a good convergence of the calculation. The structure show helical structures from Leu9 to Val15 of the TM2 (Leu69 to Val75 in Mic10) and from Asp23 to Phe26 of the TM2 (Asp83 to Phe86 in Mic10; Fig. 4.21, blue residues). The second helical domain can be extended up to residue Ala30 (Ala90 in Mic10). The helical structure from residue Asp23 to Ala30 (Asp83 to Ala90 in Mic10) does not form a rigid helical structure but it rather looks like a partially loose helical structure. These observations are in agreement with the CSI data showed in figure 4.20. However, a random coil structure is found in between the two helical domains from Val15 to Gly22 (Val75 to Gly82 in Mic10). Interestingly in this region the structure bends to allow close contacts between residues of the second helical domain Asp23 to Phe26 and residues Leu9 to Val15. Surprised by the bending helical structure obtained from the calculation, we seek for unambiguous NOE contacts using the automatic tools from CcpNmr. Unambiguous contacts between residues Ile11-Ile25, Ile11-Ala24, Trp8-Ile25 and Trp8-Tyr19 are found (Appendix 4.XII, brown). This suggests that either the two extremities are in close proximity (“helical hinge”) or that the TM2 domain forms anti-parallel dimers in micelles.

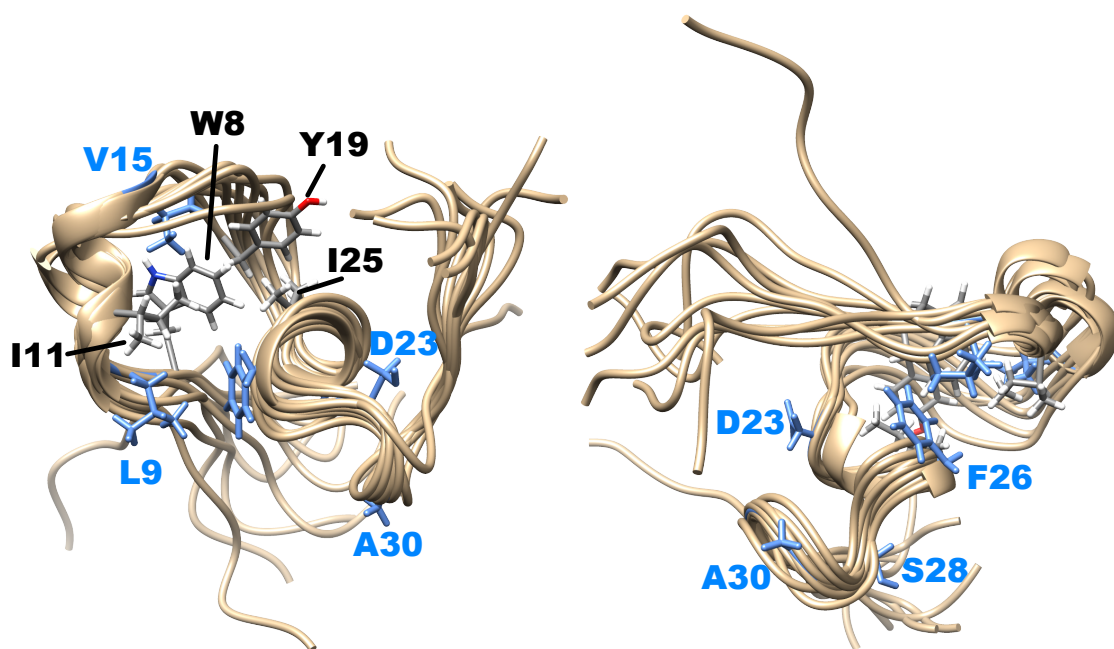


Figure 4.21. Ten best structures of the TM2 calculated by CYANA.

The structure was calculated using CYANA (235), taking only the NOE contacts as input restraints for the structure calculation. Blue residues indicates the starting amino acids of the helical structures found by CYANA (A9 to V15, D23 to F26 and S28 to A30-G31). The residues in black indicates the amino acids which have unambiguous NOE contacts (tables in appendix 4.XII). The hinge of the helix is explained by the contact between W8 and I11 with I25 and W8 with Y19. All the data was recorded with a 950 MHz Bruker spectrometer at 310 K with ~500 μ M of TM2 peptide suspended in 20 mM sodium phosphate, 150 mM NaCl and 2% DPC.

Taking all together, our data suggest that TM2 domain adopts an unexpected extended helical domain in the presence of DPC detergent from residue Leu9 to Leu32 (Leu69 to Leu92 in Mic10) which is inconsistent with the predicted *in silico* data by TMpred. However, TMpred predicts a shorter TM2 domain of 19 residues from Ala4 to Gly22 (Ala64 to Gly82 in Mic10), the experimental data is consistent with the structure prediction to a certain extent. Whereas GalaxyTBM predicts a broken but extended helical structure for the C-terminal domain of Mic10 including the TM2 domain, SIB shows a straight helical domain for the TM2 domain. A disruption of the helical structure is not expected to happen in the middle of the membranes as shown in our structure calculations (Gly14 to Asp23). It is likely an artifact of overlapping resonances, and more data would be necessary to further refine the structure of the TM2 peptide.

In order to gain insights into the length as well as the adopted conformation of the TM2 in the presence of lipids, the IGG labeled TM2 sample was reconstituted in lipid bilayers and used for solid-state NMR measurements.

3.5.4 Studies of TM2 in lipid membranes by solid-state NMR

To investigate the TM2 structure by solid-state NMR ^{13}C and ^{15}N labelled at IGG in DPhPC at 230 K and in DMPC/DMPS using a lipid protein ration (LPR) of 1 to 1 and 4 to 1 by weight, respectively (Fig. 4.22 and 4.23). Interestingly, the reconstitution in different lipid bilayers resulted in different chemical shifts in the (H)NH spectrum suggesting different protein arrangement(s) depending on the lipids. In the presence of DPhPC (Fig. 4.22, red) the observed peaks are sharper than the those in DMPC/DMPS membranes. Furthermore, the (H)CH spectrum showed strong differences of peak positions both within and between the spectra of the two different lipid compositions. The major peaks corresponding to Ile11 (Ile71 in Mic10) $\text{C}\alpha$ and $\text{C}\beta$ chemical shift could be identified in both cases by the predicted chemical shifts. However, in the presence of DMPC/DMPS, two sharp additional $\text{C}\alpha$ peaks were identified, which suggests the presence of multiple TM2 peptide conformations in our preparation (Fig. 4.23A, black). Since DMPC and DMPS lipids have phase transition temperatures of about 20°C, a (H)CH spectrum with a temperature above the phase transition was recorded. Interestingly, whereas at this higher temperature (~35°C) in the sample reconstituted with the mixture lipids (DMPC/DMPS) the isoleucine with a carbon chemical shift of at ~60 ppm is missing, the peaks at ~65 and ~72 ppm are still detectable and are sharper (Fig. 4.23B, green).

The $\text{CSI}(\partial\text{C}\alpha-\partial\text{C}\beta)$ values were calculated for the different conformational states shown by the TM2 peptide. The peak at ~60 ppm, observed in both lipid compositions, indicates a beta conformation while the other two carbon shifts, at 65 and 72 ppm, visible only in DMPC/DMPS membranes indicate a helical conformation. In addition to the conformational plasticity observed, the peak narrowing at the (H)CH recorded at high temperature suggests that TM2 peptide experiences certain mobility. Nevertheless, the mobility must be restricted because of the nature of the magnetization transfer used based on cross polarization (CP) indicating a certain order or reduced degree of mobility in the samples, such as uniaxial rotation within the membrane. The data obtained in DMPC/DMPS lipid mixtures suggest that TM2 peptide is a helical transmembrane domain that undergoes rapid uniaxial rotation in the lipid bilayer, and is sensitive to reduction of temperature, as well as to lipid composition.

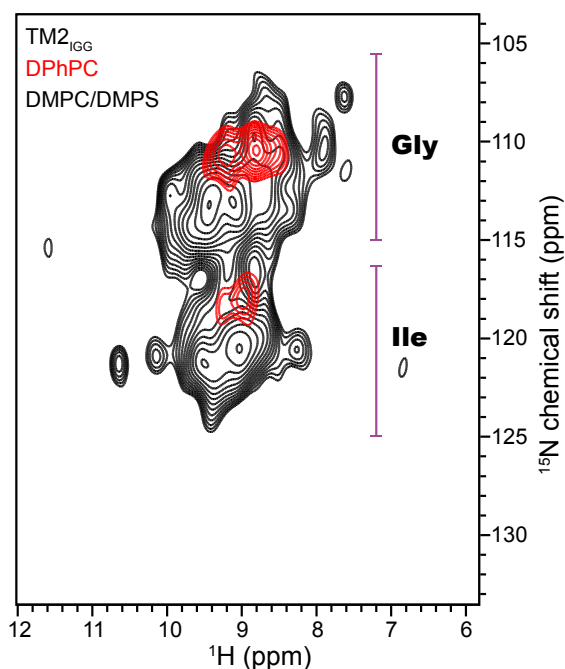


Figure 4.22. TM2 reconstituted in lipid bilayers displays inhomogeneous broadening indicative of conformational heterogeneity.

(H)NH spectra of TM2-IGG reconstituted in DPhPC (red) and in mixed lipid bilayers DMPC/DMPS (black). TM2 was reconstituted using a 1 to 1 lipid protein ratio by weight (LPR) in DPhPC while a 4 to 1 LPR by weight was used for DMPC to avoid excessive protein protein interactions.

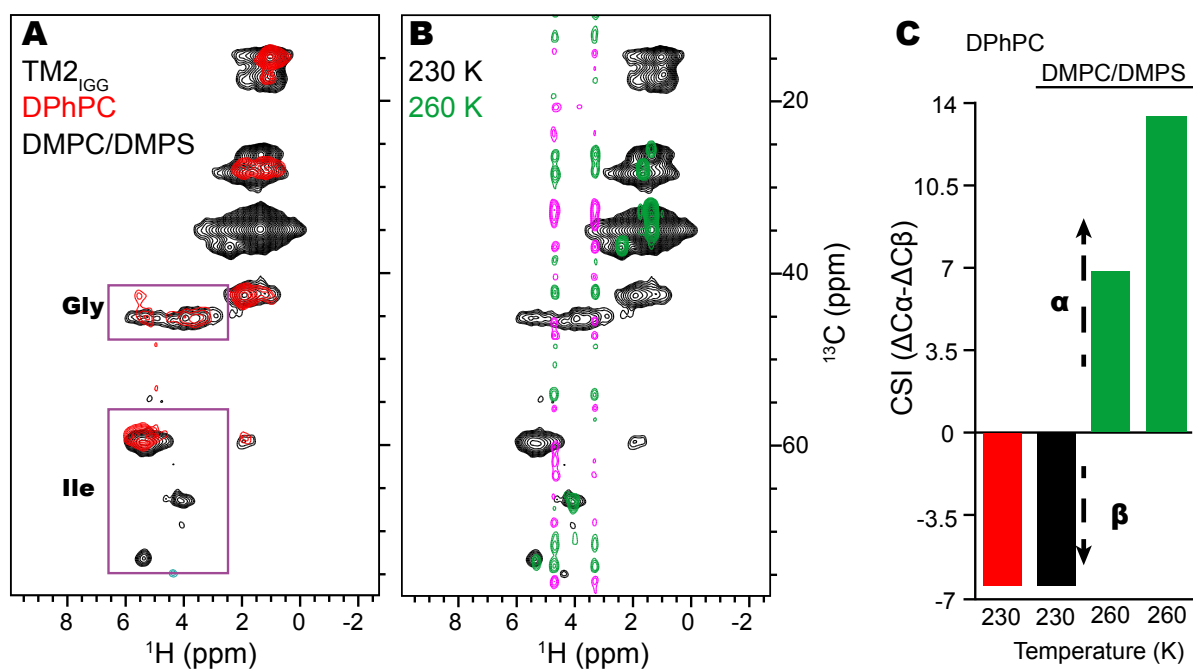


Figure 4.23. TM2 adopts different conformations in lipid bilayers.

In A, the (H)CH spectra of TM2-IGG in DPhPC (red) and DMPC/DMPS (black) are compared at 230 K (sample temperature ~15-20°C). In B, the (H)CH spectra in DMPC/DMPS at 230 K (black) and at 260 K (green) (sample temperature ~35-40°C) are shown. Panel C, shows the CSI extracted from the carbon chemical shift of the ^{13}C isoleucine label at different temperatures and lipids, 230 K (red in DPhPC and black in DMPC/DMPS) and 260 K (green in DMPC/DMPS).

Based on our results obtained in lipid bilayers, we could hypothesize that the presence of charged lipid species are necessary for the proper insertion of TM2 domain into a lipid bilayer. Alternatively, it is possible that the TM2 domain accomplish a conformational transition depending on the local environment and temperature. This dependence, as well as

the high solubility of TM2 domain in water suggests that the TM2 domain is an ‘induced helix’ that forms only under specific environmental conditions. It may be stabilized by TM1 domain (absent in these measurements) as well as other proteins of the complex. Additional measurement with different lipids varying thickness, head groups and LPR will be necessary to understand the influence of the lipids on the protein conformation.

3.5.5 No evidence of interaction between TM1 and TM2 peptides in detergent

To investigate the interaction between the TM1 and TM2-IGG peptides were mixed in a molar ratio of one to one were studied by solution NMR in the presence of DPC micelles (Fig. 4.24, black). The overlay of the spectra of TM1, TM2-IGG and TM1/TM2-IGG mixture acquired at 950 MHz in a Bruker spectrometer at 310 K are shown in figure 24 in blue, red and black, respectively. The spectrum obtained for TM1 peptide shows broad and poor signal dispersion suggesting aggregation of the TM1 peptide in these conditions. Nevertheless, we could access H α of glycine due to its particular position and the two symmetric H α peaks in the spectra. H α 's of glycine in an alpha helical conformation show a strong high field chemical shift (below 4 ppm). Here, H α glycine of TM1 peptide shows a proton chemical shift of ~3.5 ppm which is consistent with an alpha helical conformation.

When the mixture of both non labeled TM1 and TM2-IGG ¹³C,¹⁵N labeled at position Ile11, Gly12 and Gly18 (positions Ile71, Gly72 and Gly78 in Mic10) peptides was investigated, no chemical shift perturbations (CSPs) were observed on the HH-TOCSY spectrum. The absence of CSP indicates that TM1 and TM2 do not interact. However, the signal from the mixed sample showed a reduced intensity of the peaks from the TM1 peptide suggesting that there may be some modulation of the TM1 aggregation by the presence of TM2 (larger aggregates would be undetectable by solution NMR). This is also supported by the calculated glycine line width of TM1 peptide, which in the presence of TM2-IGG peptide are narrowed by ~15%. NMR line widths are related to the transverse relaxation (T_2) by $1/\pi T_2$, which is directly related to the correlation time of the molecule. Thus, the narrowing of the lines suggests a smaller particle size. However, the disappearance of certain peaks of the TM1 peptide (Fig. 4.24, black vs blue) is mostly correlated to a longer correlation time thus a higher molecular size or a dynamic regime not observed in our NMR experiments.

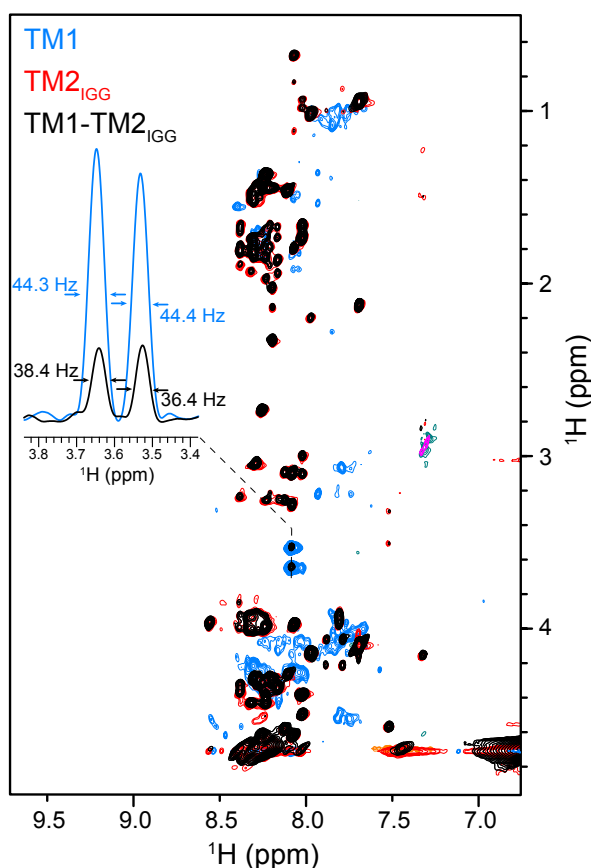


Figure 4.24. No chemical shift perturbations observed on a mixed TM1-TM2 sample.

Overlay of HH-TOCSY of TM1 (blue), TM2-IGG (TM2_{IGG}, red) and TM1-TM2_{IGG} (black). The spectra were acquired on a 950 MHz Bruker spectrometer at 310 K in 20 mM sodium phosphate, 100 mM NaCl at pH 6.5 with 2% DPC. The inset shows 1D slices of the TM1 (blue) and TM1-TM2_{IGG} (black) with their respective line widths.

It should be noted that our studies are performed in the presence of detergents that could strongly affect weak protein-protein interactions. Reconstituting the TM1/TM2-IGG mixture into lipid bilayers might improve the interaction between the TM helices and could help to decipher the structure occurring in membranes.

4. DISCUSSION

4.1 MIC10 FORMS AN INHOMOGENEOUS SAMPLE NOT SUITABLE FOR SSNMR STUDIES

To this day, neither *in vitro* nor *in vivo* data have been able to directly support the helical conformation of both Mic10's transmembrane domains. So far, only indirect evidence using bioinformatic approaches were provided to support this assumption (194, 203). As previously proposed by Barbot *et al.* and Bohnert *et al.* 2015, Mic10 from yeast adopts a hairpin like structure in the IMM with transmembrane domains of different length. This seminal work proposes a membrane bending mechanism based on the unequal length of the helical domains (194). However, no structural data was provided to back up this hypothesis. The *in silico* data obtained in this study using two different servers (SIB and GalaxyWEB) does not consistently predict neither the length and structure of the TM1 nor TM2 domains

(Fig. 4.6). This highlights the importance of carrying out structural studies on Mic10. This would allow us to determine the exact length of both transmembrane domains and understand the implication of homo-oligomerization in inducing membrane curvature.

The solid-state NMR data showed that Mic10 reconstituted in lipid bilayers forms inhomogeneous protein arrangements. The sample heterogeneity can be attributed to the formation of homo-oligomers of various size as observed on the SDS-PAGE and Western Blot (Appendix 4.V and fig. 4.7). The inhomogeneity of the sample preparation gives rise to broad lines in the solid-state NMR spectra (Fig. 4.11, 4.12 and 4.13).

4.2 MOST OF THE GLYCINES ARE FOUND IN A HELICAL CONFORMATION

The CSI measured in micelles by solution NMR provides evidence of the glycines in both of the TM domains of Mic10 being of helical conformation. (Fig. 4.20) Previous work by Barbot *et al.*, 2015; Bohnert *et al.*, 2015 associated the GxxxG motifs to the intermolecular helix - helix interaction of Mic10 (194, 203). The promoting helix-helix interaction role of GxxxG motifs have been observed in other important membrane protein such as F₁-F₀ ATP synthase (189) and have been well characterized using small peptides containing glycine rich motifs (230). This later study by Anderson *et al.* showed that the helix - helix interaction is stabilized by a combination of C α -H to C_O hydrogen bonds and Van der Waals interactions (230). In Barbot *et al.* the double mutation G50A and G52A destabilized oligomeric states of Mic10 *in vitro* as well as *in vivo* (194). The loss of Mic10 oligomerization strongly affects membrane-bending property of the protein. Both mutations (G50A and G52A), located in the TM1, are necessary to disrupt the oligomerization. Interestingly, the locations of these two mutations face the opposite direction on the surface of the helix suggesting that oligomerization of Mic10 occurs in multiple faces of the TM1 helix domain. Pfannner and co-workers introduced a series of mutations to replace most of the glycines and could show that the GxxxG motifs are responsible for the Mic10 - Mic10 oligomerization but they are not involved in the interactions between Mic10 with other MICOS components (203).

The solution NMR data on the full-length Mic10 showed that out of fourteen glycines nine exhibit a C α chemical shift above 46 ppm consistent with the expected values for a glycine in helical conformation (Fig. 4.10). This data, in combination with the solid-state NMR data of the TMs in mixed lipids (Fig. 23), strongly support the helical conformation of

Mic10 in membrane environments. Furthermore, a HH-TOCSY NMR spectrum of TM1 peptide in DPC micelles shows glycine H α protons with a chemical shift of 3.5 ppm which agrees with the expected chemical shifts of glycines normally observed in a helix conformation (Fig. 4.24). Moreover, the peak broadening together with the low chemical dispersion suggest that TM1 peptide has a strong degree of aggregation (Fig. 4.24). This observation further supports the previous observation that the aggregation is most likely promoted by the TM1 domain since the double mutation of glycine to alanine at residues 50 and 52, both located in the TM1 of Mic10, abolishes Mic10's oligomerization (194).

4.3 EVIDENCE OF THE IMPLICATION OF THE SOLUBLE DOMAINS IN THE MEMBRANE BENDING MECHANISM

Despite the fact that *in silico* data showed lower probability of oligomerization for the TM1 than for the TM2, the narrow lines on the NMR spectrum of TM2 in DPC micelles suggest that TM2 is less prone to aggregation (Fig. 4.18). Therefore, we could assign and identify helical regions of TM2 peptide with high confidence. Helical structure in TM2 is observed from residue Ile11 to Val15, from Gly18 to Ile25 and may be extended to Leu32. Interestingly, the length of the TM2 in detergents seems to be larger than the 19 residues predicted by the bioinformatic methods. This data supports the mechanism proposed by Meinecke and co-workers regarding membrane bending that the unequal length of the TM domains was responsible for membrane curvature (236). Although our data indicates a longer TM2 domain in micelles, it does not provide direct evidence on the length of the membrane spanning residues. Thus, membrane bending might also occur or be assisted by the extra helical region present on the C-terminus of Mic10. The extra helical domain (residue Ile25 to Leu32) could, instead of being part of the TM2 domain, be lying on the membrane surface thus interacting with the phospholipid head groups to induce or assist membrane curvature. An important limitation to note from our condition is the simplistic model used. We are working with short TM fragments individually. In nature, these two sequences are naturally in close proximity by a short positively charged linker (KRRK). This will constrain the distance between the TM helices in the membrane favoring the interaction between them. It will be interesting to ligate synthesized TM1 and TM2 peptides to conduct ssNMR studies with differently labelled positions in order to characterize any TM interaction that might occur.

Additionally, our enzymatic treatment results in a SDS-PAGE gel pattern with missing high molecular size bands (Fig. 4.16). These results might suggest that the accessible regions to the proteases could have an important role for the *in vitro* stabilization of Mic10His oligomers. Or alternatively, that the two transmembrane domains (TM1-TM1, TM1-TM2 or TM2-TM2) or multiple of them are in a stable interaction forming a complex with a molecular weight of ~12 kDa. Only the first hypothesis seems plausible because if the TM oligomerization hypothesis would be valid, the same band would be expected in both treatments. The importance of the soluble domains of other MICOS sub-complexes have been shown in recent studies by Zhiyin Song *et al.* regarding Mic60 complex. In this study, the authors showed that Mic19 – Mic60 interaction is sustained by the extracellular domain of the membrane protein Mic60 (202). The later study and others have shown the implication of the soluble domains of MICOS sub-units in the architecture of the IMM (199, 202). This raises the question whether the extracellular domains of Mic10 also play a role in the oligomerization or interaction with other members of MICOS.

4.4 THE SOLID-STATE NMR DATA SUGGESTS STRUCTURAL PLASTICITY OF TM2 IN MEMBRANE BILAYERS

Our solid-state NMR data obtained in conditions closer to native environments (lipid bilayers composed of DPhPC and DMPC/DMPS) indicates that TM2 domain co-exists in different conformations in the presence of lipid membranes (Fig. 4.23). In our experiments the different species can be related to the lipid transition from L β to L α phase upon temperature increase and/or the mass of lipid to protein ratio (1:1 versus 1:4 by weight). The CSI of isoleucine shows that TM2 domain adopts alpha (TM2 α) and beta (TM2 β) conformations at temperature lower than the phase transition temperature of the lipids. Whereas at higher temperatures above the phase transition of the lipids (~35°C), the beta sheet population disappears and only TM2 α is observed (Fig. 4.23).

Two scenarios may explain this phenomena happening in our sample at high temperature in the mixed membranes (DMPC/DMPS). Since the sample is packed into the rotor via a centrifugation step to pellet the sample no or little soluble protein can be expected. Additionally, the soluble protein fraction is filtered out from the ssNMR spectra by the mechanism of magnetization transfer, cross polarization, used here which is sensitive to protein motion. This suggest that either the temperature leads to a conformational transition

beta to alpha or that the beta population is interacting with the membrane surface. The interaction of TM2 β with membrane surface is reduced at higher temperatures which increases its mobility thus reducing the magnetization transfer efficiency in ssNMR via cross-polarization. An alternative situation is that the surface bound TM2 β is incorporated to the membranes once the temperature transition of lipid is reached. Although, since the DPhPC does not show a phase transition in our working temperature range, it is hard to believe that in DPhPC only membrane bound TM2 β exist and none of it spans through the membrane. However, the structure calculation of TM2 in micelles shows that the peptide forms a hinge (Fig. 4.21). This particular structure could be used by the TM2 as an anchor to bind to the lipid membrane.

Protein conformational transition occurs in different scenarios and are particularly important due to its incidence in prion like diseases or amyloid pathologies. For example, random coil, alpha or even beta structures turn into beta sheets which are then favorable for stacking together and form fibrillar structures (237). Another, more closely related, study regarding structural transitions of the SNARE proteins, Syntaxin 1 and VAMP1, has been reported (238). The authors reported an alpha to beta conformational transition of SNARE proteins modulated by the protein lipid ratio. Interestingly, SNAREs are well known proteins involved in membrane fusion processes (239). Although further studies of this process will be necessary, this modulation of the protein conformation by the lipids seems to be consistent with our data, where TM2 reconstituted in DPhPC membranes at a 1:1 lipid protein ratio by weight is only found in a beta conformation, whereas in the DMPC/DMPS with a LPR of four to one by weight both alpha and beta species co-exist. Similarly to SNARE proteins, Mic10 is also a remodeling membrane protein indispensable for the formation of CJs found in the IMM (194, 203). Other membrane proteins involved in membrane remodeling such as HIV g41 has shown the same transition phenomena from beta to an alpha helical conformation in the sample (240). This is consistent with our hypothesis regarding the conformational transition of the TM2 domain as Mic10 IMM modulator.

5. CONCLUSION

The expression and purification of the Mic10 protein was successful, and initial NMR data have led to an improved understanding of the backbone structure and the implication of GxxxG motifs in this key protein involved in the CJ architecture. Consistent with predictions, the NMR chemical shift distribution of the amide protons suggests a helical or random coil structure. However, the full-length Mic10 protein construct in micelle solution showed a low helical propensity at the glycine region with only few glycine C α chemical shifts slightly above the random-coil conformation indicating a likely helical conformation.

Even though the NMR data has not led to a complete full-length Mic10 structure neither in detergent micelles nor lipid bilayers due to the sample stability and inhomogeneity, it was still possible to show for the first time on a residue specific basis that both transmembrane domains of Mic10 exhibit helical arrangement. Interestingly, only glycines of the TM1 domain showed H α proton chemical shifts expected for a helical conformation in DPC micelles. Furthermore, the solid-state NMR spectra in lipid bilayers of the TM2 peptide showed that it adopts at least two different secondary structures as reported by the chemical shift index. These data suggest that the TM2 population can interconvert depending on temperature, as well as lipid composition. Only the alpha helical population remained visible in the NMR spectrum at high temperature, indicating an efficient protein incorporation into the membrane for this species, while it is unclear how the more extended conformation is arranged (either in, partly out of, or on the membrane).

It would be interesting to see if addition of the components of the MICOS complex helps on the stabilization of Mic10 to conduct structural studies of the complex. Previous studies showed data depletion of Mic12 played a pivotal role in the complex integrity of Mic10 sub-complex or even the whole MICOS (241, 242). Thus, we think that co-reconstituting Mic12 and Mic10 together might generate a stable and more homogeneous complex suitable for structural studies by ssNMR.

APPENDIX

Appendix 4.I. DNA and amino acid sequence of the proteins used in this thesis.

Mic10HIS:

DNA sequence:

5'atgctgactaccatcaccatcaccatcacgattacgatatcccaacgaccgaaaacctgtatttcagggcgccatggggatcctaattccgaacaagcacaacacaacaaccagegaaaagcactccctccaaggattctaacaagaatggcagctccgtgtccacgatcctcgacactaagtgggatattgtcctgtccaacatgctggcaagactgcatgggtttggcgtgggtgtgttcacctcagttatttctcaagcgcctgcatttctgtatggctaggcattggattgggtgtggaagaggctacgccgagggcgacgccatctttagatccagcgtggcctcagatcctcgaaggttaggt-3'

Protein sequence:

MSYYHHHHHH DYDIPTTENL YFQGAMGILM SEQAQTQQPAKSTPSKDSNK
NGSSVSTILD TKWDIVLSNM LVKTAMGFGV GVFTSVLFFK RRAFPVWLG
GFGVGRGYAE GDAIFRSSAG LRSSKV

Theoretical molecular weight and isoelectric point (from ProtParam):

MM(Mic10His) = 13918.78

pI = 9.3

- Mic10strep and Mic10δ's:

Mic10strep:

DNA sequence from genscript:

5'atggcctctgaacaggctcagaccagcagccggctaaatctacccgtctaaagacttaacaaaaacggttcttctgtttctaccatcctggacaccaaattgggacatcgttctgtctaacatgctggtaaaaccgctatgggttcgggtgtggtgtttcacctctgttctgttctcaaacgtcgtgctttcccggtttggctgggtatcggttcgggtgtggtcgtggttacgctgaaggtgacgctatctccgttcttctgctggtctcggttcttctaaagttggttctggttctcaccgcagttcgaaaaataa-3'

Protein sequence:

MASEQAQTQQ PAKSTPSKDS NKNSSVSTI LDTKWDIVLS NMLVKTAMGF
GVGVFTSVLF FKRRAFPVWL GIGFGVGRGY AEGDAIFRSS AGLRSSKVAA
AGSGSWHPQ FEK

Theoretical molecular weight and isoelectric point (from ProtParam):

MM(Mic10strep) = 12020.64

pI(Mic10strep) 10.09

Mic10δ10:

DNA sequence:

5'atggccaaatctaccccgctaaagactctaacaaaaacgggtctctgtttctaccatcctggacaccaaaggacatcgttctgtct
aacatgctgggttaaaccgctatgggttcgggtgttggtgtttcacctctgttctgttctcaaacgctgctttcccggttggtgggtat
cggttcgggtgttgctggttacgctgaaggtgacgctatctccggttctctgctggctcgcgttctctaaagtgggtctggttctggct
tcaccgcagttcgaaaaataa-3'

Protein sequence:

MAKSTPSKDS NKNSSVSTI LDTKWDIVLS NMLVKTAMGF GVGVFSTVLF
FKRRAFPVWL GIGFGVGRGY AEGDAIFRSS AGLRSSKVGS GSWSHQPFEK

Theoretical molecular weight and isoelectric point (from ProtParam):

$MM_{(\text{Mic10}\delta10)} = 10738.30$

$pI_{(\text{Mic10}\delta10)} = 10.28$

Mic10δ20:

DNA sequence:

5'atggccaacgggtctctgtttctaccatcctggacaccaaaggacatcgttctgtctaacatgctgggttaaaccgctatgggttcg
gtgttggtgtttcacctctgttctgttctcaaacgctgctttcccggttggtgggtatcggttcgggtgttgctggttacgctgaa
ggtgacgctatctccggttctctgctggctcgcgttctctaaagtgggtctggttctggctcaccgcagttcgaaaaataa-3'

Protein sequence:

MANGSSVSTI LDTKWDIVLS NMLVKTAMGF GVGVFSTVLF FKRRAFPVWL
GIGFGVGRGY AEGDAIFRSS AGLRSSKVGS GSWSHQPFEK

Theoretical molecular weight and isoelectric point (from ProtParam):

$MM_{(\text{Mic10}\delta20)} = 9665.13$

$pI_{(\text{Mic10}\delta20)} = 10.27$

Mic10δ30:

DNA sequence:

5'atggccaccaaatgggacatcgttctgtctaacaatgctgggtaaaaccgctatgggttcggtgttggtgtttcacctctgttctgttctc
aaacgtcgtgctttcccggtttgctgggtatcggtttcggtgttggtcgtggttacgctgaaggtgacgctatcttccgttcttctgctggtc
tgcgttcttctaaagtgggttctggttcttggctcaccgcagttcgaaaaataa-3'

Protein sequence:

MATKWDIVLS NMLVKTAMGF GVGVFTSVLF FKRRAFPVWL GIGFGVGRGY
AEGDAIFRSS AGLRSSKVGS GSWSHPQFEK

Theoretical molecular weight and isoelectric point (from ProtParam):

$MM_{(\text{Mic10}\delta30)} = 8691.10$

$pI_{(\text{Mic10}\delta30)} 10.61$

Mic10δ40:

DNA sequence:

5'atggccctggttaaaaccgctatgggttcggtgttggtgtttcacctctgttctgttcttcaaacgtcgtgctttcccggtttgctgggt
atcggtttcggtgttggtcgtggttacgctgaaggtgacgctatcttccgttcttctgctggctcgcgttcttctaaagttggttctggttcttgg
tctcaccgcagttcgaaaaataa-3'

Protein sequence:

MALVKTAMGF GVGVFTSVLF FKRRAFPVWL GIGFGVGRGY AEGDAIFRSS
AGLRSSKVGS GSWSHPQFEK

Theoretical molecular weight and isoelectric point:

$MM_{(\text{Mic10}\delta40)} = 7502.69$

$pI_{(\text{Mic10}\delta40)} 11.00$

Sequence alignment from CLUSTAL 2.1:

```
Mic10_d20 -----
Mic10_d30 -----
Mic10strep ATGGCCTCTGAACAGGCTCAGACCCAGCAGCCGGCTAAATCTACCCCGTCTAAAGACTCT
Mic10_d10 -----ATGGCCAAATCTACCCCGTCTAAAGACTCT
Mic10_d40 -----

Mic10_d20 ATGGCCAACGGTTCCTTCTGTTTCTACCATCCTGGACACCAAATGGGACATCGTTCTGTCT
Mic10_d30 -----ATGGCCACCAAATGGGACATCGTTCTGTCT
Mic10strep AACAAAAACGGTTCCTTCTGTTTCTACCATCCTGGACACCAAATGGGACATCGTTCTGTCT
Mic10_d10 AACAAAAACGGTTCCTTCTGTTTCTACCATCCTGGACACCAAATGGGACATCGTTCTGTCT
Mic10_d40 -----

Mic10_d20 AACATGCTGGTTAAAACCGCTATGGGTTTCGGTGTGGTGTTCACCTCTGTTCTGTTTC
Mic10_d30 AACATGCTGGTTAAAACCGCTATGGGTTTCGGTGTGGTGTTCACCTCTGTTCTGTTTC
Mic10strep AACATGCTGGTTAAAACCGCTATGGGTTTCGGTGTGGTGTTCACCTCTGTTCTGTTTC
Mic10_d10 AACATGCTGGTTAAAACCGCTATGGGTTTCGGTGTGGTGTTCACCTCTGTTCTGTTTC
Mic10_d40 ATGGCCTGGTTAAAACCGCTATGGGTTTCGGTGTGGTGTTCACCTCTGTTCTGTTTC
* *****

Mic10_d20 TTCAAACGTCGTGCTTTCCCGTTTGGCTGGGTATCGGTTTCGGTGTGGTTCGTGGTTAC
Mic10_d30 TTCAAACGTCGTGCTTTCCCGTTTGGCTGGGTATCGGTTTCGGTGTGGTTCGTGGTTAC
Mic10strep TTCAAACGTCGTGCTTTCCCGTTTGGCTGGGTATCGGTTTCGGTGTGGTTCGTGGTTAC
Mic10_d10 TTCAAACGTCGTGCTTTCCCGTTTGGCTGGGTATCGGTTTCGGTGTGGTTCGTGGTTAC
Mic10_d40 TTCAAACGTCGTGCTTTCCCGTTTGGCTGGGTATCGGTTTCGGTGTGGTTCGTGGTTAC
*****

Mic10_d20 GCTGAAGGTGACGCTATCTCCGTTCTTCTGCTGGTCTGCGTTCTTCTAAAGTTGGTTCT
Mic10_d30 GCTGAAGGTGACGCTATCTCCGTTCTTCTGCTGGTCTGCGTTCTTCTAAAGTTGGTTCT
Mic10strep GCTGAAGGTGACGCTATCTCCGTTCTTCTGCTGGTCTGCGTTCTTCTAAAGTTGGTTCT
Mic10_d10 GCTGAAGGTGACGCTATCTCCGTTCTTCTGCTGGTCTGCGTTCTTCTAAAGTTGGTTCT
Mic10_d40 GCTGAAGGTGACGCTATCTCCGTTCTTCTGCTGGTCTGCGTTCTTCTAAAGTTGGTTCT
*****

Mic10_d20 GGTTCTTGGTCTCACCCGCAGTTCGAAAAATAA-----
Mic10_d30 GGTTCTTGGTCTCACCCGCAGTTCGAAAAATAA-----
Mic10strep GGTTCTTGGTCTCACCCGCAGTTCGAAAAATAAGCGGCCGA
Mic10_d10 GGTTCTTGGTCTCACCCGCAGTTCGAAAAATAA-----
Mic10_d40 GGTTCTTGGTCTCACCCGCAGTTCGAAAAATAA-----
*****
```

- MBPMicX:

MBPMic12:

Mic12 DNA Sequence from genscript:

5'atgagcaaacctgggcccgcctggcgcgtagcgtgaagtggaccctgagcgtgggtgtgattggtagcgtgttctacctgtatcggtac
agcaacaacggttatatttacgaccacgatgcgacctggctgaagcaggaccaccaggtgcaagacctggttgatcgtaaagaggtgg
ttccgggcgaaaccgtaaccgtaaacctgggtggtaccgacgatggtaccgctggagccgtaccatgggcgagagcatcaaagata
tttgaacgaacaaatccgtaacagcgtggattggtatctatagctggggcaagaataagcttcatcatcatcatcactaa-3'

MBPMic12 Protein sequence (pMALc4x plasmid):

MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ
VAATGDGPD IFWAHDRFGG YAQSGLLAEI TPKAFQDKL YPFTWDAVRY
NGKLIAYPIA VEALSLIYNK DLLPNPPKTW EEIPALDKEL KAKGKSALMF
NLQEPYFTWP LIAADGGYAF KYENGGYDIK DVGVDNAGAK AGLTFLVDLI
KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK VNYGVTVLPT
FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG
LEAVNKDKPLGAVALKSYEE ELVKDPRIAA TMENAQKGEI MPNIPQMSAF
WYAVRTAVIN AASGRQTVDE ALKDAQTNSS SNNNNNNNNN NLGIEGRISE
FLVPRGSMK LGPLARSVKW TLSVGVIGSV FYLYRYSNNG YFYDHDATWL
KQDHQVQDLV DRKEVVPGET RNRKLVVTTD GTAWSRTMGE SIKDIWNEQI
RNSVDWIYSW GKNKLHHHHH H

Theoretical molecular weight and isoelectric point:

$MM_{(MBPMic12)} = 57030.36$

$pI_{(MBPMic12)} = 5.85$

MBPMic13:

Mic13 DNA sequence from genscript:

5'atgggtggcgcgtgtgtggagcctgatgcgttctctgattaagggtagcgtggcgggtggtgcggtgtatctgggttacgaccaggag
ctgctgggtccgagcgcgataagagccaggcggcgtgcaaaaagcgggcgaagtgggtccgcccgcgatgtaccagttcagccaata
tgttgccagcaaacggctctcaaatcccgaactgccggtccgcccgaagatctactttccgattcgtgacagctggaacgcgggc
attatgaccgtgatgagcgcgctgagcgttgcgccgagcaaggcgcgtgagtatagcaaaagagggttggaatatgtaa-3'

MBPMic13 Protein sequence (pMALc4x plasmid):

MKIEEGKLV I WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ
VAATGDGPD I IFWAHDRFGG YAQSGLLAEI TPKAFQDKL YPFTWDAVRY
NGKLIAYPIA VEALSLIYNK DLLPNPPKTW EEIPALDKEL KAKGKSALMF
NLQEPYFTWP LIAADGGYAF KYENKDYDIK DVGVDNAGAK AGLTFLVDLI
KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK VNYGVTVLPT
FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG
LEAVNKDKPLGAVALKSYEE ELVKDPRIAA TMENAQKGEI MPNIPQMSAF
WYAVRTAVIN AASGRQTVDE ALKDAQTNSS SNNNNNNNNN NLGIEGRISE
FLVPRGSMVA RVWSLMRFLI KGSVAGGAVY LVYDQELLGP SDKSQAALQK
AGEVVPAMY QFSQYVCQQT GLQIPQLPAP PKIYFPIRDS WNAGIMTVMS
ALSVAPSKAR EYSKEGWEYV KARTKKLHHH HHH

Theoretical molecular weight and isoelectric point:

$$MM_{(\text{MBPMic13})} = 57729.71$$

$$pI_{(\text{MBPMic13})} 6.20$$

Appendix 4.II.

