



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN



In and through the membrane by solid-state NMR: investigations of the human voltage-dependent anion channel 1 and Nup98-derived FG-domains

Dissertation

for the award of the degree

"Doctor rerum naturalium" (Dr. rer. nat.)

of the Georg-August-Universität Göttingen

within the doctoral program **Chemistry**

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

submitted by

Eszter Éva Najbauer

from

Orange, California, United States of America

Göttingen 2021

Thesis Committee

Professor Dr. Christian Griesinger

Dept. of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry

Professor Dr. Andreas Janshoff

Institute for Physical Chemistry, Georg August University Göttingen

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. Andreas Janshoff

Institute for Physical Chemistry, Georg August University Göttingen

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. Dirk Görlich

Dept. of Cellular Logistics, Max Planck Institute for Biophysical Chemistry

Professor Dr. Bert de Groot

Dept. of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry

Professor Dr. Matthew Eddy

Department of Chemistry, University of Florida, USA

Date of the oral examination: March 4, 2021

Affidavit

I hereby declare that this doctoral thesis with the title “In and through the membrane by solid-state NMR: investigations of the human voltage-dependent anion channel 1 and Nup98-derived FG-domains” (doi:10.53846/goediss-52) is based on my own scientific work, except for the sources and aids quoted.

Göttingen, January 25, 2021

I have not yet lost a feeling of wonder, and of delight, that this delicate motion should reside in all the ordinary things around us, revealing itself only to him who looks for it. I remember, in the winter of our first experiments, just seven years ago, looking on snow with new eyes. There the snow lay around my doorstep - great heaps of protons quietly precessing in the earth's magnetic field. To see the world for a moment as something rich and strange is the private reward of many a discovery.

Edward M. Purcell
Nobel Lecture, December 11, 1952

Acknowledgements

Many people have taught and supported me throughout the years of my PhD training.

First and foremost, I would like to thank **Dr. Loren Andreas** for the somewhat more than four years that I spent in his group. With unending patience, through long hours of measurement at the spectrometers, he taught all of us in the group not only about the theory of NMR, but also gave us invaluable practical knowledge. I remain astonished at the breadth and depth of his knowledge, and almost superhuman ability to follow what is going on in each project. I am very grateful for the interesting and diverse projects I had a chance to work on, from membrane proteins, through disordered hydrogels, to a macro for linear drift correction, through which I even discovered a passion for programming. Our group events outside of work – strawberry picking, movie nights, Thanksgiving dinners – will always remain fond memories.

I would like to thank **Prof. Dr. Christian Griesinger** for his guidance and generous support throughout my PhD training. The Max Planck Institute for Biophysical Chemistry has been a place of high scientific inspiration and learning.

I owe many thanks to **Prof. Dr. Andreas Janshoff** (Georg-August-Universität, Göttingen) for his generous time and serving as external committee member on my Thesis Committee.

Xizhou (Cecily) Zhang and **Marcel Forster** have been fantastic office mates and good friends, with whom we had countless coffees at the lab, and cooked many breakfasts and dinners outside of it. They were always there to cheer me up and make me laugh when I most needed it. I will miss the many trips I took with Cecily, and will be eternally grateful to her for introducing me to Chinese cuisine. Marcel's angelic patience in German conversation will also not be forgotten.

I am very thankful to **Kumar Tekwani Movellan** for all the scientific discussions and his help with the practical aspects of NMR. He was always available to discuss new developments in any project, and was a terrific colleague to have right from the start of my PhD.

Karin Giller and **Dr. Stefan Becker** have provided VDAC samples of unparalleled quality. I am also very grateful to Stefan for our regular long discussions about news in science and the world in general, and for his continued encouragement and guidance.

Dr. Sheung Chun (John) Ng has been the best collaborator anyone could wish for. John taught me all I know in the biolab, and his enthusiasm for science was infectious. I feel honored that I had a chance to work with him for three years. I won't forget the scientific coffee breaks, the many delicious recipes I had a chance to try, and the astronomy lessons.

I am grateful to **Prof. Dr. Dirk Görlich** for the exciting collaboration on FG-nucleoporins. His rigorous approach and tough questions always drove the project in the right direction.

Gitta Angerstein has kept our many spectrometers in mint condition, placed orders, and always looked after equipment shipped to Bruker to be repaired, making sure we got it back on time, which was not an easy task. Without her, the lab would not have run so smoothly. **Dr. Dirk Bockelmann** maintained a stable computer environment, and we could always turn to him in case of technical trouble.

I am grateful to the solid-state NMR group for scientific companionship and a cheerful environment – **Riza, Kai, Evgeny**, thank you! I learned many tricks of the trade in solution NMR spectroscopy from **Supriya, Christian, Hessam, Daniel, Juan Carlos, Lukas**, and **Kumar**, without which the solution measurements for FG-domains would have taken much longer.

Last, but not least, I could never have made it through all these years without the unending support of my family and my friends – thank you for everything!

Abstract

Magic-angle-spinning NMR spectroscopy is a powerful tool for the investigation of protein structure, dynamics and interactions at atomic resolution. We have applied this technique to investigate two widely different systems: the human voltage-dependent anion channel 1, a 31 kDa membrane protein, and Nup98-derived hydrogel particles that recapitulate the selectivity of the nuclear pore complex.

The voltage-dependent anion channel (VDAC) is the most abundant protein in the outer mitochondrial membrane, responsible for the transport of all ions and metabolites through the membrane. The channel derives its name from the drop in the channel's conductance upon applied voltage. Although the protein had been previously extensively studied in a detergent environment, much less was known of its structure, dynamics and interactions in a native-like lipid bilayer.

Using fast magic angle spinning (55 kHz) NMR spectroscopy on a 2D lipid crystalline preparation of human VDAC1 previously shown to yield solid-state NMR spectra of exceptional resolution, we have significantly increased the assignment of the protein from 28% to 69%, and showed that the protein's topology in a lipid bilayer is essentially the same as in a detergent environment: a 19-stranded β -barrel with an N-terminal kinked α -helix position inside the pore. Using a mutant implementing the channel's closed state, we found that dynamics appear to be a key element in the protein's gating behavior. We showed that cholesterol, previously shown to reduce the frequency of channel closure, stabilizes the barrel in comparison to the N-terminal helix, and identified three binding sites on the C-terminal barrel wall. On the other hand, through investigation of a quintuple VDAC mutant implementing the channel's closed state, we found that channel closure leads not only to destabilization of the the C-terminal barrel, but also of the $\alpha 2$ helix. We also observed an alternative mechanism for closing of the channel through steric blockage by the Bcl2-antisense oligonucleotide, G3139.

In relation to this project we developed a method to probe the environment of a membrane protein using z-z mixing under conditions of fast spinning and perdeuteration. We devised 3- and 4 dimensional pulse sequences that allowed us to determine site-specific exposure of the protein to mobile water and lipids. We also determined transfer rates from protein to lipid and within the protein to explain site-specificity.

The peak broadening caused by extensive drift during long multidimensional experiments made it necessary to find a way to compensate the effect of field drift. We found that a simple linear correction is sufficient to significantly improve spectral resolution. We achieved this using a script written in the C programming language, but directly executable from Topspin to correct for the drift occurring during 2D, 3D, and 4D experiments.

Lastly, we investigated hydrogels formed from Nup98-derived FG-domains by a combination of biophysical and structural methods. Cohesive FG (phenylalanine-glycine)-repeat-domains phase separate into a condensed phase that forms the selective permeability barrier of nuclear-pore-complexes (NPCs). Sequence complexity stemming from the lengths (~600 residues) and sequence heterogeneity of native FG domains has so far limited nanoscopic insight. Here we overcame the challenge by utilizing a perfectly repetitive GLFG_{52x12} peptide as a model FG domain, dramatically reducing chemical complexity and thus enabling measurements down to amino-acid-level resolution using a combination of solution and magic-angle spinning NMR spectroscopy methods. Several further insights were gained through microscopic observations: 1) Although phase-separated FG repeats appear hydrogel-like, the protein remains nanoscopically mobile (ns timescale mobility of the peptide chain) and disordered, lacking secondary structure in both hydrogel and solution states. 2) Increasing salt concentration not only enhanced phase separation, but also slowed down the residue-specific backbone dynamics in the hydrogel phase. 3) The change of chemical shifts/ backbone dynamics upon phase separation suggest contacts involving phenylalanine residues, which were previously shown to be essential for phase separation. However, profound changes of hydrophobic leucine were also observed, and eliminating all leucine residues in the presence of phenylalanine strongly disfavored phase separation. Consistently, we observed a lower critical solution temperature (LCST) behavior of the protein, where increasing temperatures enhanced phase separation, suggesting that phase separation is driven not by enthalpic aromatic contacts, but by entropy. We constructed a phase diagram for the GLFG_{52x12} peptide under varying salt conditions from which we could quantitatively determine the enthalpy and entropy of phase separation.

Table of Contents

1	Structure, gating, and ligand binding of the human voltage-dependent anion channel 1 in a lipid bilayer	12
1.1	Introduction	12
1.1.1	Mitochondria.....	12
1.1.2	The voltage-dependent anion channel	13
1.1.3	Solid-state NMR spectroscopy	27
1.1.4	Aims and outline	31
1.2	Materials and methods	32
1.2.1	Sample preparation	32
1.2.2	NMR spectroscopy.....	33
1.3	Results	42
1.3.1	Optimization of VDAC sample preparation	42
1.3.2	Assignment of E73VC127SC232A-hVDAC1.....	44
1.3.3	Determination of hVDAC1's topology in a lipid bilayer	52
1.3.4	Influence of the membrane environment on VDAC.....	59
1.3.5	VDAC's interaction with G3139	63
1.3.6	Investigation of VDAC's closed state through point mutations	65
1.4	Discussion	70
1.4.1	Comparison of detergent and membrane structures.....	70
1.4.2	VDAC assignments.....	70
1.4.3	VDAC's structure in a lipid bilayer	74
1.4.4	VDAC's mobility and gating behavior	80
1.4.5	Cholesterol binding to VDAC	82
1.4.6	Comparison of blockage mechanisms	83
1.5	Conclusion and outlook.....	84
1.6	References	86
2	Probing membrane protein insertion into lipid bilayers by solid-state NMR.....	101
2.1	Introduction	101
2.2	Materials and methods	106
2.3	Results and discussion.....	107
2.4	Conclusion.....	119

2.5	References	120
3	Correcting for magnetic field drift in MAS NMR	125
3.1	Introduction	125
3.2	Results	127
3.2.1	Linear phase correction of the FID	127
3.2.2	Application of drift correction	129
3.3	Conclusions	133
3.4	References	134
4	Investigating the phase separation and dynamics of Nup98-derived FG-domains	136
4.1	Introduction	136
4.1.1	Nuclear transport.....	137
4.1.2	FG-Nucleoporins.....	142
4.1.3	Models for selectivity	144
4.1.4	Phase separation in biological systems	148
4.1.5	Aims and outline	153
4.2	Materials and methods	154
4.2.1	Recombinant protein expression and purification	154
4.2.2	FG particles preparation for permeation assays and confocal laser scanning microscopy	156
4.2.3	Analysis of phase-separation by centrifugation (centrifugation assay)	156
4.2.4	Analysis of phase-separation by Dynamic Light Scattering (DLS).....	157
4.2.5	Fluorescence recovery after photobleaching (FRAP) and estimation of diffusion coefficients	157
4.2.6	NMR measurements.....	158
4.3	Results	163
4.3.1	Selectivity and spectroscopy of the <i>Tetrahymena thermophila</i> MacNup98A (TtMacNup98A) hydrogel	163
4.3.2	Simplification of spectra – design of perfectly repetitive sequences.....	166
4.3.3	High resolution ¹ H- ¹⁵ N correlation spectra and backbone assignment of the prf.GLFG _{52x12}	168
4.3.4	Nanosecond-timescale dynamics of the prf.GLFG _{52x12} hydrogel.....	172
4.3.5	Leucine’s role in phase separation	175
4.3.6	Effect of salt on phase separation and the mobility of the gel	182

4.3.7	Temperature-dependent phase separation of the prf.GLFG _{52x12}	184
4.4	Discussion	190
4.4.1	Native conformation of FG-domains	190
4.4.2	High degree of mobility in the hydrogel phase.....	191
4.4.3	Hydrophobic interactions in phase separation	192
4.4.4	Entropy-driven phase separation of FG-domains	197
4.4.5	FG-domains form a selective, mobile, hydrophobic phase.....	199
4.5	Conclusions and perspectives.....	201
4.6	References	203
5	Appendix.....	216
5.1	Chapter 1: Investigation of VDAC structure, gating, and interactions in a lipid bilayer 216	
5.1.1	Assignment of E73VC127AC232S-hVDAC1 in a DMPC lipid bilayer	216
5.1.2	Intensity values for VDAC mutants and ligand bound samples	222
5.2	Chapter 2: Probing insertion of membrane proteins into lipid bilayers	227
5.2.1	NMR pulse sequences.....	227
5.2.2	Supplementary figures	239
5.3	Chapter 3: Compensating the effect of linear field drift	241
5.3.1	Script for compensating linear field drift.....	241
5.3.2	Script for concatenation of 3D and 4D experiments.....	267
5.4	Chapter 4: Phase separation behavior of FG-domains	271
5.4.1	Amino acid sequences of engineered FG domains and reference wild-type MacNup98A domains:	271
5.4.2	Proteins and corresponding bacterial expression constructs used in this study	276
5.4.3	Composition of M9 medium with isotope labels for bacterial culture	277
5.4.4	Hydrophobic indices of FG-domains.....	278
5.4.5	Pulse sequences used in MAS NMR experiments.....	279

1 Structure, gating, and ligand binding of the human voltage-dependent anion channel 1 in a lipid bilayer

This chapter is based on the following publications, and contains identical text passages:

Eszter E. Najbauer, Stefan Becker, Karin Giller, Markus Zweckstetter, Adam Lange, Claudia Steinem, Bert L. de Groot, Christian Griesinger, Loren B. Andreas. Structure, gating and interactions of the voltage-dependent anion channel, *Eur Biophys J*, 2021, 50:159-172.

Eszter E. Najbauer, Kumar Tekwani Movellan, Karin Giller, Roland Benz, Stefan Becker, Christian Griesinger, Loren B. Andreas. Structure and Gating Behavior of the Human Integral Membrane Protein VDAC1 in a Lipid Bilayer, *J Am Chem Soc*, 2022, *in press*, DOI: <https://doi.org/10.1021/jacs.1c09848>

1.1 Introduction

1.1.1 Mitochondria

Mitochondria play a central role in eukaryotic cells, and are involved in a plethora of functions, including energy production, regulation of metabolism, and participation in signaling pathways. Mitochondria are generally thought to be of endosymbiotic origin (Martin and Mentel, 2010), which explains the existence of mitochondrial DNA and a delimiting membrane containing β -barrel integral membrane proteins. The inner- and outer mitochondrial membranes (IMM and OMM) separate the matrix of mitochondria from the cytosol, while the proteins and protein complexes embedded in these carry out oxidative phosphorylation, and ensure communication between the mitochondrion and the cell.

Cells are mostly powered by the ATP produced in mitochondria through oxidative phosphorylation (Alberts et al., 2002). In this process electrons generated from NADH are passed along by a series of respiratory enzyme complexes located in the inner mitochondrial membrane, and with the resulting energy protons are pumped across the membrane. The arising proton gradient is used as an energy source by ATP synthase to produce ATP from ADP and phosphate (Saraste, 1999). The continuous flow of ATP, ADP, small molecules, and proteins between mitochondria and the cytoplasm is ensured by integral membrane proteins in both the IMM and OMM. In the IMM, the ATP/ADP carrier (Duce and Vignais, 1965; Pfaff et al., 1965) is responsible for the exchange of nucleotides, while in the OMM it is the voltage-dependent anion channel (VDAC), classified as a porin, which allows the exchange of metabolites and ions across the OMM. There are two other large protein complexes in the OMM, the sorting and assembly machinery (SAM) (Wiedemann et al., 2003) and mitochondrial distribution and morphology (Mdm) complex. In addition, a supercomplex spanning the intermembrane space and formed by translocases of the outer (TOM) and inner (TIM) membrane (Chacinska et al., 2003) mediates translocation of all synthesized proteins from the cytoplasm to the matrix of the mitochondrion.

1.1.2 The voltage-dependent anion channel

The voltage-dependent anion channel (VDAC) is the most abundant protein in the outer mitochondrial membrane, covering up to 80% of membrane surface area in high density regions (Goncalves et al., 2007). The protein was first described in 1976 when it was isolated from mitochondria and incorporated into lipid bilayers (Schein et al., 1976). In membranes, the ~30 kDa (Zalman et al., 1980; Mannella, 1982; Roos et al., 1982) protein forms 2D crystalline arrays of aqueous pores about 3 nm in diameter (Mannella, 1982; Benz, 1994). VDAC is responsible for the permeability of the OMM to small molecules, and though the channels do not have a sharp exclusion limit, they are generally permeable to molecules up to 3-6 kDa in size (Zalman et al., 1980; Benz, 1994).

VDACs are involved in a plethora of cellular functions. The channel is the main conduit for ATP and ADP flow between the cytosol and mitochondria and controls Ca^{2+} homeostasis in mitochondria. It has been suggested that VDAC oligomerization might lead to the formation of a

mega-pore mediating the release of cytochrome *c* and other pro-apoptotic factors such as hexokinase, thus playing a key role in apoptosis (Zalk et al., 2005; Ben-Hail and Shoshan-Barmatz, 2016). So far, however, there is no high-resolution structural data proving the existence of such a molecular species. Recently, based on VDAC's interaction with mitoNEET, the channel's involvement in ferroptosis has also been suggested (Lipper et al., 2019). As the binding partner of a multitude of proteins, VDAC has been implicated in various diseases, such as Alzheimer's disease, Parkinson's disease and cancer (Caterino et al., 2017).

1.1.2.1 VDAC electrophysiology

VDAC derives its name from its characteristics displayed in reconstituted form under applied voltage, exhibiting a voltage-dependent conductance and selectivity (Schein et al., 1976). At voltages below ± 20 mV, the channels are in an anion selective (2:1 for equally mobile Cl^-/K^+) open state, exhibiting an average conductance of 4 nS in 1 M KCl (Benz, 1994). In this conductive state, VDAC is also permeable to large anions, including ATP. At increasing membrane potentials of $> \pm 40$ mV (Schein et al., 1976), conductance drops to about half (Benz, 1994), accompanied by a potential decrease in diameter (Colombini et al., 1987; Mannella and Guo, 1990). In this less conductive state, the channel remains permeable to small ions, with a moderate preference towards small cations (Benz and Brdiczka, 1992; Benz, 1994; Hodge and Colombini, 1997), but no longer permeable to ATP (Rostovtseva and Colombini, 1996, 1997). Although it is widely accepted that VDAC's open state is anion-selective, a cation-selective open state of the channel has also been observed (Pavlov et al., 2005).

There are two ways to illustrate the voltage dependency of VDAC in single channel recordings. First, the open probability P_o of a single VDAC as a function of the applied transmembrane potential U can be calculated from the corresponding current traces (**Figure 1-1A**). As shown in Figure 1B, the change in open probability as a function of applied voltage is very minor (gray curve), if all current traces are taken into account (Briones et al., 2016). If only those current traces are included, where the channel closes and re-opens within one voltage wave, as proposed by others (Teijido et al., 2012), the open probability drops below 80 % at $U = \pm 60$ mV (black curve). Second, the steady state conductance of the VDAC channel (G) relative to its conductance at low voltage (G_0) can be plotted (**Figure 1-1C**) (Noskov et al., 2016). While the open channel shows a

conductance of $G_0 = 4 \text{ nS}$, the conductance decreases in a stepwise fashion to values below 3 nS , referred to as ‘closed’ states, which are typically around 2 nS (Mertins et al., 2012), if the transmembrane potential exceeds $\pm 30 \text{ mV}$. In the example shown in **Figure 1-1A**, $G/G_0 = 1.6 \text{ nS}/4.0 \text{ nS} = 0.4$. The exact shape of the G/G_0 curve is dependent upon details of the preparation, including lipid and buffer conditions.

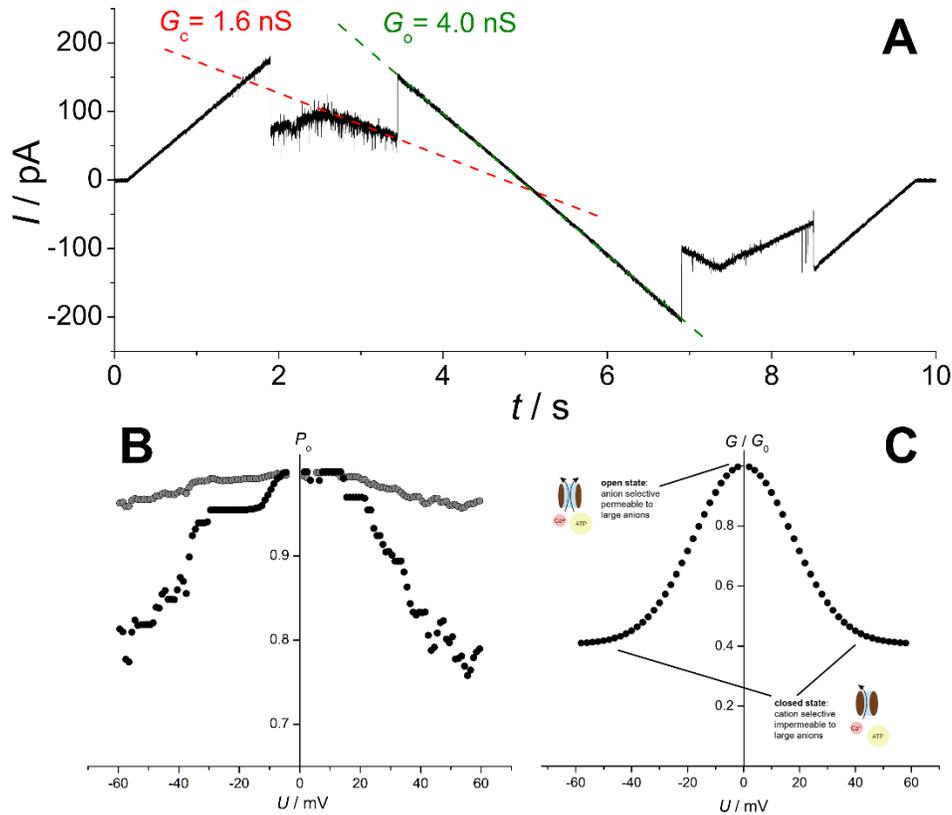


Figure 1-1: Characterization of VDAC by electrophysiology A) Representative current trace of a single hVDAC1 channel. The slopes show the difference between the open state conductance of $G_o = 4.0 \text{ nS}$ (green dotted line) and a ‘closed’ state conductance of $G_c = 1.6 \text{ nS}$ (red dotted line). B) Open probability P_o of a single hVDAC1 channel. If the open probability is calculated from all consecutively recorded voltage waves ($N = 185$), the grey curve is obtained. Considering only those traces, where the channel re-opened (18 % of all traces) lowers the open probability at higher potentials (black dots). The channel was reconstituted into a solvent-free membrane composed of DPhPC/cholesterol (9:1) bathed in 1 M KCl , 1 mM CaCl_2 , 5 mM HEPES , $\text{pH } 7.4$. 100 mHz symmetrical triangular voltage waves with amplitudes of $\pm 60 \text{ mV}$ were applied. C) Idealized voltage dependence of the steady state conductance of the VDAC channel (G) relative to the conductance at low voltage (G_0). Figures A) and B) were kindly provided by Prof. Claudia Steinem.

VDAC channels from distantly related eukaryotes have remarkably conserved biophysical properties, highlighting their universal importance (Blachly-Dyson and Forte, 2001; Saccone et al., 2003). In mammalian cells, VDAC is expressed in three isoforms. Interestingly, it is widely accepted that evolutionarily VDAC1 is the most recent of the three isoforms, VDAC3 having emerged first (Saccone et al., 2003). Both VDAC1 (Xu and Colombini, 1996; Schleiff et al., 1999) and VDAC2 insert into membranes readily, are anion selective, and have similar voltage dependence (Menzel et al., 2009), while VDAC3 seems to have a low propensity for membrane insertion and does not gate well, even at high membrane potentials of ± 80 mV (Xu et al., 1999). This suggests that VDAC3 might have a function other than altering the ion- and metabolite flow through the OMM. The three isoforms of mammalian VDAC are expressed ubiquitously, although at varying levels in different tissues, for example both VDAC2 and 3 have high expression levels in sperm outer dense fibers (Hinsch et al., 2004). Of the three isoforms, VDAC1 is by far the most abundant (Messina et al., 2012), and is also best described in literature.

1.1.2.2 VDAC structure

Initial predictions of VDAC structure based on biochemical data from mutagenesis studies, and antibody assays, as well as computational approaches all pointed to the formation of a β -barrel, but with varying number of β -strands. Estimates ranged from 12 strands (Blachly-Dyson et al., 1989; Blachly-Dyson et al., 1990), 13 (Song et al., 1998b), 16 (Depinto et al., 1991; Rauch and Moran, 1994; Casadio et al., 2002), 18 (Al Bitar et al., 2003), and up to the value eventually determined in the NMR and X-ray structures, of 19 (Forte et al., 1987) strands. Though initially debated (Forte et al., 1987), existence of an amphipathic N-terminal α -helix was widely accepted and later proven experimentally (De Pinto et al., 2007) in both detergents and lipids (Shanmugavadivu et al., 2007), however it remained unclear whether the helix forms part of the barrel wall (Blachly-Dyson et al., 1989; Blachly-Dyson et al., 1990), or is exposed to the water phase (De Pinto and Palmieri, 1992), perhaps extending away from the pore lumen (Guo et al., 1995).