Table S4.1. Minimal media M9 composition (For 1 Litre of M9).

Autoclave H ₂ O	756 ml
Trace elements	20 ml
Thiamine (20 mg/ml)	1 ml
Biotin (5 mg/ml)	3 ml
CaCl ₂ (2 M)	50 µl
MgSO ₄ ²⁻ (1 M)	2 ml
Glucose (non labelled or ¹³ C) (20% stock)	4 g (20 ml)
NH ₄ Cl or ¹⁵ NH ₄ Cl (250 mg/ml stock)	1 g (4 ml)
M9 salts (5X)	200 ml

Trace elements preparation:

Chemicals	Mass in mg for 100 ml final solution
FeSO ₄ 7H ₂ O	600
MnCl ₂ 4H ₂ O	115
CoCl ₂ 6H ₂ O	80
ZnSO ₄ 7H ₂ O	70
CuCl ₂ 2H ₂ O	30
H ₃ BO ₃	2
(NH ₄) ₆ Mo ₇ O ₂₄ 4H ₂ O	25
EDTA	500

The chemicals were added from top to bottom up till (NH₄)₆Mo₇O₂₄ 4H₂O in 20 - 30 ml of miliQ water. The solution was then kept stirring for 10 - 15 min before adding EDTA. After addition of EDTA the solution was filled up to 100 ml and keep stirring over night at room temperature. Before usage, the trace elements solution was filter sterile using a 0.35 µm filter and stored at room temperature.

M9 salts (phosphate buffer):

Chemicals	Mass (g) for 1 litre
Na ₂ HPO ₄	33.9
KH ₂ PO ₄	15
NaCl	2.5

M9 salts solution was prepared by adding the chemicals in 800 ml miliQ water. The pH was adjusted at 7.4, then the solution was fill up to 1 litter and autoclave.

Appendix 4.III.

Table S4.2. Mic10 δ s PCR primers and the expected number fo amino acids colony PCR protocol tables.

And,

Table S4.3. PCR protocol used for insert amplification

Table S4.2. Mic δ s PCR primers and the expected number of amino acids for each construct.

Name	Forward primer (melting temperature)	Reverse primer (melting temperature)	Expected protein size (aa)
Mic10 δ 10	5'TA CC ATG GC C AAATCTACCCCGTCTAAAGACTCTAACAAAAAC3' (68°C)	5'GTGCGGCCGCTTAAACCAC3' (70°C)	100
Mic10 δ 20	5'TA CC ATG GC C AACGGTTCTTCTGTGTTTCTTACCATCCTG3' (70°C)	5'GTGCGGCCGCTTAAACCAC3' (70°C)	90
Mic10 δ 30	5'TA CC ATG GC C ACCAAATGGGACATCGTTCTG3' (68°C)	5'GTGCGGCCGCTTAAACCAC3' (70°C)	80
Mic10 δ 40	5'TATA CC ATG GC C CTGGTTAAAAACCCTATGGGTTTC3' (69°C)	5'GTGCGGCCGCTTAAACCAC3' (70°C)	70

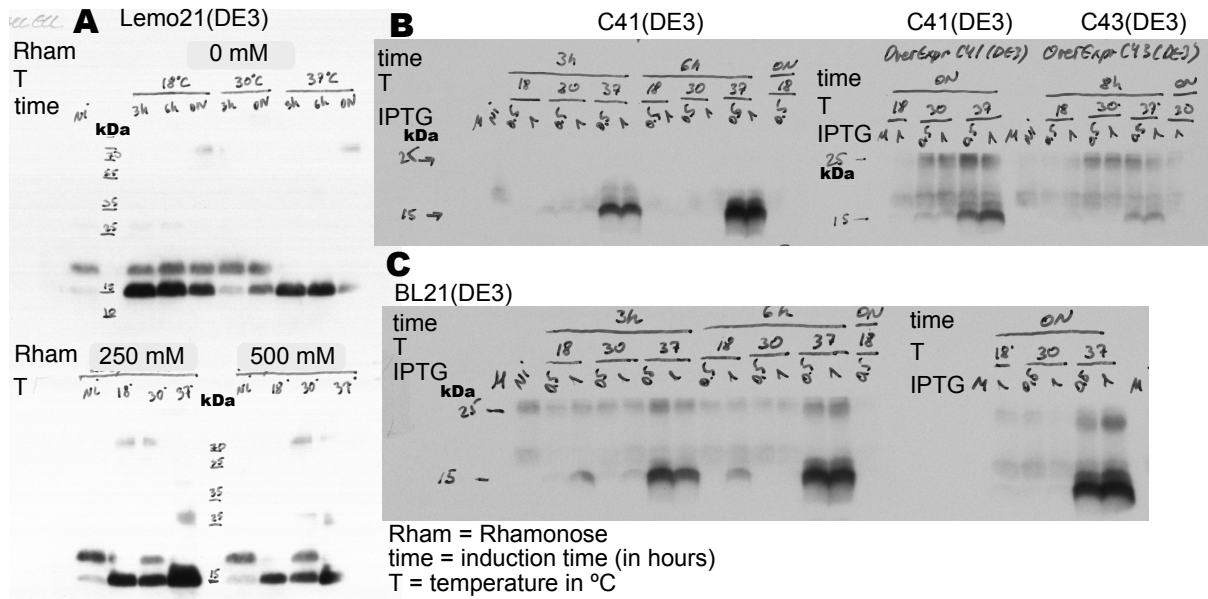
Table S4.3. PCR protocol used for insert amplification.

Step	Temperature (°C)	Time (min)	Cycles
Denaturation	95	1	25
Annealing	55	1	
Extension	72	5	
-	4	-	

Appendix 4.IV.

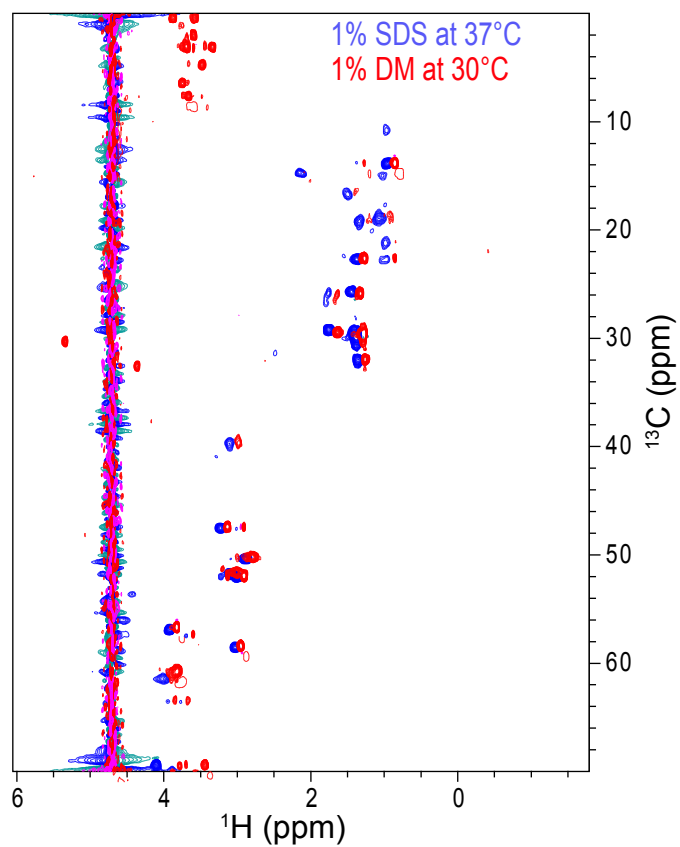
Table S4.4. Bacterial strains properties.

STRAIN	EXPRESSION SYSTEM	PROPERTIES
BL21(DE3)	T7 promotor	Cells deficient on OmpT and Lon proteases. Resistant to phage T1
C41pLysS	T7 promotor	Derived from BL21(DE3) for toxic protein
C41pLysS	T7 promotor	Derived from BL21(DE3) for toxic protein
Lemo21(DE3)	T7 promotor	Derived from BL21(DE3) for toxic protein



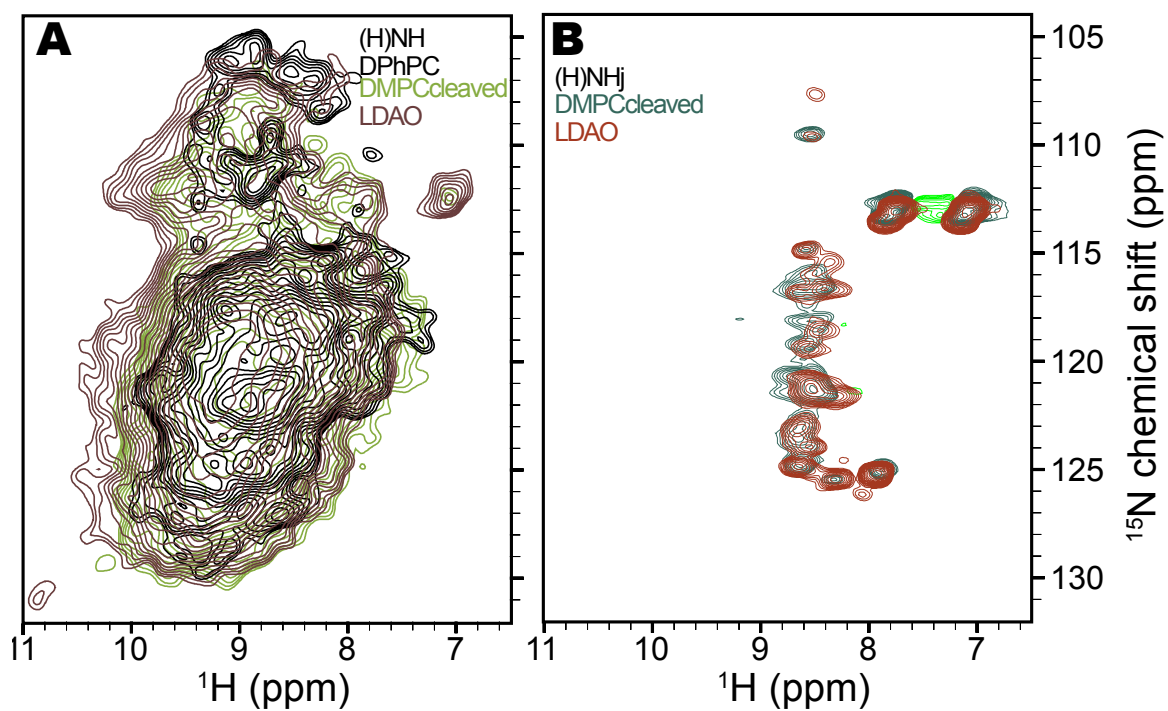
Appendix 4.V. Western blots obtained from the Mic10His test expressions.

As explained in material and methods section (2.1.3 "Test Expression of the different constructs in minimal media"), four different strains of *E. coli* were tested: Lemo21(DE3) (A), C41, C43 (B) and BL21(DE3) (C). For all the strains a wide range of conditions varying induction time (time), temperature and concentrations of IPTG and rhamnose have been tested. The best conditions were found for BL21(DE3) and Lemo21(DE3) for an overnight (~16 hours) or six hours induction respectively. The WB were performed in Meinecke's laboratory with the assistance of Dr. Daryna Tarasenko.



Appendix 4.VI. Mic10His ^{13}C -HSQC spectra obtained for different detergents.

Similar finger print ^{13}C -HSQC spectra are obtained in the presence of DM (red) and SDS (bleu) indicating that the TM2 peptide adopts a similar conformation irrespectively of the detergent used. Both spectra were recorded at 400 MHz Bruker spectrometer at the 37°C for SDS and 30°C for DDM.



Appendix 4.VII. Comparison of (H)NH spectra of the non treated and trypsin treated Mic10His.

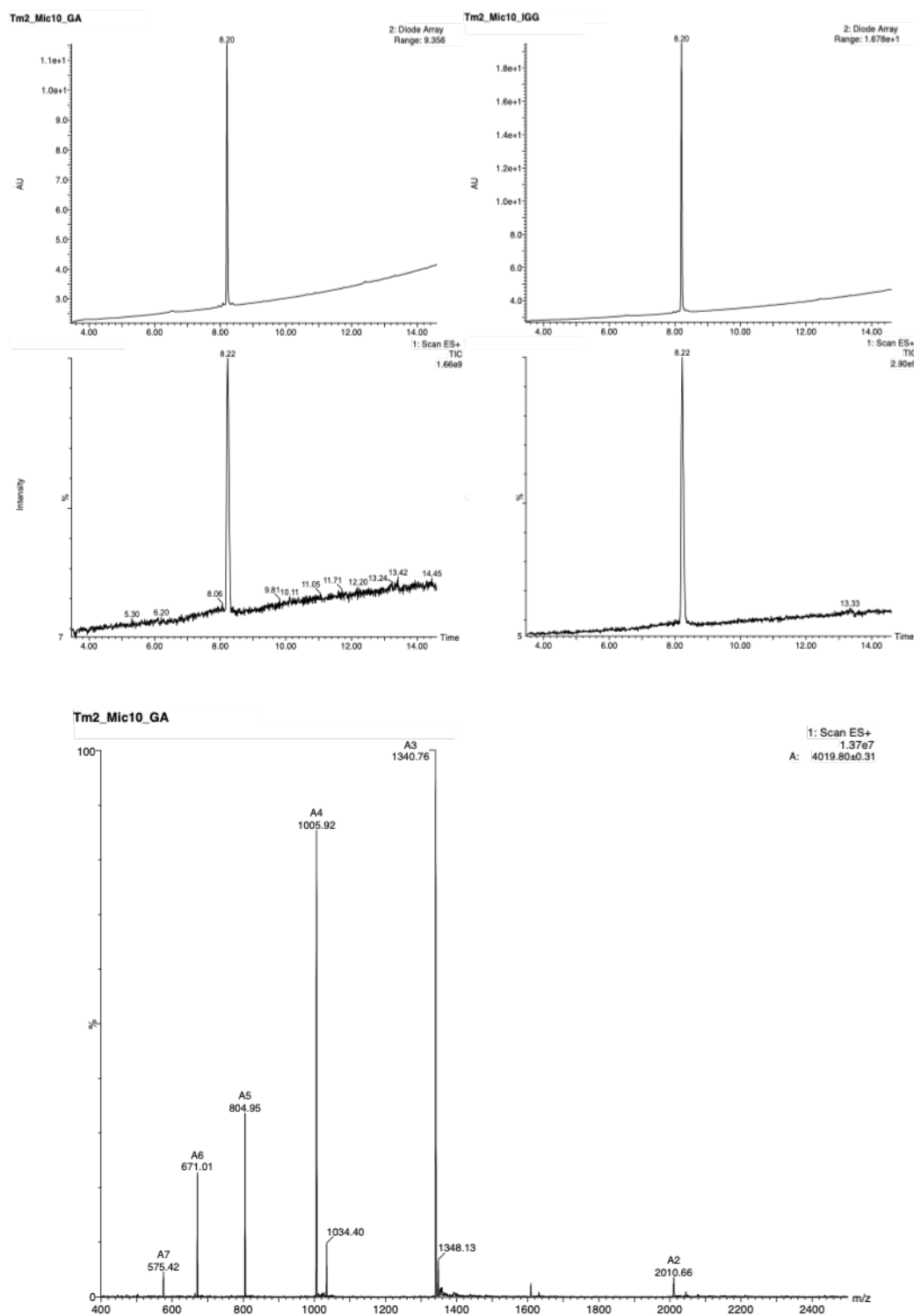
Left panel compares (H)NH spectra using cross-polarization as magnetization mechanism transfer obtained for MicHis reconstituted in DPhPC (black), LDAO (brown) and trypsin treated Mic10His in DMPC (DMPC cleaved, in green). The right panel shows the (H)NHj spectra of Mic10His in LDAO micelles (brown) and DMPC cleaved (green) recorded using INEPT based experiments.

Appendix 4.VIII. Identification of Mic10 after trypsin treatment by mass spectrometry.

The band at ~12 kDa was cut and send for mass spectrometry analysis. The regions highlighted in yellow are the identified Mic10His regions from the mass spectrometry analysis. Only 29% of sequence coverage was found.

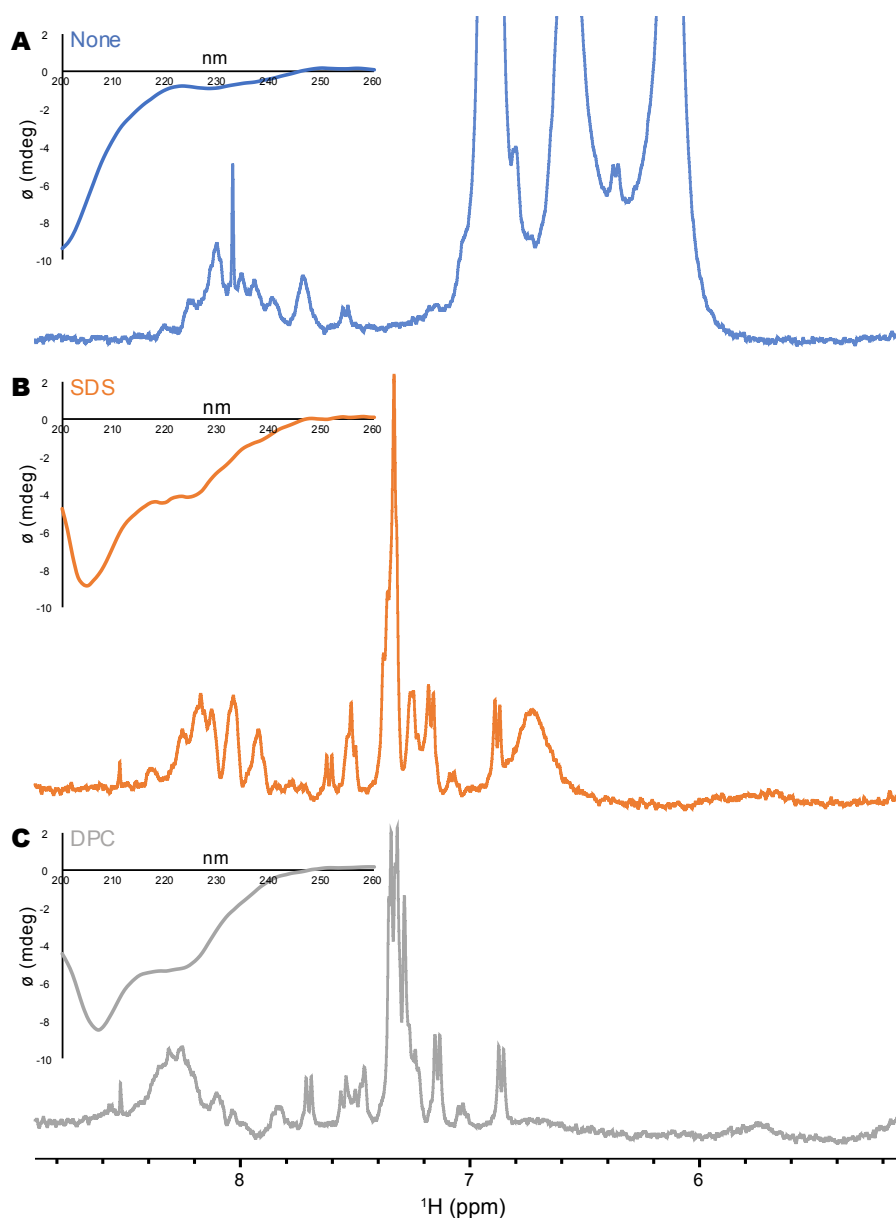
spiP1002983 (100%), 13,819.9 Da
Kumar Tekwami Mic10 seq
4 exclusive unique peptides, 4 exclusive unique spectra, 7 total spectra, 36/125 amino acids (29% coverage)

M S Y Y H H H H H H H H D Y D I P T T E N L Y F Q G A M G I L M S E Q A Q T Q Q P A K S T P S K D S N K N G S S V S T I L D T K W D I V L S N M L V K T A M G F G V G V F T S V L F F K
R R A F P V W L G I G F G V G R G Y A E G D A I F R S S A G L R S S K



Appendix 4.IX. Purity and mass control of the TM2 peptide synthesis.

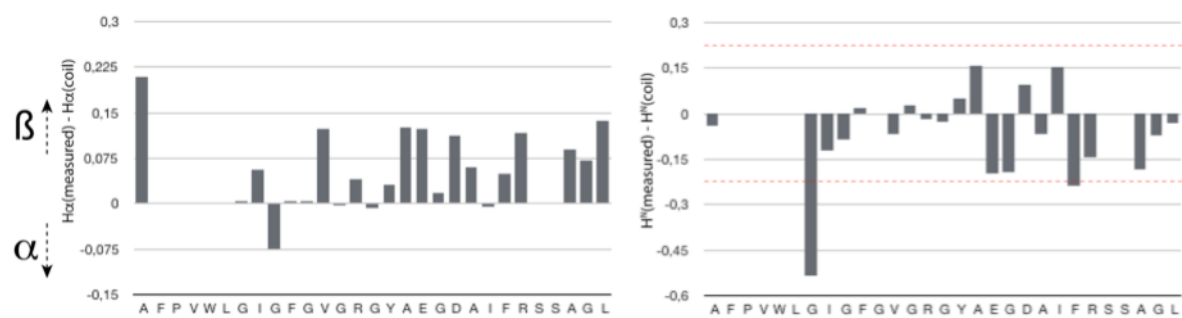
A and B show the HPLC chromatogram of the TM2 domains (TM2GA) (A) and IGG (B) labeled. Panel C, shows the mass analysis of TM2GA labeled. The data was provided by Kerstin Overkamp. Based on the amino acid composition, label amino acids and caps added the theoretical molecular weight of the TM2GA is of ~4019.6 g/mol. The obtained mass is of 4019.8 g/mol.



Appendix 4.X. Helical secondary structure of TM2 domain is induced by the presence of detergents.

The 1D NMR proton spectra were acquired for three TM2 samples, without detergents (A), in 2% SDS (B) and in 2% DPC (C) in 20 mM HEPES, 100 mM NaCl at pH 6.5. All spectra were recorded using a 400 MHz Bruker spectrometer with a three channel probe at 310 K using proton spectrum with watergate (zgpgw5 pulse sequence from the Bruker catalogue). The spectra were recorded with 256 scans and with a recycle delay of 2 seconds.

The insets show the CD spectrum of the TM2 in the same conditions as used for the NMR experiments.



Appendix 4.XI. Alpha proton (left panel) and amide (right panel) chemical shift index of TM2 in DPC micelles.

The red dashed lines indicate the threshold for establishing a conclusive interpretation regarding the adopted secondary structure. The threshold values were obtained from the CSIC web page (<http://triton.iqfr.csic.es/guide/eNMR/proteins/chempro2.html>). The random coil chemical shifts were calculated using NCIP software. Our scale of H_{α} CSI does not reach values close to the threshold for H_{α} that is at ± 0.38 ppm. The threshold for amide protons (H^N) is at -0.19 ppm for the lower and $+0.29$ ppm for the upper values with the lower CSI indicating an alpha structure and the upper a beta structure.

Appendix 4.XII. Unambiguous NOE contacts extracted using CcpNmr.

Table S4.5. Assignment table of TM2 ¹³C,¹⁵N-GA labeled.

Residue number	Amino acid	Atom	Residue number	Amino acid	Atom
16	GLY	H	15	VAL	HA
13	PHE	H	13	PHE	HA
19	TYR	H	19	TYR	HA
24	ALA	H	24	ALA	HA
24	ALA	HA	25	ILE	H
12	GLY	H	11	ILE	HA
11	ILE	HA	11	ILE	H
15	VAL	HA	15	VAL	H
25	ILE	H	25	ILE	HA
16	GLY	H	16	GLY	HA2
16	GLY	HA2	17	ARG	H
18	GLY	H	18	GLY	HA2
19	TYR	H	18	GLY	HA2
12	GLY	H	12	GLY	HA2
12	GLY	H	12	GLY	HA3
12	GLY	H	11	ILE	HB
11	ILE	H	11	ILE	HB
11	ILE	H	11	ILE	HG12
11	ILE	H	11	ILE	HG13
25	ILE	H	25	ILE	HG12
12	GLY	H	11	ILE	HG2*
11	ILE	H	11	ILE	HG2*
25	ILE	H	25	ILE	HG2*
25	ILE	H	25	ILE	HD1*
18	GLY	H	17	ARG	HA
23	ASP	H	23	ASP	HA
12	GLY	H	15	VAL	HG1*
13	PHE	HA	11	ILE	H
16	GLY	H	15	VAL	HB
11	ILE	H	11	ILE	HG13
11	ILE	H	25	ILE	HG2*
15	VAL	H	15	VAL	HG1*
16	GLY	H	15	VAL	H

16	GLY	H	15	VAL	HG1*
15	VAL	H	15	VAL	HB

Table S4.6. Assignment table of TM2 ¹³C,¹⁵N- IGG labeled.

Residue number	Amino acid	Atom	Residue number	Amino acid	Atom
11	ILE	HN	24	ALA	QB
11	ILE	HN	11	ILE	HG12
13	PHE	HB2	13	PHE	QD
19	TYR	HB2	13	PHE	QD
19	TYR	QD	19	TYR	HA
19	TYR	HB3	19	TYR	QD
19	TYR	HB2	19	TYR	QD
25	ILE	HG12	19	TYR	QD
25	ILE	HG12	19	TYR	QD
19	TYR	QD	25	ILE	QG2
19	TYR	QD	15	VAL	QG1
19	TYR	QD	15	VAL	QG1
19	TYR	QD	15	VAL	QG1
11	ILE	HG12	8	TRP	HE1
8	TRP	HE3	8	TRP	HE1
8	TRP	HE3	15	VAL	HB
25	ILE	HG12	8	TRP	HE3
8	TRP	HE3	15	VAL	QG1
8	TRP	HE3	19	TYR	QD
17	ARG	HB2	8	TRP	HE3
8	TRP	HE3	17	ARG+	HG2
24	ALA	QB	8	TRP	HE3
13	PHE	HB2	13	PHE	QD
19	TYR	HB2	13	PHE	QD
11	ILE	HG12	13	PHE	QD
13	PHE	QD	25	ILE	QD1
19	TYR	QD	19	TYR	QE

Chapter 5

Studies of M2 from influenza A in a membrane environment using proton detected solid-state NMR, DNP and DFT calculations

1. Introduction

1.1 Viruses - origin and socio-economic implications

Viruses are obligate parasites which, by definition, require a host for reproduction. For self-sufficient organisms, survival is ensured by a variety of proteins that play an essential role in maintenance of cell homeostasis and cellular division. On the other hand, intracellular parasitic entities, like viruses, do not have the necessary machinery to provide these mechanisms and therefore rely on host cell factors to perform essential processes such as transcription, translation or even cell division within a host (243).

Viruses have played an important role throughout human history in a wide range of aspects such as health, evolution and economy. The various viral epidemics encountered in the history of mankind have claimed millions of lives, putting viruses at the forefront of research. In modern times, the 1918 pandemic caused by Influenza virus was declared and caused an estimated 50 million deaths worldwide (244). This pandemic was followed by several others caused by the same virus but from a different strain (Influenza A) in 1957 and 1968 in China and Hong Kong, respectively (245). In recent decades, other viruses such as HIV, measles, SARS, MERS and others have threatened human health (246, 247). Viral threats to humankind have continued until today, with SARS-CoV2, declared pandemic by the World Health Organization (WHO) in March 2020. So far, at the time of writing, SARS - CoV2 has infected more than 50 million people and caused more than one million deaths worldwide.

Technological advancements of the last century in high-resolution techniques, allow us to study the relationship between structure and function. Despite progress in recent years, much remains to be learned to combat viruses effectively. Indeed, viruses are rapidly evolving entities, thus, serious viral infection is a near certainty in the future, unless resources are

devoted towards various strategies to control emerging strains. A clear example of the evolution of viruses and their constant threat is from the Influenza A virus, commonly known as "the flu". The flu has remained among us and is the cause of an annual seasonal epidemic which causes hundreds of thousands of deaths worldwide per year (245). The uncertainty of newly emerging strains with high lethality rates is difficult to predict accurately but will certainly occur. Thus, it is important to advance our understanding of these systems in the upcoming years to allow faster and more effective prediction, detection and treatment of viral diseases.

The scientific interest regarding virus - human symbiosis has been extended beyond the simplistic view of a pathogen - host interaction. In fact, the understanding and manipulation of the genetic material over the last century has provided a large set of tools which now enables using viruses as treatment in human diseases in so called gene editing (248). Viruses are used as vectors to transport the modified genetic material into the target cell (249). Many different types of viruses have been used to target different human diseases (250).

1.2 Viral diseases and their treatments

Nowadays, there is a clear and established relationship between virus infection and a wide range of human diseases from a mild cold to cancer or even neuropathological diseases (251-253). One of the most common infections related to cancer development is the hepatitis virus that causes liver inflammation and in ~ 80% of cases leads to a hepatocellular carcinoma with an elevated mortality rate of more than 80% within 5 years after diagnosis (254). Similarly, in more than 90% of the diagnosed cervical cancer cases there is a direct link to papillomavirus infection (253). Others such as Alphaviruses, transmitted by a mosquito bite, can lead to lethal encephalitis (255, 256). These burdens caused by viruses in human health implies that there is a major challenge today to find new treatments to prevent and treat viral infections.

In the past, there have been many ways to treat viral infection, but an even more successful public health measure has been to prevent infection with vaccination. Despite the success of vaccination, it takes time to develop efficient vaccines. And, over time, mutations on the viral genome can lead vaccine resistant strains.

Advanced methods have been successfully applied to treat and prevent viral infections such as potent specific small molecules for anti-viral treatment as well as more efficient vaccines including a wide range of epitope recognition (257, 258). High resolution molecular structure determination techniques have enabled the study of protein-drug interactions, opening new avenues for rational drug design (259). Molecular docking (259) is currently one of the methods used to develop small molecules targeting specific proteins involved in the viral life cycle. In the past decades, antiviral drugs such as adamantane derivatives and oseltamivir against Influenza have been developed and are available on the market (260).

1.3 Evolution and classification of viruses

Although there are efficient treatments for many viral infections, the threat from viral infection is not yet abolished (and may never be) due to the many viral strains circulating in humans and animals. Viruses evolve extremely fast by mostly two mechanisms; (i) the reassortment mechanism or antigenic drift (261) and (ii) mutations or antigenic shift (262). This leads to resistance towards established antiviral treatments. Antigenic drift happens by the concomitant infection by distinct viruses at the same organism (Fig. 5.1.1A). The two types of viruses can exchange genetic material and their genomic recombination leads to a new viral strain with new capacities for cross-species transmissibility, lethality and resistance to available drugs (261, 263). Another mechanism for evolution is frequent mutations occurring by error prone replication (Fig. 5.1.1B). Viral genetic material is replicated by DNA or RNA polymerases which have a variable fidelity depending on the virus (264). As shown by several studies, viruses cause a wide number of diseases not only in humans but in all living organisms. Therefore, there is an urgent need to understand and characterize non-variable regions to develop efficient treatments, which will help control virus infections, and in some cases may even lead to eradication.

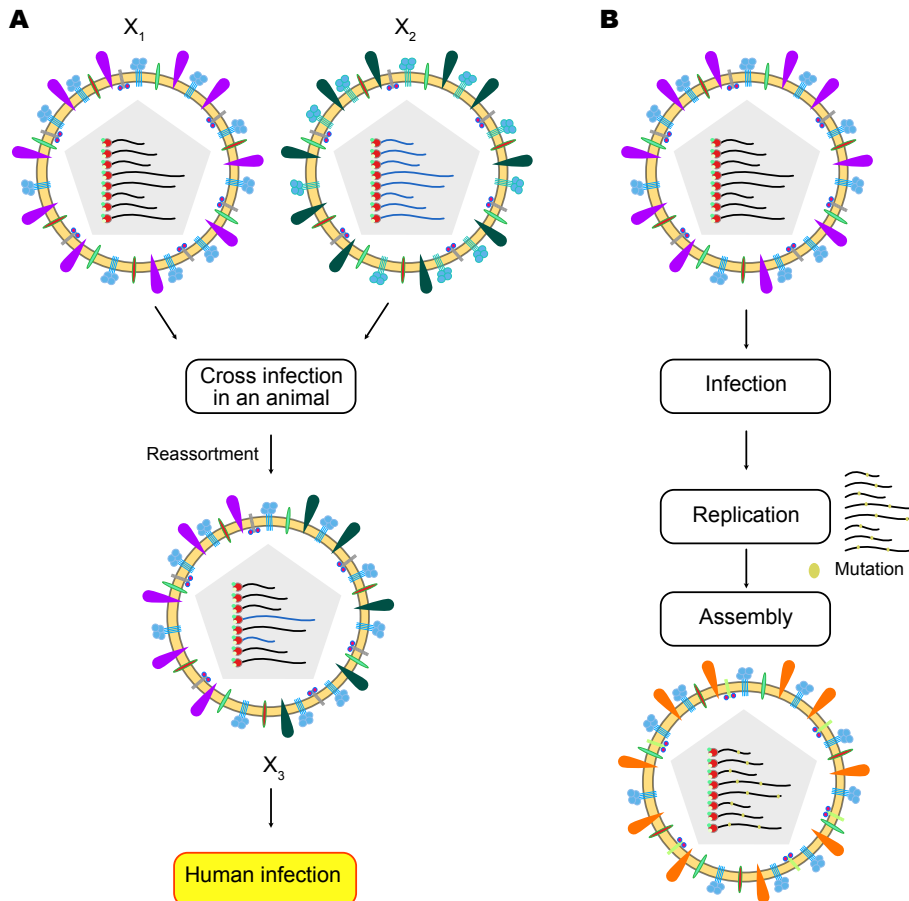


Figure 5.1.1. Mechanisms leading to antiviral resistance.

A shows the antigenic drift or reassortment, which is the recombination of the genetic material between two distinct strains of the same virus when a co-viral infection occurs, ultimately leading to a new strain of the virus. B shows the antigenic shift or mutation occurring in the genetic material of viruses during transcription.

Currently, many different classes of viruses have been catalogued. A viral classification has been established as a function of the different viral properties, commonly, a classification based on the genetic material, replication and phenotype of the virus (265). There are two main classes of viruses regarding the genetic material. On the one hand, DNA viruses where the genetic information is carried by a DNA molecule, and on the other hand, RNA viruses in which the genetic material is an RNA molecule. Within these two categories, the two groups can be subdivided into several sub-categories, e.g. viruses with double or single-stranded DNA or RNA molecules (266). In addition, viruses do not have their own protein machinery to carry out autonomous metabolism, requiring the host-cell machinery for transcription and translation of viral genetic material for producing new virions to infect other cells. The genetic material of the virus will partly determine how viruses use the host cell machinery as well as where and how the genetic material of the virus will be replicated to finally form new virions.

1.4 Influenza virus

1.4.1 Influenza: Origin, genome and proteome

Influenza A is a fast evolving virus that leads to a yearly seasonal epidemic and anti-viral resistance, incurring the need of a yearly vaccination against the flu. The flu is a well characterized viral infection that in some cases results in acute respiratory complications leading to pneumonia and causing thousands of deaths per year. Influenza is part of the *Orthomyxoviridae* family. Four distinct classes, A to D, of influenza viruses have been reported to date. Unlike other RNA viruses, influenza is a membrane coated virus whose genetic material is transported into the nucleus of the host cell for transcription (267). Influenza's genome is a negative segmented single-stranded RNA that code for at least eleven viral proteins. Its genome is composed of eight fragments, in strains A and B, or seven fragments, in strain C and D (268).

From the eleven viral proteins encoded by the RNA viral genome, three are membrane proteins (Fig. 5.1.2B, viral membrane) and eight are soluble proteins (Fig. 5.1.2B, viral interior). Within the membrane proteins, two are glycoproteins, hemagglutinin (HA) (269) and neuraminidase (NA) involved in membrane viral-host interaction (270) and in virion release. The third is the matrix 2 (M2) protein which is the less abundant protein in the membrane. M2 is required in later steps of the viral life-cycle during endosome-viral membrane fusion as well as for viral RNA release (271). Although these three proteins are the only membrane proteins encoded by the viral genome, studies by mass spectrometry have found several host membrane bound proteins such as CD9 and CD59 from the Tetraspanins family in the viral membranes indicating that the viral membrane has indeed a more complex composition (272). The soluble proteins are M1, PA, PB1, PB2, PB2-F2, NP, NS1 and NS2 (Fig. 5.1.2B).

Among the different classes of influenza, different strains have been reported and classified according to the subtypes of the HA and NA present in the membrane. To date, eighteen variants of HA (H1 to H18) and eleven variants of NA (N1 to N11) have been observed for influenza A in circulation. The subtype of HA and NA gives the name to the viral strain like for example the H1N1 which is the name designated for the first pandemic in 1918, known as the "Spanish flu".

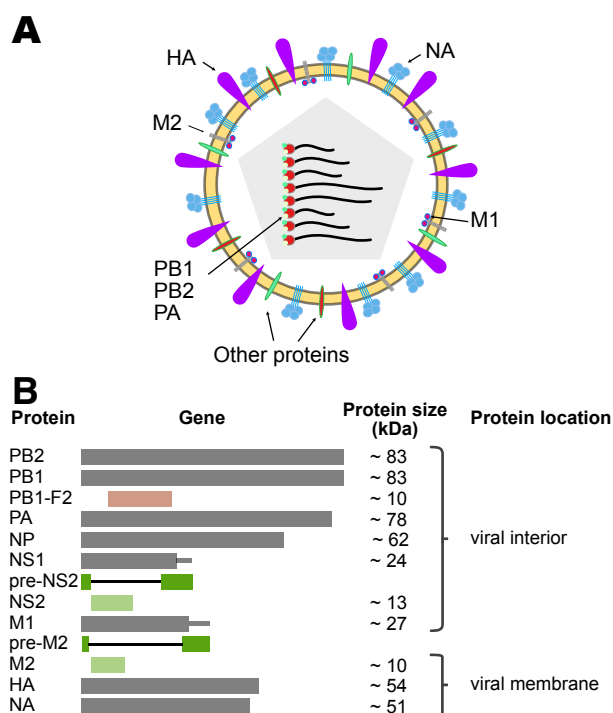


Figure 5.1.2. Schematic view of Influenza A virus structure and protein content.