Figure 1-2 shows ribbon representations of atomic resolution VDAC structures rendered with Chimera (Pettersen et al., 2004). The first of these structures were reported in 2008 using VDAC preparations in a detergent environment, using NMR spectroscopy (Hiller et al., 2008), the second

also in micelles using a combination of NMR and X-ray crystallography (Bayrhuber et al., 2008), and the third in bicelles instead of micelles using X-ray crystallography alone (Ujwal et al., 2008). All three structures unequivocally showed a 19-stranded β -barrel, with the N-terminal segment positioned inside the pore. The latter two structures resolved the N-terminal helix positioned in contact with the barrel wall. The odd number of β -strands in VDAC was quite surprising, the parallel beta sheet interaction between the first and the last strands (β 1 and β 19) were never before observed in other integral membrane proteins.

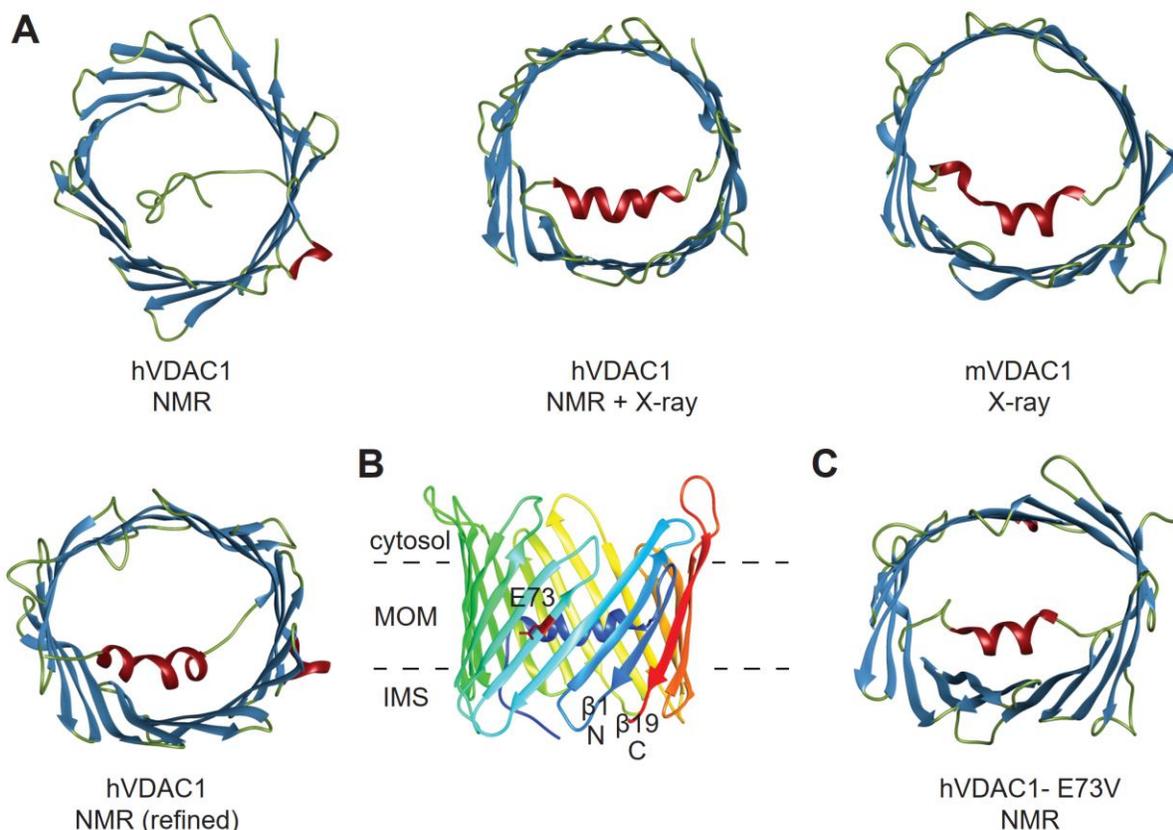


Figure 1-2. The structure of VDAC1. (A) The structure of human VDAC1 as determined by NMR spectroscopy (PDB code: 2k4t) (Hiller et al., 2008), a combination of NMR spectroscopy and X-ray crystallography (PDB code: 2jk4) (Bayrhuber et al., 2008), and the structure of mouse VDAC1 as determined by X-ray crystallography (PDB code: 3emn) (Ujwal et al., 2008). A refined NMR structure of hVDAC1 is shown bottom left (PDB code: 6tiq) (Bohm et al., 2020). The structures are colored according to secondary structural elements (helix: red, β -strand: blue, unstructured region: olive). For the NMR structures the lowest energy structure of the ensemble was selected. (B) The structure of hVDAC in side view (PDB code: 2jk4). The dashed lines indicate the insertion of the channel into the mitochondrial outer membrane. The structure is colored from blue at the N-terminus to red at the C-terminus. The parallel orientation of the first (β 1, blue) and last (β 19, red) β -strands is clearly visible. The charged sidechain of E73 pointing into the membrane is colored in red. (C) High resolution structure of E73V-hVDAC1 as determined by NMR spectroscopy (PDB code: 5jdp) (Jaremko et al., 2016). The structure is colored according to its secondary structural

elements, as described in (A). Of the structural ensemble, the structure with the lowest energy has been selected for display.

Astonishingly, a similar topology with 19 β -strands and a helix in the opening of the pore is seen in the translocase of the outer membrane (TOM40), also found recently in the OMM (Araiso et al., 2019; Tucker and Park, 2019). Slight differences in the structure occur for the N-terminal helix resolved in the NMR-X-ray hybrid structure (Bayrhuber et al., 2008) and in the structure solved exclusively by X-ray crystallography (Ujwal et al., 2008). Bayrhuber et al. found one helix spanning residues 7-17, while Ujwal et al. resolved two helices, comprised of residues 6-9 and 12-20, with L10 and G11 forming a kink between the two. In both cases, the helical N-terminus runs along the barrel wall, its position stabilized by hydrogen bonds between the N-terminus and the barrel. The residues forming the barrel have an alternating pattern of hydrophilic and hydrophobic sidechains pointing into the aqueous pore lumen and the hydrophobic environment, respectively. On the exterior, this pattern is broken only by residue E73 in β 4, for which the negatively charged glutamate sidechain faces the hydrophobic environment.

An NMR study of E73V hVDAC1 used relaxation data to determine a high-resolution structure of the protein (Jaremko et al., 2016). The E73V mutant gives spectra of higher quality, allowing more extensive assignments than wild-type VDAC (Bohm et al., 2020), but showing identical gating behavior (Queralt-Martin et al., 2019). This structure resolved two helices in the N-terminus, unambiguously showing that it was not elevated N-terminal dynamics, but rather a lack of resolution that resulted in the discrepancy of the conformation of the N-terminus in the three initial structures from 2008. Interestingly, the barrel showed a distinct elliptic deformation that was not observed in the crystal structure of the mouse variant in bicelles. It is unclear whether the E73V mutation stabilizes an existing conformation of wild-type VDAC, resulting in a predominantly elliptic barrel shape, or it is the pressure exerted by micelles that deforms the barrel, while bicelles do not exert this pressure.

The structure of VDAC2 is very similar to that of VDAC1, as shown by X-ray crystallography and solid-state NMR spectroscopy (Schredelseker et al., 2014; Gattin et al., 2015), including nearly identical dynamic behavior and conformational homogeneity (Eddy et al., 2019). Although to date,

VDAC3 has not been structurally characterized, the sample preparation and spectroscopic methods described by Eddy et al. will likely be applicable to this isoform as well (Eddy et al., 2019).

1.1.2.3 Structural studies of VDAC in a lipid bilayer

Studying VDAC in a lipid bilayer as opposed to micelles or bicelles has always been a rather challenging undertaking. Extensive efforts to determine the structure using tomography resulted in only 8.2 Å resolution of the pore (Dolder et al., 1999). The protein is also challenging to address using solid-state NMR due to the sheer size of the protein which results in a requirement for very high sample homogeneity. VDAC alone is still too small for high resolution characterization with single particle cryo-electron microscopy.

In liposome preparations, NMR data (Schneider et al., 2010) indicated that the N-terminus of human VDAC1 (hVDAC1) exists in a rigid and well-defined structure. Additionally, (^{13}C , ^{13}C) dipolar order parameters (S_{CC}) were measured for residues in the N-terminus as well as in other parts of the molecule using double-quantum (2Q) spectroscopy, a technique that is sensitive to dynamics on the pico- to millisecond timescale (Zachariae et al., 2012). Due to the considerable size of hVDAC1 and the large number of residues in β -sheet conformation, spectral overlap precluded identification of residue-specific order parameters for a large part of the molecule. To estimate overall mobility in the β -barrel, overlapping signals were analyzed to determine average order parameters. The data showed that, globally, the N-terminus is clearly not more flexible than the β -barrel on a sub-ms timescale, as might be expected if the N-terminus were primed to move under applied voltage (see voltage gating discussion below). Peak broadening or doubling, which would have indicated dynamics on slower timescales, was also not observed, further confirming the well-defined structure of the hVDAC1 N-terminus.

In the above liposome preparations, with the exception of three residues, the spectral quality was insufficient to resolve and assign barrel resonances. This emphasizes an extremely stringent requirement for preparation of homogeneous VDAC samples that may be influenced by VDAC's high propensity for oligomerization (Goncalves et al., 2007; Hoogenboom et al., 2007), resulting in a microscopically inhomogeneous sample, despite high sample purity.

Under certain conditions, in the presence of lipids, membrane proteins may form multilamellar crystalline arrays (2D crystals). Importantly, 2D crystals allow investigation of proteins in a native-like lipid bilayer environment, where they often retain full functionality, while the high microscopic order in these preparations has long been exploited by electron microscopy (Jap et al., 1990; Kuhlbrandt and Wang, 1991; Unger et al., 1997; Walz and Grigorieff, 1998), atomic force microscopy (Stahlberg et al., 2001), and solid-state NMR spectroscopy (Lewis et al., 1985; Hiller et al., 2005; Shastri et al., 2007) to gain atomic-level structural information.

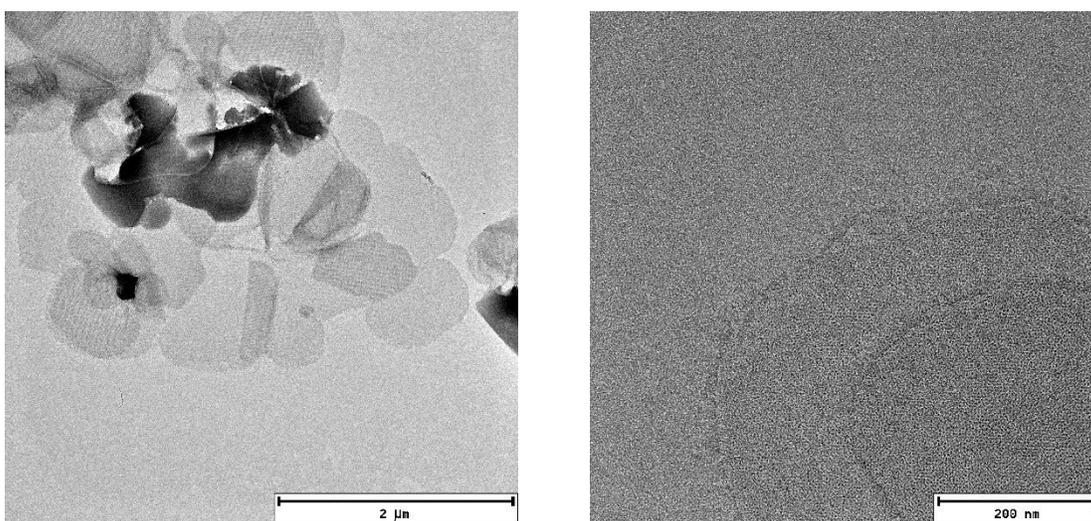


Figure 1-3: 2D crystals of VDAC. Negative stain electron microscopy images of VDAC 2D crystals at different resolutions. Horizontal bars indicate relative sizes. The lamellar structure of the crystals is clearly visible on both images. On the right, VDAC channels can be seen as small black dots.

At relatively low lipid-to-protein ratios, VDAC also forms 2D crystals (Dolder et al., 1999) of fully functional channels (**Figure 1-3**) (Eddy et al., 2012). In contrast to VDAC1 in lauryldimethylamine oxide (LDAO) micelles, the presence of cholesterol or the detergent Triton X-100 is not required for the formation of the crystalline arrays, or for the channel to be fully functional, as determined with electrophysiology. The structure of the protein remains unchanged upon 2D crystal formation in various lipids (Eddy et al., 2012).