A, pictures the architecture of the influenza virus. The virus is coated by a lipid membrane bilayer with three viral membrane proteins (HA, NA and M2). The viral interior contains eight RNA molecules together with the proteins comprised in the RNA-dependent RNA polymerase complex (PA, PB1 and PB2). In B, the different RNA strands with the transcription products are represented including the molecular size of the proteins encoded by the different genes.

1.4.2 Mechanism of Influenza infection and treatments

Mechanism of influenza A infection can be divided in five steps mainly i) attachment, ii) membrane fusion/genetic material release, which can happen concomitantly or at different times/location during the infection, iii) replication iv) assembly and v) detachment (Fig. 5.1.3). The attachment to the host cell, is driven by the HA membrane protein. HA is a ~ 63kDa protein that recognizes sialic acid moieties present on the surface of the host cell linked to glycan moieties (273, 274). The distinct forms of HA, from different influenza variants, will recognize specific sialic acid moieties which determines the species as well as the cells used by the virus for infection. In humans, the cells in the upper respiratory tract have predominantly sialic acid linked to a galactose by an α -2,6 linkage which are recognized by H1, H2 and H5. Whereas in birds sialic acid is linked to galactose via α -2,3 bond which is recognized with higher specificity by other HA variants, explaining the species specificity (275).

After attachment, the virus enters the cell via receptor mediated endocytosis mechanism (276). Later, during endosome maturation, a decrease in the endosomal pH below 5 activates M2 (277). The activated M2 conducts protons from the endosome to the viral interior. This leads to conformational changes of HA resulting in viral - endosome membrane fusion (271). Concomitantly, the acidic interior triggers viral uncoating and viral RNA release

into the cytoplasm of the host cell (278). In the cytoplasmic compartments, the viral RNA is coated by the nucleocapsid protein (NP) which contains a nuclear localization signal. This is used to transport the viral RNA into the nucleus of the host cell (279).

Once viral RNA reaches the nucleus, the negative RNA strands have to be transcribed into double stranded RNA molecules to ensure replication. Both transcription and replication are performed by the RNA-dependent RNA polymerase complex formed by PA, PB1 and PB2 proteins. For transcription to occur, a complementary sequence has to be available in order to start transcription of the RNA from 3' to 5'. This is ensured by a specific mechanism known as cap-snatching where the pre-mRNA of the host cell is sequestered by the RNP complex which cleaves the 5' and uses it as a primer for initiating the transcription of viral mRNA (280). Maturation of RNA molecules is indispensable for the export, stability and synthesis of RNAs. Like the host RNA, viral RNA is 5' methylated and has a 3' polyA tail. Thus, the viral RNA follows the same pathways as the human mRNA for transcription.

The membrane proteins HA, NA and M2 are translated and folded in the endoplasmic reticulum and are exported to the membrane via the Golgi (281, 282). Interestingly, M2 proton conduction plays a crucial role maintaining high pH values in the Golgi compartment avoiding premature maturation of HA (283). The soluble proteins M1, NEP, PBs and PA are translated in the cytoplasm and then re-localized into the nucleus where they form viral RNP complexes (vRNP) with the newly transcribed viral RNA (284). Once formed, the vRNPs are exported to the cytoplasm. Using the cytoskeleton of the host cells (microtubules), the vRNPs reach the membrane and form the new virions (275).

Finally, at the apical cell position the vRNP are assembled to the membrane proteins and budding formation is driven by matrix proteins 1 and 2 (275). Mutation studies have shown that M1 and the C-terminal domain of M2 are indispensable for the budding formation. After budding, virion release is permitted by the cleavage of sialic acid by NA (285). Even though the overall process of viral infection and replication is globally well understood there are still many open questions regarding the import and export of the viral RNA as well as the final assembly of competent virions.

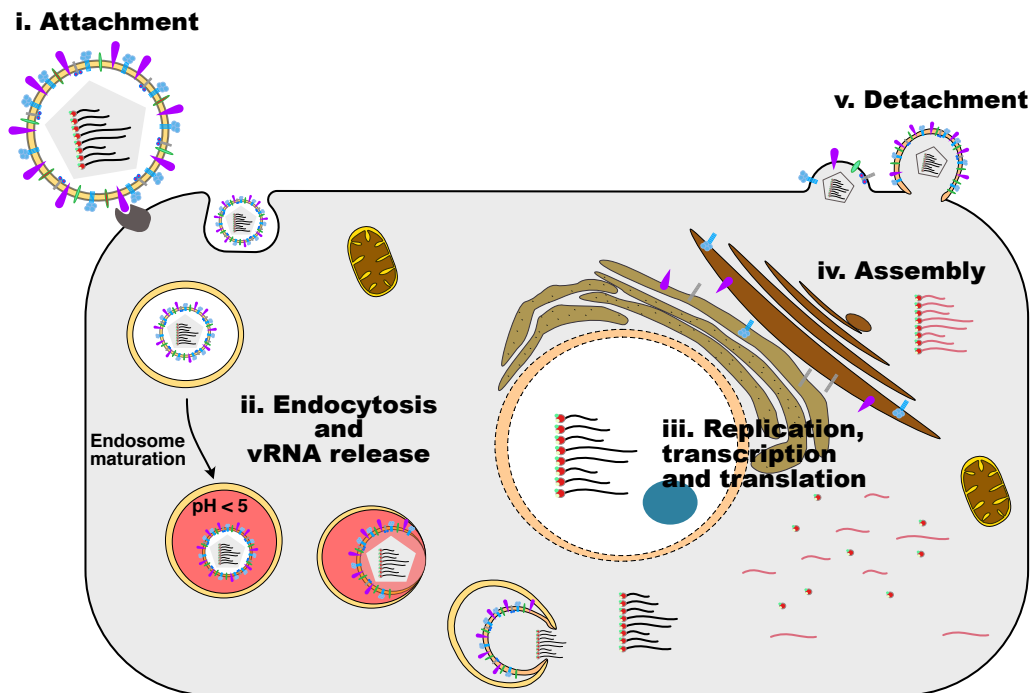


Figure 5.1.3. Influenza A life-cycle.

Influenza A infection occurs mainly in five steps: i) attachment, which depends on the host-virus recognition via specific receptor; ii) endocytosis and vRNA release, where the virus enters the host cell through a receptor mediated endosome pathway. During endosomal maturation, the activation of M2 triggers viral interior acidification followed by viral uncoating and vRNA release; iii) Replication, transcription and translation; iv) Assembly. In this step the vRNA and viral proteins are assembled to form the new and competent virions. Finally, v) Detachment is the steps of virion release.

Specific treatment of viral diseases requires a complete understanding of the viral life cycle on a biological and molecular level basis. Understanding the role of each protein at each time point during viral infection will contribute to a more efficient viral treatment. As previously mentioned, membrane coated viruses such as influenza have a complex membrane that includes specific viral membrane proteins which might play important roles during the viral life cycle. During the last century, drugs targeting neuraminidases such as Oseltamivir and Zanamivir have been used, along with the M2 channel blockers amantadine and rimantadine. However, some of the drugs for the M2 protein are no longer in use due to resistant M2 variants (286).

1.4.3 The proton channel of influenza A: Matrix 2 (M2) protein

The matrix protein 2 (M2) from influenza is a 97 residue protein that forms homotetramers in the viral membrane (287). It comprises four domains, a soluble N-terminal domain from residue 1 to 21, a transmembrane domain from 22 to 46, an amphipathic helical domain from 47 to 60 and a non-structured C-terminal domain from residue 61 to 97. Earlier

electrophysiological studies in oocytes have shown that M2 is a proton channel that is activated under acidic conditions (271, 288, 289). The acidification of the viral interior is a key step in the viral life cycle responsible for inducing membrane fusion and viral RNA (vRNA) release (271). Deletion as well as mutagenesis studies targeting the amphipathic helix and the C-terminal domain (C-terminus of the protein from residue 46 to 97) have shown the implication of M2 in budding and in the packing of vRNA. (290, 291).

Earlier mutagenesis work, by Pinto and co-workers, mutating histidine 37 (H37) to glycine, showed that H37 controls the pH-proton conduction activity of M2 (292). Further mutation and rescue studies targeting the core residue H37 have shown that H37 is the key residue imparting proton selectivity to the M2 channel (293). Titration studies of H37 sidechain have been performed and have reported distinct histidine pKa values depending on of the length of the constructs and the lipid environments used (294, 295). Although varied pKa values have been reported, all studies agreed that the +3 histidine charge state triggers channel activation (pKa ~5) (294, 296). There is also agreement in the literature that the protein arranges as a homotetramer in lipid bilayers, with a single pass TM helix. The protein arrangement is supported by a large amount of high resolution structural data obtained by X-ray and NMR for the different constructs studied. Although there is a vast structural data available, there are still many controversies regarding H37 sidechain rearrangement at different pHs, drug binding and the proton conduction mechanism.

1.4.3.1 Role of the HxxxW motif in proton conduction

In M2, the HxxxW motif located at the C-terminus of the transmembrane domain is conserved among the different strains. Mutation of residue H37 causes both loss of proton conduction by pH-dependent activation and proton selectivity (292). Additionally, it has been observed that the mutation at H37 compromises the stability of the tetrameric structure (297). While WT M2 conducts protons primarily from outward to inward of the virus, W41F and W41A mutants exhibited bi-directional conduction, providing evidence that W41 is the primary proton gate (298). A great deal of effort has been invested in studies aimed at understanding the details of the proton conduction mechanism of the M2 protein.

The combination of molecular dynamics simulations (MD) with high resolution structure and dynamics data propose two main proton conduction mechanisms. The MAS-NMR dynamics data showed a rapid reorientation of the H37 sidechain suggesting a shuttle

mechanism (294, 299). However, X-ray structures, in combination with MD simulations suggest a water wire mechanism where protons diffuse through the pool of pore water molecules process known as the Grothuss mechanism (300). It is now established that the proton conduction through the M2 protein occurs via two successive mechanisms. First, a water wire mechanism up until reaching the H37 residue and secondly a shuttling mechanism happening by the reorientation of the H37 (294, 301).

In addition to the HxxxW motif as regulator for proton selectivity, conduction and gating, residue V27 located at N-terminal of the transmembrane domain has been suggested to act as the secondary gate to enter the pore of M2. Molecular dynamics simulations have shown that the sidechain of residue V27 affects the hydrogen bonded water molecules disrupting the “water wire” which regulates the proton conduction (302).

1.4.3.2 Structural differences in the M2 tetrad

Based on atomic resolution structures deposited into the PDB, differences are observed with respect to helix tilt, the kink at G34, and the side-chain arrangement of His37.

Earlier studies using solid-state NMR in oriented samples reported different M2 transmembrane helix tilt angles within the membrane ranging from 20 to 40° depending on the sample condition (303, 304). Oriented sample experiments provide direct access to the orientation of the ¹⁵N-¹H bond with respect to the lipid bilayer and the magnetic field (305). Later, Timothy A. Cross and co-workers, use the same approach to show that in the presence of drug (amantadine) the transmembrane domain of M2 shows a kink near to residue G34 (306).

Although oriented samples gives valuable information regarding the protein secondary conformation and its relative orientation with respect to the membrane there is a lack of information regarding side-chains and tertiary or even quaternary structure that may occur. M2 is found with fourfold symmetry in solution and solid-state NMR as well as in X-ray and XFEL crystal structures at high and low pH values in the presence (Fig. 5.1.4A and B) or absence of drug (Fig. 5.1.4D and E). Interestingly, the smallest functional construct including only the transmembrane domain residues of M2 (mainly residue 22 to 46), is trapped at different states: “inward_{closed}” and “inward_{open}” (300, 307). For the inward_{closed} state, the C-terminus of the transmembrane helix is relatively closely packed with the opposite helices,

with a diagonal distance of ~ 17 Å (R45-C β to R45C β) (Fig. 5.1.4B and F). In contrast, for the inward_{open}, the C-terminus is widely open with a distance of ~ 27 Å to the opposite helices (Fig. 5.1.4A and F). Interestingly, structural data in lipid cubic phase at room temperature in the “inward_{open}” state shows in both, high and low pH, an intermediate diagonal distance of ~ 23 Å (Fig. 5.1.4D and E). In both states, the N-terminus show a similar opening of the transmembrane helix with an average distance of ~ 8 Å from opposite helices. The different states (“inward_{closed}” and “inward_{open}”) are in equilibrium and have been related to the sample preparation conditions including pH, which varies the protonation state of H37, drug binding and even lipid composition (304, 308). Indeed, at low pH (pH < 5), the histidine is positively charged (+3 to +4), thus electrostatic repulsion takes place leading to an opening of the C-terminus of the transmembrane domain.

Until very recently, high resolution structural data from solid and solution NMR has only been reported at high pH leading to similar structures as the inward_{closed} state with the transmembrane helices closely packed (~ 20 Å from 45R-C β -A to 45R-C β -D, see figure 5.1.4E) (22, 23, 309). Recently, a study combining solution NMR and X-ray crystallography has used different detergents to stabilize M2 in different states (310).

Interestingly, the crystallographic data in both high and low pH show that H37 sidechains are facing the ring of the neighboring H37 forming the so-called “box conformation” (307, 311) (Fig. 5.1.4A, B, D and E, black boxes). In this conformation, water molecules have been proposed to bridge between the neighboring histidine sidechains. Similarly structure has been proposed by ssNMR data using a similar M2 construct reconstituted in viral membranes that shows a H δ 1N of H37 proton chemical shifts just above 10 ppm (22).

On the other hand, using a longer construct referred to as the conductance domain (CD from residue 22 to 60), Cross and co-workers found an unexpected low field proton chemical shift in DOPC membranes for the H δ 1N above 15 ppm using ¹⁵N based solid-state NMR measurement. This extreme low field proton chemical shifts in proteins have been previously reported for protons involved in hydrogen bonds suggesting the presence of a low barrier hydrogen bond at the histidine sidechain protons (312) (Fig. 5.1.4C, black box). The inter imidazole-imidazole hydrogen bond opposes the C4 symmetry shown by the crystallographic data. Therefore, a dimer of dimer structure (C2 symmetry) with parallel histidine - histidine

arrangement was proposed (Fig. 5.1.4C) (312). Other solid-state NMR studies using the same construct but different lipid bilayers showed similar H37 δ 1 proton chemical shifts in wild type as well as in the S31N mutants (313). Further evidences were provided in 2015, with the first proton detected solid-state NMR structure of the resistant M2 S31N (309). In this work Andreas *et al.* assigned two sets of peaks to the M2 polypeptide chain indicating that M2 has a two fold symmetry rather than a four fold symmetry forming what is referred to as “dimer of dimers” in DPhPC membranes (309).

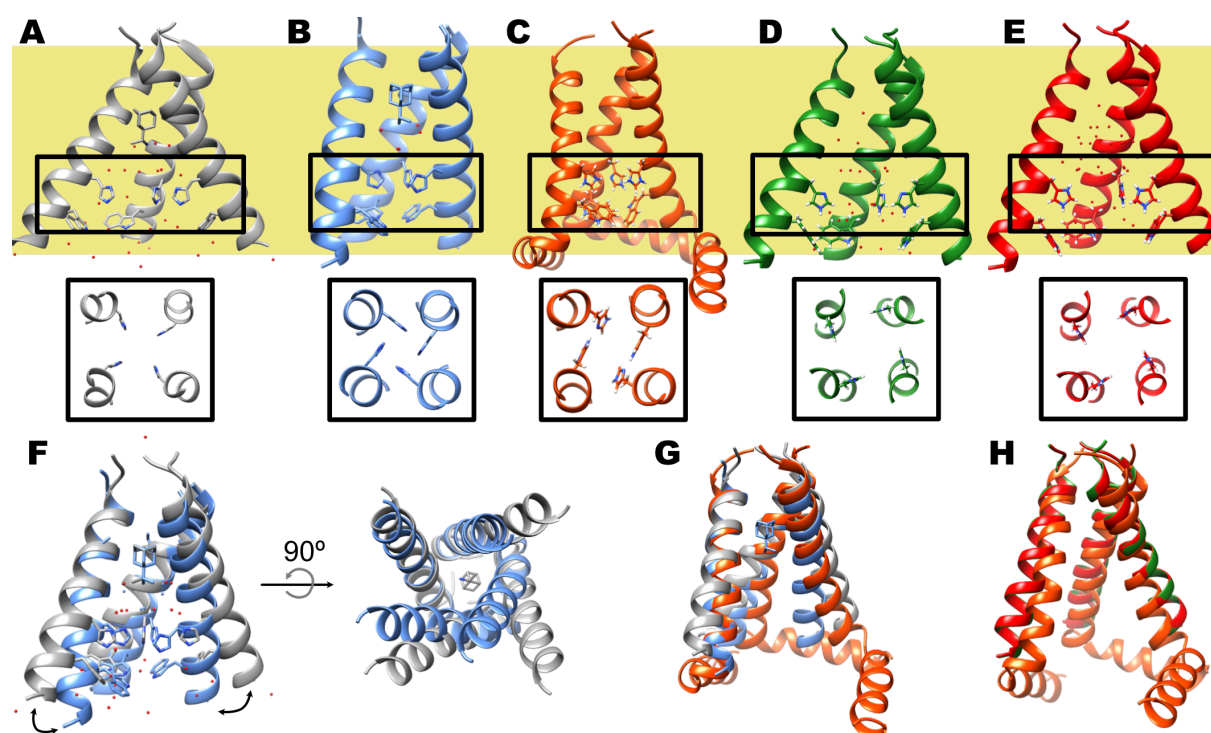


Figure 5.1.4. Comparison of M2 structures and H37 sidechain arrangements.

Both A and B show crystal structures in lipid cubic phase of M2 bound to rimantadine at low pH (gray, pdb: 6BOC) and high pH (blue, pdb: 6BKL). In C, the solid-state NMR structure using the CD construct (pdb: 2L0J) is shown. D and E shows the XFEL structures at room temperature of the apo M2 at high (green, pdb: 5TTC) and low (red, pdf: 5JOO) pH, both in the “Inward_{open}”. Black boxes in A, B, C, D and E show histidine 37 sidechain arrangement in the different conditions. In F, both structures at high “Inward_{closed}” and low “Inward_{open}” pH are compared showing the different states and the opening of the helices at the C-terminus at low pH. G shows the comparison between the X-ray and ssNMR structures showing the similar transmembrane arrangement for the ssNMR and the “Inward_{closed}” state. H, shows the comparison of the apo M2 structures from ssNMR (orange) and the room temperature XFEL structures at high (green) and low (red) pH. In the represented structures (A to E, F, G and H) the front helix of the tetramer has been removed using chimera to show the sidechain residues His37 and Trp41 that face the pore of M2 as well as the water molecules that are located in the pore of M2.

4.2.3 Insights regarding drug binding

In the past, amantadine drugs (289, 314) and derivatives have been successfully used to block M2 proton conduction. However, the fast mutation of influenza A viruses as well as

natural variation in existing strains has led to resistance to the amantadine derivatives (315, 316). The binding site of amantadine was first reported by solution NMR data in the presence of 1,2-diheptanoyl-sn-glycero-3-phosphocholine detergent (DHPC). It showed the presence of four external binding sites for amantadine near R45 at the amphipathic helix (23). Based on the NMR chemical shifts, it was suggested that amantadine has an allosteric mechanism blocking proton conduction by prohibiting pore opening. However, the lack of drug resistance while mutating the residues at the amphipathic helix including R45 contradicted the proposed amantadine inhibitory mechanism.

Later, X-ray crystallography and solid-state NMR in lipid bilayers showed that amantadine and derivatives are bound to the pore of M2 blocking proton conduction (22, 317). Nonetheless, the studies performed in membrane environments with ssNMR by Hong and co-workers showed evidence of an external binding site in large excess of drug concentration (4 to 1 excess of drug respect to the tetramer) (22). Similar to Hong's group findings, the external binding site was observed in DPhPC membranes using DNP (318). The difference of binding between the solution and solids data performed in lipid bilayers suggests that there is a clear influence of the environments for drug binding in M2. Further studies showed that the presence of cholesterol as well as the length of the construct influence the binding properties of the protein (318, 319).

Nowadays, all available structural data has served as an input for docking studies that have led to potential new drugs via rational drug development (320, 321). Despite having vast data regarding the structural features of M2 (22, 300, 309, 317), many open questions remain concerning molecular details which could heavily influence the development of new drug candidates. Why does the pore binding not occur in the presence of DHPC detergent? What is the conformation of the histidine sidechain in lipid bilayers at different pHs for the conductance domain? Is lipid composition more important than the construct length in perturbing the M2 conformational arrangement?

Goals:

This chapter is focused on revealing atomic details of the membrane protein M2, which is naturally found in influenza A. Herein, a combination of biophysical and computational methods were used to shed light on several remaining questions regarding the structure and drug interactions of M2 in a membrane lipid environment. To begin with we address the question regarding the conformational arrangement of the histidine sidechain using proton detected solid-state NMR at 100 kHz MAS (part 5.3) Using the same methodology, aim for solving the structure of the conductance domain of a fully protonated M2 construct (M2CD comprising residue 18 to 60) in lipid bilayers (DPhPC) (part 5.4). With the final goal to obtain full assignment of the residues H37 and W41. To continue, the *de novo* protein assignment of the fully protonated protein was used to investigate the binding kinetics of rmt to the M2 conductance domain by temperature controlled NMR measurements (part 5.5). Finally, we investigate the pore water cluster using the combination of low temperature measurements in solid-state NMR and dynamic nuclear polarization as well as NMR chemical shift calculation by density functional theory (DFT) (part 5.6).

This chapter has been written as a cumulative thesis format. Hence, parts of this chapter have been adapted from manuscripts that are published, in preparation or submitted.

The citations are as follows:

Part 5.3. Imidazole–Imidazole Hydrogen Bonding in the pH-Sensing Histidine Sidechains of Influenza A M2 (Kumar Tekwani Movellan, Melanie Wegstroth, Kerstin Overkamp, Andrei Leonov, Stefan Becker, and Loren B. Andreas*)

<https://doi.org/10.1021/jacs.9b10984>

Part 5.4. *De novo* structure of a fully protonated M2 in DPhPC lipids bilayers

Part 5.5. Characterization of the non specific rimantadine binding (to be submitted)

Part 5.6. Identification of the pore water in M2 by ssNMR (to be submitted)

2. Materials and Methods

Here, we used the conductance domain (CD) of M2 Udon from influenza A which is composed of 43 residues of the original sequence from residue 18 to 60. Cysteines 19 and 50 from the original sequence have been replaced by serine, as in previous studies resulting in: RS²⁰NDSSDPLVVAA³⁰SIIGILHLI⁴⁰LWILDRLFFKS⁵⁰IYRFFEHLGL⁶⁰K. The M2_{CD} contains two domains, the transmembrane domain (TM) of 19 residues from residue 25 to 43 and an amphipathic helix at the C-terminus from residue 44 to 60. The vector containing the wild type M2_{CD} sequence used in this work was kindly provided by Prof. Dr. James Chou Harvard Medical School (23). The expression, purification and reconstitution protocols used were previously reported (322) and they are briefly detailed below.

2.1 Expression and purification

M2_{CD} sequence expression was carried in *E. coli* BL21(DE3) strain using a double tagged sequence with a His (9x) and TrpLE fusion tags at the N-terminal of the amino acid sequence. Protein expression was conducted in minimal media with ¹³C-glucose and ¹⁴N-NH₄Cl supplemented with centrum vitamins (309). The expression was induced once OD_{600nm} reaches 0.75 with 150 μM of isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 37°C. After ~16 hours, the cells were harvested and resuspended in 50 mM Tris Buffer with 150 mM NaCl at pH 7.8 before rupturing them with the the emulsiflex at 1000 psi 3 times. The protein was purified using denaturing conditions from from inclusion bodies. After washing 3 times the inclusion bodies, the pellet was solubilised in a 6 M guanidinium buffer solution. Solubilised inclusion bodies were loaded on a nickel affinity column (Ni-NTA) and eluted with 400 mM imidazole at pH 6.8. The elution fractions were collected and pooled together before inducing protein precipitation by dialysis against water. The precipitant was suspended in 70 percent formic acid and TrpLE was cleaved with an excess of cyanide bromide and lyophilised. Finally, the sample was resuspended in hexafluoro isopropanol: formic acid: water (2:1:1) and passed through a C4 reverse phase HPLC column using a linear gradient elution from 5% 2-propanol to acetonitrile/2-propanol/water (58:37:5).

2.2 Refolding and reconstitution

The pure protein was then solubilized and refolded in 40 mM phosphate, 30 mM glutamate, 0.03% sodium azide at pH 7.8 containing 2% octyl-β-glucoside (OG from Sigma).

The reconstitution was performed in either protonated or deuterated DPhPC (95% d78-phytanoyl 50% d at the alpha position, methyl-d9-choline; from fbreagents). Deuterated DPhPC was dissolved in water/OG solution and mixed with the refolded protein in a 1 to 1 by weight. OG was removed by dialysis against a free detergent buffer (40 mM phosphate, 30 mM glutamate, 0.02% sodium azide at pH 7.8). The dialysis was carried out for at least one week with successive changes of buffer until a white precipitated appeared.

2.3 Drug binding assays

Two protocols for binding were tested: one where the content of a packed 1.3 mm rotor (containing ~1 mg of protein) was taken up in a 300 μ l solution containing 40 mM rimantadine (rmt). The incubation was performed for several days to weeks at 4°C. The second binding method tested was the addition of rmt powder into the rotor, as previously reported (313). For the direct powder addition, 0.2 mg of rmt were packed in a 1.3 mm rotor, and ~0.5 to 1 mg of M2 was added. The incubation temperature was kept between ~4 to 10°C.

2.4 Re-Solubilization of membrane inserted M2 in DHPC

Two distinct M2 samples were re-solubilized in 300 mM 1,2-diheptanoyl-sn-glycero-3-phosphocholine (DHPC), an unbound M2 and a 40 mM rmt pore bound M2. Rmt binding was tracked by ssNMR to ensure pore binding before re-solubilization in 300 mM DHPC detergents in 40 mM phosphate, 30 mM glutamate and 0.04% sodium azide at pH 7.8. Insoluble material from both samples were pelleted using a top bench centrifuge at 4°C for 15 min, then the solution was filled into a 3 mm NMR tube with 10% D2O.

2.5 Solution NMR, solid-state NMR and DNP

2.5.1 Solution NMR

¹⁵N-HSQC trosy based spectra were recorded for both re-solubilized M2 samples (apo and rmt bound) in a 600 MHz Bruker spectrometer equipped with a 5 mm four channel cryoprobe. In both cases the temperature was set to 312 K. The indirect ¹⁵N dimension was recorded up to 61 ms (180 points) with a spectral width of 35 ppm and the direct ¹H dimension was sampled up to 131 ms (2K points) using 13 ppm spectral width. 32 scans and 128 scans were used for either drug free or bound sample, respectively. Both spectra were processed using apodization with a squared cosine bell function in both dimensions with 512

and 170 points on the proton and nitrogen dimensions, respectively. Both dimensions were zero filled with 4 k points in proton and 1 k points in nitrogen.

2.5.2 Solid-state NMR (ssNMR)

Solid-state NMR spectra were recorded in three different Bruker spectrometers, at 800, 850 and 950 MHz. Both 800 and 850 MHz instruments were equipped with a 1.3 mm MAS probes either a narrow bore 3 channel or a 4 channel wide bore probe, respectively. The 950 MHz spectrometer was equipped with a 4 channel 0.7 mm MAS Bruker probe. The table below lists detailed acquisition parameters for all the spectra recorded including temperature, spectral width, number of points, number of scans and sample.

Table 5.2.1. Solid-state NMR experiments and parameters.

Experiment	(H)NH	(H)NH	(H)NH	(H)CH	hNH in	hNH water
Transfer 1	HN-CP	HN-CP	HN-CP	HC-CP	HN-CP	HN-CP
Max RF Field (kHz) (Ramp in H)	149/20 (100 to 80)	149/20 (100 to 80)	149/20 (100 to 80)	139/29 (100 to 80)	149/20 (100 to 80)	142/30 (100 to 80)
Time (ms)*	1	3	6	1.2	1.25	4
Transfer 2	NH-CP	NH-CP	NH-CP	CH-CP	N - N inept	NH-CP
Max RF Field (kHz) (Ramp in H)	30/137 (80 to 100)	30/137 (80 to 100)	30/137 (80 to 100)	29/127 (80 to 100)	-	30/140 (80 to 100)
Time (ms)*	0.5	3	6	0.35	Different times used	4
Transfer 3	-	-	-	-	NH-CP	HN-CP HC-CP
Max RF Field (kHz) (Ramp in H)	-	-	-	-	30/140 (80 to 100)	140/30 (HN) 140/33 (HC) (80 to 100)
Time (ms)*	-	-	-	-	0.5	0.5 (HN) 0.2 (HC)
sw (t1) (ppm)	52	300	300	30	500	500
Acq.time (t1) (ms)	40	18	18	36	5	5
sw (t2) (ppm)	52	52	52	30	52	52
Acq.time (t2) (ms)	40	41	41	36	41	41
¹ H decoupling	sltpm	sltpm	sltpm	sltpm	sltpm	sltpm
Field (kHz)	22	22	22	22	22	22
Inter-scan delay (s)	1.1	1.1	1.1	1.1	1.1	1.1
Number of scans	4	40	96	4	-	-
Measurement time (h)	1.25	14	34	1.25	-	-

Experiment	(H)CaNH	(HCo)Ca(Co)NH	(H)CoNH	(H)Co(Ca)NH	(H)Ca)Cb(CaCo)NH
Transfer 1	HC-CP	HC-CP	HC-CP	HC-CP	HC-CP
Max RF Field (kHz) (Ramp in H)	132/29 (100 to 80)	131/29 (100 to 80)	121/22	121/22 (85 to 100)	119/22 (85 to 100)
Time (ms)*	1.4	2.7	2.9	2.9	0.8
Transfer 2	CN-CP	CO-CA-CO INEPT	CN-CP	CO-CA INEPT	CA-CB-CA INEPT
Max RF Field (kHz) (Ramp in N)	61/42 (100 to 80)	-	69/42 (Tan)	-	-
Time (ms)*	9.8	T2 Ca and Co optimised	14	4.2 (T2 optimised)	4.5 (T2 optimised)
Transfer 3	NH-CP	CN-CP	NH-CP	CN-CP	CN-CP
Max RF Field (kHz) (Ramp)	31/126 (80 to 100 in H)	62/42 (Tan in N)	31/130 (80 to 100 in H)	69/42 (80 to 100 in N)	69/42 (80 to 100 in N)
Time (ms)*	0.3	11.2	0.5	14	14
Transfer 4	-	NH-CP	-	NH-CP	NH-CP
Max RF Field (kHz) (Ramp in H)	-	31/133 (80 to 100)	-	31/122 (80 to 100)	31/122 (80 to 100)
Time (ms)*	-	0.5	-	0.5	0.5
Transfer 5	-	-	-	-	-
Max RF Field (kHz) (Ramp)	-	-	-	-	-
Time (ms)*	-	-	-	-	-
sw (t1) (ppm)	30	35	30	30	30
Acq.time (t1) (ms)	17	7	16.6	16.6	16.6
sw (t2) (ppm)	30	35	20	20	60
Acq.time (t2) (ms)	9	7	16.7	16.7	8
sw (t3) (ppm)	52	100	52	52	52
Acq.time (t3) (ms)	40	20.5	20.4	20.5	20.5
¹ H decoupling	sltpm	sltpm	sltpm	sltpm	sltpm
Field (kHz)	22	22	22	22	22
Interscan delay (s)	1.1	1.1	1.1	1.1	1.1
Number of scans	4	20	4	16	16
Measurement time (h)	13	60	23	92	138
Experiment	(H)NCaHa	(H)COCAHa	(H)Ca)Cb(Ca)NH		
Transfer 1	HN-CP	HC-CP	HC-CP		
Max RF Field (kHz) (Ramp in H)	139/31 (80 to 100)	133/29 (100 to 80)	120/22 (85 to 100)		
Time (ms)*	1.5	3.5	0.75		
Transfer 2	NC-CP	CO-CA INEPT	CA-CB INEPT		

Max RF Field (kHz) (Ramp)	71/42	-	-
Time (ms)*	13	4.16 (T2 optimised)	4.3
Transfer 3	CH-CP	CH-CP	CN-CP
Max RF Field (kHz) (Ramp)	30/125 (100 to 85 in H)	29/137 (80 to 100 in H)	69/42 (tan in N)
Time (ms)*	0.35	0.25	13
Transfer 4	-	-	NH-CP
Max RF Field (kHz) (Ramp in H)	-	-	31/122 (100 to 85)
Time (ms)*	-	-	0.5
sw (t1) (ppm)	30	52	30
Acq.time (t1) (ms)	15.2	40	15
sw (t2) (ppm)	30	30	60
Acq.time (t2) (ms)	6.8	9	10
sw (t3) (ppm)	52	15	52
Acq.time (t3) (ms)	20.4	16	15
¹ H decoupling	sltpm	sltpm	sltpm
Field (kHz)	23	22	23
Interscan delay (s)	1.1	1.1	1.1
Number of scans	4	4	12
Measurement time (h)	13	21	118

Experiment	H(H)NH	H(H)CH	(H)C(HH)CH	(H)X(HH)XH
Transfer 1	HN-CP	HC-CP	HC-CP	HC-CP HN-CP
Max RF Field (kHz) (Ramp in H)	150/30 (100 to 80)	169/51 (100 to 85)	167/51 (85 to 100)	149(H)/33(C)/ 31(N) (80 to 100)
Time (ms)*	1	1.2	0.35	1 (HC-CP) 1.5 (HN-CP)
Transfer 2	NH-CP	CH-CP	CH-CP	CH-CP NH-CP
Max RF Field (kHz) (Ramp in N)	30/137 (80 to 100)	51/158 (85 to 100)	51/156 (100 to 85)	138H)/32(C)/ 31(N) (100 to 80)
Time (ms)*	0.4	0.35	0.2	1 (HC-CP) 1.5 (HN-CP)
Transfer 3	-	-	-	HC-CP HN-CP

Max RF Field (kHz) (Ramp in H)	-	-	-	149(H)/33(C)/ 31(N) (80 to 100)
Time (ms)*	-	-	-	1 (HC-CP) 1.5 (HN-CP)
Transfer 4	-	-	-	CH-CP NH-CP
Max RF Field (kHz) (Ramp in H)	-	-	-	138H)/32(C)/ 31(N) (100 to 80)
Time (ms)*	-	-	-	0.3 (HC-CP) 0.45 (HN-CP)
Mixing (ms)*	-	-	-	0.5
sw (t1) (ppm)	22	15	15	290
Acq.time (t1) (ms)	4	4	4	4.5
sw (t2) (ppm)	52	30	30	268
Acq.time (t2) (ms)	40	36	36	4.5
sw (t3) (ppm)	-	-	-	30
Acq.time (t3) (ms)	-	-	-	36
¹ H decoupling	sltpm	sltpm	sltpm	sltpm
Field (kHz)	22	22	22	22
Interscan delay (s)	1.1	1.1	1.1	1.1
Number of scans	4	2	4	2
Measurement time (h)	26	36	54	122

All the spectra were processed using a squared cosine apodization function. The processing parameters are indicated in the captions or the main text for more specific explanation when needed.

2.5.3 Dynamic nuclear polarization (DNP)

M2 samples were transferred from 1.3 mm rotors (~1 mg of M2 protein) to 3.2 mm or 2.5 mm rotors for DNP purposes. Prior to packing the samples were mixed with ‘DNP juice’ composed of a mixture of glycerol:D2O:water (60:30:10) and 4 - 10 mM of AMuPOL (polarization agent). Spectra were recorded in a 600 MHz Bruker spectrometer with 395 GHz gyrotron. Two different DNP probes were used, a 3.2 mm Bruker DNP probe and a 2.5 mm Phoenix probe. In the table below, all the spectra recorded with the respective set ups are presented.

Table 5.2.2. DNP spectra performed with a 3.2 mm Bruker DNP probe.

Experiment	(H)NH	(H)NHdep	(H)NHde p ref	(H)N	(H)C
Transfer 1	HN-CP	HN-CP	HN-CP	HN-CP	HC-CP
Max RF Field (kHz) (Ramp in H)	149/20 (100 to 80)	142/30 (100 to 80)	142/30 (100 to 80)	149/20 (100 to 80)	139/29 (100 to 80)
Time (ms)*	1	4	4	1.25	1.2
Transfer 2	NH-CP	NH-CP	NH-CP	N - N inept	CH-CP
Max RF Field (kHz) (Ramp in H)	30/137 (80 to 100)	30/140 (80 to 100)	30/140 (80 to 100)	-	29/127 (80 to 100)
Time (ms)*	0.5	4	4	Different times used	0.35
Transfer 3	-	HN-CP HC-CP	HN-CP HC-CP	NH-CP	-
Max RF Field (kHz) (Ramp in H)	-	140/30 (HN) 140/33 (HC) (80 to 100)	140/30 (HN) 140/33 (HC) (80 to 100)	30/140 (80 to 100)	-
Time (ms)*	-	0.5 (HN) 0.2 (HC)	0.5 (HN) 0.2 (HC)	0.5	-
sw (t1) (ppm)	52	500	500	500	30
Acq.time (t1) (ms)	40	5	5	5	36
sw (t2) (ppm)	52	52	52	52	30
Acq.time (t2) (ms)	40	41	41	41	36
¹ H decoupling	sltppm	sltppm	sltppm	sltppm	sltppm
Field (kHz)	22	22	22	22	22
Interscan delay (s)	1.1	1.1	1.1	1.1	1.1
Number of scans	4	-	-	-	4
Measurement time (h)	1.25	-	-	-	1.25

Table 5.2.3 DNP spectra performed with a 2.5 mm Phoenix DNP probe.

Experiment	(H)NH	(H)NHdep	(H)NHde p ref	(H)N	(H)C
Transfer 1	HN-CP	HN-CP	HN-CP	HN-CP	HC-CP
Max RF Field (kHz) (Ramp in H)	149/20 (100 to 80)	142/30 (100 to 80)	142/30 (100 to 80)	149/20 (100 to 80)	139/29 (100 to 80)
Time (ms)*	1	4	4	1.25	1.2
Transfer 2	NH-CP	NH-CP	NH-CP	N - N inept	CH-CP
Max RF Field (kHz) (Ramp in H)	30/137 (80 to 100)	30/140 (80 to 100)	30/140 (80 to 100)	-	29/127 (80 to 100)
Time (ms)*	0.5	4	4	Different times used	0.35
Transfer 3	-	HN-CP HC -CP	HN-CP HC -CP	NH-CP	-
Max RF Field (kHz) (Ramp in H)	-	140/30 (HN) 140/33 (HC) (80 to 100)	140/30 (HN) 140/33 (HC) (80 to 100)	30/140 (80 to 100)	-
Time (ms)*	-	0.5 (HN) 0.2 (HC)	0.5 (HN) 0.2 (HC)	0.5	-
sw (t1) (ppm)	52	500	500	500	30
Acq.time (t1) (ms)	40	5	5	5	36
sw (t2) (ppm)	52	52	52	52	30
Acq.time (t2) (ms)	40	41	41	41	36
¹ H decoupling	sltppm	sltppm	sltppm	sltppm	sltppm
Field (kHz)	22	22	22	22	22
Interscan delay (s)	1.1	1.1	1.1	1.1	1.1
Number of scans	4	-	-	-	4
Measurement time (h)	1.25	-	-	-	1.25

2.6 Structure calculations of fully protonated M2

Structure calculations were carried out using CYANA an automated software for assigning NMR spectra (323). All spectra used for structure calculation were recorded using a four channel 0.7 mm MAS probe on a 950 MHz spectrometer spinning at 100 kHz at 260 K (sample temperature 5-10°C). To reduce bias on the assignment of structural data, the peak lists from (H)NH, (HH)NH, (HH)CH and (H)X(HH)XH spectra were used as input to CYANA without manual assignment. In addition to the peak list, manual restraints such as the imidazole - imidazole hydrogen bond between dimers, the contact between Nε2 from H37 to Nε W41 and the H-bonds for helical arrangement predicted by TALOS were added in order to initially form a helical structure and thus improving convergence of the tetramer.

Furthermore, constraints on the top and bottom of the transmembrane domain were manually added (Appendix 5.II) in order to drive the formation of the tetramer. Several (H)NH spectra were used changing the cross polarization time from 3 to 6 ms.

2.7 DENSITY FUNCTIONAL THEORY CALCULATIONS

Density functional theory calculations were carried out using Gaussian 16 software (324). To reduce the calculation time, the M2 structure was truncated at its N and C-terminus. Only the residues between 34G to 41W were kept. Different truncated quaternary M2 structures were submitted to DFT calculation. On the one hand, the dimer structure referred to as “dimer” (two M2 monomer units) keeping the imidazole-imidazole hydrogen bond. On the other hand, the C2 tetramer structure (four M2 monomers) is referred to as “tetramer”. The dimer and tetramer structures were designed from our own structure calculations. Structure optimizations were performed using a step wise approach increasing the accuracy of the method used at each step. Briefly, the accuracy was started from a semi-empirical method using PM6 to DFT using hybrid functionals Becke Three-Parameter Hybrid Functionals (B3LYP) (325) and a gradient corrected method mPW91 (326). In all cases all the atoms except histidine 37 sidechain were immobilized. Once the structure was fully optimized, a water molecule for the dimer (two for the tetramer) were explicitly placed in close proximity to the free N δ 1 at the histidine. The same systematic approach for optimization was used from PM6 to DFT using either B3LYP or mPW91 with 631G basis sets. For the NMR calculation, only half of the tetramer (a dimer) was used to reduce the computing power needed for NMR calculation. NMR calculations were performed using a variety of basis sets from 631G to 6311G++2d,p. To terminate, polarizable continuum model (PCM) to include water solvation was used with the best basis set matching experimental data.

3. Identification and characterization of an imidazole – imidazole hydrogen bond in M2 from influenza A

3.1 Abstract

The arrangement of histidine sidechains in influenza A M2 tetramer determines their pKa values, which define controlled proton conduction critical to the virus lifecycle. Both water associated and hydrogen bonded imidazole–imidazolium histidine quaternary structures have been proposed, based on crystal structures and NMR chemical shifts, respectively (311, 312). Yet, no direct evidence of the imidazole - imidazole hydrogen bond has been reported. NMR has a particular strength regarding covalent or hydrogen bond characterization, due to the resulting J-coupling. Indeed, two bonded atoms have a physical property known as J-coupling, and the coupling strength is directly related to the distance and the geometry between the two interacting spins. J-couplings have been extensively used in solution NMR to directly access hydrogen bond strength. Here we show, using the conduction domain construct of M2 in lipid bilayers, that the imidazole rings are hydrogen bonded even at a pH of 7.8 in the neutral charge state. An intermolecular 8.9 ± 0.3 Hz ${}^2\text{h}J_{\text{NN}}$ hydrogen bond is observed between H37B- N_ϵ and H37A- N_δ recorded in a fully protonated sample with 100 kHz magic-angle spinning. Interestingly, the hydrogen bond could not be detected in a sample with rimantadine bound to the pore.

3.2 Results and discussion

We measured M2₁₈₋₆₀ in 1,2-diphytanoyl-sn-glycero-3-phosphatidylcholine (DPhPC) bilayers at pH 7.8, as in previous work (313). This construct has been dubbed the ‘conductance domain’ since it has been shown to recapitulate the proton conduction rates and drug sensitivity properties of the full-length protein in liposome flux assays (327). In this construct, a C₂ symmetric referred to as dimer of dimers arrangement (313) at high pH results in two H37 H_{e2} chemical shifts, at 12 and 14.5 ppm (295), herein indexed as ‘A’ and ‘B’, respectively. The question remains whether the fourfold symmetry is broken by hydrogen bonded dimers or whether only water is the hydrogen bonding partner. Although the H37 proton chemical shifts are indicative of normal hydrogen bonds, a dimer arrangement in the neutral charge state would be expected to persist in the +2 state, since the dimer further

stabilizes a positive charge, according to *ab initio* quantum chemical calculations on imidazole dimers (328) and in the core of the M2 tetramer (329). Such calculations show that a positive charge strengthens the dimer interaction and might lead to a low barrier hydrogen bond.

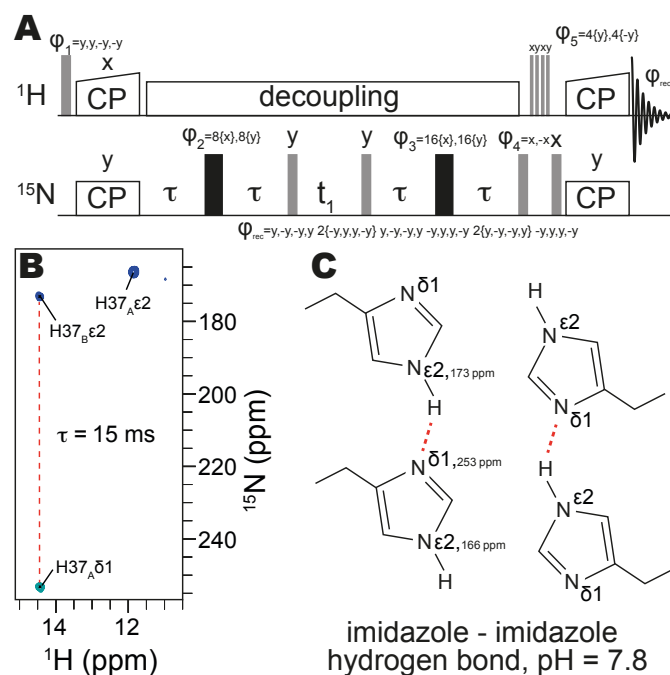


Figure 5.3.1. Measurement of $^2\text{hJ}_{\text{NHN}}$ hydrogen bonding in H37 imidazole dimers within influenza M2.

The pulse sequence is shown in a). Cross-polarization (CP) is used to establish ^{15}N polarization. A homonuclear out-and-back INEPT period follows to record the chemical shift of the J-coupled nitrogen. Following water suppression, CP returns signal to protons for detection. The spectrum in b) was recorded with τ of 15 ms, and clearly shows a negative peak indicative of an intermolecular J-coupling, and a C_2 symmetric tetramer at H37, as shown schematically in c).

We observed an imidazole-imidazole $^2\text{hJ}_{\text{NN}}$ -coupling using a homonuclear INEPT (330) period on the nitrogen channel, combined with cross polarization (CP) for detection of the attached proton (Fig. 5.3.1). When a homonuclear J-coupling is present, an additional peak occurs at the ^{15}N frequency of the coupled spin, with a buildup of intensity following the well-known relation, $\sin^2(2\pi Jt)e^{-\frac{\tau}{T_2}}$. Since the original peak follows $\cos^2(2\pi Jt)e^{-\frac{\tau}{T_2}}$ (Fig. 5.3.2B), normalization by the total signal results in a single parameter fit to $\sin^2(2\pi Jt)$. We measured a N_{δ_1} - N_{ϵ_2} J-coupling of 8.9 ± 0.3 Hz (Fig. 5.3.2A), which was unambiguously assigned to an intermolecular interaction (Fig. 5.3.3). This intermolecular N-H- N $^2\text{hJ}_{\text{NN}}$ coupling occurs for the most strongly downfield shifted proton at 14.5 ppm. No homonuclear J-coupling could be detected for the other N_{ϵ_2} , indicating that its attached H_{ϵ_2} at 12.1 ppm is likely hydrogen bonded to oxygen.

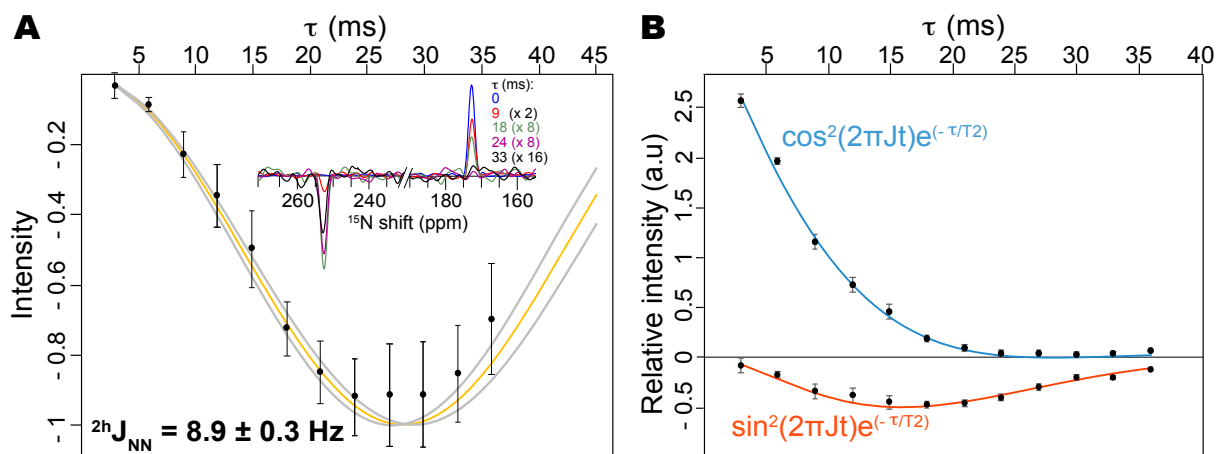


Figure 5.3.2. Quantification of the intermolecular $^2hJ_{\text{NHN}}$ J-coupling.

In the inset from A, slices of the 2D spectrum at the proton frequency of 14.5 ppm are shown for the indicated mixing times. In A, the experimental data (points) are shown with 2σ error bars accounting for random spectral noise. Relaxation was accounted for by dividing each intensity at 254 ppm by the total signal magnitude of the slice. The best fit (orange) resulted in a coupling strength of 8.9 ± 0.3 Hz. The curves in grey indicate the error at twice the standard deviation, σ , as estimated with a Monte Carlo approach and considering random spectral noise. The first point was acquired with 8 scans (1.5 hours) while the last point required 128 scans (26 hours) due to transverse relaxation. B shows the decay of the in-phase term from the $\text{N}\epsilon 2$ (blue curve) and the negative build up of the anti-phase term observed in $\text{N}\delta 1$.

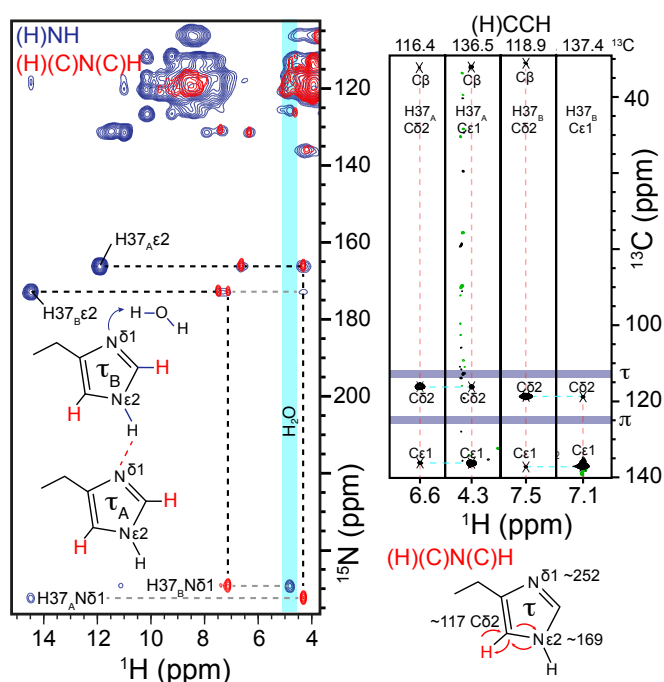


Figure 5.3.3. Histidine-water contact and assignment of H37 tautomer state.

In blue, an NH correlation spectrum shows magnetization transfer from a non-protonated imidazole nitrogen at ~ 250 ppm to water (4.85 ppm) using 6 ms of CP. In red, the nitrogen and carbon resonances are assigned by out-and-back 1-bond CP transfer $(\text{H})(\text{C})\text{N}(\text{C})\text{H}$. The $\text{H}\delta 2$ is correlated only with $\text{N}\epsilon 2$ in this magnetization transfer scheme, which resonates at ~ 170 ppm and establishes that all histidine residues in the channel are in the τ tautomer. Magnetization transfers are indicated by curved arrows. The d and e carbon assignments were confirmed in a RFDR-based $(\text{H})\text{CCH}$ spectrum (black).

To assess the tautomeric states of histidine, we recorded a CP based out-and-back $(\text{H})(\text{C})\text{N}(\text{C})\text{H}$ spectrum. The use of 1-bond transfer times (1.6 ms) ensures that the $\delta 2$ proton is connected with only the $\epsilon 2$ nitrogen. Correlations are found at 165 and 172 ppm, corresponding to the two protonated imidazole nitrogen resonances observed in the $(\text{H})\text{NH}$

spectrum and establishes *de novo* that both signals are from τ tautomers. The C δ_2 resonance of H37_B is found midway between the shifts observed in crystals of τ and π tautomers of Histidine (Fig. 5.3.3, purple)(331), highlighting the need to independently establish the tautomeric states.

That both water-associated and hydrogen bonded dimer conformations of H37 exist under different conditions suggests that these conformations are both relatively stable. The main differences in sample preparation that result in these different structures are the nature of the membrane mimetic, and the length of the construct. The fact that the $^2\text{hJ}_{\text{NN}}$ coupling is observed in the functional ‘conductance domain’ construct in lipid bilayers suggests that it is a relevant state, and that the proton affinities of the dimer control the interconversion to a conducting channel. This does not rule out the possibility that other quaternary structures may also lead to a proton current in virus particles, although a different pKa would be expected at H37. This is in line with the large range of pKa values reported for M2 (294-296, 332).

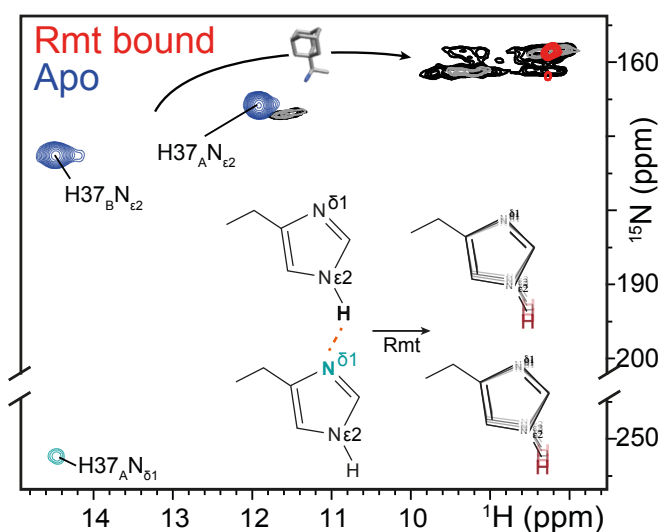


Figure 5.3.4. Chemical shift changes in the histidine side-chain upon addition of the drug rimantadine (rmt).

The pulse sequence of fig. 5.3.1 with τ of 6 ms (blue, red) and (H)NH spectra with 25 ms (grey) or 200 ms (black) of ^{15}N exchange during the water suppression. A 3-5 ppm change is observed in the drug bound spectrum (red) and no $^2\text{hJ}_{\text{NHN}}$ J-coupling was observed. Instead, the imidazole NH peaks are broadened, and the peaks at 9 ppm are in exchange. The (H)NH spectra were acquired at 250 K and 80 kHz MAS to reduce the temperature by $\sim 10^\circ\text{C}$ to slow exchange.

Interestingly, with the inhibitor rimantadine bound to the pore, the exchangeable histidine protons shift upfield by about 3 ppm, and no $^2\text{hJ}_{\text{NN}}$ coupling could be detected (Fig. 5.3.4). This disruption of the hydrogen bonding interaction explains the difference in H37 pKa in the drug bound state (333) and lends support to the hypothesis that imidazole dimers are functionally important. The simplest explanation for inhibitor efficacy is pore blockage preventing the passage of hydronium ions. And yet, significant chemical shift perturbations were detected widely over transmembrane residues (313, 334). This suggested that the

inhibitors have far reaching effects, modifying the conformational distribution of the protein. However, it has been difficult to connect a specific structural change to the chemical shift perturbations. It is now clear that in the conductance domain construct, rimantadine affects the structure of the channel by impacting the hydrogen bonding of the important functional residue H37.

3.3 Conclusion

In conclusion, through the measurement of a ${}^2\text{hJ}_{\text{NN}}$ coupling, we confirmed the existence of imidazole–imidazole dimers in the M2 protein from influenza. Such a configuration was proposed to stabilize positive charge in the tetrameric channel. However, direct evidence of the interaction was not previously reported, and crystal structures of M2 showed an alternate structure, leading to controversy over whether the dimeric histidine arrangement exists at all in M2. We have solved this controversy through NMR measurements of M2 in lipid bilayers at high pH where we observe a neutral charge state at the functional H37 residue. The coupling strength is consistent with a normal hydrogen bonding interaction. Binding of the drug rimantadine resulted in breaking off this hydrogen bond. It remains to be seen whether evidence can be found that this geometry persists in the important +2 charge state, where imidazole–imidazolium dimers have been proposed, and whether such a state results in a normal hydrogen bond or a low barrier hydrogen bond.

4. Structure determination of M2 in membranes by proton detected NMR under ultra fast MAS

4.1 Abstract

As of today, only a handful of structures have been solved by proton detected solid-state NMR in 100% back change samples (335) or fully protonated samples (81). Solution NMR necessitates a detergent environment and high deuteration levels for accessing site specific information on large membrane proteins, rendering the technique limited to small molecular size. On the other hand, solid-state NMR, although applicable to larger proteins suffers from relaxation due to dipolar coupling and CSA that deteriorates both the spectral resolution and sensitivity. To ameliorate spectral resolution, ssNMR uses sample rotation at the “magic angle” of 54.74° from the B_0 field.

Previous ssNMR measurements use carbon and nitrogen detection with high proton power decoupling. This limits structural information because protons are not accessible. In addition, these measurements are time consuming due to long coherence life time of the detected nuclei (336, 337). In the last decades, with an increase of the spinning speeds up to 55 kHz, proton detection has become available for protonated samples with limited proton resolution (82) and for largely deuterated samples with exquisite resolution and improved sensitivity (30, 338). However, perdeuterated samples give limited structural information and often require several samples with specific labelling to gain sufficient information to perform high quality structure calculations.

Recent hardware developments have resulted in spinning speeds up to 110 kHz using rotor sizes 0.7 mm of diameter. Despite the small rotor diameter and the small sample quantity (half a milligram), there is limited loss in sensitivity that can be less than the variation in probe performance (339), making this approach suitable for investigating fully protonated samples in membrane environments.

Here, the structure of the conductance domain (residue 18 to 60) of the matrix 2 (M2) protein from Influenza A has been solved using a single fully protonated ^1H , ^{13}C , ^{15}N sample using proton detected ssNMR at ultra fast MAS. M2 is a 97 residue transmembrane protein which has been found as a homo-tetramer in membranes (271). The protein has a largely unstructured N-terminus (residue 1 to 22), facing the outside of the virus particle, a single

helical transmembrane domain (TM from residue 22 to 43), an amphipathic helix (residue 44 to 60) and an unstructured C-terminal domain that interacts with matrix protein 1 (M1).

4.2 Results and discussion

^1H , ^{13}C , ^{15}N M2 protein was packed in a 0.7 mm rotor and spun at 100 kHz at a 950 MHz NMR spectrometer in order to elucidate a de novo structure. Since full protonation gives access to protons in the alpha position as well as in the sidechains, we recorded the full set of proton detected experiments recently reported by Pintacuda and co-workers (338). The experiments and parameters used for assignment spectra and distance measurements are listed in the materials and methods section (5.2.5 and 5.2.6). Shortly, (H)NH, (H)CANH, (H)(CO)CA(CO)NH, (H)CONH, (H)CO(CA)NH, (H)(CA)CB(CA)NH, (H)(CO)CB(CO)NH were used for assignment proposes in combination with H(H)NH using 3 and 6 ms of proton - proton mixing and (H)X(HH)XH (X denotes ^{15}N and ^{13}C) as structural restraints.

Figure 5.4.1 shows the strips extracted from (H)CANH, (H)(CO)CA(CO)NH, (H)CONH and (H)CO(CA)NH using CcpNmr. The line widths obtained in our preparation are ~300 Hz for $\text{H}\alpha$, ~120 Hz for $\text{C}\alpha$, ~210 Hz H^{N} and ~60 Hz N^{H} . With these narrow lines, we were able to assign two sets of peaks for the backbone resonances of the M2 tetramer supporting the dimer of dimer structure previously reported (322, 328).

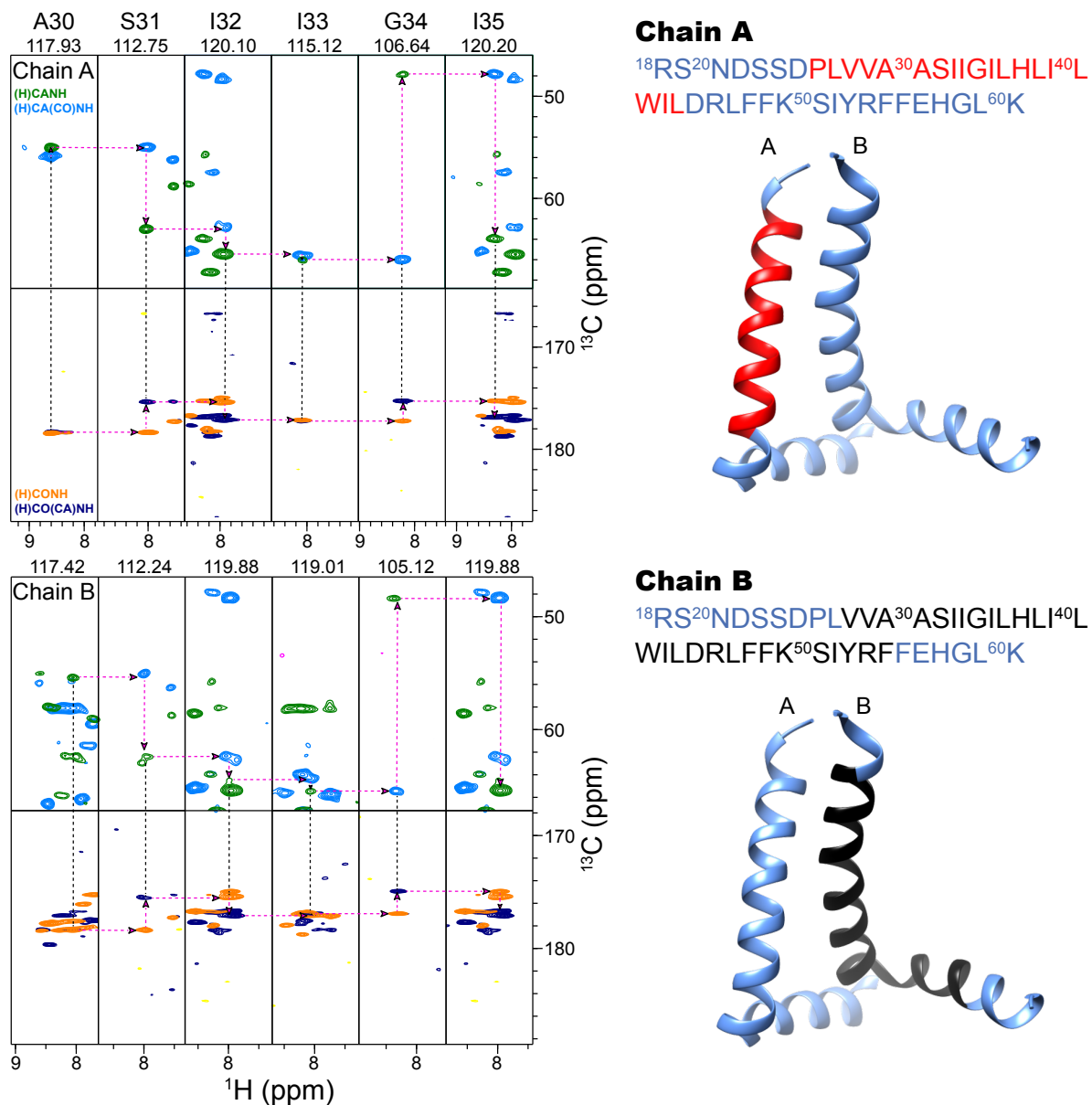


Figure 5.4.1. Backbone walking assignment spectra of a fully protonated $^{13}\text{C},^{15}\text{N}$ M2 at 100 kHz MAS. Linking assignments spectra, (H)CANH (green), (H)CA(CO)NH (blue), (H)CONH (orange) and (H)CO(CA)NH (dark blue), with the assignments of the residues A30 to I35 are shown for both chains. Both sets of peaks for Chain A (top panel) and Chain B (bottom panel) were assigned. On the right hand, the assigned residues are represented in red (Chain A, top) and black (Chain B, bottom) on the M2 structure (pdb: 2L0J).

The sample allows the complete $^1\text{H},^{13}\text{C}$ and ^{15}N assignment of key functional residues H37 and W41 in the lipid bilayer (Fig. 5.4.2, A and D). Sequential assignment was improved by using both (H)COCAHA and (H)NCAHA spectra which allows linking through the $\text{H}\alpha$ (Fig. 5.4.2 E to G). Figure 5.4.2 shows the full set of experiments measured on a 950 MHz Bruker spectrometer with a 0.7 mm probe at 100 kHz MAS. In A the (H)NH spectrum with the assignment of the HN of the imidazole histidine sidechains and the W41 sidechains is shown. Figure 5.4.2 E to H show the assignment of the $\text{H}\alpha$ of Ala30_A and His37_B. The

assignment of the W41_A, W41_B, H37_A and H37_B sidechain is shown on the (H)NH (Fig. 5.4.2A, inset) and (H)CH (Fig. 5.4.2D).

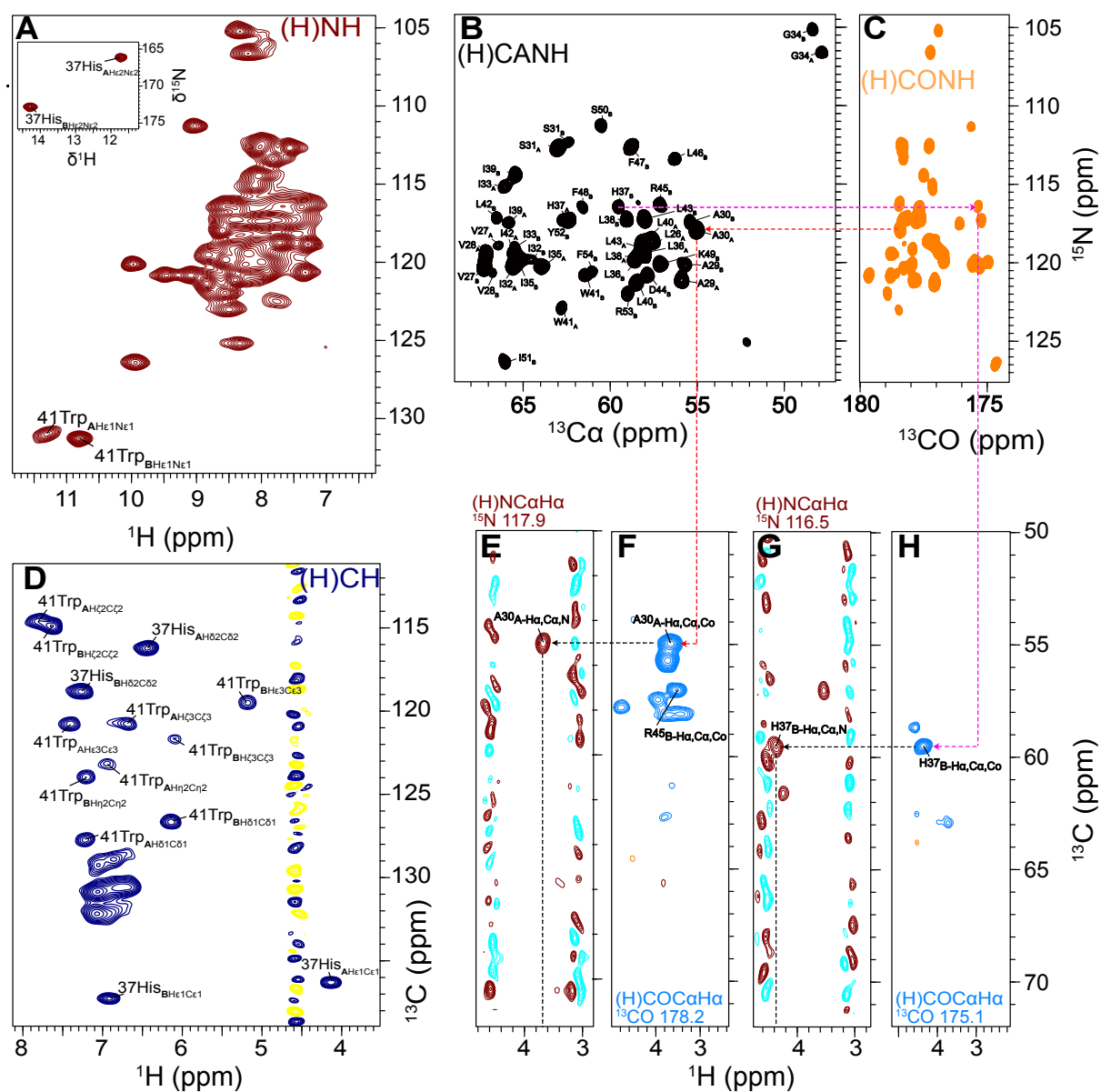


Figure 5.4.2. Solid-state NMR assignments of a fully protonated M2.

In A, the (H)NH spectrum with the assignment of 37His and 41Trp is shown. B and C show the $^{13}\text{C},^{15}\text{N}$ assignment of M2 from the 2D projection of the (H)CANH and (H)CONH, respectively. The aromatic region of the (H)CH with the assignment for the residue W41 and H37 are shown in D. F to I show the H α assignment of residue A30_A and H37_B from the (H)COCAHA (blue) and (H)NCAHA (brown) spectra. All the spectra were recorded at 100 kHz MAS with a 0.7 mm 4 channel Bruker probe in a 950 MHz Bruker spectrometer at 265 K.

To determine the sidechain assignments a 3D (H)CCH was recorded using 1.6 ms of carbon RFDR mixing. Due to repeated residues and the dimer of dimers architecture of the protein, the sidechain assignments are still ambiguous for certain residues (full assignment list

is provided in appendix 5.I). Figure 5.4.3 shows the structure of M2 from only proton detected data of the single sample at 100 kHz MAS.

Table 5.4.1. Spectra used for the structure calculation.

Spectrum	Mixing	Mixing time (ms)
(H)NH	Cross polarization	3 and 6
(H)C(HH)CH	RFDR	0.5
(H)N(HH)NH	RFDR	0.5
(H)X(HH)XH	RFDR	0.5

The time share ($^{13}\text{C}/^{15}\text{N}$) (H)X(HH)XH allows for recording 4 spectra in one ((H)C/N(HH)C/NH) concomitantly (340, 341). Since M2 forms a tetrameric structure in the lipid bilayer and we did not have any unambiguous inter-helical contact, manual restraints had to be added at the N and C-terminus of the protein to ensure tetramer arrangement (Appendix 5.II, table S5.5). The distances entered were determined based on the previous high resolution M2 data available (pdb: 2l0j). The recently reported imidazole - imidazole hydrogen bond using the same sample conditions (342), was manually entered (Appendix 5.II, table S5.4), and TALOS (343) restraints were used based on $C\alpha$, $C\beta$, H^N and N^H chemical shift information (Appendix 5.II, table S5.2 and S5.3). The long HN CPs of 6 and 3 ms showed the existence of contacts between sidechains of H37_{He2Ne2} and W41_{He1Ne1}. Those contacts were used to complement the N and C-terminal restraints.

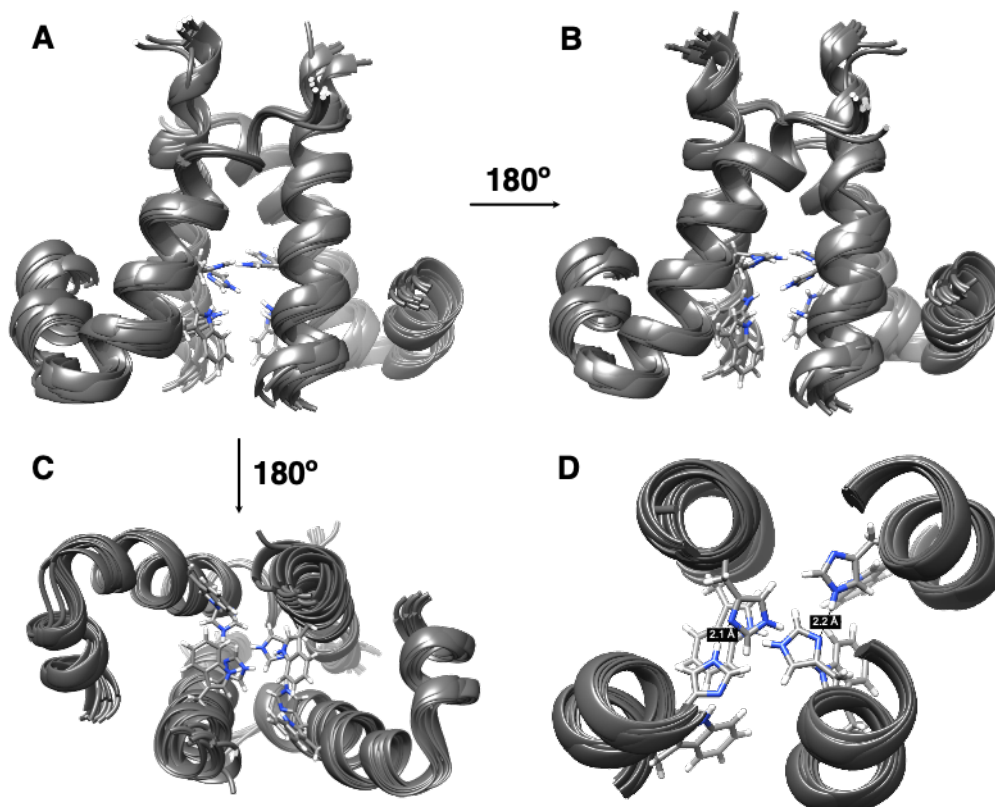


Figure 5.4.3. De novo structure calculation of M2 from influenza A.

In all the panels the ten best structures calculated from CYANA are shown. A and B show the full tetramer rotated 180° with respect to each other. C shows the C-terminus of the tetramer and D shows the top view from residue 34 to 41 with the H37 and W41 sidechains included.

To evaluate our structure calculation, independently of the NMR data and the manual restraints used, we calculate the structure of M2 based uniquely on the amino acid sequence with CS-Rosetta. CS-Rosetta allows a structure model prediction based on sequence alignment and permit to exclude the existing homologue structures for calculation (344). Figure 5.4.4 shows the comparison between the two best structural models obtained with Rosetta (Fig. 5.4.4A to D, cyan = best hit; orange = second best hit) and our structure calculated from CYANA (Fig. 5.4.4E to G). Although the four helical bundles are found by the CS-Rosetta run, the structures still show some degree of freedom at the TM domain, as can be observed in the helix close to residue G34 (red circle) where the TM is disrupted. This is consistent with the previously reported kink at residue G34 by PISEMA data (303). In contrast to the crystal structures, the structure from CS-Rosetta shows a close packing at the C-termini of the tetramer, similar to the reported ssNMR structures (23, 309, 312).

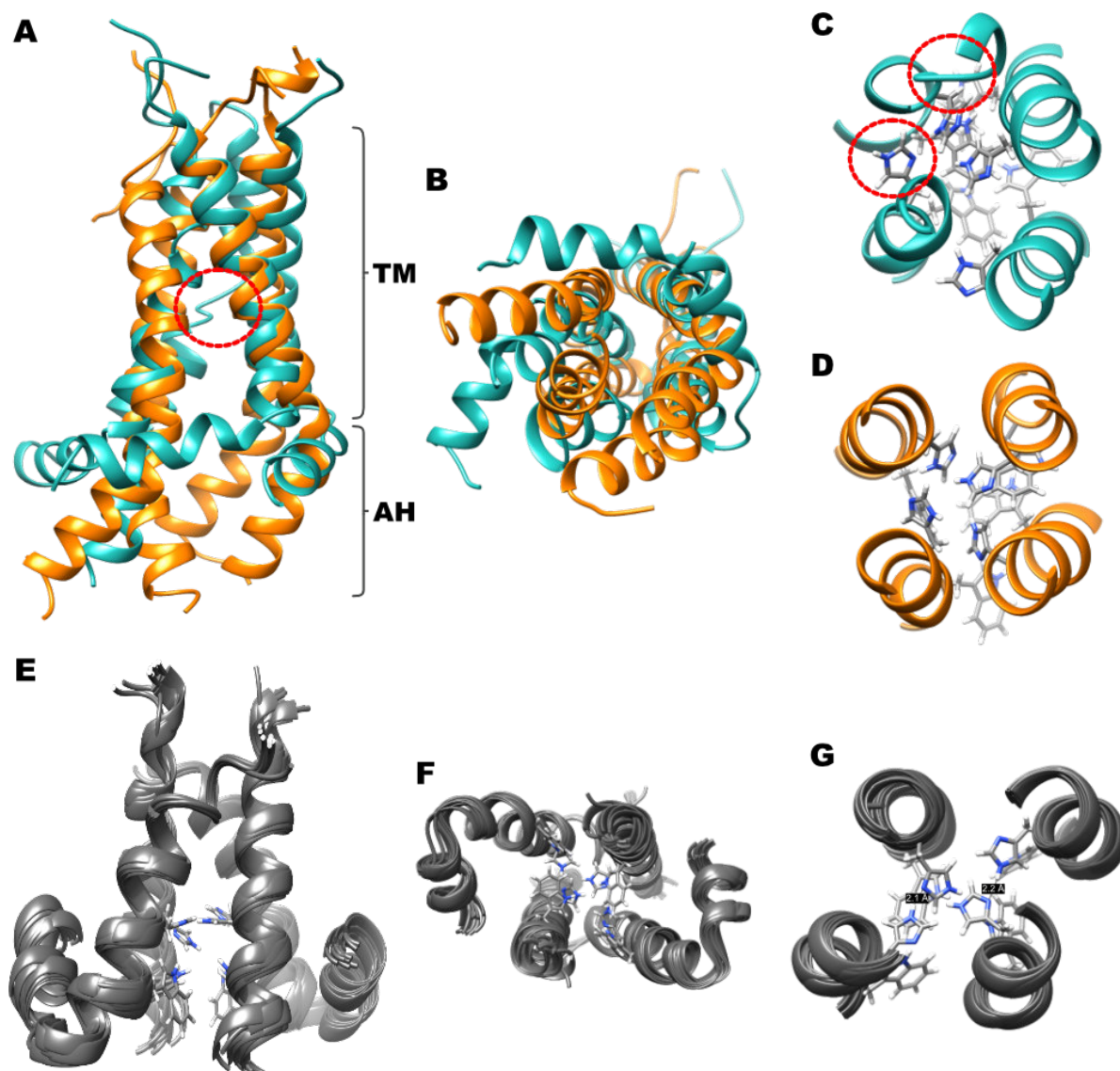


Figure 5.4.4. Comparison of the M2 structure built by homology or *de novo* calculation.