The 2D crystalline preparations yield excellent quality solid-state NMR spectra (heteronuclear linewidths <0.5 ppm), and – perhaps due to a reduction of local mobility by tight packing – also facilitate the assignment of loop regions. Using carbon detection and several samples with different

isotopic labeling, it was possible to assign 88 residues, including most of the α -helix, as well as residues from β -strands 5, 6, 9, 13, 18, and 19 (Eddy et al., 2015b). The use of both uniformly ^{13}C , ^{15}N -labeled VDAC as well as several inverse labeled samples was key to reduce ambiguity for assignment of beta strands.

The structure of both high and low conductance states of VDAC has long been controversial. It has been argued that while published structures of the protein show the existence of a 19-stranded β -barrel, functional studies point to a barrel formed from 13 β -strands and an α -helix (Colombini, 2012). Secondary chemical shifts calculated from assignments in a lipid bilayer can give some insight into the secondary structure in its functional open state. To date, this information confirms the presence of several beta sheets matching those of the 19-stranded structure, yet information is incomplete due to the lack of solid-state NMR assignments for many strands. With the exception of a long-range contact between A14 C_β in the N-terminal helix and S193 C_β (Eddy et al., 2015a), the information available on the 3D structure of VDAC in a lipid bilayer is limited to chemical shifts, which does not define the β -sheet arrangement.

In order to mimic a lipid bilayer environment in solution, preparation of VDAC in nanodiscs has also been established that yield high resolution NMR spectra (Raschle et al., 2009), and later further improved by employing covalently circularized nanodiscs (Wagner et al., 2017). While these samples yield excellent quality 2D spectra, recording higher dimensionality experiments necessary for sequential assignment may be difficult due to fast transverse relaxation rates, resulting in low efficiency magnetization transfer, or a requirement for elevated temperatures.

1.1.2.4 VDAC gating mechanism

The mechanism of VDAC's voltage gating is not yet understood, though several models exist. Early electrophysiology studies on VDAC reconstituted in lipid bilayers suggested changes in the channel diameter upon gating, from ~ 3 nm to 1.8 nm (Zimmerberg and Parsegian, 1986; Colombini et al., 1987). A series of mutagenesis and biotinylation studies identified several residues throughout the protein sequence to influence gating behavior (Blachly-Dyson et al., 1990; Thomas et al., 1993; Song et al., 1998b), and this observation led to the proposal of a gating model, in which VDAC gating is accomplished by a complex rearrangement of the protein involving the

movement of a large, positively charged “voltage sensor” region of the protein out of the membrane, resulting in a smaller pore (Thomas et al., 1993; Song et al., 1998a).

After the publication of VDAC’s high resolution structure in 2008 (Figure 2), several models for gating were proposed based on the fact that the N-terminal α -helix is located inside the pore, by the barrel wall, at a key position to regulate gating. A possibility suggested by Hiller et al. was that the N-terminus adopts different conformations upon gating (Hiller et al., 2008), possibly including unwinding of the helix (**Figure 1-4C**) (Zachariae et al., 2012). Another suggestion was that upon gating the entire helix moves to the center of the channel, thus obstructing the flux of ions and metabolites (**Figure 1-4A**). In this case modulators, such as NADH would close the channel by binding to the interaction site between the helix and the barrel wall, disrupting the hydrogen bonding pattern and dislodging the helix (Ujwal et al., 2008). A recent study has shown however, that reduction of channel conductance upon NADH binding occurs through NADH sterically blocking the pore, during which the helix conformation is essentially unchanged. This is in contrast to one of the proposed mechanisms of channel closure upon voltage gating, the basis for which is increased mobility of the N-terminus, in particular the α 2 helix, as dynamics and electrophysiology measurements on cross-linked VDAC mutants suggest (Bohm et al., 2020).

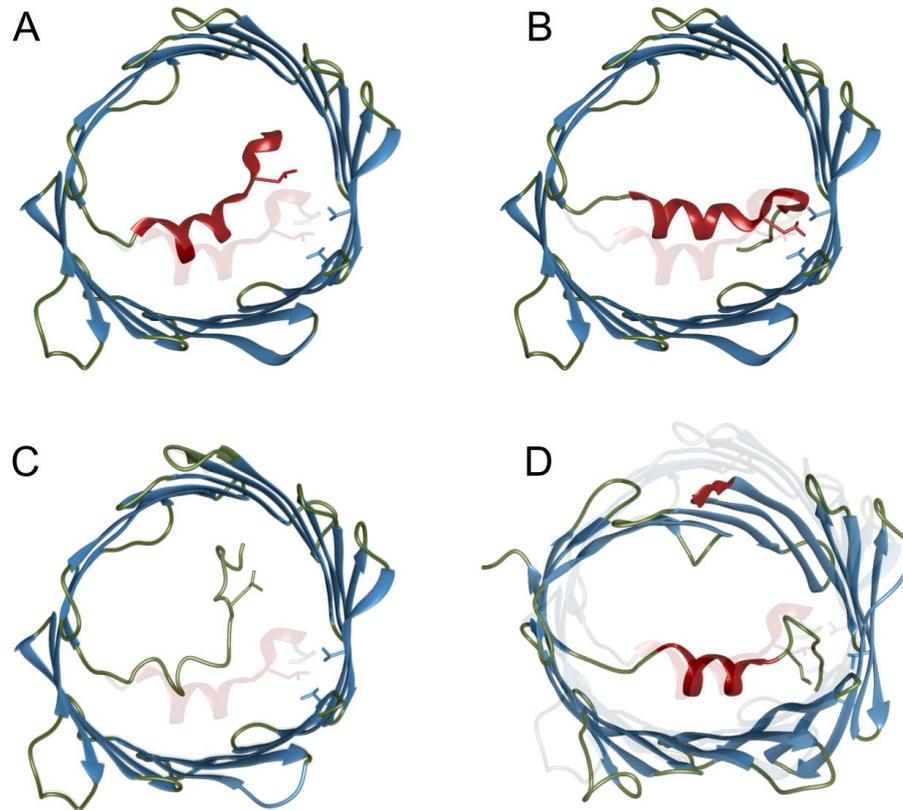


Figure 1-4: Models for VDAC's voltage gating mechanism. A) Voltage gating by movement of the helix to the center of the pore (Ujwal et al., 2008). B) Voltage gating through dislodging of the $\alpha 2$ -helix from the barrel wall (Hiller and Wagner, 2009). C) Unwinding of the helix upon gating (Zachariae et al., 2012). D) Voltage gating upon elliptic deformation the barrel without dislodging of the helix (Villinger et al., 2010; Zachariae et al., 2012). The depictions were generated by manual modification of the VDAC structures deposited under pdb ID 3emn, 5jdp and 2k4t.

These electrophysiology studies investigating cross-linked VDAC mutants yield further insight into the role of the N-terminus. Affixing the very end of the N-terminus results in loss of symmetric voltage response (V3C-K119C), while cross-linking the $\alpha 2$ helix to the barrel wall (A14C-S193C) locks the channel into a permanently open state potentially providing support for the gating mechanism illustrated in **Figure 1-4B** (Mertins et al., 2012). Notably, a cross-link between L10 in the kink between the two helices and the barrel wall (L10C-A170C) does not prevent the channel from closing under applied voltage, which suggests that the N-terminus does not completely move away from the barrel wall upon gating (Tejjido et al., 2012). This excludes the gating models illustrated in **Figure 1-4A** and C.

Electrophysiology studies have also shown that the voltage-dependence of VDAC gating is modulated by the lipid composition of the surrounding membrane (Rostovtseva et al., 2006). Because the N-terminal α -helix is positioned inside the pore and does not directly contact the membrane, additional regions within the β -barrel are likely to contribute to the voltage sensitivity of VDAC. A combination of NMR dynamics studies and molecular dynamics (MD) simulations performed by Villinger et al. showed that the N-terminal six β -strands, which contain several residues important for gating undergo μ s-ms motion (Villinger et al., 2010). The motions result in deformation of the barrel including the elliptical barrel of the E73V mutant (**Figure 1-4D**), which can influence the diffusion of small ions and metabolites through the channel. MD simulations in combination with solid-state NMR spectroscopy further showed the occurrence of barrel deformations and concomitantly a decrease of channel conductance (Briones et al., 2016).

These barrel deformations, leading almost to the faltering of the barrel, are most drastically observed upon removal of the N-terminal helix (Schneider et al., 2010). MD simulations predict significant displacement of charged helical residues in the N-terminus, especially in the K12-K20 region, as well as residues in β -sheets 1, 3, 7 and residues 104-107 and 266-268 located in cytoplasmic loops (Briones et al., 2016). At least partial detachment of the α -helix from the barrel wall might therefore result in faltering of the barrel. Displacement of barrel residues upon voltage gating was further supported by surface-enhanced infrared absorption and electrochemical impedance spectroscopy (Kozuch et al., 2014). The combined data support a gating model in which transport of small molecules and metabolites is regulated via elliptic deformations of the VDAC β -barrel coupled to detachment of the C-terminal part of the N-terminal α -helix

There are still many controversies regarding the mechanism of gating. While a recent publication finds E73 is not involved in the gating (Queralt-Martin et al., 2019), NMR spectroscopy finds a clear reduction of barrel motion in locations that are involved in the barrel faltering, and older electrophysiological measurements on the E73Q mutant confirm reduced voltage dependence (Zaid et al., 2005). Reconciling the partially contradictory information regarding the gating behavior of VDAC, and understanding the role of the N-terminus will likely require structural studies in a native-like detergent-free environment, where detergents cannot disrupt helix-barrel contacts, as well as further MD simulations.

1.1.2.5 VDAC interactions

VDAC regulates cellular metabolism and the exchange of ATP, ADP, phosphate, Ca^{2+} and various small molecules between mitochondria and the cytoplasm not only through its gating behavior but also through its interactions with a large variety of molecules, making it an important checkpoint in controlling various cellular processes.

Mitochondrial function and ATP production is largely dependent on mitochondrial Ca^{2+} concentration (Gunter and Sheu, 2009). As the primary avenue for the transport of Ca^{2+} and ATP through the mitochondrial outer membrane, VDAC is a key player in controlling mitochondrial Ca^{2+} - homeostasis and cellular function (Shoshan-Barmatz et al., 2018a). VDAC binds (Gincel et al., 2001) and allows flow of Ca^{2+} in both its open and closed states (Tan and Colombini, 2007), and this binding has been shown to influence molecular plasticity, possibly altering the channel's gating behavior (Ge et al., 2016). ATP molecules reversibly bind to VDAC as they pass through the channel (Florke et al., 1994), and a low-affinity interaction site on the N-terminal helix (Yehezkel et al., 2007) and the adjacent β -barrel residues (Villinger et al., 2014), as well as the mechanism for permeation have been identified (Choudhary et al., 2014). Other nucleotides, including GTP and UTP share a common binding site with ATP, and partially overlap with β -NADH binding sites previously identified by NMR spectroscopy (Hiller et al., 2008; Villinger et al., 2014).

VDAC also binds a variety of other small molecules, whose function is still not fully understood. Cholesterol has been shown to co-purify with VDAC at a 5:1 ratio (De Pinto et al., 1989), and cholesterol binding has been suggested to be necessary for VDAC to achieve full function (Popp et al., 1995), though it does not seem to influence basic channel properties (Queralt-Martin et al., 2019). At least two potential cholesterol binding sites in detergent micelles have been identified using NMR spectroscopy (Hiller et al., 2008), and a docking study found these two patches to comprise five distinct binding sites (Weiser et al., 2014). Using photo-affinity labeling and mass spectrometry, cholesterol was found to bind to four of these sites, as well as a site near E73, which could potentially have implications for channel mobility, and thus, gating. Other sterols, such as the neurosteroid allopregnanolone also largely share cholesterol's binding sites (Cheng et al., 2019).

VDAC's opening/closing is an important checkpoint in cellular metabolism, and is regulated through a large variety of protein-protein interactions as reviewed recently (Caterino et al., 2017). VDAC exhibits pro-apoptotic activity by interacting with several proteins from the apoptosis-related Bcl2 protein family. Through its interaction with Bax and Bak, VDAC has been suggested to participate in the release of cytochrome *c* into the cytosol and the activation of the apoptotic cascade (Shimizu et al., 2000; Tsujimoto and Shimizu, 2000), while the anti-apoptotic Bcl-x_L protein closes the pore by direct binding (Shimizu et al., 1999). Hexokinase 1 binding to VDAC may suppress apoptosis by modulating VDAC activity and controlling the channel switching between off and on states (Dubey et al., 2016; Caterino et al., 2017). VDAC's inhibition by tubulin could influence ATP trafficking, and even induce a switch, known as the Warburg effect, between oxidative phosphorylation and glycolysis in cancer cells (Maldonado et al., 2013).