In A to D, the two best structures from CS-Rosetta are represented in different views. The higher scored structure is represented in orange and the second best in sky blue. A shows the front view of both CS-Rosetta structures. In B, the C-terminal view of the tetramer in both structures is shown. In C and D the top view of both structures is shown and the sidechains of H37 and W41 residues are represented. From E to F, the 10 best solid-state NMR are shown. The front view and the view from the C-terminus of the tetramer are shown in E and F, respectively. G shows the top view from residue G34 to W41. The red circles in C point out the bizarre arrangement of the W41 sidechain predicted by CS-Rosetta.

Despite the similar arrangement between the different methods used for structure determination, Rosetta shows a distinct sidechain arrangement regarding the important residue H37. Interestingly, the H37 sidechain points towards the TM interface and lipid bilayer in the best model (Fig. 5.4.4C, red circle). The second best structure (in orange) shows a more extended amphipathic helix inconsistent with the amphipathic domain laying on the membrane surface (312). Additionally, contrary to the available structural data, the extended structure shows a W41 pointing towards the TM interface and lipids. Indeed, It is well

established that W41 and H37 sidechains are pointing towards the pore of M2 acting as gate and proton selectivity filter, respectively (300, 311, 345).

The solid-state NMR structure shows a widely opened C-terminal region and a disrupted N-terminal domain for one of the helices. Beside these discrepancies, the calculated structure is in agreement with the CS-Rosetta and the previous NMR structures reported. Although a refinement of the structure is necessary it is a good starting point with the recorded data and shows the possibilities of using ultra fast MAS in fully protonated proteins for structure determination.

4.3 Conclusion

Here we reported a de novo structure calculation for the M2 proton channel from Influenza A. Full assignment of the important residues W41 and H37 are reported and the structure was calculated by using data exclusively from a $^1\text{H},^{13}\text{C},^{15}\text{N}$ M2 sample at 100 kHz MAS. The distance restraints were recorded in about a month and with only ~0.5 mg of sample, hence showing the ample possibilities offered by this technique to assess atomic resolution for membrane proteins in terms of time and protein yields.

5. Non-specific binding of rimantadine

5.1 Abstract

The drug rimantadine (rmt) binds to two different sites in the M2 protein from influenza A, a peripheral site and a pore site, which is also the primary site of efficacy. It remains enigmatic that pore binding of rmt did not occur under certain sample conditions such as DHPC micelles, as well as in a mixture of lipids selected to match the viral membrane. Here by using controlled temperature binding experiments, at ~4 degrees, we were able to trap an intermediate state of M2 occurring before pore binding. These data explain how thermodynamic limitations are the origin of the lack of pore binding in the presence of DHPC. Two effects are responsible, namely the kinetics of binding and changes in the surface properties of the tetramer upon rmt pore binding. We tracked both specific and non-specific effects via chemical shift perturbation using ultra-fast magic-angle spinning of 55 to 100 kHz and fully protonated M2 reconstituted in DPhPC bilayers. Slow pharmacological binding

kinetics allowed us to characterize the non-specific binding in the absence of pore binding. Binding kinetic measurements showed non-Arrhenius behavior suggesting a cooperative effect in the binding kinetics. Pore binding requires high activation energies (~117 kJ/mol) consistent with rearrangement of the water pore, hydrogen bond disruption and structural rearrangement. Pore binding causes a change in the physicochemical properties on the M2 surface such that solubilization in DHPC is hampered.

5.2 Results and discussion

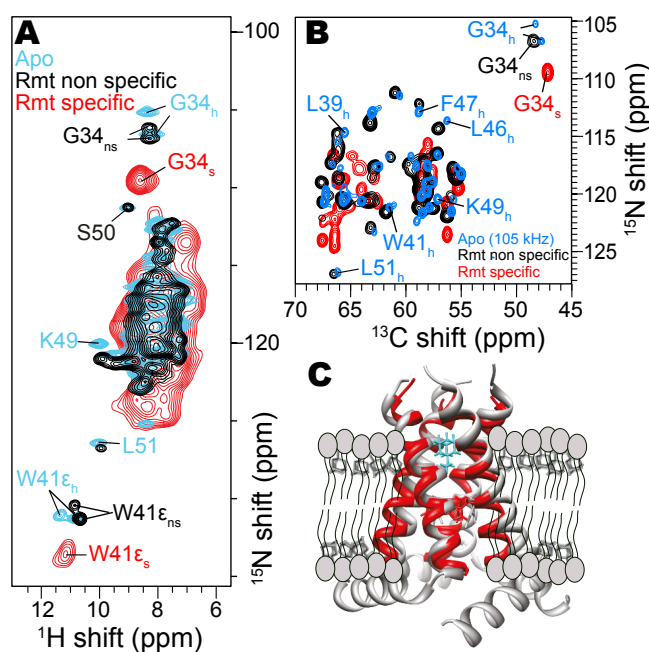


Figure 5.5.1. Chemical shift changes in M2 induced by non-specific binding with 40 mM rmt.

In A, 2D (H)NH spectra of M2 show the small changes between apo (blue), and non-specific binding (black) and the larger changes that accompany pore binding in (red). In B, the $^{13}\text{C}^{15}\text{N}$ projections of (H)CANH spectra show chemical shift perturbation of the $\text{C}\alpha$ and N^{H} backbone. A large excess of 16 molecules of rmt drug was used per tetramer. In C, the pore binding location (cyan) is shown on the conductance model, pdb 2L0J in grey and on the bound TM structure 6BKL in red.

The M2₁₈₋₆₀ protein was reconstituted in deuterated DPhPC and then the sample was incubated with 40 mM rmt at 4 °C. In contrast to previous reports, only small chemical shift changes (Fig. 5.5.1A, black) are observed in the presence of drug (Fig. 5.5.1A, cyan) in (H)NH magic-angle spinning (MAS) spectra. The previous studies in phosphocholine lipid bilayers showed large chemical shift perturbation (CSP > 2 ppm for ^{15}N and > 1 ppm for ^{13}C) throughout the TM region residues 27 to 42 upon pore binding (313, 319, 334). This apparent discrepancy is explained by slow kinetics of pore binding at 4°C, and is qualitatively consistent with the fact that glycerol was observed to kinetically prevent pore binding (318). Even after several weeks of incubation, pore binding was not observed, while both pore-bound (specific) and pore-unbound (pore-apo) populations were present after a week of incubation at 20 °C. The two populations can be seen clearly for G34, with two distinct ^{15}N chemical shifts at 106.5 ppm (Fig. 5.5.1B, G34_{ns} black) and at 109.5 ppm (Fig. 5.5.1B, G34_s

red), corresponding to pore-apo and pore-bound states, respectively (Fig. 5.5.1B). We refer to pore binding as specific since it is the primary site of inhibition (317), while the non-specific effects of the drug refer to any other changes observed in the spectrum.

Kinetic control allowed characterization of non-specific CSP. While apo assignments were carried out at 100 kHz (Appendix 5.I), for the non-specific bound we reduced the spinning to 80 kHz MAS with a 0.7 mm rotor to reach a sample temperature of ~5 °C to prevent pore binding during acquisition. Figure 5.5.2 shows the excellent resolution obtained under these conditions for a fully protonated sample, which allowed us to transfer apo assignments to the non-specific bound state using 3D triple resonance spectra. Consistent with previous reports (309, 313, 328), two sets of peaks in the TM region indicate a dimer of dimers structure. The CSP was calculated for each residue as

$$\sqrt{\frac{1}{3}(\Delta\delta_H^2 + 0.14\Delta\delta_N^2 + 0.3\Delta\delta_C^2)}.$$

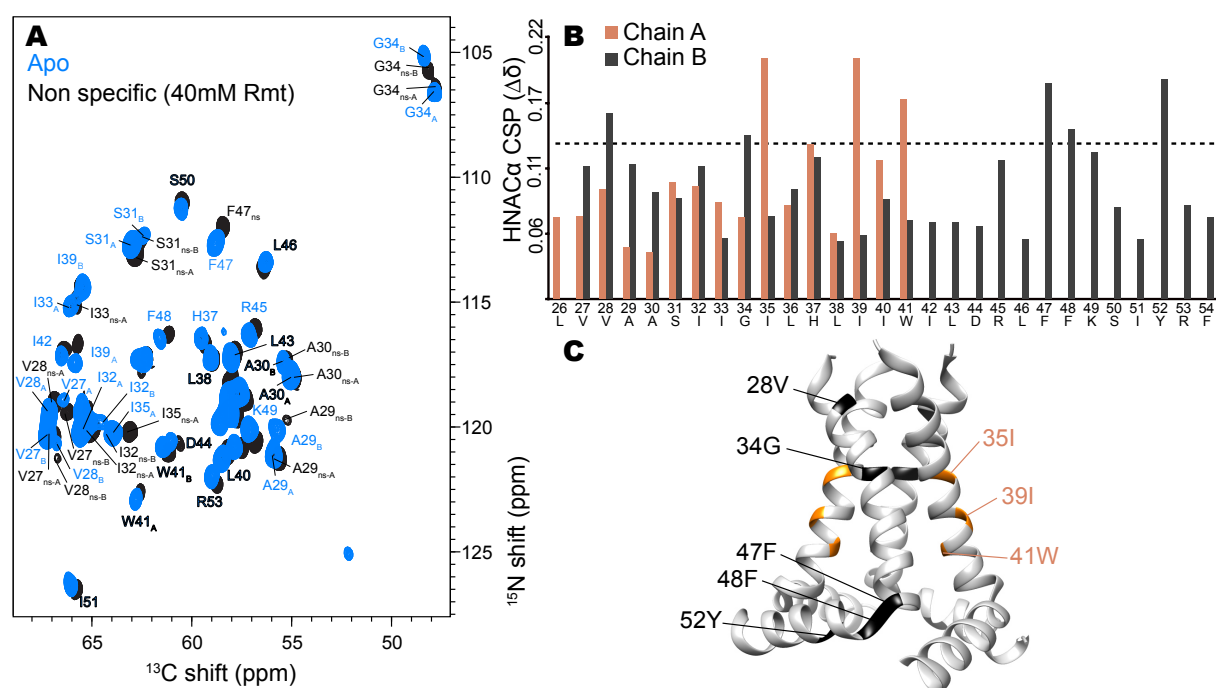


Figure 5.5.2. Non-specific CSP.

The CN projections of (H)CANH spectra are shown in A with apo in blue and non-specific in black. In B, the CSP is shown per residue by combining H^N , $C\alpha$, and N chemical shifts as described in the text for both A and B chains of the dimer. In C, residues with CSP greater than 2 sigma (~0.16 ppm) are represented on the structure of pdb 2L0J with the color corresponding to the chain where the residue is affected (chain A in orange and chain B in black).

In the TM region, chain A shows the largest perturbation toward the C-terminus, from G34 to W41 with a cluster of large CSP for residues I35, I39 and W41. Chain B shows the largest CSP for residues V28 and G34 located at the N-terminus of the TM. The non-specific

CSP is much smaller than observed for pore binding, which causes > 2 ppm CSP across the TM residues. This is indicative of conformational changes for the pore-bound state (306, 313, 334). The non-specific CSP is likely due to a combination of changes in the lipid properties as the hydrophobic drug partitions to the membrane and to association at the peripheral binding site, nearby residue R45, identified previously (23). This is also consistent with DNP data at cryogenic temperature using TEDOR experiments showing contacts between rmt and D44 and R45 (318). Consistently, for non-specific bound state, we observed elevated $C\alpha$ CSP for residues F47, F48, and Y52 near this peripheral site. While the CSP for non-specific binding are small, they are distributed across the helices and include the pore facing residue G34, which is sensitive to membrane conditions and can form a kink upon pore binding (306). This emphasizes a degree of sensitivity of the channel to its environment, a topic explored before for the TM construct (308). A detailed plot of CSPs for each of H^N , N, and $C\alpha$, is shown in figure 5.5.3.

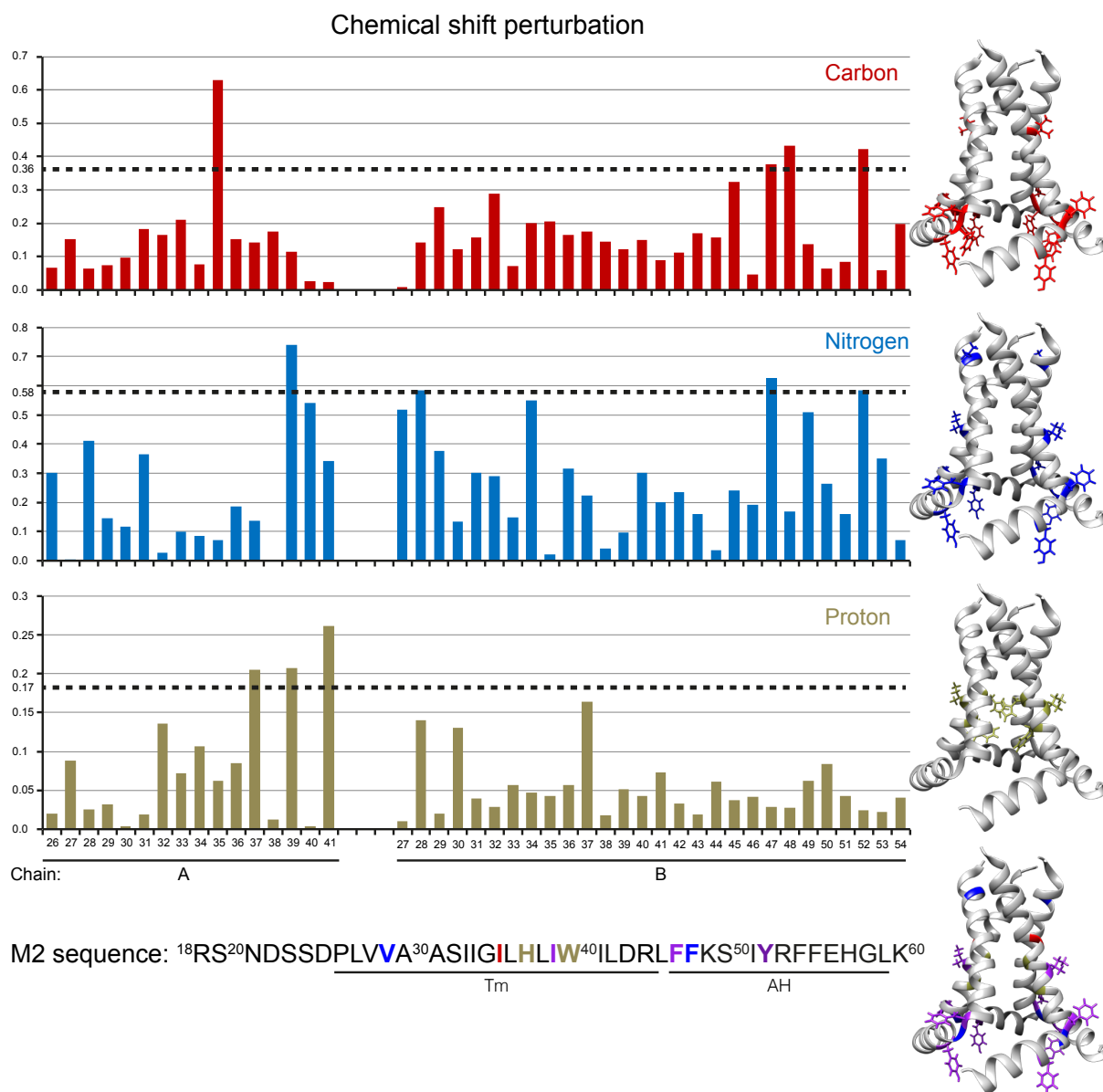


Figure 5.5.3. Individual chemical shift perturbation observed for rmt non-specific binding in M2.

In red are represented the carbon chemical shift perturbation (CSP). In blue are shown the nitrogen CSP and in gold the proton CSP.

Residues H37 and W41 are key functional residues controlling conduction in response to pH. We recently showed that there is an imidazole - imidazole hydrogen bond in the apo state at pH 7.8 and that the hydrogen bond is broken upon rmt pore binding (342). H37 N-H- - N hydrogen bonding persists at a lower pH of 6.2 in a full-length construct (346). Non-specific binding causes small CSP at the H37N ϵ , and the imidazole - imidazole hydrogen bond was still present (Fig. 5.5.4). This indicates that the non-specific effects and peripheral binding likely have at most a minor influence to the pore. Since our pore-bound preparations are also likely bound peripherally, however, we cannot exclude that the two binding locations act cooperatively.

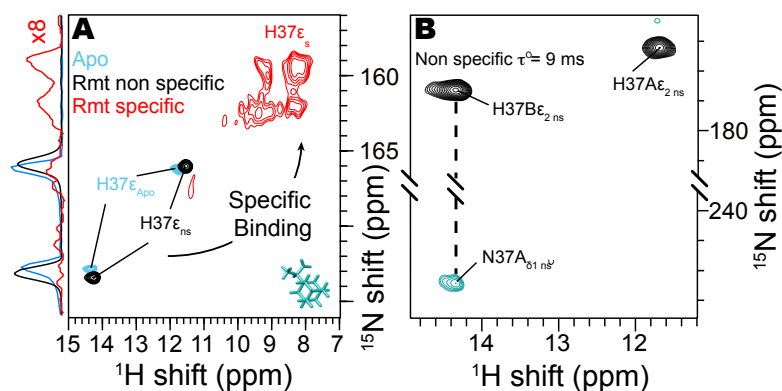


Figure 5.5.4. Histidine side-chain CSP and hydrogen bonding.

Non-specific changes (cyan) are small compared with specific pore binding (red). The left panel shows (H)NH spectra of the three sample conditions. The right panel shows via evolution of the homonuclear J coupling that imidazole-imidazole hydrogen bonding occurs. Data were recorded at a 950 MHz spectrometer.

To kinetically characterize the binding process, the rotor containing 40 mM rmt bound non-specifically, previously incubated at 4 degrees, was sealed and incubated at different temperatures 25°C, 40 °C and 55°C in a water bath. Binding was tracked by recording (H)NH spectra periodically and replacing the sample in the bath. Binding occurred over several hours or days depending on the incubation temperature, and was quantified with the intensities of three well-resolved residues, H37Nε2 (for both chains A and B) and G34_s. Figure 5.5.5 shows the combined intensities of both H37Nε2 A and B. Surprisingly, at temperatures of 20 °C (Fig. 5.5.5, blue) and 40 °C (Fig. 5.5.5, green) the curves do not follow an exponential decay, but rather proceed more slowly at first, indicating kinetic cooperativity. A possible explanation for this behavior is that neighboring tetramers may interact in the dense and highly concentrated NMR sample, such that binding is accelerated when a neighboring tetramer is already bound. Interestingly, it has been observed that M2 clusters in the membrane. The clustering of M2 has been suggested to play a role in membrane curvature (347, 348) but might also explain the unusual binding kinetics observed.

To further characterize the kinetics, a first rate kinetics approximation was used for extrapolating the energy barrier using the Arrhenius equation, $k = A e^{\frac{-E_a}{RT}}$ with k the kinetic rate, E_a the energy activation, R the gas constant and T the temperature. Figure 5.5.5 shows the Arrhenius plot of $\ln(k)$ as a function of $1000/T$ where the slope gives direct access to the activation energy (slope = $-E_a/R$). Approximating the rates based on initial points until 50% decay of intensity, an activation energy of about ~113 to 119 kJ.mol⁻¹ is estimated, depending on whether the H37Nε2 or G34 peaks are measured, respectively. This is a relatively high activation energy, similar to the barrier encountered in enzymatic processes, that will be consistent with the observation of substantial helical repacking. It suggests that the pore is not

easily accessible to the drug in the apo state, which is closed at the N-terminus by residue V27. That has been referred to as a secondary gate (302).

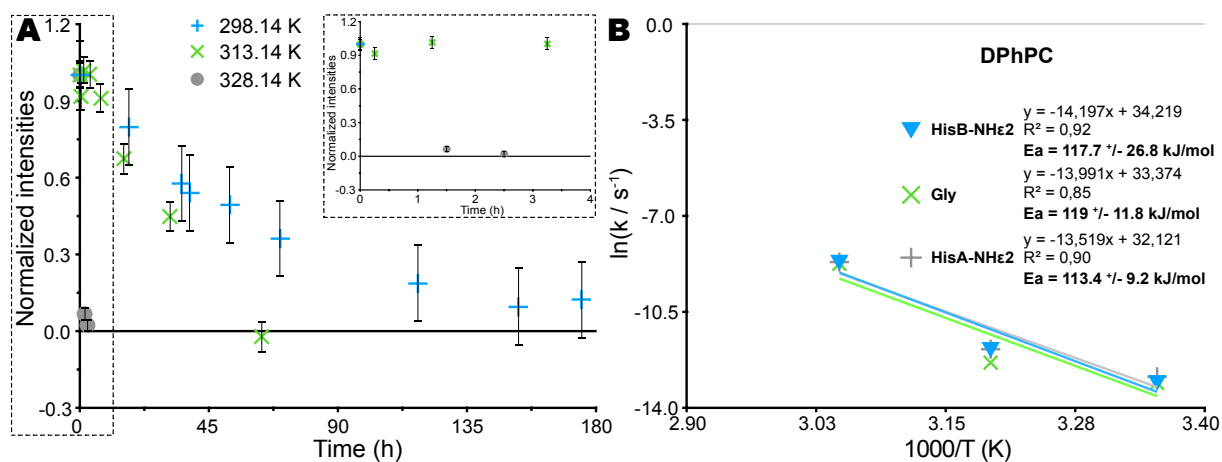


Figure 5.5.5. Kinetics of rmt binding.

A shows the disappearance of characteristic peaks (H37Nε2 and G34N) of the apo (H)NH spectrum at three temperatures. B shows fits of the initial rates to the Arrhenius equation and extraction of the activation energy (E_a). The samples were prepared with 40 mM rmt at 4°C and then incubated by placing the rotor in a water bath at 25°C, 40°C, or 55°C. Two samples were used for the measurements. WT M2 was used at 40 and 55°C and H57Y M2 (a point mutation in the amphipathic domain) was used for the 25°C measurements. Identical chemical shifts were observed for both samples.

Studies on similar systems, where a pore contains a cluster of organized water molecules, have shown that the reorganization of water molecules necessitates ~3-fold higher activation energies than bulk water (349), but this is still only ~30 kJ/mol (350). X-ray structures suggest that during pore binding, rmt displaces water molecules found at the N-terminus of the pore (300, 351). In addition to the water displacement, rmt pore binding disrupts the imidazole - imidazole hydrogen bonding (342) yet hydrogen bonding accounts for an energy of generally less than 10 kJ/mol (352). Therefore larger collective motions such as helix repacking and transient opening of the N-terminus are likely to be responsible for the bulk of the activation energy.

Since we observed slow kinetics in lipid bilayer samples, we thought that this might be a limiting factor for drug binding in DPhPC micelles, and that pore-bound samples in micelles could be prepared by first binding rmt in lipids, and then solubilizing in detergent. A solution of 300 mM DPhPC micelles was prepared and used to solubilize rmt-bound M2 from the DPhPC lipid preparation. While some protein was solubilized, a large precipitate remained (Fig. 5.5.6A). A second sample of M2 without rmt was fully solubilized in DPhPC. Solution ^{15}N -HSQC spectra were recorded for both supernatants (Fig. 5.5.6B), and both

matched the pore-apo M2 spectrum previously reported by Chou and co-workers (23). Much lower sensitivity was observed in the supernatant from the bound sample, suggesting that only the remaining pore-apo protein could be solubilized. This suggests that pore binding has a strong impact on the protein surface via repacking of the helices that affects the solubility in micellar conditions. The fact that a chimeric protein containing C-terminal residues from Influenza B M2 could be solubilized with drug in the pore suggests that these effects may depend on C-terminal residues. Indeed, the chimeric protein displayed large chemical shift changes upon binding for many residues, including those of the C-terminus, in particular at A40 (353).

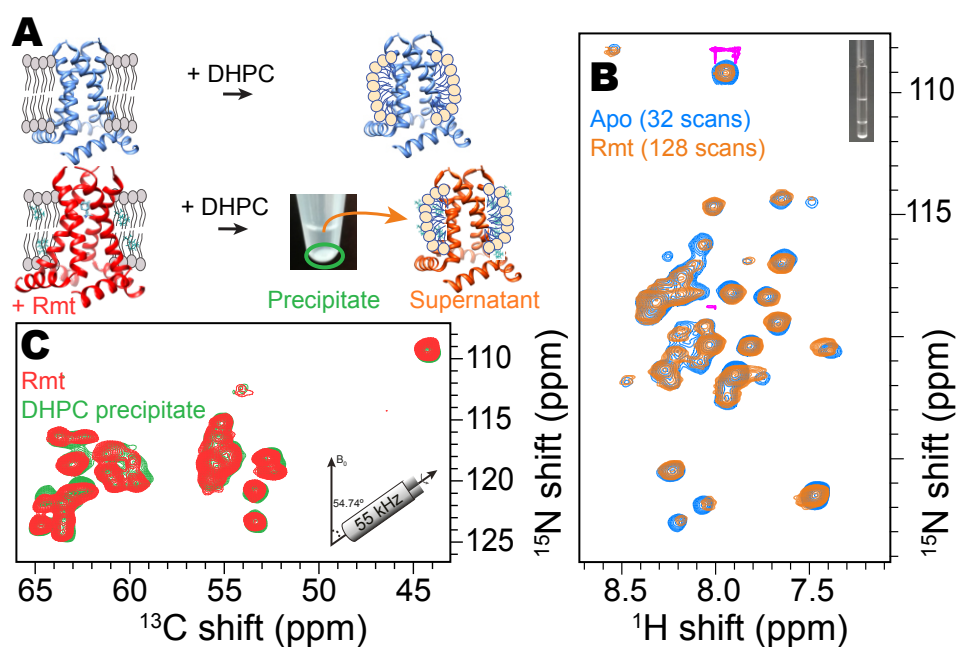


Figure 5.5.6. Pore binding results in channel restructuring.

In A, a schematic shows how the M2 tetramer was either dissolved directly with DHPC detergent (top, blue spectrum in B) or first bound with rmt and then extracted with DHPC leaving a supernatant (orange spectrum in B) and an insoluble pellet (green spectrum). In B, the ^{15}N -HSQC spectrum is shown for the soluble fraction after DHPC addition for the apo (blue) and pore bound (orange) samples. In C) the MAS spectrum of the pellet is compared with the DPhPC lipid spectrum (red). Panel B) shows spectra recorded at a Bruker 600 MHz spectrometer using a cryoprobe. Panel C), shows projections of proton detected 3D (H)CANH spectra acquired at an 800 MHz spectrometer using 55 kHz MAS and a 1.3 mm rotor.

To our surprise, the pellet from the bound sample showed nearly the same spectrum as the drug bound sample in lipid bilayers (Fig. 5.5.6C), suggesting that lipids were not removed by detergent. This indicates that bound tetramers likely pack closely together in such a way that prevents removal of lipids. This is also consistent with the observation of non-exponential

kinetic behavior. Clearly there is a change in the physicochemical surface properties of the M2 tetramer upon drug binding.

5.3 Conclusion

In summary, we showed using proton detection in combination with ultra-fast MAS, that at low temperature M2 can be kinetically trapped in a pore-apo state. This allowed measurement of non-specific CSP and determination of binding kinetics with real time NMR. The results suggested a change in tetramer packing upon rmt drug binding. The change in packing happening upon pore binding, impacts the aggregation propensity in DHPC micelles, which explains the lack of pore bound protein in solution NMR samples. The significant kinetic barrier to binding highlights the importance of kinetics when considering development of new inhibitors targeting the pore of M2.

6. Pore bound water to histidine 37 sidechain in influenza A M2

6.1 Abstract

Water plays an important role in many biological processes including enzymatic reaction, folding and even conduction mechanisms. Proton conductance through the M2 water channel is a crucial step for release of viral RNA into the host cell during Influenza A infection. Despite the high resolution structure available on the proton channel M2 from influenza A, only two water populations have been observed in reconstituted lipid bilayers samples. NMR studies reported these two populations of water as associated water with a chemical shift of ~ 4.8 ppm and bulk water at ~ 4.6 ppm. According to high structural resolution data in combination with molecular dynamics simulations, proton conductance in M2 from influenza A is driven by histidine 37 (H37). Using controlled temperature proton detected ssNMR measurements at ultra fast MAS and dynamic nuclear polarization at cryogenic temperature, we report bound water near residue H37 in the M2 conductance domain, comprising residue 18 to 60, in a lipid environment. This appears at a chemical shift of 11 ppm at proton and 249 ppm at nitrogen corresponding to the H37_{Bδ1} side-chain. DNP experiments show that bound water is not an intermediate exchange state and has slow exchange with bulk water. Combining NMR data with DFT calculations, we show that the

bound water forms a hydrogen bond to the $N_{\delta 1}$ of H37_B. Additionally, the DFT calculation supports the recently reported imidazole-imidazole hydrogen bond, which is 1.9 Å in length. Furthermore, this work emphasizes the strength of proton detection under ultra fast MAS to address key features of channel membrane proteins. Atomic details on the pore water are indispensable to understand the proton conduction mechanism occurring throughout these channel proteins.

6.2 Results and discussion

By measuring CP based (H)NH spectra at low temperature, set temperature of 260 K corresponding to a sample temperature of ~5 to 10°C, ssNMR at 80 kHz MAS, we observed an HN peak with proton resonance at 11 ppm correlated to nitrogen 249 ppm. Previous work assigned the $N_{\delta 1}$ at 249 ppm to His_{37B} (342). A down-field chemical shift of 11 ppm ¹H is not commonly observed in proteins. Since no other correlation with the resonance at 11 ppm was found in (H)CH and (H)NH, We therefore strongly suspected a water molecule in close proximity to $N_{\delta 1B}$. Previous studies by Hong *et al.* and others have reported chemical shifts of water protons in membrane systems at ~4.6 ppm for bulk water and ~4.8 ppm or even 5.2 ppm for membrane-associated water (H₂O_a)(354, 355).

We therefore seek to identify the moiety to which the proton belongs using filtered cross-polarization (CP) experiments. Figure 5.6.1 shows the (H)NH spectra recorded either with an additional CP which acts as a filter before the acquisition of the signal referred as “back-filtered CP” ((H)NH_{dep}, Fig. 5.6.1A) for proton covalently bound to either ¹³C or ¹⁵N, or without the back-filtered CP ((H)NH_{ref}, Fig. 5.6.1B). The intensity of transferred magnetization through CP is directly related to the distance between the interacting atoms by $1/r^3$. The signal intensity for protons covalently bound to a ¹⁵N or ¹³C (a distance below 2Å) will undergo fast decay, strongly attenuating all backbone or sidechain protons, except hydroxyl groups. Water protons will be only weakly attenuated, since water protons are bound to oxygen. The intensity decay of the backbone protons (H^N and H^C) as well as the associated water (H₂O_a) at 4.8 ppm was tracked with the filtered CP experiment (Fig. 5.6.1A) and compared to the decay observed for the proton at ~11 ppm (H- -N_{37Bδ1}), assigned to pore bound water (H₂O_b). Indeed, a faster signal decay was observed for HN_{37Aε2}, HN_{37Bε2} and H^C

resonances compared with that measured for H_2O_b and the associated water (H_2O_a) resonances (Fig. 5.6.1E).

Similar protein sites, the $\text{HN}_{37A\epsilon 2}$ and $\text{HN}_{37B\epsilon 2}$ peaks showed different signal decay with about 85% and 93%, respectively. The difference on the decay observed between the two $\text{HN}_{\epsilon 2}$ resonances might be explained by the difference in the local environment of these two sites. Whereas, the faster decayed signal site, $\text{HN}_{37B\epsilon 2}$, forms a hydrogen bond with $\text{N}_{37A\delta 1}$ at the measured pH of 7.8 (342), the slowly decaying site, $\text{HN}_{37A\epsilon 2}$, is “free”. In $\text{HN}_{37B\epsilon 2}$ the proton is “shared” by both $\text{N}_{\epsilon 2}$ and $\text{N}_{\delta 1}$ within a 2-3 Å distance which could explain the extra signal loss. As expected for the H_2O_a , since there is no sink for magnetization loss, the full signal is retained (Fig. 5.6.1E). While, only about 61% of the signal is retained for $\text{H}-\text{N}_{37B\delta 1}$ (Fig. 5.6.1E). This attenuated signal decay indicates that the proton of $\text{H}-\text{N}_{37B\delta 1}$ at ~11 ppm could come from a water molecule in close proximity to the free $\text{N}_{\delta 1}$ of H_{37B} .

Indeed, pore water clusters have been previously observed in high-resolution X-ray diffraction data (300). Unlike the crystallographic data where only at low pH the water molecules are close enough to interact with the $\text{N}_{\delta 1}$ to form a hydrogen bond (300), our data suggest an extended and structured water network coming close to $\text{N}_{\delta 1}$ even at high pH.

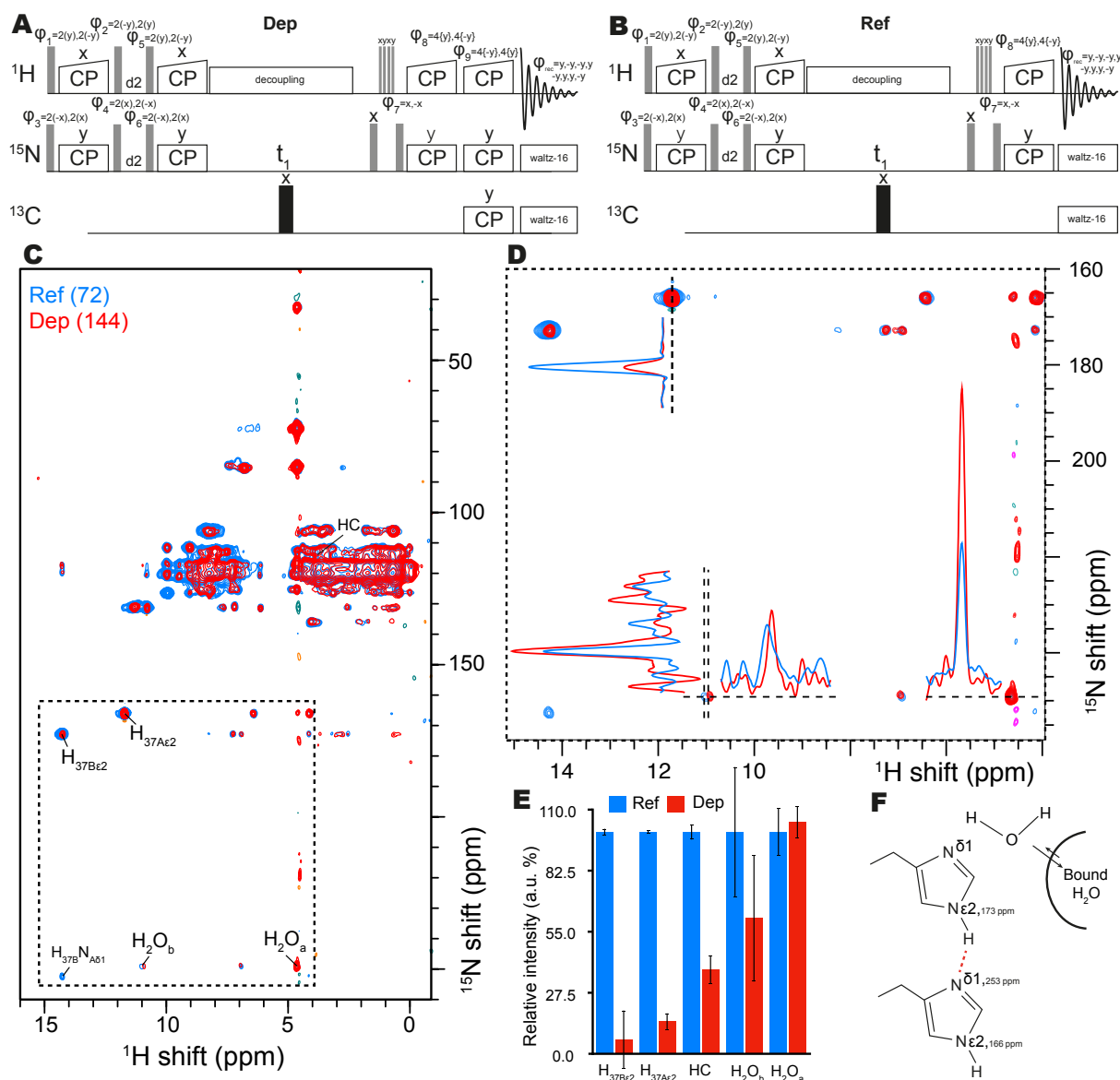


Figure 5.6.1. Identification of water close to N δ 1B of His37.

A and B show the pulse sequences used for identifying the chemical nature of the proton chemical shift observed using an inverse back CP for depolarization. In C, 2D (H)NH spectra obtained using sequence A (red) and B (blue) with 500 μs and 200 μs of depolarization CP contact time for ^{15}N and ^{13}C , respectively. In D, an expansion showing the histidine sidechain peaks is represented with 1D slices of the different peaks. In E, the intensities extracted from the 2D spectra are plotted, histidine HN sidechains (A and B), HC and HN peaks (identified as protons coming from water) with and without depolarization CP. In F, a schematic interpretation of the data obtained, with the bound water that have a proton chemical shift at ~ 4.8 and the more shifted water that shows a proton chemical shift at ~ 11 ppm. All the data were recorded on a 0.7 mm 4 channel Bruker probe with 100 kHz MAS on a 950 MHz Bruker spectrometer at $\sim 10^\circ\text{C}$ (260K). The depolarisation experiments were measure with 144 scans while the reference spectra used 44 scans.

In recent years direct detection of protons in membrane systems in a near-native physiological context has been possible by the ultra-fast MAS in ssNMR. To the best of our knowledge there is no previous report of water molecules presenting such downfield proton chemical shift in these conditions. Hans-Heinrich Limbach and co-workers report water proton

chemical shifts up to 19 ppm for the water bound to the catalytic centre in a synthetic system mimicking the catalytic center of carboanhydrase (356). In this case, the water molecule is doubly hydrogen bonded, on one side the oxygen with Zn and on the other with a nitrogen atom. This clearly shows the sensitivity of the water proton chemical shift regarding its local chemical environment. By analogy, this suggests that in M2 there is a pore water (H_2O_b) that forms a stable hydrogen bond to $N_{\delta 1}$ of the imidazole H37_B. The close proximity with the N_{δ} would lead to a strong deshielding effect on the proton which could explain the 11 ppm proton chemical shift observed.

Although we have observed a downfield shift of the proton, which indicates a hydrogen-bonded water molecule, this is in contrast to the crystallographic data at high pH (8) (300) and the long CP contact times required during the magnetization transfer from the ^{15}N to the H_2O_b . Different factors can account for the low transfer magnetization efficiency obtained in the ssNMR experiments such as the limited resident lifetime of the water molecules at this position, which will limit the NMR detection due to rapid exchange with other water moieties. Or a relatively long distance between the water molecule and the $N_{\delta 1}$, which will require a long spin-lock for an efficient magnetization transfer to occur. In any case, in both situations, low sensitivity is expected since the number of water molecules at this location is restricted by the C2 symmetry arrangement of the M2 tetramer which leaves only two $N_{\delta 1}$ s free to accept the proton from the H_2O_b molecule. In addition to the limited number of H_2O_b , there is a limited number of possibilities to observed water clusters by NMR. Water can only be detected by either protons (1H) or isotope oxygen 17 (^{17}O) which has low natural abundant of 0.037%.

In order to overcome these limitations, rapid exchange and lack of sensitivity imposed by the conventional room temperature solid-state NMR measurements, we took advantage of dynamic nuclear polarization (DNP). The new hardware and methodological development, in combination with the cryogenic temperature used in DNP quenching exchange processes, finally allows proton detection with sufficient resolution for certain studies (357, 358). So, overall, this makes DNP a suitable approach to study functional, low populated and in fast exchanging water clusters on proton channels such as M2. Here, we combine a relatively fast

DNP rotational speed of 24 kHz with a sample temperature of ~ 90 K which allows the freezing of dynamic sites revealing the ensemble of the water populations in the sample.