VDAC is a promising drug target in the therapy of neurodegenerative and cardiovascular diseases associated with mitochondrial dysfunction, as well as cancer as recently reviewed (Magri et al., 2018). For cancer therapy, drug candidates rely mainly on inducing apoptosis by direct blockage of the channel or through promoting VDAC oligomerization, however many of these drug candidates suffer from lack of selectivity, difficult delivery or high toxicity. A particularly promising molecule that has entered clinical trials for leukemia (O'Brien et al., 2007) and breast cancer (Moulder et al., 2008) is the 18-mer phosphorothioate oligonucleotide G3139 (also known as Genasense or oblimersen). It was developed as an antisense oligonucleotide, complementary to the first six bases of the anti-apoptotic protein Bcl2- mRNA, however it has been shown to selectively bind to VDAC, blocking the channel (Lai et al., 2006; Tan et al., 2007). The mechanism of blockage is unknown, though the kinetics indicate at least a partial entry into the pore.

VDAC potentially forms important interactions with other integral membrane proteins collocated in the OMM. One such protein, the translocator protein (TSPO) has received attention as a marker for oxidative stress and inflammation. A series of PET ligands has been developed that bind TSPO with nanomolar affinity (Veenman et al., 2016), one of which was used to stabilize the structure for determination in detergent micelles (Jaremko et al., 2014). In lipid bilayers, it was possible to probe the influence of cholesterol and observe the equilibrium between monomer and dimeric forms of the protein (Jaipuria et al., 2017). Both VDAC (Shoshan-Barmatz et al., 2018b) and TSPO (Repalli, 2014) are upregulated in Alzheimer's, a serious neurodegenerative disease. While there

is evidence for VDAC-TSPO interaction in mouse models of Alzheimer's (Oakley et al., 2006), to date the details of this interaction remain unknown. Due to the apparent importance of these two proteins in apoptosis, their interactions have been proposed as a potential drug target for future development (Veenman and Gavish, 2006; Veenman et al., 2007).

1.1.3 Solid-state NMR spectroscopy

In the last three decades, structural biology has undergone a true revolution, with the number of available macromolecular structures deposited in the Protein Data Bank growing exponentially, and the number of new structures deposited each year now exceeding 10000. The vast majority of these structures was determined by either X-ray crystallography (~89%), NMR spectroscopy (~8%), and the recently emerging cryo-electron microscopy (cryo EM, ~3%). As a result of advances in hardware (high-field magnets, sensitive probes) and methods (pulse sequences (Pervushin et al., 1997; Tugarinov and Kay, 2003), labeling schemes), NMR spectroscopy has proven a powerful technique for determining atomic-resolution structure and dynamics of proteins up to 1 MDa. Contrary to X-ray crystallography and cryo-EM, NMR spectroscopy does not need protein crystals, or cryogenic conditions for measurement, but can investigate protein structure and functional dynamics in a native-like environment. Solid-state NMR (ssNMR) complements other techniques especially well when it comes to studying insoluble proteins such as amyloid fibrils, membrane proteins, protein-ligand complexes, and even large molecular assemblies.

1.1.3.1 Solid-state NMR spectroscopy and proton detection

Nuclear magnetic resonance (NMR) spectroscopy relies on the phenomenon that when non-zero spin atomic nuclei are placed into a static magnetic field, their energy levels split up, and transitions between these levels can occur upon interaction with a radiofrequency magnetic field. These transition frequencies (determined by Larmor frequencies and chemical shifts) are characteristic not only of the type of the nucleus (Larmor frequency) but also of the distribution of electrons around it (chemical shift). Chemical shifts thus give insight into the chemical environment of nuclei, and allow identification of each nucleus in a molecule. The splitting pattern of the energy levels (and thus the splitting pattern of each signal in a spectrum) is determined by direct (through-space) dipolar or indirect (through-bond) J-coupling interactions of nuclei with each other, as well

as the interaction of the nuclear quadrupole moment with electric field gradients (possible, but rarely exploited in biological samples). These effects may be entirely independent of orientation compared to an external magnetic field (J-coupling), have some dependency (chemical shift, with the isotropic chemical shift δ and chemical shift anisotropy (CSA) components, due to non-spherical distribution of electrons around the nucleus), or be entirely determined by orientation (dipolar couplings, determined by the mutual orientation of two nuclei relative to each other). Although the orientation-dependent interactions may offer valuable steric information in some cases, due to their large magnitude (several kHz), they may severely broaden peaks and obscure spectral features.

In solution, all orientation-dependent interactions are averaged by the fast isotropic tumbling of molecules, however in solid samples this averaging must be done artificially. This is achieved by spinning the sample around an axis inclined at 54.74° (the magic angle) with respect to the static magnetic field (magic angle spinning, MAS). Interactions much smaller in magnitude than the spinning frequency will be averaged out completely. For biological samples at moderate spinning frequencies of up to ~ 25 kHz, these typically include CSA, as well as homonuclear dipolar interactions of low- γ ^{13}C and ^{15}N nuclei with other nuclei. Homonuclear dipolar interactions between high- γ ^1H -s tend to be more than 100 kHz, and their averaging is incomplete under these conditions.

For this reason, ssNMR has traditionally mostly relied on the detection of heteronuclei (^{13}C , ^{15}N). However, these experiments require large amounts (tens of mg-s) of sample, and (due to the low gyromagnetic ratio (γ) of the detected nucleus) are relatively insensitive. To average the dipolar coupling of ^{13}C or ^{15}N with ^1H -s, high power radiofrequency decoupling is required, which leads to significant sample and coil heating.

In the past decade, advances in hardware (fast MAS probes) and sample preparation (dilution of a dense proton network by deuteration and back-exchange) have made it possible to detect (high- γ) ^1H -s, which has several advantages over the detection of heteronuclear frequencies. All else being equal, the detection of protons boosts sensitivity by factors of $(\gamma_{^1\text{H}}/\gamma_{^{13}\text{C}})^{3/2} = 8$ and $(\gamma_{^1\text{H}}/\gamma_{^{15}\text{N}})^{3/2} = 31$ compared to ^{13}C - and ^{15}N -detected experiments (Paulson et al., 2003), respectively, significantly reducing the amount of sample needed, while still allowing for higher sensitivity. In addition, protons as a third nucleus beside ^{13}C and ^{15}N decrease the degeneracy of

chemical shifts, a considerable bottleneck in the assignment of large proteins. The absence of high power proton decoupling during the pulse sequences allows using shorter recycle delays (fast repetition of scans), which are now determined by the relatively short longitudinal relaxation time of ^1H -s rather than the duty cycle of the probe. Proton detected experiments in the solid-state are analogous in concept to solution NMR experiments used for sequential assignment.

Until recently however, at the available moderate spinning frequencies (up to 60 kHz), dilution of the dense proton network was necessary to measure well-resolved proton detected spectra (Zheng et al., 1993; Reif et al., 2001; Paulson et al., 2003; Chevelkov et al., 2006; Andreas et al., 2015a). In the majority of cases this is achieved by perdeuteration of the sample, followed by the back-exchange of the mobile (mostly amide) protons (Akbey et al., 2010; Lewandowski et al., 2011; Fricke et al., 2017). To gain access to aliphatic protons, various labeling schemes have been developed, selectively introducing alpha- or sidechain protons through the use of labeled metabolic precursors (Tugarinov and Kay, 2003; Kainosho et al., 2006; Isaacson et al., 2007; Linser et al., 2014; Movellan et al., 2019).

The higher resolution due to the ability to access proton chemical shifts upon ^1H detection at fast MAS, thus gaining another dimension also exploited in structure determination. At 24 kHz MAS, perdeuteration and ^1H - ^1H back-exchange in 25% H_2O -75% D_2O buffer, correlation between ^1H -s of up to 13 Å in distance were observed for microcrystalline chicken α -spectrin SH3 domain using ^1H - ^1H RFDR (Linser et al., 2011). At very fast MAS (111 kHz) the dipolar-based BASS-SD scheme could correlate ^1H -s 5-6 Å in distance in fully protonated microcrystalline GB1 (Jain et al., 2017). This distance limit enables the identification of cross- β -strand contacts between H^{N} -s (~ 3 Å), as well as i - $i+2$ (~ 4 - 4.5 Å), and even i - $i+3$ (~ 5 Å) contacts in α -helices.

Recently, probes capable of MAS frequencies of up to 111 kHz have become commercially available, with custom built probes capable of spinning up to 140 kHz (Lin et al., 2018; Penzel et al., 2019). Measurement in these probes requires < 1 mg of sample, and the fast spinning eliminates the need for deuteration, proton linewidths being as low as 99 Hz in microcrystalline ubiquitin at 126 kHz (Penzel et al., 2019). This makes it possible to easily access not only the protein backbone, but also sidechains without the use of special isotopic labeling schemes, which greatly facilitates structure determination. The efficiency of this approach has been demonstrated through the

assignment (Stanek et al., 2016; Lalli et al., 2017; Vasa et al., 2019) and structure determination of protein assemblies as large as a 28 kDa dimer in a 2.5 MDa viral capsid (Andreas et al., 2016).

1.1.4 Aims and outline

The goal of this chapter is to explore the structure and interactions of hVDAC1 in a lipid bilayer and to gain insight into the channel's gating mechanism.

In the first part of the chapter (Sections 1.3.1-1.3.2), we investigate sample quality and obtain extensive assignments for the protein using a combination automated and manual assignments. These serve as a basis for structural determination and further residue-specific investigations. In Section 1.3.3. we determine the protein's topology from dipolar recoupling and MTSL spin labeling experiments.

Next, we explore the influence of the membrane environment on VDAC, in particular the channel's interaction with cholesterol (Section 1.3.4). We also study VDAC's interaction with G3139, a phosphorothioate oligonucleotide previously shown to block the channel (Section 1.3.5.).

Lastly, by introducing two mutations to the linker between the N-terminal α -helix and the barrel, we gain access to the protein's closed state. Comparing peak intensities indicative of motion in the channel's open and closed states, we study the mechanism of channel closure (Section 1.3.6) and gain insight into dynamics.

1.2 Materials and methods

1.2.1 Sample preparation

1.2.1.1 Protein expression

The expression of ^2H , ^{13}C , ^{15}N -labeled and fully back-exchanged human VDAC1 with a C-terminal His₆-tag was based on the protocols of Malia and Wagner (Malia and Wagner, 2007) and Eddy et al. (Eddy et al., 2012), with several modifications. Briefly, the protein was expressed in minimal medium. Depending on the specific labeling scheme, the medium was supplemented with ^{15}N ammonium chloride and ^{13}C D-glucose as nitrogen and carbon source. For producing perdeuterated protein, the bacteria were stepwise adapted to 100% D₂O and the expression culture contained 100% D₂O. A 1 L shaking culture was grown at 37°C to an OD₆₀₀ of 0.7-0.8 and expression was induced with 1 mM IPTG. Cells were harvested 6-7 hours after induction. In case of perdeuterated cultures the cells were harvested 20 hours after induction. After cell lysis inclusion bodies were isolated, dissolved in denaturing buffer (8 M urea, 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 20 mM Imidazol, 0.5 mM PMSF, 5 mM DTT), loaded on a 10 ml Ni²⁺ agarose column (Macherey-Nagel) and purified by elution with the same buffer supplemented with 250 mM imidazole. Fractions containing VDAC1 protein were combined and dialyzed overnight against 4 L precipitation buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 5 mM DTT). The precipitated protein was dissolved 25 mM sodium phosphate pH 7.0, 6 M guanidinium hydrochloride, 100 mM NaCl, 1 mM EDTA, 5 mM DTT and further purified by denaturing gel filtration on a SD75 16/60 gel filtration column (GE Healthcare). Subsequently the protein concentration was adjusted to 5 mg/ml and then refolded by dropwise dilution into a 10-fold volume of 4°C refolding buffer (25 mM NaPi, pH 7.0, 100 mM NaCl, 1 mM EDTA, 1% LDAO (Anatrace), 5 mM DTT). The refolded protein was dialyzed against a 20-fold volume of 4°C cold 25 mM sodium phosphate, pH 7.0, 1 mM EDTA, 0.1% LDAO, 5 mM DTT), loaded on a 5 ml SP XL-cation exchange column and eluted with a 60 ml linear gradient to 1 M NaCl in the same buffer. Subsequently, folded VDAC1 protein was separated from misfolded or aggregated protein by a further gel filtration on a SD200 16/60 column equilibrated with 25 mM sodium phosphate pH 7.0, 1 mM EDTA, 150 mM NaCl, 0.1% LDAO, 5 mM DTT.

Importantly, no reducing agent was used at any stage for purifying cysteine-free mutants of VDAC1. For MTSL labeling of cysteine mutants of the protein, the pellet of the precipitation step (see above paragraph) was washed with cold dialysis buffer containing no reducing agent. All subsequent purification steps were performed in the absence of reducing agents. After dissolving the pellet in the denaturing gel filtration buffer, MTSL was added in 10-fold molar excess and the solution was incubated for 2 hours before the following gel filtration step. MTSL was once more added in 3-fold excess before refolding of VDAC1.

1.2.1.2 2D-crystallization of VDAC1

The fractions containing the folded protein were combined and concentrated to about 3 mg/ml. 2D crystals were then prepared according to Eddy et al (Eddy et al., 2012). Briefly, the purified protein was dialyzed overnight at 4 °C against 4 L of 50 mM Tris, pH 8.0, 0.6% (w/v) n-octyl-POE (Bachem), 5 mM DTT. Lyophilized DMPC was dissolved in cold 50 mM Tris pH 8.9, 1% n-octyl-POE to a final concentration of 0.5 mg/ml. After adjusting the concentration of the dialyzed VDAC1 protein to 1 mg/ml, equal volumes of the protein and the lipid solution were mixed and dialyzed at room temperature against 4 L of 10 mM MES, pH 6.5, 150 mM NaCl, 20 mM MgCl₂, and 4 mM DTT, with 4 exchanges of the dialysis buffer over the course of 2 days. After the last buffer change the sample was dialyzed for 6 more days at room temperature. During this time crystals appeared. The crystals were finally collected by ultracentrifugation and transferred to a solid-state NMR rotor.

The same protocol was used for preparing the G21V/G23V/E73V/C127A/C232A quintuple and the G21V/G23V/E73V/C127A/V143A/L150A/C232S septuple mutants. Site directed mutagenesis are described in detail in the thesis of Christian Schmidt (Schmidt, 2017).

1.2.2 NMR spectroscopy

All magic-angle spinning (MAS) spectra were recorded on a narrow-bore Bruker 800 MHz spectrometer in a 3-channel probe (¹H, ¹³C, ¹⁵N) at 55 kHz spinning frequency in a 1.3 mm rotor, with the VT gas flow was set to 900 l/h at 250 K. The only exception were the hcaCBcaNH and

hcaCBcacoNH spectra, which were measured on a narrow-bore Bruker 950 MHz spectrometer in a 4-channel 0.7 mm probe at 90.909 kHz MAS. For these two experiments the VT gas flow was set to 300 l/h at 270 K. Typical pulse lengths of 2.5 μs (^1H), 3.1 μs (^{15}N) and 4 μs (^{13}C) were used for all measurements. Heteronuclear magnetization transfers were implemented using cross-polarization. During evolution periods, 12.5 kHz TPPM decoupling (Bennett et al., 1995) was used on ^1H , and 10 kHz waltz-16 decoupling (Shaka et al., 1983a; Shaka et al., 1983b) on both heteronuclei. Water suppression was achieved with the MISSISSIPPI (Zhou and Rienstra, 2008) scheme applied at 13.75 kHz. Recycle delays were set to 0.8-1.0 s. The 4D HhCANH experiment (Najbauer et al., 2019) was recorded using 3.55% non-uniform sampling, with the NOE mixing time set to 75 ms. The spectra were referenced to 4.7 ppm water chemical shift, assuming a sample temperature of 30°C based on external calibration. For the multidimensional experiments (3D and above), typical acquisition times in the indirect dimensions were 20 ms on ^{15}N , 10 ms on ^{13}C , and 5.3 ms in the indirect ^1H dimension. The experimental parameters are summarized in **Table 1-1**. Addition of spectra followed a correction of magnetic field drift using the Topspin AU macro described in Chapter 3, and by Najbauer and Andreas (Najbauer and Andreas, 2019) For processing, all points in the indirect dimensions were used, in the direct ^1H dimension data from the first 10 ms was taken. The data were processed in Topspin 3.5pl7 using a squared sine window function with a phase shift of $\pi/2$ and analyzed in Sparky.

Table 1-1: Experiments recorded and their experimental parameters

Spectrum	(H)NH	(H)CANH	(H)(CO)CA(CO)NH	(H)CONH	(H)CO(CA)NH
Scans per point	48	36	28	16	64
Experimental time / hours	14	151	117.5	34.8	139.8
Field	800	800	800	800	800
Spinning frequency (Hz)	55000	55000	55000	55000	55000
Sweep width (t_1) / ppm	120.0 (N)	33.2 (N)	33.2 (N)	40.0 (N)	40.0 (N)
Max indirect evolution (t_1) / ms	52.6 (N)	21.9 (N)	21.9 (N)	22.2 (N)	22.2 (N)

Spectrum	(H)NH	(H)CANH	(H)(CO)CA(CO)NH	(H)CONH	(H)CO(CA)NH
Sweep width (t_2) / ppm	30.0 (H)	32.7 (C)	32.7 (C)	16.0 (C)	16.0 (C)
Max indirect evolution (t_2) / ms	54.9 (H)	9.7 (C)	9.7 (C)	22.1 (C)	22.1 (C)
Sweep width (t_3) / ppm		30.0 (H)	30.0 (H)	30.0 (H)	30.0 (H)
Max indirect evolution (t_3) / ms		21.3 (H)	21.3 (H)	21.3 (H)	21.3 (H)
Transfer I	HN (dipolar)	HCA (dipolar)	HCO (dipolar)	HCO (dipolar)	HCO (dipolar)
^1H field / kHz	90	90	90	87.83	91.3
X field / kHz	18.1	38.2	38.22	37.235	41.15
Shape	80 - 100% linear (H)	80 - 100% linear (H)	80 - 100% linear (H)	80 - 100% linear (H)	80 - 100% linear (H)
Carrier ^{13}C		53.7	173.3	173.8	173.8
Time / ms	1	6	6	5	5
Transfer II	NH (dipolar)	CAN (dipolar)	COCA (scalar)	CON (dipolar)	COCA (scalar)
^1H field / kHz	84.4				
^{13}C field / kHz		23.31		22.886	
^{15}N field / kHz	18.1	29.83		30.16	
Carrier ^{13}C		53.7	173.3 and 53.3	173.8	173.8 and 53.3
Shape	100-80% linear (H)	63-94% tangential (N)		63-94% tangential (N)	
Time / ms	0.6	8	9.1 (1st step)	12	7.6 (1st step) and 6.2 (second step)
Transfer III		NH (dipolar)	CON (dipolar)	NH (dipolar)	CAN (dipolar)
^1H field / kHz		84.43		84.8	
^{13}C field / kHz			23.31		23.36
^{15}N field / kHz		39.71	29.83	39.707	30.83
Carrier ^{13}C			173.3		53.3
Shape		100 - 80% linear (H)	63-94% tangential (N)	100 - 80% linear (H)	63-94% tangential (N)

Spectrum	(H)NH	(H)CANH	(H)(CO)CA(CO)NH	(H)CONH	(H)CO(CA)NH
Time / ms		0.6	8	0.6	10
Transfer IV			NH (dipolar)		NH (dipolar)
¹ H field / kHz			83.43		86.5
¹⁵ N field / kHz			39.71		39.71
Shape			100 - 80% linear (H)		100 - 80% linear (H)
Time / ms			0.6		0.6

Spectrum	(H)(CA)CB(CA)NH	(H)(CA)CB(CA)(CO)NH	(H)N(CA)(CO)NH	(H)N(CO)(CA)NH
Scans per point	8	14	32	28
Experimental time / hours	163.8	286.7	157.2	137.6
Field	950	950	800	800
Spinning frequency (Hz)	90909	90909	55000	55000
Sweep width (<i>t</i> ₁) / ppm	40.0 (N)	40.0 (N)	36.0 (N)	36.0 (N)
Max indirect evolution (<i>t</i> ₁) / ms	20.8 (N)	20.8 (N)	23.0 (N)	23.0 (N)
Sweep width (<i>t</i> ₂) / ppm	80.5 (C)	80.5 (C)	36.0 (N)	36.0 (N)
Max indirect evolution (<i>t</i> ₂) / ms	10.0 (C)	10.0 (C)	22.6 (N)	22.6 (N)
Sweep width (<i>t</i> ₃) / ppm	52.6 (H)	52.6 (H)	30.0 (H)	30.1 (H)
Max indirect evolution (<i>t</i> ₃) / ms	41.0 (H)	41.0 (H)	21.3 (H)	21.3 (H)
Transfer I	HCA (dipolar)	HCA (dipolar)	HN (dipolar)	HN (dipolar)
¹ H field / kHz	72.9	72.9	84.91	85.18
X field / kHz	15.3	15.3	37.3	37.3
Shape	85 - 100% linear (H)	85 - 100% linear (H)	80 - 100% linear (H)	80 - 100% linear (H)
Carrier ¹³ C	53.7	53.7		

Spectrum	(H)(CA)CB(CA)NH	(H)(CA)CB(CA)(CO)NH	(H)N(CA)(CO)NH	(H)N(CO)(CA)NH
Time / ms	4	4	1	1
Transfer II	CACB (scalar, out-and-back)	CACB (scalar, out-and-back)	NCA (dipolar)	NCO(dipolar)
¹³ C field / kHz			23.67	25.4
¹⁵ N field / kHz			28.02	27.89
Carrier ¹³ C	39.7	39.7	53.3	173.3
Shape			63-94% tangential (N)	63-94% tangential (N)
Time / ms	10.4	10.4	13	16
Transfer III	CAN (dipolar)	CACO (scalar)	CACO (scalar)	COCA (scalar)
¹ H field / kHz				
¹³ C field / kHz	58.8			
¹⁵ N field / kHz	41.85			
Carrier ¹³ C	30.1	173.3 and 53.3	53.3 and 173.3	173.3 and 53.3
Shape	76-96% tangential (N)			
Time / ms	19	5.7 ms (1st, simultaneous with CB), 8.2 2nd transfer	7.5 (1st step) and 6.7 (2nd step)	7.5 (1st step) 6.7 (2nd step)
Transfer IV	NH (dipolar)	CON (dipolar)	CON (dipolar)	CAN (dipolar)
¹ H field / kHz	110.1			
¹³ C field / kHz		60.4	24.21	25.07
¹⁵ N field / kHz	35	32.3	35.71	35.87(???)
Carrier ¹³ C		173.7	173.3	53.3
Shape	80 - 100% linear (H)	76-96% tangential (N)	63-94% tangential (N)	63-94% tangential (N)
Time / ms	0.7		12	16
Transfer V		NH (dipolar)	NH (dipolar)	NH (dipolar)
¹ H field / kHz		110.1	80.325	82.23
¹⁵ N field / kHz		35	37.3	37.3
Shape		80 - 100% linear (H)	100 - 80% linear (H)	100 - 80% linear (H)
Time / ms		0.7	0.6	0.7

Spectrum	(H)COCANH	(H)(CO)CACONH	(H)COCA(N)H	HN(H)(H)NH
NUS%	2.86	3.21		7.87
Scans per point	16	6	28	
Experimental time / hours	149.9	90.5	141.4	
Field	800	800	800	800
Spinning frequency (Hz)	55000	55000	55000	55000
Sweep width (t_1) / ppm	16.0 (CO)	34.0 (N)	16.0 (CO)	5.5 (H)
Max indirect evolution (t_1) / ms	22.1 (CO)	22.1 (N)	22.1 (CO)	4.0 (H)
Sweep width (t_2) / ppm	32.7 (CA)	34.0 (CA)	32.7 (CA)	33.0 (N)
Max indirect evolution (t_2) / ms	9.7 (CA)	13.0 (CA)	9.7 (CA)	22.0 (N)
Sweep width (t_3) / ppm		15.0 (CO)	30.1 (H)	33.0 (N)
Max indirect evolution (t_3) / ms		12.9 (CO)	21.3 (H)	22.0 (N)
Sweep width (t_4) / ppm	30.1 (H)	30.0 (H)		30.0 (H)
Max indirect evolution (t_4) / ms	21.3 (H)	21.3 (H)		21.3 (H)
Transfer I	HCO (dipolar)	HCO (dipolar)	HCO (dipolar)	HN (dipolar)
^1H field / kHz	91.277	94.87	76.85	96
X field / kHz	41.154	38.44	41.32	39.06
Shape	80 - 100% linear (H)	80 - 100% linear (H)	80 - 100% linear (H)	80 - 100% linear (H)
Carrier ^{13}C	173.3	173.3	173.3	
Time / ms	5	6	6	1
Transfer II	COCA (scalar, SCT)	COCA (scalar, out-and-back)	COCA (scalar, SCT)	NH (dipolar)
^1H field / kHz				91.03
^{15}N field / kHz				39.06

Spectrum	(H)COCANH	(H)(CO)CACONH	(H)COCA(N)H	HN(H)(H)NH
Carrier ¹³ C	173.3 then 53.3	173.3 and 53.3	173.3 then 53.3	
Shape				100 - 80% linear (H)
Time / ms	7.6 (1st transfer) and 6.7 (2nd transfer)	7.5	7.2 (1st transfer) 5.4 (2nd transfer)	0.6
Transfer III	CAN (dipolar)	CON (dipolar)	CAN (dipolar)	HH (RFDR)
¹ H field / kHz				
¹³ C field / kHz	23.36	24.38	25.2	
¹⁵ N field / kHz	30.83	29.83	28.76	
Carrier ¹³ C	53.3	173.3	53.3	
Shape	63-94% tangential (N)	63-94% tangential (N)	63-94% tangential (N)	
Time / ms	10	8	8	2.3
Transfer IV	NH (dipolar)	NH (dipolar)	NH (dipolar)	HN (dipolar)
¹ H field / kHz	86.5	88.99	76.48	96
¹⁵ N field / kHz	39.71	39.71	38.47	39.06
Shape	100 - 80% linear (H)	100 - 80% linear (H)	100 - 80% linear (H)	80 - 100% linear (H)
Time / ms	0.6	0.6	0.6	1
Transfer V				NH (dipolar)
¹ H field / kHz				91.03
¹⁵ N field / kHz				39.06
Shape				100 - 80% linear (H)
Time / ms				0.6

Automated assignments were obtained from the FLYA module of the CYANA software package (Schmidt and Guntert, 2012). Tolerances were set to 0.07 ppm for ¹H and 0.4 ppm for heteronuclei. For each of the 10 runs, the initial population was set to 50, and 15000 iterations were performed. A resonance was counted as confident, when it converged to the same assignment in at least 80% of the runs.