Earlier studies have reported that a substantial decrease of pH occurs in phosphate buffers upon liquid nitrogen freezing (359). This is a concern for M2 due to the pH dependence involving the charge state of H37. We, therefore, tested the buffer pH upon freezing using a color solution and we found that the pH shift ~ 7 pH (Appendix 5.III). Taking in consideration the previous reported pKa's (295), this data suggests that H37 is a mixture of neutral and +2 charge state in our DNP conditions. At this pH, H_2O_b studies are still relevant since the inter imidazole-imidazole H-bond is preserved (346).

The applied 24 kHz MAS DNP gives enough sensitivity and proton resolution to distinguish between the $HN_{\epsilon 2B}$, $HN_{\epsilon 2A}$ and H_2O_b protons (Appendix 5.IV). Figure 5.6.2A shows an expansion of the H37 sidechain region of the $(H)NH_{ref}$ (blue) and $(H)NH_{dep}$ (red) spectra. Whereas the room temperature spectrum needed ~ 2 days of data collection, the DNP spectrum needed only ~ 0.5 days, despite an under packed rotor (~ 1 mg of protein). The inset in figure 5.6.2A shows the 1D proton slices of the $(H)NH$ spectra. The pore bound water peak shows the exact same 1H chemical shift in both the DNP and the room temperature $(H)NH$ spectra. This suggests that the H_2O_b is not on an intermediate exchange regime with neither bulk water nor H_2O_a . If this were the case, an average chemical shift between the bulk water or the H_2O_a chemical shift will be observed for the H_2O_b . At the cryogenic temperatures used in DNP water exchange is hampered. We therefore hypothesize that the exchange between bound water and the bulk water is slow at room temperature. This suggests that the resident time of bound water is at least on the ms time scale or longer since it is observed by ssNMR experiments.

Similar to the room temperature measurements, at 90 K, the retained signal of H_2O_b is significantly higher than the signal from protons covalently bound to ^{15}N atoms for the filtered CP experiment. Only $\sim 25\%$ of the signal is retained for both $HN_{\epsilon 2A}$ and $HN_{\epsilon 2B}$ in contrast to the $\sim 50\%$ signal retained for the H_2O_b (Fig. 5.6.2B and appendix 5.IV). This data further supports the room temperature experiments where we assign the 11 ppm proton chemical shift to a “bound water” molecule located near the “free” H37 $N_{\delta 1B}$. Although the

fast MAS ssNMR and DNP data suggest that there is a water molecule in proximity to the free $N_{\delta 1}$ of H37, the data does not provide direct information regarding the pore bound water location.

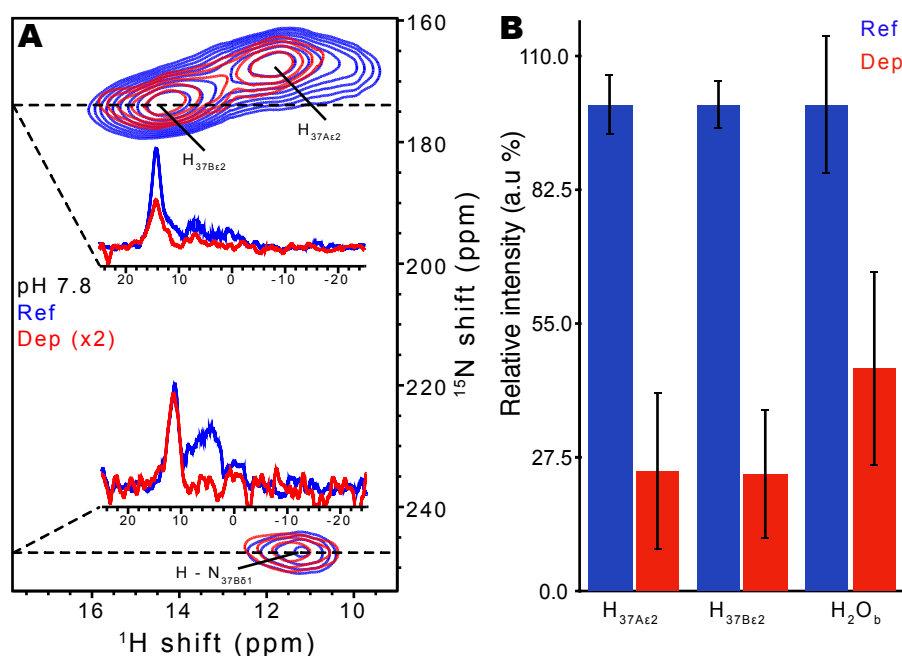


Figure 5.6.2. Water - imidazole cluster revealed under DNP conditions.

In A, overlay of the 90 Kelvin (H)NH water edited DNP spectra using a 2.5 mm four channel Phoenix probe at 24 kHz MAS with back CP off (blue) and 200 ms back CP on (red) of M2 H57Y at pH 7.8 (~7 if we take the value from the pH color indicator). Insets show the 1D proton slices of the sidechain $H_{37B\delta 1}$ and the water nitrogen cross polarization ($H - N_{37B\delta 1}$). In B intensity plots of H_{37Be2} , H_{37Ae2} and $H - N_{37B\delta 1}$ at pH 7.8 (~7). The signal to noise was used for statistical analysis which is here reported using two sigma deviation. The spectral noise and the intensities were extracted using CcpNmr.

To assess the pore water location, we performed DFT calculations to calculate the theoretical NMR isotropic chemical shift of a water molecule explicitly positioned near the $N_{37B\delta 1}$ sidechain. To accelerate the DFT and NMR calculations, only residues G34 to W41 were used. Two starting structures were used, either a dimer or the tetramer structure (Appendix 5.V). The geometry optimization was carried out stepwise from low to high accuracy of calculation theory (PM6 to BLYP3). A single water molecule (two for the tetramer) was added on an explicit position near to the $N_{37B\delta 1}$ after geometry optimization. The system was then re-optimized following the same step wise approach. As expected, the water molecule approaches the $N_{\delta 1}$ of the imidazole ring, however, the water molecule is found in two different positions (Fig. 5.6.3A and B). The water is either in the middle of the pore (Fig. 5.6.3A) with the second water proton pointing up towards the N-terminus of the

protein (pore facing water) or bridging between imidazole N_{37Bδ1} and the backbone NH of glycine 34 with both hydrogen moieties from the water molecule forming a hydrogen bond (bridging water) (Fig. 5.6.3B). In both cases, one of the hydrogen atoms from the water moiety is at a distance of ~1.6 Å from the N_{δ1} of H37_B, consistent with a hydrogen bond distance. In nearly perfect agreement with the measured shift, ~11 ppm, the calculated proton NMR chemical shift for the bridging water molecule is 10.6 ppm (Fig. 5.6.3D), which strongly supports the assignment. For the other water position the δH_{calc} is 8.5 ppm which deviates about 2.5 ppm from the measured value (~11 ppm). Taking a closer look, both positions showed a hydrogen bond length of ~1.6 Å with respect to the H-N_{δ1}, which could not explain the 2 ppm difference observed in the calculation. However, major electronic differences regarding the two molecular orientations are perceptible. In fact, this situation can be explained by the dipole moment and shielding effects on the proton by the neighboring atoms.

Free water proton gives an unexpected δH_{calc} of 0.9 ppm in contrast to the hydrogen bonded proton that gives a δH_{calc} of 4.6 ppm close to the expected chemical shift for bulk water. This difference can be explained by the DFT conditions used for the calculation, the pore facing molecule will be interacting with other water molecules filling the pore and here the calculations are performed in a vacuum. To address whether the water is either in the “bridging” or in the “pore facing” situation we decided to use solvated DFT calculations by including water in the polarizable continuum model (PCM) (Appendix 5.VI). Interestingly, the pore facing proton calculated chemical shift is still far, 1.5 ppm, from the expected close bulk water chemical shift of ~4.6 ppm. Nevertheless, the PCM model adds a dielectric potential over the whole structure, but does not account for other water molecules being a continuum hydrogen bonded network (water wire). With our current data, none of the water positions could be excluded. Despite this we confirm that a water molecule is hydrogen bonded to the free nitrogen of the histidine sidechain even at high pH values (~7). Both either “bridging water” or “pore facing water” have been supported in previous studies. Similar to our data, a bridging water molecule from the N_{37Bδ1} with the carboxyl group of G34 has been observed (301, 360). Although, contrary to our data, at high pH (8), the XFEL shows that the distance of the nearest water cluster to the H37N_{δ1} at high pH (pH 8) is inconsistent with the formation of a hydrogen bond (300).

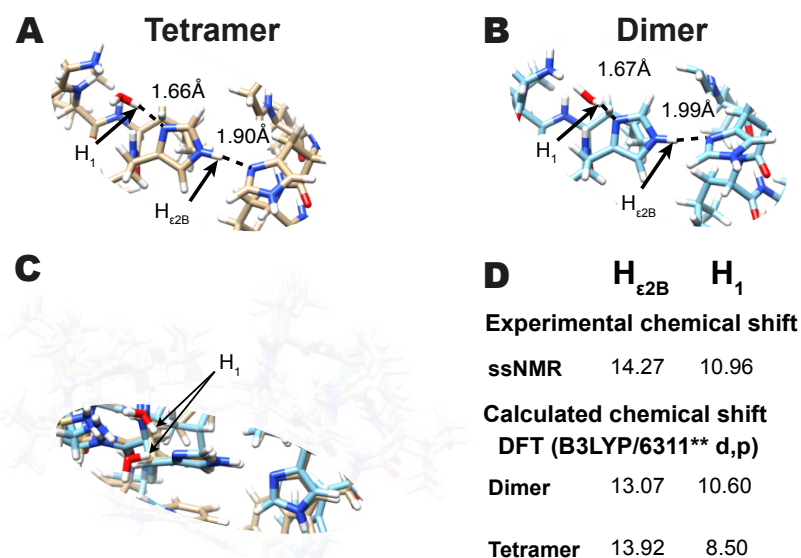


Figure 5.6.3. Chemical shifts of the structured water in the core of M2 calculated by density functional theory.

In A and B, the water location is shown after DFT geometry optimization for the full tetramer and a dimer, starting from an NMR structure. (Appendix Table S1 for the assignments and restraints used for CYANA calculation of the NMR structure). The unusual proton shifts of water, and the N-H—N involved proton are labeled, as H_1 and $H_{\epsilon 2}$, respectively. C shows an overlay of both tetramer (A) and dimer (B) to show the different location of the pore bound water. Panel D compares calculated and experimental chemical shifts.

In addition to the water location, the DFT NMR calculations further supports the recent published the $NH_{37B\epsilon 2}$ and $N_{37A\delta 1}$ hydrogen bond (342, 346). In fact, the calculated proton chemical shifts for $NH_{37B\epsilon 2}$ are 13.07 and 13.97 ppm for the dimer and the tetramer structure in the presence of the water molecule, respectively. This is consistent with the measured 1H chemical shift value of ~ 14.27 ppm.

6.3 Conclusion

To conclude, the combination of theoretical DFT calculations and NMR measurements has allowed localization of a pore bound water molecule in close proximity to the key residue histidine 37 of M2. In addition, the close agreement between DFT calculated and experimental chemical shifts further support the imidazole-imidazole hydrogen bond with a length of ~ 1.9 Å. Since water can only be detected in NMR with 1H or ^{17}O , it is difficult to identify water in complex biological molecules since the proton chemical shift cannot be easily dispersed with a heteronucleus; ^{17}O is the only possibility but it comes with tradeoffs due to the quadrupolar nature of the isotope. The combination of fast MAS and dipolar filtering overcomes these challenges, resulting in efficient detection of narrow proton resonances, and the ability to identify three distinct water chemical shifts in a membrane

protein sample. In addition, MAS rates above 20 kHz with DNP enables high sensitivity at cryogenic temperature where exchange processes are quenched. We expect that further advances in the spinning rate will result in much narrower proton resonances in future DNP installations. The wide chemical shift range of bound water results in an ideal application for DFT calculations, which nearly perfectly matched the experimental chemical shifts for a converged tetrameric structure of M2. We hope that this work, which sheds light on the water arrangement in the pore the M2 proton channel at a high physiological pH will contribute in the design of small compounds targeting M2 as therapeutics.

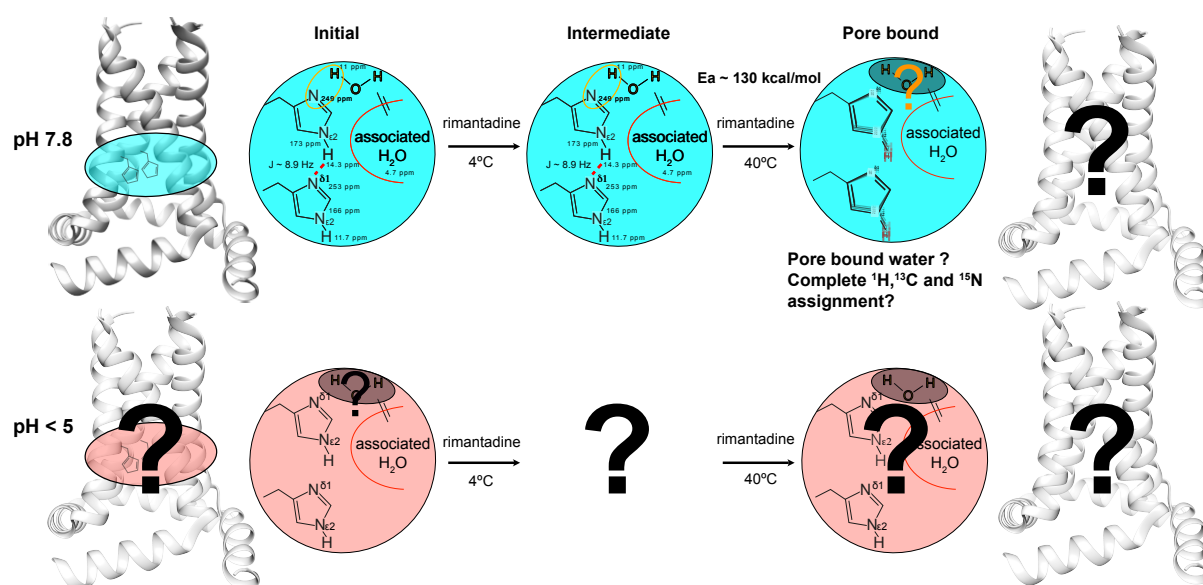
7. Conclusion

In this chapter, a de novo structure of the Influenza A M2 protein is reported (Fig. 5.4.3), including details of the inter imidazole-imidazole hydrogen bonded histidine 37 side-chains (Fig. 5.3.2), and a bound water molecule (Fig. 5.6.1). In addition, we determined the kinetic properties of drug binding using real time NMR (Fig. 5.5.5). The data are from a single fully protonated sample using proton detected ultra-fast MAS (100 kHz), complemented by DNP data and DFT calculations. The employment of faster MAS, which sufficiently averages the proton dipolar couplings in the sample and allows efficient proton decoupling using low radio frequency irradiations, has enabled us to measure a ${}^2J_{\text{NN}}$ coupling strength by INEPT transfer experiments in ssNMR (Fig. 5.3.2). The 8.9 Hz ${}^2J_{\text{NN}}$ coupling measured in the M2₁₈₋₆₀ construct is consistent with a dimer of dimers arrangement of the tetramer previously suggested by Timothy A. Cross and co-workers (312). The fast acquisition and the strict control of the sample temperature provides access to direct kinetic measurements using proton detection ssNMR. Kinetic measurements showed a cooperative binding mechanism upon rmt pore binding with a high energy barrier of ~ 117 kJ/mol (Fig. 5.5.5). The high activation barrier together with the insoluble pore bound M2 in DHPC micelles suggest a change in the physicochemical properties of the tetramer surface which seems to promote the tetramer - tetramer interactions (Fig. 5.5.6). This is consistent with the M2 clustering recently reported by Tim and co-workers (348). Low temperature measurements allow the identification of a water moiety with a proton chemical shift at ~ 11 ppm bound to the free $N_{\delta 1}$ of histidine 37B by using depolarization CP experiments. Finally, DFT calculations made it possible to locate the water molecule and to determine the length of the imidazole - imidazole hydrogen bond (~ 2 Å) (Fig. 5.6.3). Overall, the combination of these techniques improved our atomic understanding regarding the arrangement of key residue His37 of the M2 protein and its partners. We could show the location of the water in the vicinity of the His37, the imidazole-imidazole hydrogen bond and the nonspecific binding of rimantadine.

This data provides further structural information on the M2 protein that will help to understand drug binding mechanisms for future strains of M2 that will arise due to mutation and reassortment. Further studies on the conductance domain or even the full-length M2 at low pH are necessary to understand the histidine arrangement and drug binding mechanism in

the high conduction state of M2. Additionally, DFT as well as low temperature NMR measurement will be crucial to reveal the function of the H₂O_b in the conduction state. Is the bridging water assisting the histidine turn over? Does the bridging water get pushed out in order to allow the His37 to shuttle? How does the water cluster assist conformational changes upon drug binding? Many questions remain open to fully understand the drug binding mechanism, and answering them could help to further improve the development of new and efficient drugs.

Yet, a high resolution structure of the fully bound rmt-M2 complex has not been achieved by solid-state NMR in lipid bilayers. This achievement will definitely increase considerably our understanding regarding the changes on the physicochemical properties M2 upon binding and could perhaps relate the binding to a clustering process of the protein in the membranes.



Summary figure showing the solved and remaining questions of the proton channel M2 in lipid bilayers.

Top panel shows a summary of the work performed at pH 7.8 during this thesis. Using the conductance domain of M2 (residue 18 to 60), we detected and characterized an imidazole-imidazole hydrogen-bonded between H37N_{ε2}B and H37N_{δ1}A and h-bonded water at position H37N_{δ1}B. Additionally, we identify an intermediate state in the presence of rimantadine (rmt) for which the pore bonded water and the imidazole-imidazole hydrogen bond are retained. We tracked the rmt pore binding and determined that rmt pore binding has a high energy barrier (E_a ~117 kcal/mol). Upon pore binding, the imidazole-imidazole h-bond is broken and large chemical shifts are observed consistent with a large structural rearrangements. The structure of the pore drug bound M2 at high pH (7.8) is still to be solved. The bottom panel shows the low pH condition. So far the only available data with the conductance domain by ssNMR at low pH (< 5) is the absence of the imidazole-imidazole hydrogen bond. Many questions remain to be answered starting from the structure of the conductance domain construct at the activated state of the protein (low pH).

Appendix

Appendix 5.I. M2 assignments

Table S5.1. ¹H, ¹³C and ¹⁵N chemical shift assignment Helix A.

Amino acid sequence	Atom	Chain A	Chain B
25	N	*	*
25	CD	51.124	*
25	CA	65.497	*
25	HA	4.078	*
25	CB	32.819	*
25	HB2	2.176	*
25	HB3	1.690	*
25	QB	*	*
25	CG	27.608	*
25	HG2	1.952	*
25	HG3	2.010	*
25	QG	*	*
25	HD2	3.937	*
25	HD3	3.785	*
25	QD	*	*
25	C	177.164	*
26	N	118.649	*
26	H	7.710	*
26	CA	57.607	*
26	HA	3.907	*
26	CB	40.408	*
26	HB2	1.684	*
26	HB3	0,593	*
26	QB	*	*
26	CG	26.331	*
26	HG	1.091	*
26	QD1	-0,83	*
26	QD2	*	*
26	CD1	23.937	*
26	HD11	*	*
26	HD12	*	*

Amino acid sequence	Atom	Chain A	Chain B
26	HD13	*	*
26	CD2	*	*
26	HD21	*	*
26	HD22	*	*
26	HD23	*	*
26	QQD	*	*
26	C	178.259	*
27	N	120.324	118.821
27	H	8.157	7.266
27	CA	67.245	66.470
27	HA	3.198	3.318
27	CB	31.536	31.630
27	HB	2.243	2.099
27	QG1	0,887	0,927
27	QG2	0,761	*
27	CG1	22.840	23.020
27	HG11	*	*
27	HG12	*	*
27	HG13	*	*
27	CG2	21.623	21.523
27	HG21	*	*
27	HG22	*	*
27	HG23	*	*
27	QQG	*	*
27	C	178.737	*
28	N	119.423	120.700
28	H	8.223	7.880
28	CA	67.127	66.913
28	HA	3.345	3.408
28	CB	31.495	31.578
28	HB	2.215	1.917
28	QG1	0,874	0,875
28	QG2	0,437	0,755
28	CG1	23.139	22.335
28	HG11	*	*
28	HG12	*	*

Amino acid sequence	Atom	Chain A	Chain B
28	HG13	*	*
28	CG2	21.636	*
28	HG21	*	*
28	HG22	*	*
28	HG23	*	*
28	QQG	*	*
28	C	177.835	177.982
29	N	121.120	120.082
29	H	8.400	8.250
29	CA	55.910	55.776
29	HA	3.713	3.719
29	QB	1.415	1.303
29	CB	18.531	18.501
29	HB1	*	*
29	HB2	*	*
29	HB3	*	*
29	C	178.411	178.325
30	N	117.916	117.391
30	H	8.601	8.057
30	CA	55.037	55.485
30	HA	3.649	3.683
30	QB	1.254	1.298
30	CB	18.581	18.386
30	HB1	*	*
30	HB2	*	*
30	HB3	*	*
30	C	178.324	178.393
31	N	112.691	112.161
31	H	8.007	7.956
31	CA	63.024	62.378
31	HA	3.787	3.963
31	CB	63.433	62.063
31	HB2	3.788	3.923
31	HB3	4.146	3.683
31	QB	*	*
31	OG	*	*

Amino acid sequence	Atom	Chain A	Chain B
31	HG	*	*
31	C	175.346	175.502
32	N	120.139	119.837
32	H	7.910	7.983
32	CA	65.581	64.490
32	HA	3.461	3.487
32	CB	37.746	37.203
32	HB	1.777	1.872
32	QG2	0,631	0,652
32	CG2	17.387	17.579
32	HG21	*	*
32	HG22	*	*
32	HG23	*	*
32	CG1	29.635	29.474
32	HG12	1.723	1.766
32	HG13	0,829	0,627
32	QG1	*	*
32	QD1	0,584	0,621
32	CD1	13.765	13.763
32	HD11	*	*
32	HD12	*	*
32	HD13	*	*
32	C	177.137	177.083
33	N	115.096	118.987
33	H	8.093	8.061
33	CA	66.055	65.422
33	HA	3.630	3.400
33	CB	37.319	37.555
33	HB	1.738	1.766
33	QG2	0,645	0,626
33	CG2	17.904	17.353
33	HG21	*	*
33	HG22	*	*
33	HG23	*	*
33	CG1	29.753	29.507
33	HG12	1.734	1.618

Amino acid sequence	Atom	Chain A	Chain B
33	HG13	0,741	0,808
33	QG1	*	*
33	QD1	0,546	0,634
33	CD1	13.572	13.455
33	HD11	*	*
33	HD12	*	*
33	HD13	*	*
33	C	177.205	176.911
34	N	106.614	105.172
34	H	8.193	8.310
34	CA	47.830	48.308
34	HA2	4.307	4.021
34	HA3	3.551	3.376
34	QA	*	*
34	C	175.235	174.928
35	N	120.262	119.955
35	H	8.314	7.977
35	CA	63.882	65.245
35	HA	3.737	3.546
35	CB	37.420	37.553
35	HB	1.778	1.789
35	QG2	0,627	0,746
35	CG2	17.488	18.036
35	HG21	*	*
35	HG22	*	*
35	HG23	*	*
35	CG1	29.966	29.719
35	HG12	0,641	0,877
35	HG13	1.799	1.655
35	QG1	*	*
35	QD1	0,576	0,568
35	CD1	13.557	13.542
35	HD11	*	*
35	HD12	*	*
35	HD13	*	*
35	C	176.846	176.684

Amino acid sequence	Atom	Chain A	Chain B
36	N	119.375	119.797
36	H	8.207	8.545
36	CA	58.230	58.454
36	HA	3.806	3.792
36	CB	41.701	41.832
36	HB2	1.742	1.744
36	HB3	1.359	1.361
36	QB	*	*
36	CG	30.191	30.247
36	HG	1.716	1.525
36	QD1	*	1.742
36	QD2	*	0,633
36	CD1	26.803	26.856
36	HD11	*	*
36	HD12	*	*
36	HD13	*	*
36	CD2	*	23.110
36	HD21	*	*
36	HD22	*	*
36	HD23	*	*
36	QQD	*	*
36	C	177.564	177.663
37	N	117.285	116.491
37	H	7.998	8.361
37	CA	62.343	59.558
37	HA	3.624	4.358
37	CB	32.203	31.203
37	HB2	2.964	2.923
37	HB3	2.682	*
37	QB	*	*
37	CG	*	*
37	ND1	252.639	*
37	CD2	116.227	118.742
37	HD1	*	*
37	CE1	136.282	137.228
37	NE2	166.110	172.784

Amino acid sequence	Atom	Chain A	Chain B
37	HD2	6.434	7.260
37	HE1	4.130	6.910
37	C	176.748	175.214
38	N	119.527	117.268
38	H	8.078	7.711
38	CA	58.171	59.024
38	HA	3.378	3.527
38	CB	40.642	42.469
38	HB2	1.557	2.119
38	HB3	0,565	1.396
38	QB	*	*
38	CG	26.507	26.591
38	HG	*	1.385
38	QD1	1.179	0,645
38	QD2	*	*
38	CD1	21.769	22.950
38	HD11	*	*
38	HD12	*	*
38	HD13	*	*
38	CD2	*	*
38	HD21	*	*
38	HD22	*	*
38	HD23	*	*
38	QQD	*	*
38	C	178.370	177.485
39	N	117.393	114.446
39	H	8.261	8.026
39	CA	65.795	65.472
39	HA	3.311	3.391
39	CB	37.442	37.571
39	HB	1.820	1.831
39	QG2	0,644	0,653
39	CG2	17.327	17.384
39	HG21	*	*
39	HG22	*	*
39	HG23	*	*

Amino acid sequence	Atom	Chain A	Chain B
39	CG1	29.606	29.659
39	HG12	1.800	1.762
39	HG13	0,920	0,835
39	QG1	*	*
39	QD1	0,564	0,575
39	CD1	13.610	13.698
39	HD11	*	*
39	HD12	*	*
39	HD13	*	*
39	C	177.077	177.056
40	N	118.846	121.280
40	H	7.729	8.823
40	CA	58.214	58.491
40	HA	3.839	3.845
40	CB	40.725	41.359
40	HB2	1.693	2.164
40	HB3	0,637	*
40	QB	*	*
40	CG	30.110	26.836
40	HG	1.686	*
40	QD1	1.639	1.665
40	QD2	*	0,628
40	CD1	26.699	25.618
40	HD11	*	*
40	HD12	*	*
40	HD13	*	*
40	CD2	26.800	22.959
40	HD21	*	*
40	HD22	*	*
40	HD23	*	*
40	QQD	*	*
40	C	178.352	178.703
41	N	123.000	120.805
41	H	8.508	9.460
41	CA	62.783	61.473
41	HA	3.750	3.654

Amino acid sequence	Atom	Chain A	Chain B
41	CB	28.130	27.473
41	HB2	3.341	2.503
41	HB3	2.546	*
41	QB	*	*
41	CG	*	*
41	CD1	127.718	126.563
41	CD2	*	*
41	CE3	120.720	119.417
41	CE2	*	*
41	NE1	130.961	131.228
41	HD1	7.210	6.145
41	HE3	7.392	*
41	CZ3	120.717	121.646
41	CZ2	114.559	114.847
41	HE1	11.282	10.779
41	HZ3	6.673	6.075
41	CH2	123.168	123.900
41	HZ2	7.812	7.614
41	HH2	6.923	7.182
41	C	178.527	178.143
42	N	119.515	117.149
42	H	8.812	7.835
42	CA	65.605	66.579
42	HA	3.040	3.230
42	CB	37.620	37.638
42	HB	1.811	1.873
42	QG2	0,599	0,670
42	CG2	17.048	17.506
42	HG21	*	*
42	HG22	*	*
42	HG23	*	*
42	CG1	29.388	29.610
42	HG12	0,820	1.831
42	HG13	0,448	0,632
42	QG1	*	*
42	QD1	0,594	0,630

Amino acid sequence	Atom	Chain A	Chain B
42	CD1	13.634	13.804
42	HD11	*	*
42	HD12	*	*
42	HD13	*	*
42	C	177.924	177.743
43	N	119.131	117.246
43	H	8.476	8.441
43	CA	58.150	58.071
43	HA	*	3.835
43	CB	*	41.717
43	HB2	*	2.042
43	HB3	*	1.382
43	QB	*	*
43	CG	*	26.654
43	HG	*	1.671
43	QD1	*	*
43	QD2	*	*
43	CD1	*	22.938
43	HD11	*	*
43	HD12	*	*
43	HD13	*	*
43	CD2	*	*
43	HD21	*	*
43	HD22	*	*
43	HD23	*	*
43	QQD	*	*
43	C	*	179.626
44	N	*	120.827
44	H	*	9.066
44	CA	*	57.821
44	HA	*	4.761
44	CB	*	42.688
44	HB2	*	2.571
44	HB3	*	2.878
44	QB	*	*
44	CG	*	*

Amino acid sequence	Atom	Chain A	Chain B
44	OD1	*	*
44	OD2	*	*
44	C	*	178.417
45	N	*	116.307
45	H	*	8.530
45	CA	*	57.082
45	HA	*	3.515
45	CB	*	30.472
45	HB2	*	1.656
45	HB3	*	1.535
45	QB	*	*
45	CG	*	25.154
45	HG2	*	1.018
45	HG3	*	0,771
45	QG	*	*
45	CD	*	43.206
45	HD2	*	*
45	HD3	*	*
45	QD	*	*
45	NE	*	*
45	HE	*	*
45	CZ	*	*
45	NH1	*	*
45	HH11	*	*
45	HH12	*	*
45	NH2	*	*
45	HH21	*	*
45	HH22	*	*
45	QH2	*	*
45	C	*	178.289
46	N		113.321
46	H		7.795
46	CA		56.308
46	HA		3.796
46	CB		42.254
46	HB2		0,994

Amino acid sequence	Atom	Chain A	Chain B
46	HB3		1.437
46	QB		*
46	CG		26.683
46	HG		*
46	QD1		*
46	QD2		*
46	CD1		*
46	HD1		*
46	HD1		*
46	HD1		*
46	CD2		*
46	HD2		*
46	HD2		*
46	HD2		*
46	QQD		*
46	C		177.261
47	N		112.622
47	H		7.539
47	CA		58.771
47	HA		4.616
47	CB		41.510
47	HB2		2.496
47	HB3		*
47	QB		*
47	QD		*
47	QE		*
47	QR		*
47	CG		*
47	CD1		*
47	HD1		*
47	CE1		*
47	HE1		*
47	CZ		*
47	HZ		*
47	CE2		*
47	HE2		*

Amino acid sequence	Atom	Chain A	Chain B
47	CD2		*
47	HD2		*
47	C		175.323
48	N		116.440
48	H		7.374
48	CA		61.539
48	HA		4.216
48	CB		39.503
48	HB2		3.521
48	HB3		3.131
48	QB		*
48	QD		*
48	QE		*
48	QR		*
48	CG		*
48	CD1		*
48	HD1		*
48	CE1		*
48	HE1		*
48	CZ		*
48	HZ		*
48	CE2		*
48	HE2		*
48	CD2		*
48	HD2		*
48	C		175.579
49	N		120.104
49	H		9.956
49	CA		57.147
49	HA		3.323
49	CB		30.366
49	HB2		1.675
49	HB3		1.025
49	QB		*
49	CG		25.281
49	HG2		0,994

Amino acid sequence	Atom	Chain A	Chain B
49	HG3		*
49	QG		*
49	CD		29.297
49	HD2		*
49	HD3		*
49	QD		*
49	CE		*
49	HE2		*
49	HE3		*
49	QE		*
49	NZ		*
49	HZ1		*
49	HZ2		*
49	HZ3		*
49	QZ		*
49	C		175.633
50	N		111.291
50	H		9.020
50	CA		60.519
50	HA		4.428
50	CB		66.293
50	HB2		3.512
50	HB3		3.632
50	QB		*
50	OG		*
50	HG		*
50	C		174.659
51	N		126.449
51	H		9.927
51	CA		66.079
51	HA		3.164
51	CB		38.119
51	HB		1.164
51	QG2		-0,014
51	CG2		17.184
51	HG2		*

Amino acid sequence	Atom	Chain A	Chain B
51	HG2		*
51	HG2		*
51	CG1		28.621
51	HG1		1.470
51	HG1		0,413
51	QG1		*
51	QD1		0,468
51	CD1		13.806
51	HD1		*
51	HD1		*
51	HD1		*
51	C		176.078
52	N		117.337
52	H		7.911
52	CA		62.800
52	HA		3.854
52	CB		36.611
52	HB2		2.575
52	HB3		1.885
52	QB		*
52	QD		*
52	QE		*
52	QR		*
52	CG		*
52	CD1		*
52	HD1		*
52	CE1		*
52	HE1		*
52	CZ		*
52	CE2		*
52	HE2		*
52	CD2		*
52	HD2		*
52	OH		*
52	HH		*
52	C		178.934

Amino acid sequence	Atom	Chain A	Chain B
53	N		121.953
53	H		8.545
53	CA		58.909
53	HA		3.994
53	CB		29.381
53	HB2		1.335
53	HB3		*
53	QB		*
53	CG		25.241
53	HG2		0,990
53	HG3		*
53	QG		*
53	CD		42.646
53	HD2		2.713
53	HD3		*
53	QD		*
53	NE		*
53	HE		*
53	CZ		*
53	NH1		*
53	HH1		*
53	HH1		*
53	NH2		*
53	HH2		*
53	HH2		*
53	QH2		*
53	C		177.941
54	N		120.659
54	H		7.314
54	CA		61.023
54	HA		*
54	CB		*
54	HB2		*
54	HB3		*
54	QB		*
54	QD		*

Amino acid sequence	Atom	Chain A	Chain B
54	QE		*
54	QR		*
54	CG		*
54	CD1		*
54	HD1		*
54	CE1		*
54	HE1		*
54	CZ		*
54	HZ		*
54	CE2		*
54	HE2		*
54	CD2		*
54	HD2		*
54	C		178.051

* DENOTES NO ASSIGNMENT

Appendix 5.II. Structure calculation

Table S5.2. Upper and lower distance used in Cyana for Helix A

Residue number	Amino acid	Atom	Residue number	Amino acid	Atom	Distance (lol)	Distance (upl)
27	VAL	O	31	SER	H	1.80	2.00
27	VAL	O	31	SER	N	2.70	3.00
28	VAL	O	32	ILE	H	1.80	2.00
28	VAL	O	32	ILE	N	2.70	3.00
29	ALA	O	33	ILE	H	1.80	2.00
29	ALA	O	33	ILE	N	2.70	3.00
30	ALA	O	34	GLY	H	1.80	2.00
30	ALA	O	34	GLY	N	2.70	3.00
31	SER	O	35	ILE	H	1.80	2.00
31	SER	O	35	ILE	N	2.70	3.00
32	ILE	O	36	LEU	H	1.80	2.00
32	ILE	O	36	LEU	N	2.70	3.00
33	ILE	O	37	HIS	H	1.80	2.00
33	ILE	O	37	HIS	N	2.70	3.00
34	GLY	O	38	LEU	H	1.80	2.00
34	GLY	O	38	LEU	N	2.70	3.00
35	ILE	O	39	ILE	H	1.80	2.00
35	ILE	O	39	ILE	N	2.70	3.00
36	LEU	O	40	LEU	H	1.80	2.00
36	LEU	O	40	LEU	N	2.70	3.00
37	HIS	O	41	TRP	H	1.80	2.00
37	HIS	O	41	TRP	N	2.70	3.00
38	LEU	O	42	ILE	H	1.80	2.00
38	LEU	O	42	ILE	N	2.70	3.00
39	ILE	O	43	LEU	H	1.80	2.00
39	ILE	O	43	LEU	N	2.70	3.00
40	LEU	O	44	ASP	H	1.80	2.00
40	LEU	O	44	ASP	N	2.70	3.00

Table S5.3. Upper and lower distance used in Cyana for Helix B

Residue number	Amino acid	Atom	Residue number	Amino acid	Atom	Distance (lo)	Distance (upl)
127	VAL	O	131	SER	H	1.80	2.00
127	VAL	O	131	SER	N	2.70	3.00
128	VAL	O	132	ILE	H	1.80	2.00
128	VAL	O	132	ILE	N	2.70	3.00
129	ALA	O	133	ILE	H	1.80	2.00
129	ALA	O	133	ILE	N	2.70	3.00
130	ALA	O	134	GLY	H	1.80	2.00
130	ALA	O	134	GLY	N	2.70	3.00
131	SER	O	135	ILE	H	1.80	2.00
131	SER	O	135	ILE	N	2.70	3.00
132	ILE	O	136	LEU	H	1.80	2.00
132	ILE	O	136	LEU	N	2.70	3.00
133	ILE	O	137	HIST	H	1.80	2.00
133	ILE	O	137	HIST	N	2.70	3.00
134	GLY	O	138	LEU	H	1.80	2.00
134	GLY	O	138	LEU	N	2.70	3.00
135	ILE	O	139	ILE	H	1.80	2.00
135	ILE	O	139	ILE	N	2.70	3.00
136	LEU	O	140	LEU	H	1.80	2.00
136	LEU	O	140	LEU	N	2.70	3.00
137	HIST	O	141	TRP	H	1.80	2.00
137	HIST	O	141	TRP	N	2.70	3.00
138	LEU	O	142	ILE	H	1.80	2.00
138	LEU	O	142	ILE	N	2.70	3.00
139	ILE	O	143	LEU	H	1.80	2.00
139	ILE	O	143	LEU	N	2.70	3.00
140	LEU	O	144	ASP	H	1.80	2.00
140	LEU	O	144	ASP	N	2.70	3.00
141	TRP	O	145	ARG	H	1.80	2.00
141	TRP	O	145	ARG	N	2.70	3.00
142	ILE	O	146	LEU	H	1.80	2.00
142	ILE	O	146	LEU	N	2.70	3.00
143	LEU	O	147	PHE	H	1.80	2.00
143	LEU	O	147	PHE	N	2.70	3.00
148	PHE	O	152	TYR	H	1.80	2.00

Residue number	Amino acid	Atom	Residue number	Amino acid	Atom	Distance (lol)	Distance (upl)
148	PHE	O	152	TYR	N	2.70	3.00
149	LYS	O	153	ARG	H	1.80	2.00
149	LYS	O	153	ARG	N	2.70	3.00
150	SER	O	154	PHE	H	1.80	2.00
150	SER	O	154	PHE	N	2.70	3.00
151	ILE	O	155	PHE	H	1.80	2.00
151	ILE	O	155	PHE	N	2.70	3.00
152	TYR	O	156	GLU	H	1.80	2.00
152	TYR	O	156	GLU	N	2.70	3.00

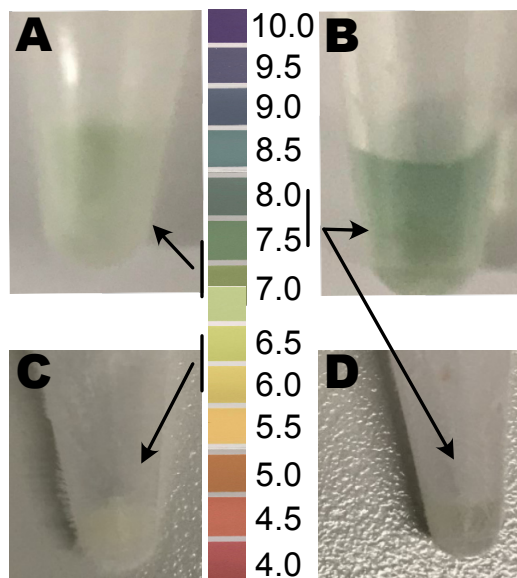
Helix B has been define as starting at residue 127 instead of 27 in Cyana.

Table S5.4. Entry of the imidazole-imidazole hydrogen bond for Cyana.

Residue number	Amino acid	Atom	Residue number	Amino acid	Atom	Distance (lol)	Distance (upl)
137	HIST	HE2	37	HIST	ND1	1.50	2.00
137	HIST	NE2	37	HIST	ND1	2.80	3.00
337	HIST	HE2	237	HIST	ND1	1.50	2.00
337	HIST	NE2	237	HIST	ND1	2.80	3.00

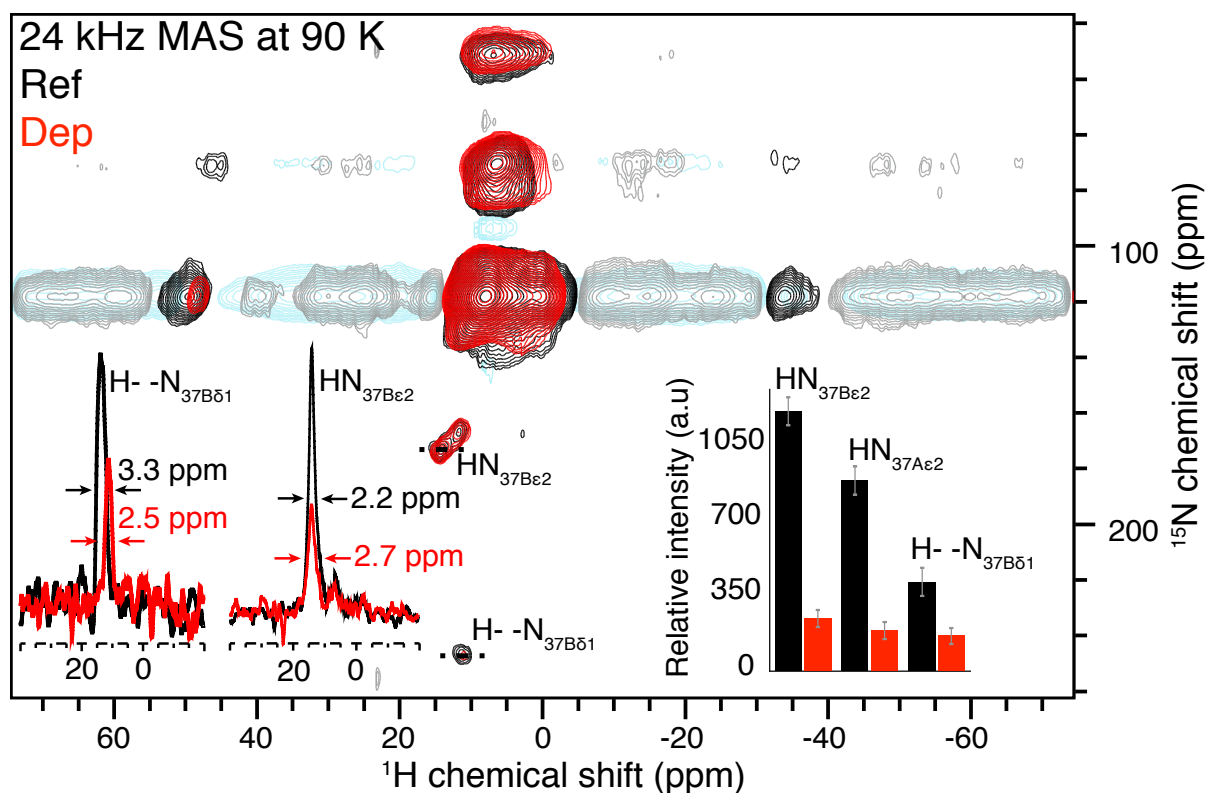
Table S5.5. Manual restraints for helping tetramer formation.

Residue number	Amino acid	Atom	Residue number	Amino acid	Atom	Distance (lol)
27	VAL	CA	127	VAL	CA	10.00
27	VAL	CA	327	VAL	CA	10.00
27	VAL	CA	227	VAL	CA	14.00
327	VAL	CA	127	VAL	CA	14.00
45	ARG	CB	245	ARG	CB	21.00
145	ARG	CB	345	ARG	CB	21.00
37 H	IST	HE2	341	TRP	NE1	5.00
237	HIST	HE2	141	TRP	NE1	5.00



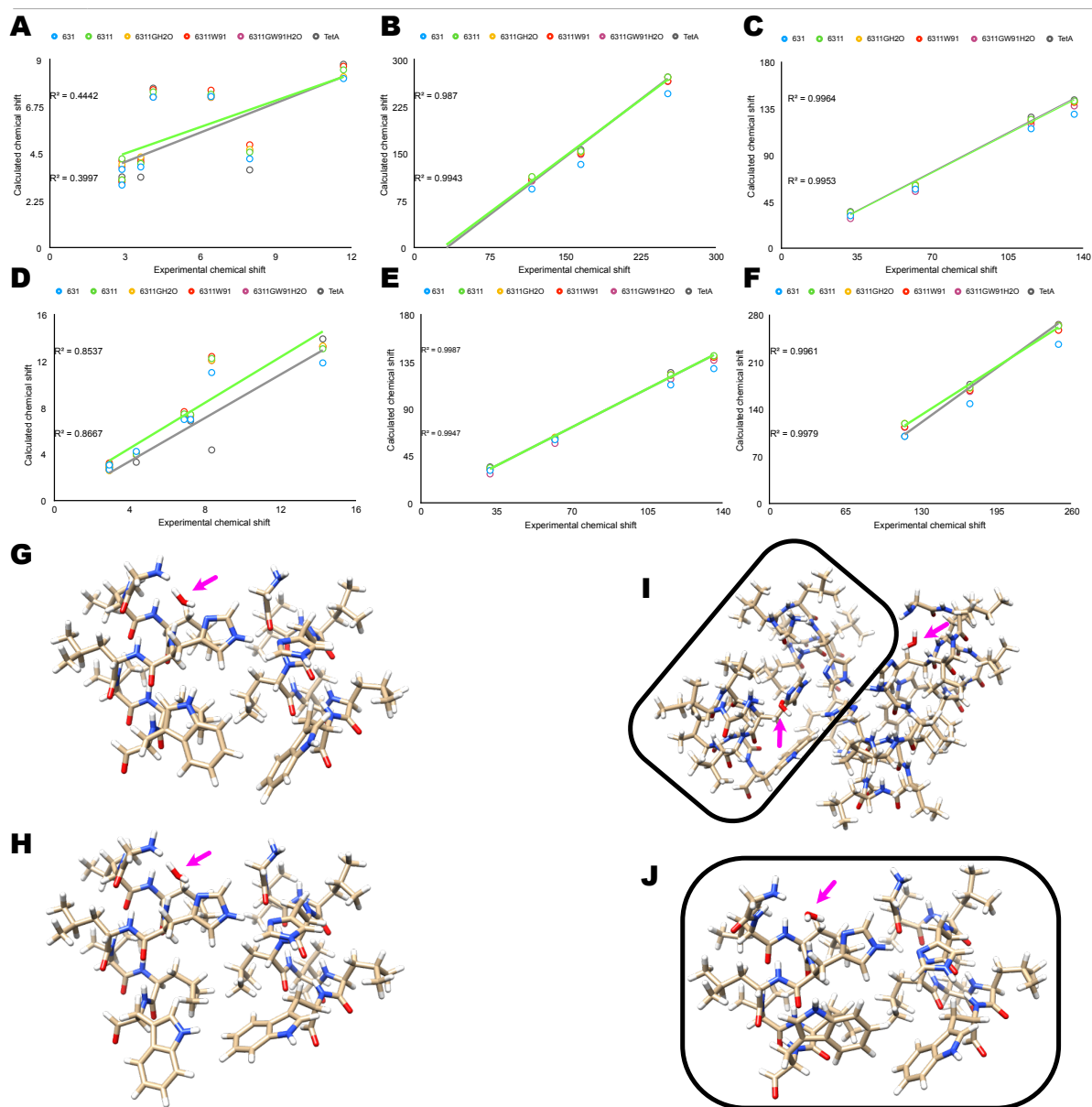
Appendix 5.IV. pH at the DNP conditions.

Buffer pH control upon liquid nitrogen freezing for DNP. The pH shifts for DNP sample containing ~1 mg of reconstituted protein resuspended in 24 μ l of DNP juice are shown for frozen sample (A) and solution (B). C and D show the pH of the DNP solution with 6 μ l of buffer solution instead of protein containing solution of the frozen and liquid samples, respectively.



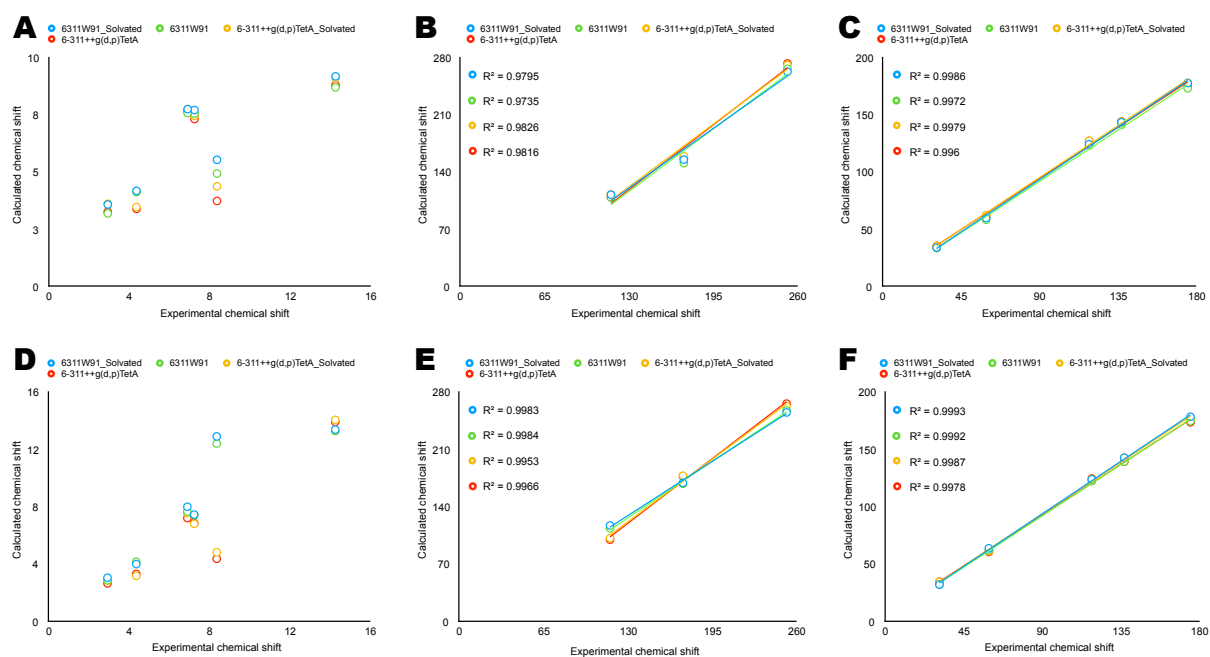
Appendix 5.III. 2D (H)NH DNP at 90 K using 24 kHz MAS.

The spectra were recorded using a 600 MHz DNP Bruker spectrometer with a gyrotron of 369 GHz. The temperature was set at 90 K and sample spun at 24 kHz MAS. The data was measured with a four channel phoenix probe with a 2.5 mm rotor.



Appendix 5.V. DFT structures after geometry optimization.

Correlation between experimental and the calculated NMR chemical shift. A to C shows the correlation of the histidine 37 chemical shifts of chain A for protons (A), nitrogens (B) and carbons (C). Histidine 37 of chain B is shown in D, E and F for protons, nitrogens and carbons, respectively. Only two linear fitting have been plotted corresponding to the same level of theory for the different structures used for the NMR calculations (J coming from the tetramer and G the dimer). Green, yellow and grey are the NMR shifts calculated with B3LYP functional and 6311G**2d,p as basis set for two slightly different pathways during geometry optimization using the dimer structure leading to G and H structure call int1 and H₂O, respectively. The grey named TetA is the NMR calculation from the dimer structure from the optimization of the tetramer structure (I). For NMR calculation only half of the tetramer was selected (black box I). J, shows the dimer from the tetramer used for the NMR calculation. Only this dimer structure (J) was used because the water position was different from the other two structures using the dimer structures (G and H) for the whole geometry optimization process. The pink arrows in G, H, I and J indicate the water position found after geometry optimization. The geometry optimization and NMR calculation were also carried out using a different functional mPW1PW91 with 6-311G**2d,p as basis set, here noted as W91 (red and purple).



Appendix 5.VI. NMR calculations using solvation (water with PCM).

Correlation between experimental and the calculated NMR chemical shift from the dimer and the TetA. A to C shows the correlation of the histidine 37 chemical shifts of chain A for protons (A), nitrogens (B) and carbons (C). Histidine 37 of chain B is shown in D, E and F for protons, nitrogens and carbons, respectively.

Bibliography

1. B. Alberts *et al.*, Molecular biology of the cell. (2018).
2. P. C. Champe, R. A. Harvey, D. R. Ferrier, *Biochemistry* (Lippincott Williams & Wilkins, 2005).
3. M. H. Levitt, Spin dynamics: basics of nuclear magnetic resonance (John Wiley & Sons, 2013).
4. M. J. Duer *et al.*, The Basics of Solid-State NMR Essential Techniques for Spin-½ Nuclei Dipolar Coupling: Its Measurement and Uses Quadrupole Coupling: Its Measurement and Uses Shielding and Chemical shift NMR Techniques for Studying Molecular Motion in Solids Molecular Structure Determination: Applications in Biology. *Solid-State NMR Spectroscopy Principles and Applications* <https://doi.org/10.1002/9780470999394.ch1> <https://doi.org/10.1002/9780470999394.ch2> <https://doi.org/10.1002/9780470999394.ch3> <https://doi.org/10.1002/9780470999394.ch4> <https://doi.org/10.1002/9780470999394.ch5> <https://doi.org/10.1002/9780470999394.ch6> <https://doi.org/10.1002/9780470999394.ch7>, 1-72 (2001).
5. J. Keeler, Understanding NMR spectroscopy (John Wiley & Sons, 2011).
6. S. K. Gupta, P. Shukla, Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications. *Crit Rev Biotechnol* **36**, 1089-1098 (2016).
7. P. Buckel, Recombinant proteins for therapy. *Trends Pharmacol Sci* **17**, 450-456 (1996).
8. C. Rabert, D. Weinacker, A. Pessoa, Jr., J. G. Farias, Recombinants proteins for industrial uses: utilization of *Pichia pastoris* expression system. *Braz J Microbiol* **44**, 351-356 (2013).
9. L. Pinzi, G. Rastelli, Molecular Docking: Shifting Paradigms in Drug Discovery. *Int J Mol Sci* **20** (2019).
10. H. Saibil, Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol* **14**, 630-642 (2013).
11. F. Chiti, C. M. Dobson, Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366 (2006).
12. M. Sastry, C. A. Bewley, P. D. Kwong, Effective isotope labeling of proteins in a mammalian expression system. *Methods Enzymol* **565**, 289-307 (2015).
13. C. Structural Genomics *et al.*, Protein production and purification. *Nat Methods* **5**, 135-146 (2008).
14. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Breaking the next Cryo-EM resolution barrier – Atomic resolution determination of proteins! *bioRxiv* 10.1101/2020.05.21.106740, 2020.2005.2021.106740 (2020).
15. D. Lyumkis, Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem* **294**, 5181-5197 (2019).
16. A. Rivera-Calzada, M. Carroni, Editorial: Technical Advances in Cryo-Electron Microscopy. *Front Mol Biosci* **6**, 72 (2019).
17. S. Y. Liao, M. Lee, T. Wang, I. V. Sergeyev, M. Hong, Efficient DNP NMR of membrane proteins: sample preparation protocols, sensitivity, and radical location. *J Biomol NMR* **64**, 223-237 (2016).
18. I. Pritisanac, J. M. Wurz, T. R. Alderson, P. Guntert, Automatic structure-based NMR methyl resonance assignment in large proteins. *Nat Commun* **10**, 4922 (2019).
19. A. M. Seddon, P. Curnow, P. J. Booth, Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* **1666**, 105-117 (2004).

20. J. Birch *et al.*, The fine art of integral membrane protein crystallisation. *Methods* **147**, 150-162 (2018).
21. F. Hagn, M. L. Nasr, G. Wagner, Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR. *Nature Protocols* **13**, 79-98 (2018).
22. S. D. Cady *et al.*, Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* **463**, 689-692 (2010).
23. J. R. Schnell, J. J. Chou, Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **451**, 591-595 (2008).
24. L. A. Baker, M. Baldus, Characterization of membrane protein function by solid-state NMR spectroscopy. *Curr Opin Struct Biol* **27**, 48-55 (2014).
25. B. J. Wylie, M. P. Bhate, A. E. McDermott, Transmembrane allosteric coupling of the gates in a potassium channel. *Proc Natl Acad Sci U S A* **111**, 185-190 (2014).
26. T. Schubeis *et al.*, A beta-barrel for oil transport through lipid membranes: Dynamic NMR structures of AlkL. *Proc Natl Acad Sci U S A* **117**, 21014-21021 (2020).
27. S. Penzel *et al.*, Spinning faster: protein NMR at MAS frequencies up to 126 kHz. *J Biomol NMR* **73**, 19-29 (2019).
28. R. Verardi, N. J. Traaseth, L. R. Masterson, V. V. Vostrikov, G. Veglia, Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Adv Exp Med Biol* **992**, 35-62 (2012).
29. D. M. LeMaster, Deuterium labelling in NMR structural analysis of larger proteins. *Q Rev Biophys* **23**, 133-174 (1990).
30. D. H. Zhou *et al.*, Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy. *Angew Chem Int Ed Engl* **46**, 8380-8383 (2007).
31. U. Akbey *et al.*, Optimum levels of exchangeable protons in perdeuterated proteins for proton detection in MAS solid-state NMR spectroscopy. *J Biomol NMR* **46**, 67-73 (2010).
32. J. R. Lewandowski *et al.*, Enhanced Resolution and Coherence Lifetimes in the Solid-State NMR Spectroscopy of Perdeuterated Proteins under Ultrafast Magic-Angle Spinning. *The Journal of Physical Chemistry Letters* **2**, 2205-2211 (2011).
33. V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, L. E. Kay, Cross-correlated relaxation enhanced ^1H - ^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **125**, 10420-10428 (2003).
34. P. Gans *et al.*, Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl* **49**, 1958-1962 (2010).
35. M. Kainosho, T. Tsuji, Assignment of the three methionyl carbonyl carbon resonances in *Streptomyces subtilisin* inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution. *Biochemistry* **21**, 6273-6279 (1982).
36. D. Lacabanne, B. H. Meier, A. Bockmann, Selective labeling and unlabeled strategies in protein solid-state NMR spectroscopy. *J Biomol NMR* **71**, 141-150 (2018).
37. S. Asami, B. Reif, Proton-Detected Solid-State NMR Spectroscopy at Aliphatic Sites: Application to Crystalline Systems. *Accounts of Chemical Research* **46**, 2089-2097 (2013).
38. V. Chevelkov, K. Rehbein, A. Diehl, B. Reif, Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew Chem Int Ed Engl* **45**, 3878-3881 (2006).

39. J. R. Lewandowski *et al.*, Enhanced resolution and coherence lifetimes in the solid-state NMR spectroscopy of perdeuterated proteins under ultrafast magic-angle spinning. *The Journal of Physical Chemistry Letters* **2**, 2205-2211 (2011).
40. S. P. Brown, Applications of high-resolution ¹H solid-state NMR. *Solid State Nucl Magn Reson* **41**, 1-27 (2012).
41. S. Wang, V. Ladizhansky, Recent advances in magic angle spinning solid state NMR of membrane proteins. *Prog Nucl Magn Reson Spectrosc* **82**, 1-26 (2014).
42. L. B. Andreas, T. Le Marchand, K. Jaudzems, G. Pintacuda, High-resolution proton-detected NMR of proteins at very fast MAS. *J Magn Reson* **253**, 36-49 (2015).
43. V. Tugarinov, L. E. Kay, Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125**, 13868-13878 (2003).
44. R. L. Isaacson *et al.*, A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J Am Chem Soc* **129**, 15428-15429 (2007).
45. A. Velyvis, A. M. Ruschak, L. E. Kay, An economical method for production of (²H,¹³C)CH₃-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One* **7**, e43725 (2012).
46. M. Kainosho *et al.*, Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52-57 (2006).
47. S. Asami, P. Schmieder, B. Reif, High resolution ¹H-detected solid-state NMR spectroscopy of protein aliphatic resonances: access to tertiary structure information. *J Am Chem Soc* **132**, 15133-15135 (2010).
48. V. Agarwal, B. Reif, Residual methyl protonation in perdeuterated proteins for multi-dimensional correlation experiments in MAS solid-state NMR spectroscopy. *J Magn Reson* **194**, 16-24 (2008).
49. S. Asami, K. Szekely, P. Schanda, B. H. Meier, B. Reif, Optimal degree of protonation for (¹H) detection of aliphatic sites in randomly deuterated proteins as a function of the MAS frequency. *J Biomol NMR* **54**, 155-168 (2012).
50. T. Yamazaki, H. Tochio, J. Furui, S. Aimoto, Y. Kyogoku, Assignment of Backbone Resonances for Larger Proteins Using the ¹³C-¹H Coherence of a ¹H α -, ²H-, ¹³C-, and ¹⁵N-Labeled Sample. *Journal of the American Chemical Society* **119**, 872-880 (1997).
51. F. Hishinuma, K. Izaki, H. Takahashi, Effects of glycine and D-amino acids on growth of various microorganisms. *Agricultural and Biological Chemistry* **33**, 1577-1586 (1969).
52. S. K. Bardaweel, D-amino Acids: Prospects for new therapeutic agents. *Journal of Medical and Bioengineering Vol 3* (2014).
53. J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wuthrich, Uniform and residue-specific ¹⁵N-labeling of proteins on a highly deuterated background. *J Biomol NMR* **29**, 289-297 (2004).
54. E. S. O'Brien *et al.*, Improving yields of deuterated, methyl labeled protein by growing in H₂O. *J Biomol NMR* **71**, 263-273 (2018).
55. F. Lohr, V. Katsemi, J. Hartleib, U. Gunther, H. Ruterjans, A strategy to obtain backbone resonance assignments of deuterated proteins in the presence of incomplete amide ²H/¹H back-exchange. *J Biomol NMR* **25**, 291-311 (2003).
56. S. H. Kim, B. L. Schneider, L. Reitzer, Genetics and regulation of the major enzymes of alanine synthesis in *Escherichia coli*. *J Bacteriol* **192**, 5304-5311 (2010).
57. D. Rudman, A. Meister, Transamination in *Escherichia coli*. *J Biol Chem* **200**, 591-604 (1953).

58. G. S. Hossain *et al.*, L-Amino acid oxidases from microbial sources: types, properties, functions, and applications. *Appl Microbiol Biotechnol* **98**, 1507-1515 (2014).
59. M. Z. Sun *et al.*, Biochemical, functional and structural characterization of Akbu-LAAO: a novel snake venom L-amino acid oxidase from *Agkistrodon blomhoffii ussurensis*. *Biochimie* **92**, 343-349 (2010).
60. J. T. Nuutinen, E. Marttinen, R. Soliymani, K. Hilden, S. Timonen, L-Amino acid oxidase of the fungus *Hebeloma cylindrosporum* displays substrate preference towards glutamate. *Microbiology (Reading)* **158**, 272-283 (2012).
61. G. A. Lazar, J. R. Desjarlais, T. M. Handel, De novo design of the hydrophobic core of ubiquitin. *Protein Sci* **6**, 1167-1178 (1997).
62. J. Pauli, B. van Rossum, H. Forster, H. J. de Groot, H. Oschkinat, Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the alpha-spectrin SH3 domain. *J Magn Reson* **143**, 411-416 (2000).
63. V. Chevelkov *et al.*, Differential line broadening in MAS solid-state NMR due to dynamic interference. *Journal of the American Chemical Society* **129**, 10195-10200 (2007).
64. M. Dolder *et al.*, Crystallization of the human, mitochondrial voltage-dependent anion-selective channel in the presence of phospholipids. *J Struct Biol* **127**, 64-71 (1999).
65. M. T. Eddy *et al.*, Lipid dynamics and protein-lipid interactions in 2D crystals formed with the beta-barrel integral membrane protein VDAC1. *J Am Chem Soc* **134**, 6375-6387 (2012).
66. A. Bax, D. G. Davis, MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance (1969)* **65**, 355-360 (1985).
67. P. Zhou, L. J. Sun, V. Dotsch, G. Wagner, G. L. Verdine, Solution structure of the core NFATC1/DNA complex. *Cell* **92**, 687-696 (1998).
68. V. Massey, B. Curti, On the reaction mechanism of *Crotalus adamanteus* L-amino acid oxidase. *J Biol Chem* **242**, 1259-1264 (1967).
69. H. C. Dunathan, L. Davis, P. G. Kury, M. Kaplan, The stereochemistry of enzymatic transamination. *Biochemistry* **7**, 4532-4537 (1968).
70. D. Wellner, Stereospecificity of enzymatic formation and oxidation of glycine. *Biochemistry* **9**, 2307-2310 (1970).
71. T. Yoshimura, K. H. Jhee, K. Soda, Stereospecificity for the hydrogen transfer and molecular evolution of pyridoxal enzymes. *Biosci Biotechnol Biochem* **60**, 181-187 (1996).
72. K. V. Loscha, G. Otting, Biosynthetically directed (²H) labelling for stereospecific resonance assignments of glycine methylene groups. *J Biomol NMR* **55**, 97-104 (2013).
73. J. Medeiros-Silva *et al.*, (¹H)-Detected Solid-State NMR Studies of Water-Inaccessible Proteins In Vitro and In Situ. *Angew Chem Int Ed Engl* **55**, 13606-13610 (2016).
74. K. Xue *et al.*, Limits of Resolution and Sensitivity of Proton Detected MAS Solid-State NMR Experiments at 111 kHz in Deuterated and Protonated Proteins. *Sci Rep* **7**, 7444 (2017).
75. J. Stanek *et al.*, NMR Spectroscopic Assignment of Backbone and Side-Chain Protons in Fully Protonated Proteins: Microcrystals, Sedimented Assemblies, and Amyloid Fibrils. *Angew Chem Int Ed Engl* **55**, 15504-15509 (2016).
76. D. Cala-De Paepe *et al.*, Is protein deuteration beneficial for proton detected solid-state NMR at and above 100 kHz magic-angle spinning? *Solid State Nucl Magn Reson* **87**, 126-136 (2017).
77. R. Schneider *et al.*, The native conformation of the human VDAC1 N terminus. *Angew Chem Int Ed Engl* **49**, 1882-1885 (2010).

78. M. T. Eddy *et al.*, Magic angle spinning nuclear magnetic resonance characterization of voltage-dependent anion channel gating in two-dimensional lipid crystalline bilayers. *Biochemistry* **54**, 994-1005 (2015).
79. M. T. Eddy *et al.*, Lipid bilayer-bound conformation of an integral membrane beta barrel protein by multidimensional MAS NMR. *J Biomol NMR* **61**, 299-310 (2015).
80. M. E. Ward *et al.*, Proton-detected solid-state NMR reveals intramembrane polar networks in a seven-helical transmembrane protein proteorhodopsin. *J Am Chem Soc* **133**, 17434-17443 (2011).
81. L. B. Andreas *et al.*, Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc Natl Acad Sci U S A* **113**, 9187-9192 (2016).
82. D. H. Zhou *et al.*, Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning. *J Am Chem Soc* **129**, 11791-11801 (2007).
83. M. E. Ward *et al.*, Proton detection for signal enhancement in solid-state NMR experiments on mobile species in membrane proteins. *J Biomol NMR* **63**, 375-388 (2015).
84. P. Fricke *et al.*, Backbone assignment of perdeuterated proteins by solid-state NMR using proton detection and ultrafast magic-angle spinning. *Nat Protoc* **12**, 764-782 (2017).
85. A. Bockmann, E. Guittet, Determination of fast proton exchange rates of biomolecules by NMR using water selective diffusion experiments. *FEBS Lett* **418**, 127-130 (1997).
86. V. Agarwal, R. Linser, U. Fink, K. Faelber, B. Reif, Identification of hydroxyl protons, determination of their exchange dynamics, and characterization of hydrogen bonding in a microcrystallin protein. *J Am Chem Soc* **132**, 3187-3195 (2010).
87. D. M. LeMaster, F. M. Richards, NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteration. *Biochemistry* **27**, 142-150 (1988).
88. D. A. Torchia, S. W. Sparks, A. Bax, Delineation of α -helical domains in deuterated Staphylococcal nuclease by 2D NOE NMR spectroscopy. *Journal of the American Chemical Society* **110**, 2320-2321 (1988).
89. P. Lundstrom, D. F. Hansen, P. Vallurupalli, L. E. Kay, Accurate measurement of alpha proton chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy. *J Am Chem Soc* **131**, 1915-1926 (2009).
90. P. Vallurupalli, D. F. Hansen, P. Lundstrom, L. E. Kay, CPMG relaxation dispersion NMR experiments measuring glycine 1H alpha and 13C alpha chemical shifts in the 'invisible' excited states of proteins. *J Biomol NMR* **45**, 45-55 (2009).
91. T. A. Cross, M. Sharma, M. Yi, H. X. Zhou, Influence of solubilizing environments on membrane protein structures. *Trends Biochem Sci* **36**, 117-125 (2011).
92. L. Frey, N. A. Lakomek, R. Riek, S. Bibow, Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics. *Angew Chem Int Ed Engl* **56**, 380-383 (2017).
93. V. Kurauskas *et al.*, How Detergent Impacts Membrane Proteins: Atomic-Level Views of Mitochondrial Carriers in Dodecylphosphocholine. *J Phys Chem Lett* **9**, 933-938 (2018).
94. J. S. Retel *et al.*, Structure of outer membrane protein G in lipid bilayers. *Nat Commun* **8**, 2073 (2017).
95. O. Saurel *et al.*, Local and Global Dynamics in Klebsiella pneumoniae Outer Membrane Protein a in Lipid Bilayers Probed at Atomic Resolution. *J Am Chem Soc* **139**, 1590-1597 (2017).
96. A. Lesage, A. Bockmann, Water-protein interactions in microcrystalline crh measured by 1H-13C solid-state NMR spectroscopy. *J Am Chem Soc* **125**, 13336-13337 (2003).

97. A. Lesage *et al.*, Polarization transfer over the water-protein interface in solids. *Angew Chem Int Ed Engl* **47**, 5851-5854 (2008).
98. A. Bockmann *et al.*, Characterization of different water pools in solid-state NMR protein samples. *J Biomol NMR* **45**, 319-327 (2009).
99. G. Comellas, L. R. Lemkau, D. H. Zhou, J. M. George, C. M. Rienstra, Structural intermediates during alpha-synuclein fibrillogenesis on phospholipid vesicles. *J Am Chem Soc* **134**, 5090-5099 (2012).
100. D. Huster, K. Arnold, K. Gawrisch, Investigation of lipid organization in biological membranes by two-dimensional nuclear Overhauser enhancement spectroscopy. *The Journal of Physical Chemistry B* **103**, 243-251 (1999).
101. D. Huster, X. L. Yao, M. Hong, Membrane protein topology probed by H-1 spin diffusion from lipids using solid-state NMR spectroscopy. *Journal of the American Chemical Society* **124**, 874-883 (2002).
102. G. J. Gallagher, M. Hong, L. K. Thompson, Solid-state NMR spin diffusion for measurement of membrane-bound peptide structure: gramicidin A. *Biochemistry* **43**, 7899-7906 (2004).
103. C. Ader *et al.*, Structural rearrangements of membrane proteins probed by water-edited solid-state NMR spectroscopy. *J Am Chem Soc* **131**, 170-176 (2009).
104. T. Wang, H. Jo, W. F. DeGrado, M. Hong, Water Distribution, Dynamics, and Interactions with Alzheimer's beta-Amyloid Fibrils Investigated by Solid-State NMR. *J Am Chem Soc* **139**, 6242-6252 (2017).
105. E. K. Paulson *et al.*, High-sensitivity observation of dipolar exchange and NOEs between exchangeable protons in proteins by 3D solid-state NMR spectroscopy. *J Am Chem Soc* **125**, 14222-14223 (2003).
106. H. Van Melckebeke *et al.*, Probing water accessibility in HET-s(218-289) amyloid fibrils by solid-state NMR. *J Mol Biol* **405**, 765-772 (2011).
107. C. Fernandez, C. Hilty, G. Wider, P. Guntert, K. Wuthrich, NMR structure of the integral membrane protein OmpX. *J Mol Biol* **336**, 1211-1221 (2004).
108. A. Lange, S. Luca, M. Baldus, Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *J Am Chem Soc* **124**, 9704-9705 (2002).
109. W. Luo, M. Hong, Conformational changes of an ion channel detected through water-protein interactions using solid-state NMR spectroscopy. *J Am Chem Soc* **132**, 2378-2384 (2010).
110. Z. L. Jia, L. L. Zhang, Q. Chen, E. W. Hansen, Proton spin diffusion in polyethylene as a function of magic-angle spinning rate. A phenomenological approach. *Journal of Physical Chemistry A* **112**, 1228-1233 (2008).
111. N. P. Wickramasinghe *et al.*, Nanomole-scale protein solid-state NMR by breaking intrinsic 1HT1 boundaries. *Nat Methods* **6**, 215-218 (2009).
112. M. Huber *et al.*, A proton-detected 4D solid-state NMR experiment for protein structure determination. *Chemphyschem* **12**, 915-918 (2011).
113. M. J. Knight *et al.*, Fast resonance assignment and fold determination of human superoxide dismutase by high-resolution proton-detected solid-state MAS NMR spectroscopy. *Angew Chem Int Ed Engl* **50**, 11697-11701 (2011).
114. R. Linser *et al.*, Proton-detected solid-state NMR spectroscopy of fibrillar and membrane proteins. *Angew Chem Int Ed Engl* **50**, 4508-4512 (2011).

115. T. Kobayashi *et al.*, Study of intermolecular interactions in the corrole matrix by solid-state NMR under 100 kHz MAS and theoretical calculations. *Angew Chem Int Ed Engl* **52**, 14108-14111 (2013).
116. Y. Nishiyama, M. Malon, Y. Ishii, A. Ramamoorthy, 3D (1)(5)N/(1)(5)N/(1)H chemical shift correlation experiment utilizing an RFDR-based (1)H/(1)H mixing period at 100 kHz MAS. *J Magn Reson* **244**, 1-5 (2014).
117. K. K. Kumashiro *et al.*, A novel tool for probing membrane protein structure: Solid-state NMR with proton spin diffusion and X-nucleus detection. *Journal of the American Chemical Society* **120**, 5043-5051 (1998).
118. W. Luo, M. Hong, A 1D sensitivity-enhanced 1H spin diffusion experiment for determining membrane protein topology. *Solid State Nucl Magn Reson* **29**, 163-169 (2006).
119. M. G. Jain *et al.*, Selective (1)H-(1)H Distance Restraints in Fully Protonated Proteins by Very Fast Magic-Angle Spinning Solid-State NMR. *J Phys Chem Lett* **8**, 2399-2405 (2017).
120. T. S. Schwarzer, M. Hermann, S. Krishnan, F. C. Simmel, K. Castiglione, Preparative refolding of small monomeric outer membrane proteins. *Protein Expr Purif* **132**, 171-181 (2017).
121. T. J. Malia, G. Wagner, NMR structural investigation of the mitochondrial outer membrane protein VDAC and its interaction with antiapoptotic Bcl-xL. *Biochemistry* **46**, 514-525 (2007).
122. M. Bayrhuber *et al.*, Structure of the human voltage-dependent anion channel. *Proc Natl Acad Sci U S A* **105**, 15370-15375 (2008).
123. S. Hiller *et al.*, Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* **321**, 1206-1210 (2008).
124. T. Goddard, K. D. SPARKY, San Francisco: Univ. of California, 2008. *SPARKY, NMR Assignment and Integration Software. Version 3*.
125. A. W. Hing, S. Vega, J. Schaefer, Transferred-echo double-resonance NMR. *Journal of Magnetic Resonance* **96**, 205-209 (1992).
126. A. Lends *et al.*, Direct amide (15)N to (13)C transfers for solid-state assignment experiments in deuterated proteins. *J Biomol NMR* **72**, 69-78 (2018).
127. M. Veshkort, R. G. Griffin, Proton-driven spin diffusion in rotating solids via reversible and irreversible quantum dynamics. *J Chem Phys* **135**, 134509 (2011).
128. J. J. Wittmann *et al.*, Accelerating proton spin diffusion in perdeuterated proteins at 100 kHz MAS. *J Biomol NMR* **66**, 233-242 (2016).
129. J. Clauss, K. Schmidt-Rohr, H. W. Spiess, Determination of domain sizes in heterogeneous polymers by solid-state NMR. *Acta Polymerica* **44**, 1-17 (1993).
130. A. Fick, On liquid diffusion. *Journal of Membrane Science* **100**, 33-38 (1995).
131. S. Macura, R. Ernst, Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Molecular Physics* **41**, 95-117 (1980).
132. C. Oster *et al.*, Characterization of Protein-Protein Interfaces in Large Complexes by Solid-State NMR Solvent Paramagnetic Relaxation Enhancements. *J Am Chem Soc* **139**, 12165-12174 (2017).
133. R. Ujwal *et al.*, The crystal structure of mouse VDAC1 at 2.3 Å resolution reveals mechanistic insights into metabolite gating. *Proc Natl Acad Sci U S A* **105**, 17742-17747 (2008).
134. M. Jaremko *et al.*, High-Resolution NMR Determination of the Dynamic Structure of Membrane Proteins. *Angew Chem Int Ed Engl* **55**, 10518-10521 (2016).

135. A. Waterhouse *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).
136. H. Hong, D. R. Patel, L. K. Tamm, B. van den Berg, The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. *J Biol Chem* **281**, 7568-7577 (2006).
137. D. S. Touw, D. R. Patel, B. van den Berg, The crystal structure of OprG from *Pseudomonas aeruginosa*, a potential channel for transport of hydrophobic molecules across the outer membrane. *PLoS One* **5**, e15016 (2010).
138. R. Horst, P. Stanczak, K. Wuthrich, NMR polypeptide backbone conformation of the *E. coli* outer membrane protein W. *Structure* **22**, 1204-1209 (2014).
139. I. Kucharska, P. Seelheim, T. Edrington, B. Liang, L. K. Tamm, OprG Harnesses the Dynamics of its Extracellular Loops to Transport Small Amino Acids across the Outer Membrane of *Pseudomonas aeruginosa*. *Structure* **23**, 2234-2245 (2015).
140. J. Dev *et al.*, Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* **353**, 172-175 (2016).
141. R. Linser, U. Fink, B. Reif, Probing surface accessibility of proteins using paramagnetic relaxation in solid-state NMR spectroscopy. *J Am Chem Soc* **131**, 13703-13708 (2009).
142. S. Cohen, A. M. Valm, J. Lippincott-Schwartz, Interacting organelles. *Curr Opin Cell Biol* **53**, 84-91 (2018).
143. C. P. Satori *et al.*, Bioanalysis of eukaryotic organelles. *Chem Rev* **113**, 2733-2811 (2013).
144. H. E. Grecco, M. Schmick, P. I. Bastiaens, Signaling from the living plasma membrane. *Cell* **144**, 897-909 (2011).
145. R. E. Lawrence, R. Zoncu, The lysosome as a cellular centre for signalling, metabolism and quality control. *Nat Cell Biol* **21**, 133-142 (2019).
146. D. S. Schwarz, M. D. Blower, The endoplasmic reticulum: structure, function and response to cellular signaling. *Cell Mol Life Sci* **73**, 79-94 (2016).
147. E. Gomes, J. Shorter, The molecular language of membraneless organelles. *J Biol Chem* **294**, 7115-7127 (2019).
148. J. P. Abrahams, A. G. Leslie, R. Lutter, J. E. Walker, Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**, 621-628 (1994).
149. M. T. Ryan, N. J. Hoogenraad, Mitochondrial-nuclear communications. *Annu Rev Biochem* **76**, 701-722 (2007).
150. F. Vogel, C. Bornhovd, W. Neupert, A. S. Reichert, Dynamic subcompartmentalization of the mitochondrial inner membrane. *J Cell Biol* **175**, 237-247 (2006).
151. I. Scott, R. J. Youle, Mitochondrial fission and fusion. *Essays Biochem* **47**, 85-98 (2010).
152. J. W. Taanman, The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* **1410**, 103-123 (1999).
153. S. Campello, L. Scorrano, Mitochondrial shape changes: orchestrating cell pathophysiology. *EMBO Rep* **11**, 678-684 (2010).
154. Y. Chaban, E. J. Boekema, N. V. Dudkina, Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. *Biochim Biophys Acta* **1837**, 418-426 (2014).
155. J. B. Spinelli, M. C. Haigis, The multifaceted contributions of mitochondria to cellular metabolism. *Nat Cell Biol* **20**, 745-754 (2018).