1.2.2.1 Influence of membrane environment and ligand binding studies

To investigate VDAC in a membrane environment with native-like composition, porcine brain extract was purchased from Avanti, and for reconstitution a lipid to protein ratio (n/n) of 26:1 was used. The hNH spectrum was recorded with the parameters specified in **Table 1-1**, evolution times were 21.3 ms on ^1H and 35.9 ms on ^{15}N .

To test cholesterol binding, a 26:1 lipid to protein ratio (n/n) 2D crystalline sample was prepared with d_{54} -DMPC, and cholesterol was added at a 1:5 VDAC to cholesterol ratio (n/n). The HhCANH spectrum was recorded with the CP-based pulse sequence described in Chapter 2, and in Najbauer et al. (Najbauer et al., 2019), replacing the H-N and N-C CP steps with one H-CA CP step. The CP conditions are identical to those described in **Table 1-1** for the hCANH experiment. The employed NOE mixing time was 75 ms. The spectrum was recorded using 3.55% non-uniform sampling with 5.2 ms indirect ^1H , 10.1 ms ^{13}C , 20.1 ms ^{15}N , and 21.3 ms direct ^1H evolution times, averaging a total of 38 scans.

G3139 (Sigma Aldrich) was added to the reconstituted membrane protein sample resuspended in 200 μl of sample buffer at an estimated 1:1.5 molar ratio. hNH, hCANH, and hcoCAcoNH spectra were recorded with the parameters described in **Table 1-1**, in the hNH spectrum evolution times were 21.3 ms on ^1H and 40.0 ms on ^{15}N .

All spectra were recorded under the previously described conditions used for the assignment experiments.

1.2.2.2 Chemical shift perturbations

All chemical shift perturbations were calculated from a weighted average of chemical shift differences for the various nuclei, using weighting factors from literature (Williamson, 2013):

$$CSP = \sqrt{\frac{1}{3}(\delta_{1,^1\text{H}} - \delta_{2,^1\text{H}})^2 + \left(0.3(\delta_{1,^{13}\text{C}} - \delta_{2,^{13}\text{C}})\right)^2 + \left(0.17(\delta_{1,^{15}\text{N}} - \delta_{2,^{15}\text{N}})\right)^2}$$

Chemical shift perturbations were plotted onto either the topology model of the E73VC127AC232S mutant, or the lowest energy structure from structural ensemble in the PDB entry (5jdp) (Jaremko et al., 2016) with the Chimera software (Pettersen et al., 2004).

1.2.2.3 Relative intensities

Relative intensity values were determined for the H/D exchange experiments, for MTSL-labeled samples, and for comparison of the closed-state (G21V/G23V/E73V/C127A/C232S) mutant to the open-state (E73V/C127A/C232S) mutant. Intensities were obtained from maximal peak heights measured in the hCANH spectra. The spectral parameters for all measurements were identical to those listed in **Table 1-1**. Due to spectral overlap, exact peak heights could not be determined for residues A8, Y22, L29, E59-Y62, Y67, F71, K74, T86, E88, D89, K96, L97, S101, L125, M129, D130, L142, L150, M155, L180, V184, D186, F190, I194, V206, L208, A212, N216, I221, S234, Q249, K256, S260, L262, E280.

1.3 Results

1.3.1 Optimization of VDAC sample preparation

Initial VDAC samples used in several studies performed at the department (Schneider et al., 2010; Briones et al., 2016) were prepared as liposomes. Even with the use of perdeuteration and fast magic angle spinning, no individual peaks could be resolved in the hNH spectra, thus the exact ^1H linewidth could not be measured. (**Figure 1-5**)

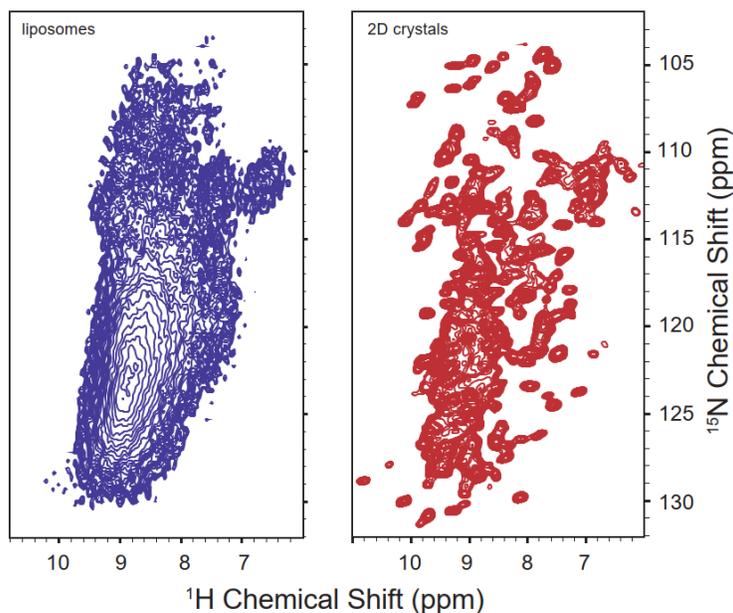


Figure 1-5: Spectral quality of liposome and 2D crystalline samples. hNH spectra were recorded of wild-type and E73VC127SC232A-hVDAC1 in liposomes and 2D crystals, respectively (the mutations are not expected to dramatically influence spectral quality). Both samples were uniformly $^2\text{H}^{13}\text{C}^{15}\text{N}$ -labeled, and measured at an 800 MHz spectrometer at 55 kHz MAS. The improvement of spectral quality in the 2D crystalline sample (red) is clearly visible.

It was clear however, that liposome samples only enable the use of heteronuclear (e.g. ^{13}C) detection. Using the protocol developed by Dolder et al (Dolder et al., 1999), and adapted for ssNMR measurements by Eddy et al. (Eddy et al., 2012), a 2D crystalline sample of E73VC127SC232A-hVDAC1 was prepared. The three mutations were introduced to improve sample stability (preventing oligomerization through the removal of the two cysteines), and to improve spectral quality. These mutations, however, do not influence VDAC's gating behavior: Aram et al. had previously shown that the two cysteines in VDAC are not required for voltage

gating (Aram et al., 2010), and the same has recently been shown for the E73 residue as well (Queralt-Martin et al., 2019). Upon investigation of the sample by negative stain electron microscopy, a lamellar structure was observed, showing that our preparation reproduced the microscopically ordered layers described by Eddy et al. in their preparation (Eddy et al., 2012).

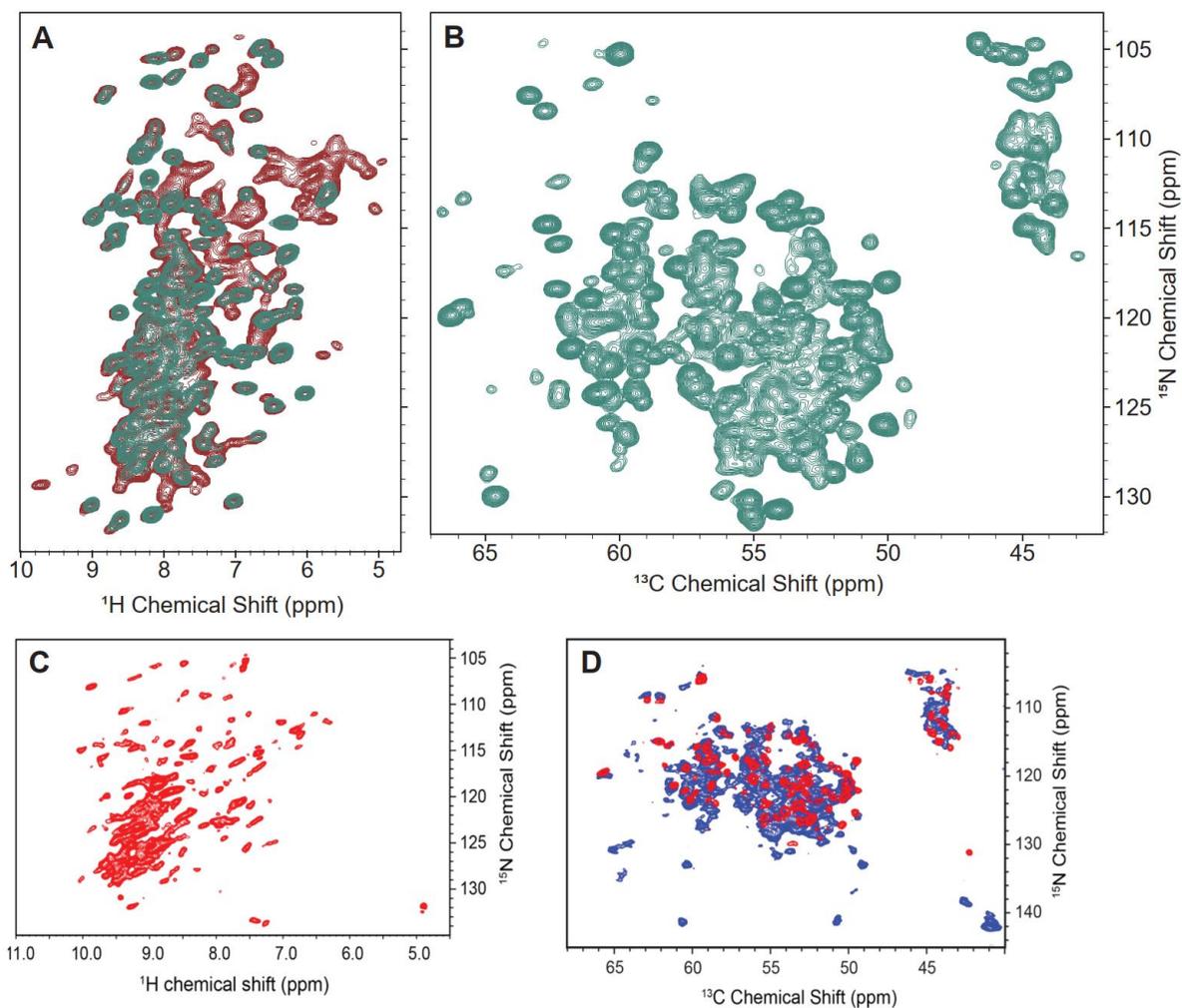


Figure 1-6: Projections of the hCANH spectrum and comparison of spectral quality. A) Overlay of the ^1H - ^{15}N (dark red) and the ^1H - ^{15}N projection of the hCANH spectrum (turquoise). B) ^{13}C - ^{15}N projection of the hCANH spectrum. Spectra shown in panel A) and B) were recorded on $[\text{u-}^1\text{H}, ^{13}\text{C}, ^{15}\text{N}]$ E73VC127SC232A-hVDAC1 in DMPC 2D lipid crystals at 55 kHz MAS at 800 MHz. C) Proton-detected ^1H - ^{15}N correlation measured by Eddy et al. D) Overlay of ^{13}C - ^{15}N projection of the hCANH spectrum, and the 2D ^{15}N - ^{13}C TEDOR spectrum of hVDAC1, measured by Eddy et al. Spectra shown on panels C) and D) were measured on $[\text{u-}^1\text{H}, ^{13}\text{C}, ^{15}\text{N}]$ hVDAC1 in DMPC 2D lipid crystals at 60 kHz MAS at 1 GHz. Panels C) and D) have been reproduced from (Eddy et al., 2015b).

Peaks are well-dispersed in the hNH spectrum (**Figure 1-6**), ranging from $^1\text{H}^{\text{N}}$ chemical shifts of 6 ppm to 10.5 ppm, with signals of α -helical residues characteristically appearing at lower shifts than those of the β -barrel, confirming that the protein is in its folded state in our 2D crystalline preparation. Multidimensional spectra of the lipid crystalline sample showed exceptional resolution, the best heteronuclear linewidths being <0.4 ppm, and ^1H linewidths <0.15 ppm (**Figure 1-6**). The spectral quality was thus high enough to attempt a *de novo*, proton-detected assignment strategy, which had previously proven to be a successful approach on membrane proteins of similar size (OmpG – 34 kDa (Retel et al., 2017), AlkL – 24 kDa (Schubeis et al., 2020a; Schubeis et al., 2020b)).