156. B. Mesmin, Mitochondrial lipid transport and biosynthesis: A complex balance. *J Cell Biol* **214**, 9-11 (2016).
157. D. R. Green, J. C. Reed, Mitochondria and apoptosis. *Science* **281**, 1309-1312 (1998).
158. J. Godiker *et al.*, QIL1-dependent assembly of MICOS complex-lethal mutation in C19ORF70 resulting in liver disease and severe neurological retardation. *J Hum Genet* **63**, 707-716 (2018).
159. M. Milone, L. J. Wong, Diagnosis of mitochondrial myopathies. *Mol Genet Metab* **110**, 35-41 (2013).
160. T. N. Nguyen, B. S. Padman, M. Lazarou, Deciphering the Molecular Signals of PINK1/Parkin Mitophagy. *Trends Cell Biol* **26**, 733-744 (2016).
161. V. S. Van Laar, S. B. Berman, T. G. Hastings, Mic60/mitofilin overexpression alters mitochondrial dynamics and attenuates vulnerability of dopaminergic cells to dopamine and rotenone. *Neurobiol Dis* **91**, 247-261 (2016).
162. A. Zeharia *et al.*, Mitochondrial hepato-encephalopathy due to deficiency of QIL1/MIC13 (C19orf70), a MICOS complex subunit. *Eur J Hum Genet* **24**, 1778-1782 (2016).
163. E. M. Mejia, G. M. Hatch, Mitochondrial phospholipids: role in mitochondrial function. *J Bioenerg Biomembr* **48**, 99-112 (2016).
164. E. Muro, G. E. Atilla-Gokcumen, U. S. Eggert, Lipids in cell biology: how can we understand them better? *Mol Biol Cell* **25**, 1819-1823 (2014).
165. S. E. Horvath, G. Daum, Lipids of mitochondria. *Prog Lipid Res* **52**, 590-614 (2013).
166. J. Dudek, Role of Cardiolipin in Mitochondrial Signaling Pathways. *Front Cell Dev Biol* **5**, 90 (2017).
167. N. Ikon, R. O. Ryan, Cardiolipin and mitochondrial cristae organization. *Biochim Biophys Acta Biomembr* **1859**, 1156-1163 (2017).
168. N. Wiedemann, N. Pfanner, Mitochondrial Machineries for Protein Import and Assembly. *Annu Rev Biochem* **86**, 685-714 (2017).
169. A. K. S. Camara, Y. Zhou, P. C. Wen, E. Tajkhorshid, W. M. Kwok, Mitochondrial VDAC1: A Key Gatekeeper as Potential Therapeutic Target. *Front Physiol* **8**, 460 (2017).
170. W. Kuhlbrandt, Structure and function of mitochondrial membrane protein complexes. *BMC Biol* **13**, 89 (2015).
171. K. M. Davies, C. Anselmi, I. Wittig, J. D. Faraldo-Gomez, W. Kuhlbrandt, Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. *Proc Natl Acad Sci U S A* **109**, 13602-13607 (2012).
172. P. Paumard *et al.*, The ATP synthase is involved in generating mitochondrial cristae morphology. *EMBO J* **21**, 221-230 (2002).
173. G. Perkins *et al.*, Electron tomography of neuronal mitochondria: three-dimensional structure and organization of cristae and membrane contacts. *J Struct Biol* **119**, 260-272 (1997).
174. M. Strauss, G. Hofhaus, R. R. Schroder, W. Kuhlbrandt, Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. *EMBO J* **27**, 1154-1160 (2008).
175. M. W. Gray, G. Burger, B. F. Lang, The origin and early evolution of mitochondria. *Genome Biol* **2**, REVIEWS1018 (2001).
176. A. J. Roger, S. A. Munoz-Gomez, R. Kamikawa, The Origin and Diversification of Mitochondria. *Curr Biol* **27**, R1177-R1192 (2017).
177. S. A. Munoz-Gomez, J. G. Wideman, A. J. Roger, C. H. Slamovits, The Origin of Mitochondrial Cristae from Alphaproteobacteria. *Mol Biol Evol* **34**, 943-956 (2017).

178. M. van der Laan, S. E. Horvath, N. Pfanner, Mitochondrial contact site and cristae organizing system. *Curr Opin Cell Biol* **41**, 33-42 (2016).
179. F. Wollweber, K. von der Malsburg, M. van der Laan, Mitochondrial contact site and cristae organizing system: A central player in membrane shaping and crosstalk. *Biochim Biophys Acta Mol Cell Res* **1864**, 1481-1489 (2017).
180. C. A. Wurm, S. Jakobs, Differential protein distributions define two sub-compartments of the mitochondrial inner membrane in yeast. *FEBS Lett* **580**, 5628-5634 (2006).
181. S. Stoldt *et al.*, Spatial orchestration of mitochondrial translation and OXPHOS complex assembly. *Nat Cell Biol* **20**, 528-534 (2018).
182. L. Pellegrini, L. Scorrano, A cut short to death: Parl and Opa1 in the regulation of mitochondrial morphology and apoptosis. *Cell Death Differ* **14**, 1275-1284 (2007).
183. L. Scorrano *et al.*, A distinct pathway remodels mitochondrial cristae and mobilizes cytochrome c during apoptosis. *Dev Cell* **2**, 55-67 (2002).
184. S. E. Horvath *et al.*, Role of membrane contact sites in protein import into mitochondria. *Protein Sci* **24**, 277-297 (2015).
185. N. Wiedemann *et al.*, Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature* **424**, 565-571 (2003).
186. S. Cogliati, J. A. Enriquez, L. Scorrano, Mitochondrial Cristae: Where Beauty Meets Functionality. *Trends Biochem Sci* **41**, 261-273 (2016).
187. R. Quintana-Cabrera, A. Mehrotra, G. Rigoni, M. E. Soriano, Who and how in the regulation of mitochondrial cristae shape and function. *Biochem Biophys Res Commun* **500**, 94-101 (2018).
188. L. Colina-Tenorio, P. Horten, N. Pfanner, H. Rampelt, Shaping the mitochondrial inner membrane in health and disease. *J Intern Med* **287**, 645-664 (2020).
189. G. Arselin *et al.*, The GxxxG motif of the transmembrane domain of subunit e is involved in the dimerization/oligomerization of the yeast ATP synthase complex in the mitochondrial membrane. *Eur J Biochem* **270**, 1875-1884 (2003).
190. N. Pfanner *et al.*, Uniform nomenclature for the mitochondrial contact site and cristae organizing system. *J Cell Biol* **204**, 1083-1086 (2014).
191. R. M. Zerbes *et al.*, Role of MINOS in mitochondrial membrane architecture: cristae morphology and outer membrane interactions differentially depend on mitofilin domains. *J Mol Biol* **422**, 183-191 (2012).
192. S. Hoppins *et al.*, A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *J Cell Biol* **195**, 323-340 (2011).
193. K. von der Malsburg *et al.*, Dual role of mitofilin in mitochondrial membrane organization and protein biogenesis. *Dev Cell* **21**, 694-707 (2011).
194. M. Barbot *et al.*, Mic10 oligomerizes to bend mitochondrial inner membranes at cristae junctions. *Cell Metab* **21**, 756-763 (2015).
195. S. A. Munoz-Gomez *et al.*, Ancient homology of the mitochondrial contact site and cristae organizing system points to an endosymbiotic origin of mitochondrial cristae. *Curr Biol* **25**, 1489-1495 (2015).
196. M. Harner *et al.*, The mitochondrial contact site complex, a determinant of mitochondrial architecture. *EMBO J* **30**, 4356-4370 (2011).
197. G. B. John *et al.*, The mitochondrial inner membrane protein mitofilin controls cristae morphology. *Mol Biol Cell* **16**, 1543-1554 (2005).

198. D. Tarasenko *et al.*, The MICOS component Mic60 displays a conserved membrane-bending activity that is necessary for normal cristae morphology. *J Cell Biol* **216**, 889-899 (2017).
199. M. Hessenberger *et al.*, Regulated membrane remodeling by Mic60 controls formation of mitochondrial crista junctions. *Nat Commun* **8**, 15258 (2017).
200. A. K. Alkhaja *et al.*, MINOS1 is a conserved component of mitofilin complexes and required for mitochondrial function and cristae organization. *Mol Biol Cell* **23**, 247-257 (2012).
201. M. Darshi *et al.*, ChChd3, an inner mitochondrial membrane protein, is essential for maintaining crista integrity and mitochondrial function. *J Biol Chem* **286**, 2918-2932 (2011).
202. J. Tang *et al.*, Sam50-Mic19-Mic60 axis determines mitochondrial cristae architecture by mediating mitochondrial outer and inner membrane contact. *Cell Death Differ* **27**, 146-160 (2020).
203. M. Bohnert *et al.*, Central role of Mic10 in the mitochondrial contact site and cristae organizing system. *Cell Metab* **21**, 747-755 (2015).
204. H. Rampelt *et al.*, Assembly of the Mitochondrial Cristae Organizer Mic10 Is Regulated by Mic26-Mic27 Antagonism and Cardiolipin. *J Mol Biol* **430**, 1883-1890 (2018).
205. W. P. Russ, D. M. Engelman, The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **296**, 911-919 (2000).
206. I. Arnold, K. Pfeiffer, W. Neupert, R. A. Stuart, H. Schagger, Yeast mitochondrial F1F0-ATP synthase exists as a dimer: identification of three dimer-specific subunits. *EMBO J* **17**, 7170-7178 (1998).
207. S. Brunner, V. Everard-Gigot, R. A. Stuart, Some of the yeast F1Fo-ATP synthase forms homodimers. *J Biol Chem* **277**, 48484-48489 (2002).
208. N. Chester, D. R. Marshak, Dimethyl sulfoxide-mediated primer Tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Anal Biochem* **209**, 284-290 (1993).
209. M. E. Woodman, C. R. Savage, W. K. Arnold, B. Stevenson, Direct PCR of Intact Bacteria (Colony PCR). *Curr Protoc Microbiol* **42**, A 3D 1-A 3D 7 (2016).
210. P. Y. Lee, J. Costumbrado, C. Y. Hsu, Y. H. Kim, Agarose gel electrophoresis for the separation of DNA fragments. *J Vis Exp* 10.3791/3923 (2012).
211. J. Marley, M. Lu, C. Bracken, A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71-75 (2001).
212. A. A. Al-Tubuly, "SDS-PAGE and Western Blotting" in Diagnostic and Therapeutic Antibodies, A. J. T. George, C. E. Urch, Eds. (Humana Press, Totowa, NJ, 2000), 10.1385/1-59259-076-4:391, pp. 391-405.
213. P. Gravel, "Protein Blotting by the Semidry Method" in The Protein Protocols Handbook, J. M. Walker, Ed. (Humana Press, Totowa, NJ, 2009), 10.1007/978-1-59745-198-7_59, pp. 621-629.
214. R. B. Pepinsky, Selective precipitation of proteins from guanidine hydrochloride-containing solutions with ethanol. *Analytical Biochemistry* **195**, 177-181 (1991).
215. T. Horigome, H. Sugano, A rapid method for removal of detergents from protein solution. *Anal Biochem* **130**, 393-396 (1983).
216. G. A. Signorell, T. C. Kaufmann, W. Kukulski, A. Engel, H. W. Remigy, Controlled 2D crystallization of membrane proteins using methyl-beta-cyclodextrin. *J Struct Biol* **157**, 321-328 (2007).

217. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-2890 (2006).
218. G. Metz, X. L. Wu, S. O. Smith, Ramped-Amplitude Cross Polarization in Magic-Angle-Spinning NMR. *Journal of Magnetic Resonance, Series A* **110**, 219-227 (1994).
219. W. F. Vranken *et al.*, The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696 (2005).
220. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327 (2015).
221. M. Ikeda, M. Arai, T. Okuno, T. Shimizu, TMPDB: a database of experimentally-characterized transmembrane topologies. *Nucleic Acids Res* **31**, 406-409 (2003).
222. J. Ko, H. Park, C. Seok, GalaxyTBM: template-based modeling by building a reliable core and refining unreliable local regions. *BMC Bioinformatics* **13**, 198 (2012).
223. A. Senes, M. Gerstein, D. M. Engelman, Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* **296**, 921-936 (2000).
224. W. Parker, P. S. Song, Protein structures in SDS micelle-protein complexes. *Biophys J* **61**, 1435-1439 (1992).
225. K. Pervushin, R. Riek, G. Wider, K. Wuthrich, Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-12371 (1997).
226. D. S. Wishart, B. D. Sykes, F. M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651 (1992).
227. R. Simbeni, L. Pon, E. Zinser, F. Paltauf, G. Daum, Mitochondrial membrane contact sites of yeast. Characterization of lipid components and possible involvement in intramitochondrial translocation of phospholipids. *J Biol Chem* **266**, 10047-10049 (1991).
228. D. Schneider, D. M. Engelman, Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J Mol Biol* **343**, 799-804 (2004).
229. M. G. Teese, D. Langosch, Role of GxxxG Motifs in Transmembrane Domain Interactions. *Biochemistry* **54**, 5125-5135 (2015).
230. S. M. Anderson, B. K. Mueller, E. J. Lange, A. Senes, Combination of Calpha-H Hydrogen Bonds and van der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. *J Am Chem Soc* **139**, 15774-15783 (2017).
231. A. A. Polyansky *et al.*, PREDDIMER: a web server for prediction of transmembrane helical dimers. *Bioinformatics* **30**, 889-890 (2014).
232. M. Chevallet, S. Luche, T. Rabilloud, Silver staining of proteins in polyacrylamide gels. *Nat Protoc* **1**, 1852-1858 (2006).
233. B. M. Baynes, D. I. Wang, B. L. Trout, Role of arginine in the stabilization of proteins against aggregation. *Biochemistry* **44**, 4919-4925 (2005).
234. K. Tamiola, B. Acar, F. A. Mulder, Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* **132**, 18000-18003 (2010).
235. P. Guntert, Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-378 (2004).
236. M. Barbot, M. Meinecke, Reconstitutions of mitochondrial inner membrane remodeling. *J Struct Biol* **196**, 20-28 (2016).

237. F. Chiti, C. M. Dobson, Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* **86**, 27-68 (2017).
238. W. Yassine *et al.*, Reversible transition between alpha-helix and beta-sheet conformation of a transmembrane domain. *Biochim Biophys Acta* **1788**, 1722-1730 (2009).
239. J. Han, K. Pluhackova, R. A. Bockmann, The Multifaceted Role of SNARE Proteins in Membrane Fusion. *Front Physiol* **8**, 5 (2017).
240. K. Sackett, M. J. Nethercott, Z. Zheng, D. P. Weliky, Solid-state NMR spectroscopy of the HIV gp41 membrane fusion protein supports intermolecular antiparallel beta sheet fusion peptide structure in the final six-helix bundle state. *J Mol Biol* **426**, 1077-1094 (2014).
241. V. Guarani *et al.*, QIL1 is a novel mitochondrial protein required for MICOS complex stability and cristae morphology. *Elife* **4** (2015).
242. R. M. Zerbes, P. Hoss, N. Pfanner, M. van der Laan, M. Bohnert, Distinct Roles of Mic12 and Mic27 in the Mitochondrial Contact Site and Cristae Organizing System. *J Mol Biol* **428**, 1485-1492 (2016).
243. V. Chinchar, REPLICATION OF VIRUSES. *Encyclopedia of Virology* 10.1006/rwvi.1999.0245, 1471-1478 (1999).
244. J. K. Taubenberger, D. M. Morens, 1918 Influenza: the mother of all pandemics. *Emerg Infect Dis* **12**, 15-22 (2006).
245. A. S. Fauci, Seasonal and pandemic influenza preparedness: science and countermeasures. *J Infect Dis* **194 Suppl 2**, S73-76 (2006).
246. E. de Wit, N. van Doremalen, D. Falzarano, V. J. Munster, SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* **14**, 523-534 (2016).
247. J. Liu *et al.*, A comparative overview of COVID-19, MERS and SARS: Review article. *Int J Surg* **81**, 1-8 (2020).
248. X. Chen, M. A. Goncalves, Engineered Viruses as Genome Editing Devices. *Mol Ther* **24**, 447-457 (2016).
249. M. A. Kay, J. C. Glorioso, L. Naldini, Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med* **7**, 33-40 (2001).
250. K. Lundstrom, Viral Vectors in Gene Therapy. *Diseases* **6** (2018).
251. M. E. McLaughlin-Drubin, K. Munger, Viruses associated with human cancer. *Biochim Biophys Acta* **1782**, 127-150 (2008).
252. P. S. Moore, Y. Chang, Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nat Rev Cancer* **10**, 878-889 (2010).
253. A. Morales-Sanchez, E. M. Fuentes-Panana, Human viruses and cancer. *Viruses* **6**, 4047-4079 (2014).
254. D. Lavanchy, Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Viral Hepat* **11**, 97-107 (2004).
255. H. Budka, Neuropathology of human immunodeficiency virus infection. *Brain Pathol* **1**, 163-175 (1991).
256. M. Ludlow *et al.*, Neurotropic virus infections as the cause of immediate and delayed neuropathology. *Acta Neuropathol* **131**, 159-184 (2016).
257. F. Krammer, P. Palese, Advances in the development of influenza virus vaccines. *Nat Rev Drug Discov* **14**, 167-182 (2015).
258. S. Rewar, D. Mirdha, P. Rewar, Treatment and Prevention of Pandemic H1N1 Influenza. *Ann Glob Health* **81**, 645-653 (2015).

259. A. A. T. Naqvi, T. Mohammad, G. M. Hasan, M. I. Hassan, Advancements in Docking and Molecular Dynamics Simulations Towards Ligand-receptor Interactions and Structure-function Relationships. *Curr Top Med Chem* **18**, 1755-1768 (2018).
260. R. R. Razonable, Antiviral drugs for viruses other than human immunodeficiency virus. *Mayo Clin Proc* **86**, 1009-1026 (2011).
261. D. Vijaykrishna, R. Mukerji, G. J. Smith, RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. *PLoS Pathog* **11**, e1004902 (2015).
262. S. Duffy, L. A. Shackelton, E. C. Holmes, Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**, 267-276 (2008).
263. J. N. Mandl *et al.*, Reservoir host immune responses to emerging zoonotic viruses. *Cell* **160**, 20-35 (2015).
264. M. C. Zambon, The pathogenesis of influenza in humans. *Rev Med Virol* **11**, 227-241 (2001).
265. C. International Committee on Taxonomy of Viruses Executive, The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat Microbiol* **5**, 668-674 (2020).
266. S. Durmus, K. O. Ulgen, Comparative interactomics for virus-human protein-protein interactions: DNA viruses versus RNA viruses. *FEBS Open Bio* **7**, 96-107 (2017).
267. F. Krammer *et al.*, Influenza. *Nat Rev Dis Primers* **4**, 3 (2018).
268. N. M. Bouvier, P. Palese, The biology of influenza viruses. *Vaccine* **26 Suppl 4**, D49-53 (2008).
269. M. J. Gething, K. McCammon, J. Sambrook, Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. *Cell* **46**, 939-950 (1986).
270. S. J. Gamblin *et al.*, The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838-1842 (2004).
271. L. J. Holsinger, D. Nichani, L. H. Pinto, R. A. Lamb, Influenza A virus M2 ion channel protein: a structure-function analysis. *J Virol* **68**, 1551-1563 (1994).
272. M. L. Shaw, K. L. Stone, C. M. Colangelo, E. E. Gulcicek, P. Palese, Cellular proteins in influenza virus particles. *PLoS Pathog* **4**, e1000085 (2008).
273. S. A. Wharton, The role of influenza virus haemagglutinin in membrane fusion. *Microbiol Sci* **4**, 119-124 (1987).
274. F. Ciampor, D. Cmarko, J. Cmarkova, E. Zavodska, Influenza virus M2 protein and haemagglutinin conformation changes during intracellular transport. *Acta Virol* **39**, 171-181 (1995).
275. D. Dou, R. Revol, H. Ostbye, H. Wang, R. Daniels, Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Front Immunol* **9**, 1581 (2018).
276. T. Sakai, S. I. Nishimura, T. Naito, M. Saito, Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. *Sci Rep* **7**, 45043 (2017).
277. S. L. Zebedee, R. A. Lamb, Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *J Virol* **62**, 2762-2772 (1988).
278. L. H. Pinto, R. A. Lamb, The M2 proton channels of influenza A and B viruses. *J Biol Chem* **281**, 8997-9000 (2006).
279. W. W. Wu, Y. H. Sun, N. Pante, Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. *Virology* **4**, 49 (2007).

280. C. De Vlugt, D. Sikora, M. Pelchat, Insight into Influenza: A Virus Cap-Snatching. *Viruses* **10** (2018).
281. T. J. Bos, A. R. Davis, D. P. Nayak, NH₂-terminal hydrophobic region of influenza virus neuraminidase provides the signal function in translocation. *Proc Natl Acad Sci U S A* **81**, 2327-2331 (1984).
282. J. D. Hull, R. Gilmore, R. A. Lamb, Integration of a small integral membrane protein, M2, of influenza virus into the endoplasmic reticulum: analysis of the internal signal-anchor domain of a protein with an ectoplasmic NH₂ terminus. *J Cell Biol* **106**, 1489-1498 (1988).
283. E. Alvarado-Facundo *et al.*, Influenza virus M2 protein ion channel activity helps to maintain pandemic 2009 H1N1 virus hemagglutinin fusion competence during transport to the cell surface. *J Virol* **89**, 1975-1985 (2015).
284. A. J. Einfeld, G. Neumann, Y. Kawaoka, At the centre: influenza A virus ribonucleoproteins. *Nat Rev Microbiol* **13**, 28-41 (2015).
285. J. L. McAuley, B. P. Gilbertson, S. Trifkovic, L. E. Brown, J. L. McKimm-Breschkin, Influenza Virus Neuraminidase Structure and Functions. *Front Microbiol* **10**, 39 (2019).
286. M. Hussain, H. D. Galvin, T. Y. Haw, A. N. Nutsford, M. Husain, Drug resistance in influenza A virus: the epidemiology and management. *Infect Drug Resist* **10**, 121-134 (2017).
287. R. A. Lamb, S. L. Zebedee, C. D. Richardson, Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Cell* **40**, 627-633 (1985).
288. L. H. Pinto, L. J. Holsinger, R. A. Lamb, Influenza virus M2 protein has ion channel activity. *Cell* **69**, 517-528 (1992).
289. C. Wang, K. Takeuchi, L. H. Pinto, R. A. Lamb, Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. *J Virol* **67**, 5585-5594 (1993).
290. M. F. McCown, A. Pekosz, The influenza A virus M2 cytoplasmic tail is required for infectious virus production and efficient genome packaging. *J Virol* **79**, 3595-3605 (2005).
291. M. L. Grantham, S. M. Stewart, E. N. Lalime, A. Pekosz, Tyrosines in the influenza A virus M2 protein cytoplasmic tail are critical for production of infectious virus particles. *J Virol* **84**, 8765-8776 (2010).
292. C. Wang, R. A. Lamb, L. H. Pinto, Activation of the M2 ion channel of influenza virus: a role for the transmembrane domain histidine residue. *Biophys J* **69**, 1363-1371 (1995).
293. P. Venkataraman, R. A. Lamb, L. H. Pinto, Chemical rescue of histidine selectivity filter mutants of the M2 ion channel of influenza A virus. *J Biol Chem* **280**, 21463-21472 (2005).
294. F. Hu, K. Schmidt-Rohr, M. Hong, NMR detection of pH-dependent histidine-water proton exchange reveals the conduction mechanism of a transmembrane proton channel. *J Am Chem Soc* **134**, 3703-3713 (2012).
295. M. T. Colvin, L. B. Andreas, J. J. Chou, R. G. Griffin, Proton association constants of His 37 in the Influenza-A M218-60 dimer-of-dimers. *Biochemistry* **53**, 5987-5994 (2014).
296. J. Hu *et al.*, Histidines, heart of the hydrogen ion channel from influenza A virus: toward an understanding of conductance and proton selectivity. *Proc Natl Acad Sci U S A* **103**, 6865-6870 (2006).
297. K. P. Howard, J. D. Lear, W. F. DeGrado, Sequence determinants of the energetics of folding of a transmembrane four-helix-bundle protein. *Proc Natl Acad Sci U S A* **99**, 8568-8572 (2002).
298. Y. Tang, F. Zaitseva, R. A. Lamb, L. H. Pinto, The gate of the influenza virus M2 proton channel is formed by a single tryptophan residue. *J Biol Chem* **277**, 39880-39886 (2002).
299. F. Hu, W. Luo, M. Hong, Mechanisms of proton conduction and gating in influenza M2 proton channels from solid-state NMR. *Science* **330**, 505-508 (2010).

300. J. L. Thomaston *et al.*, XFEL structures of the influenza M2 proton channel: Room temperature water networks and insights into proton conduction. *Proc Natl Acad Sci U S A* **114**, 13357-13362 (2017).
301. C. Wei, A. Pohorille, Activation and proton transport mechanism in influenza A M2 channel. *Biophys J* **105**, 2036-2045 (2013).
302. M. Yi, T. A. Cross, H. X. Zhou, A secondary gate as a mechanism for inhibition of the M2 proton channel by amantadine. *J Phys Chem B* **112**, 7977-7979 (2008).
303. C. Tian, P. F. Gao, L. H. Pinto, R. A. Lamb, T. A. Cross, Initial structural and dynamic characterization of the M2 protein transmembrane and amphipathic helices in lipid bilayers. *Protein Sci* **12**, 2597-2605 (2003).
304. C. Li, H. Qin, F. P. Gao, T. A. Cross, Solid-state NMR characterization of conformational plasticity within the transmembrane domain of the influenza A M2 proton channel. *Biochim Biophys Acta* **1768**, 3162-3170 (2007).
305. F. M. Marassi, S. J. Opella, NMR structural studies of membrane proteins. *Current opinion in structural biology* **8**, 640-648 (1998).
306. J. Hu *et al.*, Backbone structure of the amantadine-blocked trans-membrane domain M2 proton channel from Influenza A virus. *Biophys J* **92**, 4335-4343 (2007).
307. J. L. Thomaston *et al.*, Inhibitors of the M2 Proton Channel Engage and Disrupt Transmembrane Networks of Hydrogen-Bonded Waters. *J Am Chem Soc* **140**, 15219-15226 (2018).
308. F. Hu, W. Luo, S. D. Cady, M. Hong, Conformational plasticity of the influenza A M2 transmembrane helix in lipid bilayers under varying pH, drug binding, and membrane thickness. *Biochim Biophys Acta* **1808**, 415-423 (2011).
309. L. B. Andreas *et al.*, Structure and Mechanism of the Influenza A M218-60 Dimer of Dimers. *J Am Chem Soc* **137**, 14877-14886 (2015).
310. J. L. Thomaston *et al.*, X-ray Crystal Structure of the Influenza A M2 Proton Channel S31N Mutant in Two Conformational States: An Open and Shut Case. *J Am Chem Soc* **141**, 11481-11488 (2019).
311. R. Acharya *et al.*, Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc Natl Acad Sci U S A* **107**, 15075-15080 (2010).
312. M. Sharma *et al.*, Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer. *Science* **330**, 509-512 (2010).
313. L. B. Andreas, M. T. Eddy, R. M. Pielak, J. Chou, R. G. Griffin, Magic angle spinning NMR investigation of influenza A M2(18-60): support for an allosteric mechanism of inhibition. *J Am Chem Soc* **132**, 10958-10960 (2010).
314. K. C. Duff, R. H. Ashley, The transmembrane domain of influenza A M2 protein forms amantadine-sensitive proton channels in planar lipid bilayers. *Virology* **190**, 485-489 (1992).
315. A. J. Hay, A. J. Wolstenholme, J. J. Skehel, M. H. Smith, The molecular basis of the specific anti-influenza action of amantadine. *EMBO J* **4**, 3021-3024 (1985).
316. R. A. Bright, D. K. Shay, B. Shu, N. J. Cox, A. I. Klimov, Adamantane resistance among influenza A viruses isolated early during the 2005-2006 influenza season in the United States. *JAMA* **295**, 891-894 (2006).
317. A. L. Stouffer *et al.*, Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* **451**, 596-599 (2008).
318. L. B. Andreas *et al.*, Dynamic nuclear polarization study of inhibitor binding to the M2(18-60) proton transporter from influenza A. *Biochemistry* **52**, 2774-2782 (2013).

319. S. Cady, T. Wang, M. Hong, Membrane-dependent effects of a cytoplasmic helix on the structure and drug binding of the influenza virus M2 protein. *J Am Chem Soc* **133**, 11572-11579 (2011).
320. J. Wang *et al.*, Molecular dynamics simulation directed rational design of inhibitors targeting drug-resistant mutants of influenza A virus M2. *J Am Chem Soc* **133**, 12834-12841 (2011).
321. J. Wang *et al.*, Structure and inhibition of the drug-resistant S31N mutant of the M2 ion channel of influenza A virus. *Proc Natl Acad Sci U S A* **110**, 1315-1320 (2013).
322. L. B. Andreas, M. T. Eddy, J. J. Chou, R. G. Griffin, Magic-angle-spinning NMR of the drug resistant S31N M2 proton transporter from influenza A. *J Am Chem Soc* **134**, 7215-7218 (2012).
323. E. Schmidt, P. Guntert, A new algorithm for reliable and general NMR resonance assignment. *J Am Chem Soc* **134**, 12817-12829 (2012).
324. M. J. Frisch *et al.* (2016) Gaussian 16 Rev. B.01. (Wallingford, CT).
325. C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B Condens Matter* **37**, 785-789 (1988).
326. C. A. a. V. Barone, Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J Chem Phys* **108** (1997).
327. R. M. Pielak, J. R. Schnell, J. J. Chou, Mechanism of drug inhibition and drug resistance of influenza A M2 channel. *Proc Natl Acad Sci U S A* **106**, 7379-7384 (2009).
328. T. V. Can *et al.*, Magic angle spinning and oriented sample solid-state NMR structural restraints combine for influenza a M2 protein functional insights. *J Am Chem Soc* **134**, 9022-9025 (2012).
329. H. Dong, M. Yi, T. A. Cross, H. X. Zhou, Ab initio calculations and validation of the pH-dependent structures of the His37-Trp41 quartet, the heart of acid activation and proton conductance in the M2 protein of Influenza A virus. *Chem Sci* **4**, 2776-2787 (2013).
330. G. A. Morris, R. Freeman, Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* **101**, 760-762 (1979).
331. S. Li, M. Hong, Protonation, tautomerization, and rotameric structure of histidine: a comprehensive study by magic-angle-spinning solid-state NMR. *J Am Chem Soc* **133**, 1534-1544 (2011).
332. Y. Miao, R. Fu, H. X. Zhou, T. A. Cross, Dynamic Short Hydrogen Bonds in Histidine Tetrad of Full-Length M2 Proton Channel Reveal Tetrameric Structural Heterogeneity and Functional Mechanism. *Structure* **23**, 2300-2308 (2015).
333. J. Hu, R. Fu, T. A. Cross, The chemical and dynamical influence of the anti-viral drug amantadine on the M2 proton channel transmembrane domain. *Biophys J* **93**, 276-283 (2007).
334. S. D. Cady, M. Hong, Amantadine-induced conformational and dynamical changes of the influenza M2 transmembrane proton channel. *Proc Natl Acad Sci U S A* **105**, 1483-1488 (2008).
335. V. Agarwal *et al.*, De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy. *Angew Chem Int Ed Engl* **53**, 12253-12256 (2014).
336. F. Castellani *et al.*, Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98-102 (2002).

337. A. Lange *et al.*, A concept for rapid protein-structure determination by solid-state NMR spectroscopy. *Angew Chem Int Ed Engl* **44**, 2089-2092 (2005).
338. E. Barbet-Massin *et al.*, Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J Am Chem Soc* **136**, 12489-12497 (2014).
339. T. Schubeis, T. Le Marchand, L. B. Andreas, G. Pintacuda, (1)H magic-angle spinning NMR evolves as a powerful new tool for membrane proteins. *J Magn Reson* **287**, 140-152 (2018).
340. Y. Xu, D. Long, D. Yang, Rapid data collection for protein structure determination by NMR spectroscopy. *J Am Chem Soc* **129**, 7722-7723 (2007).
341. R. Linser, B. Bardiaux, V. Higman, U. Fink, B. Reif, Structure calculation from unambiguous long-range amide and methyl 1H-1H distance restraints for a microcrystalline protein with MAS solid-state NMR spectroscopy. *J Am Chem Soc* **133**, 5905-5912 (2011).
342. K. T. Movellan *et al.*, Imidazole-Imidazole Hydrogen Bonding in the pH-Sensing Histidine Side Chains of Influenza A M2. *J Am Chem Soc* **142**, 2704-2708 (2020).
343. Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56**, 227-241 (2013).
344. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526-531 (2004).
345. J. K. Williams, Y. Zhang, K. Schmidt-Rohr, M. Hong, pH-dependent conformation, dynamics, and aromatic interaction of the gating tryptophan residue of the influenza M2 proton channel from solid-state NMR. *Biophys J* **104**, 1698-1708 (2013).
346. R. Fu, Y. Miao, H. Qin, T. A. Cross, Observation of the Imidazole-Imidazolium Hydrogen Bonds Responsible for Selective Proton Conductance in the Influenza A M2 Channel. *J Am Chem Soc* **142**, 2115-2119 (2020).
347. J. J. Madsen, J. M. A. Grime, J. S. Rossman, G. A. Voth, Entropic forces drive clustering and spatial localization of influenza A M2 during viral budding. *Proc Natl Acad Sci U S A* **115**, E8595-E8603 (2018).
348. J. Paulino, X. Pang, I. Hung, H. X. Zhou, T. A. Cross, Influenza A M2 Channel Clustering at High Protein/Lipid Ratios: Viral Budding Implications. *Biophys J* **116**, 1075-1084 (2019).
349. A. Chernyshev, S. Cukierman, Thermodynamic view of activation energies of proton transfer in various gramicidin A channels. *Biophys J* **82**, 182-192 (2002).
350. R. Pomès, B. Roux, Free energy profiles for H⁺ conduction along hydrogen-bonded chains of water molecules. *Biophysical Journal* **75**, 7 (1998).
351. J. L. Thomaston *et al.*, High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proc Natl Acad Sci U S A* **112**, 14260-14265 (2015).
352. T. Steiner, The hydrogen bond in the solid state. *Angew Chem Int Ed Engl* **41**, 49-76 (2002).
353. R. M. Pielak, K. Oxenoid, J. J. Chou, Structural investigation of rimantadine inhibition of the AM2-BM2 chimera channel of influenza viruses. *Structure* **19**, 1655-1663 (2011).
354. M. Hong, K. J. Fritzsche, J. K. Williams, Hydrogen-bonding partner of the proton-conducting histidine in the influenza M2 proton channel revealed from 1H chemical shifts. *J Am Chem Soc* **134**, 14753-14755 (2012).
355. C. Shi *et al.*, Structure and Dynamics of the Rhomboid Protease GlpG in Liposomes Studied by Solid-State NMR. *Journal of the American Chemical Society* **141**, 17314-17321 (2019).

356. S. B. Lesnichin, I. G. Shenderovich, T. Muljati, D. Silverman, H. H. Limbach, Intrinsic proton-donating power of zinc-bound water in a carbonic anhydrase active site model estimated by NMR. *J Am Chem Soc* **133**, 11331-11338 (2011).
357. Y. Su, L. Andreas, R. G. Griffin, Magic angle spinning NMR of proteins: high-frequency dynamic nuclear polarization and (1)H detection. *Annu Rev Biochem* **84**, 465-497 (2015).
358. K. Jaudzems *et al.*, Dynamic Nuclear Polarization-Enhanced Biomolecular NMR Spectroscopy at High Magnetic Field with Fast Magic-Angle Spinning. *Angew Chem Int Ed Engl* **57**, 7458-7462 (2018).
359. D. L. Williams-Smith, R. C. Bray, M. J. Barber, A. D. Tsopanakis, S. P. Vincent, Changes in apparent pH on freezing aqueous buffer solutions and their relevance to biochemical electron-paramagnetic-resonance spectroscopy. *Biochem J* **167**, 593-600 (1977).
360. M. Yi, T. A. Cross, H. X. Zhou, Conformational heterogeneity of the M2 proton channel and a structural model for channel activation. *Proc Natl Acad Sci U S A* **106**, 13311-13316 (2009).

LIST of PUBLICATIONS:

- Zhang, Xizhou*; Forster, Marcel*; Nimerovsky, Evgeny; Tekwani Movellan, Kumar; Andreas, Loren (under revision)
Transferred-Rotational-Echo DOuble Resonance (TREDOR)
*denotes equal contribution
- Tobias Schubeis, Tanguy Le Marchand, Wojciech Kopec, Kumar Tekwani Movellan, Jan Stanek, Tom Schwarzer, Kathrin Castiglione, Bert L. de Groot, Guido Pintacuda, Loren B Andreas (2020).
A β -barrel for oil transport through lipid membranes: Dynamic NMR structures of AlkL. Proceedings of the National Academy of Sciences Aug 2020, 202002598; DOI: 10.1073/pnas.2002598117
- Tekwani Movellan, K., Wegstroth, M., Overkamp, K., Leonov, A., Becker, S., & Andreas, L. B. (2020). Imidazole-imidazole hydrogen bonding in the pH sensing Histidine sidechains of Influenza A M2. *Journal of the American Chemical Society*, 142, 6, 2704-2708.
- Movellan, K. T., Najbauer, E. E., Pratihari, S., Salvi, M., Giller, K., Becker, S., & Andreas, L. B. (2019). Alpha protons as NMR probes in deuterated proteins. *Journal of biomolecular NMR*, 1-11.
- Najbauer, E. E.*, Movellan, K. T.*, Schubeis, T., Schwarzer, T., Castiglione, K., Giller, K., ... & Andreas, L. B. (2019). Probing Membrane Protein Insertion into Lipid Bilayers by Solid-State NMR. *ChemPhysChem*, 20(2), 302-310.
*denotes equal contribution
- Gervais, V., Muller, I., Mari, P. O., Mourcet, A., Movellan, K. T., Ramos, P., ... & Czaplicki, G. (2018). Small molecule-based targeting of TTD-A dimerization to control TFIIH transcriptional activity represents a potential strategy for anticancer therapy. *Journal of Biological Chemistry*, 293(39), 14974-14988.
- Grohe, K., Movellan, K. T., Vasa, S. K., Giller, K., Becker, S., & Linser, R. (2017). Non-equilibrium hydrogen exchange for determination of H-bond strength and water accessibility in solid proteins. *Journal of biomolecular NMR*, 68(1), 7-17.