1.3.2 Assignment of E73VC127SC232A-hVDAC1

In order to achieve an assignment as complete as possible of VDAC in lipids, a combination of 3- and 4-dimensional assignment experiments were recorded on a perdeuterated, 100% H^{N} back-exchanged 2D crystalline sample of E73VC127SC232A-hVDAC1, following the proton-detected strategy described in several previous publications. (Barbet-Massin et al., 2014; Fricke et al., 2017). A full list of the experiments recorded and the experimental parameters is shown in **Table 1-1**.

The “basic 6” 3D experiments routinely used for a backbone walk (hCANH, hcoCAcoNH, hCOcaNH, hCONH, hcaCBcaNH, hcaCBcacoNH) (**Figure 1-7**) were measured as described by Barbet-Massin et al (Barbet-Massin et al., 2014). Under conditions of fast magic angle spinning (>55 kHz), and perdeuteration of the protein, transverse relaxation becomes increasingly slow ($T_2^*_{\text{CO}}=38.0$ ms, $T_2^*_{\text{CA}}=21.5$ ms). Although dipolar-based cross-polarization is still efficient for heteronuclear transfers ($^1\text{H}^{15}\text{N}$, $^1\text{H}^{13}\text{C}$, $^{15}\text{N}^{13}\text{C}$ CPs), J-based INEPT transfers were used for all homonuclear ^{13}C - ^{13}C transfers. The use of J-transfers also allows the implementation of useful elements found in solution NMR sequences. Out-and-back type sequences avoid evolving fast-decaying transverse magnetization during the echo periods of coherence transfer, resulting in higher sensitivity (Barbet-Massin et al., 2013). Constant-time (or if longer indirect evolution times are needed, semi-constant time) design of the pulse sequence can also be used, which eliminates the homogeneous component of transverse relaxation during indirect evolution, while also

allowing for homonuclear decoupling. **Figure 1-8** shows the implementation of out-and-back, as well as semi-constant time pulse sequence elements in solid-state NMR assignment sequences.

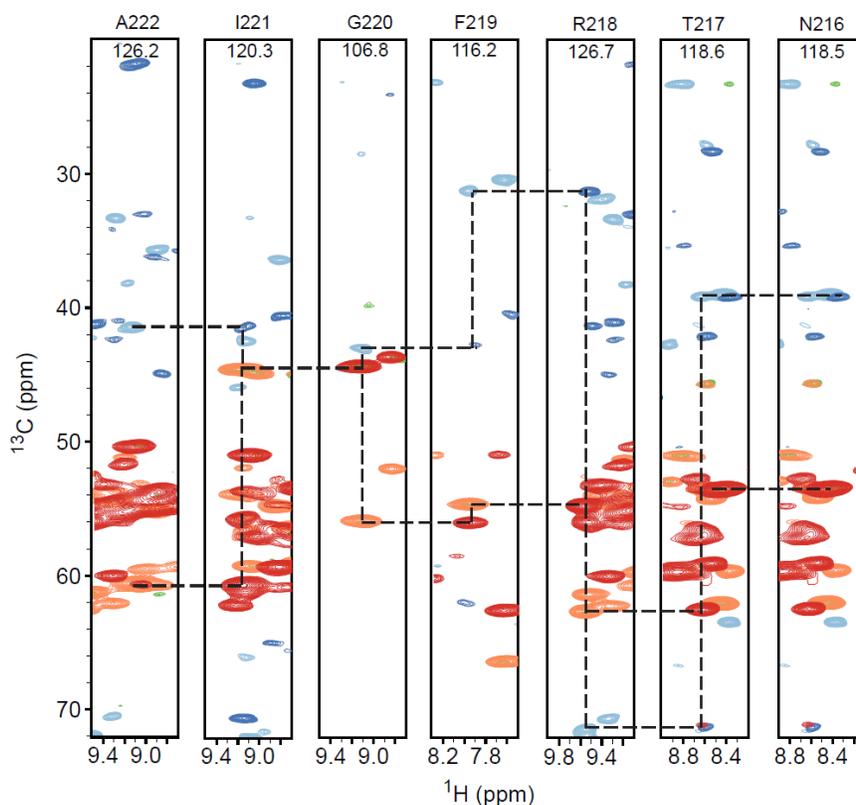


Figure 1-7: Backbone walk on $u[{}^2\text{H}, {}^{13}\text{C}, {}^{15}\text{N}]\text{-E73VC127A/C232S-hVDAC1}$. Red shows strips from the hCANH, orange from the hcoCAcoNH, dark blue from the hcaCBcaNH, and light blue from the hcaCBcacoNH spectrum. The depicted strips belong to residues in $\beta 15$, starting from the preceding loop.

hNcocaNH (Xiang et al., 2015) and hNcacoNH (Andreas et al., 2015b) spectra were recorded to unambiguously confirm residue links, especially since the lack of chemical shift correlation (as can often be observed for chemical shifts within the same residue) reduces signal overlap in these spectra.

Assignment experiments hCOCANH and hcoCAcoNH were also expected to significantly reduce spectral overlap by resolving signals in 4 dimensions (Fraga et al., 2017). The pulse sequence used for recording the hCOCANH spectrum is shown on **Figure 1-8**.

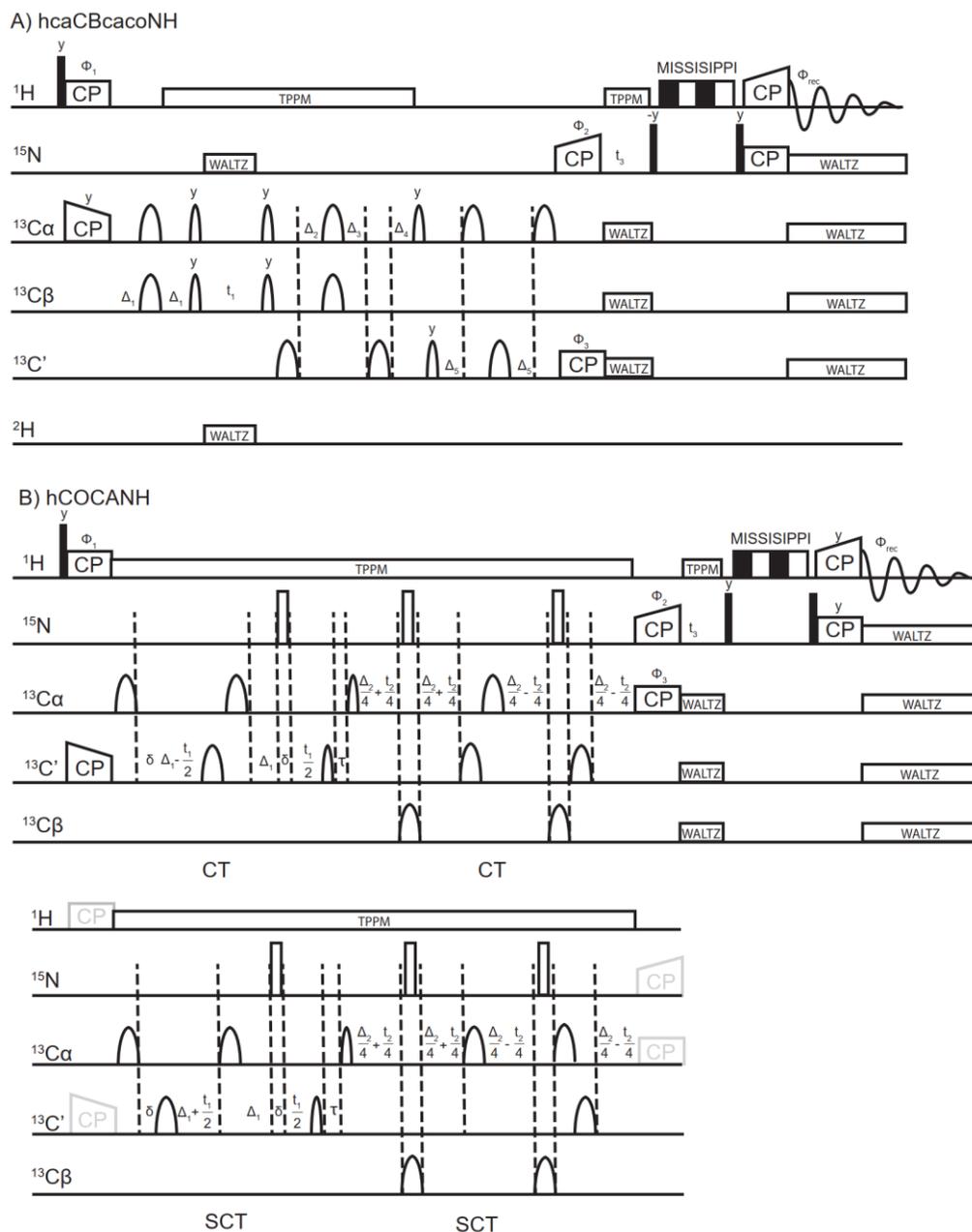


Figure 1-8: Assignment experiment pulse sequences using a combination of dipolar- and J-based transfer elements, optimal under conditions of fast MAS and perdeuteration. Full and broad rectangles indicate hard 90° and 180° pulses, respectively. Narrow and broad half-ellipses stand for selective excitation and inversion pulses, respectively. All pulses are phase 0, unless indicated otherwise. INEPT delays were calculated to optimize for sensitivity, taking transverse relaxation into account. A) The 3D hcaCBcacoNH pulse sequence. (Barbet-Massin et al., 2014) In the out-and-back element magnetization is not transferred fully to $C\beta$, but rather evolved in an antiphase term. In the second INEPT transfer step, delays Δ_2 and Δ_3 are set such that a simultaneous $C\beta$ - $C\alpha$, and $C\alpha$ -CO transfer can be accomplished. Phase cycling is as follows. $\phi_1 = (x)_4(-x)_4$, $\phi_2 = (-x)_2(x)_2$, $\phi_3 = y-y$, and $\phi_{rec} = -x x x-x x-x-x x$. B) The 4D hCOCANH pulse sequence. If ^{13}C indirect evolution times are longer than allowed by the constant time sequence, the semi-constant time block (shown below the sequence) is executed instead. Phase cycling is as follows: $\phi_1 = (x)_4(-x)_4$, $\phi_2 = (-x)_2(x)_2$, $\phi_3 = x-x$, and $\phi_{rec} = -y y y-y y-y-y y$. The pulse sequence was written independently from, and was thus similar, but not identical to the 4D sequence published by Fraga et al. (Fraga et al., 2017)

Since these experiments require relatively large sweep widths in all three indirect dimensions, as well as long indirect acquisition times due to the long coherence times of magnetization on the heteronuclei ($T_2'_{CO}=38.0$ ms, $T_2'_{CA}=21.5$ ms), non-uniform sampling was used to achieve optimal resolution in all dimensions. Adding a fourth indirect dimension to a 3D experiment results in a decrease of signal to noise by a factor of $\sqrt{2}$, however, and depending on the reconstruction algorithm, artifacts can appear in the spectra. In order to correlate CA and CO nuclei within the same residue at high sensitivity and to avoid reconstruction artifacts, the hCOCA_nH 3D experiment was recorded (Fricke et al., 2017).

A selectively H^α-labeled sample (Movellan et al., 2019) was used to introduce yet another nucleus to confirm assignments. In the hNCAHA and hNcoCAHA sequences, residue linking is similar to the backbone walk of the hCANH and hcoCAcoNH experiments, but instead of H^N, H^α is detected.

1.3.2.1 Automated assignment

Peak lists from the spectra listed above and in **Table 1-1** were used as an input for the FLYA algorithm of the CYANA software package. FLYA uses a genetic algorithm to assign proteins in an automated manner, avoiding the bias of human spectral evaluation (Schmidt and Guntert, 2012). Although the software had been written for the assignment of proteins in solution, magnetization transfer pathways used in solid-state protein NMR experiments can easily be defined. FLYA results are invaluable in the assignment of large proteins, however the obtained results must be confirmed, and can usually be extended manually.

To test for the usefulness of the information provided by different spectra in the assignment of membrane proteins, different subsets of assignment spectra were used as an input to FLYA. Peak lists from only the six 3D experiments linking C_α, C', and C_β (“basic six”) yielded confident assignments for 74 residues, the algorithm performing particularly well for the majority of the α-helix (residues 5-21), as well as residues from β-strands either containing many residues with distinct chemical shift values (β2) or strands with particularly large chemical shift dispersion of the peaks (e.g. β15, β16). (Assignments where the residue was part of an assigned stretch of at least 2 residues with at least two assigned backbone atoms were considered confident.)

The most assignments were gained by including information from either 4D hCOCANH and hcoCACoNH spectra or the hNcocaNH and hNcacoNH spectra. Including information from either of these two spectral pairs besides the “basic six” experiments roughly doubled the number of confident assignments: 150 and 137 residues were assigned, respectively. Little additional information was to be gained by using peak lists from additional spectra, achieving 163 confident assignments in total, when including peak lists from 13 different experiments. The assignment levels achieved by using different subsets of spectra are shown in **Table 1-2**.

Table 1-2: Confident assignments obtained with the FLYA algorithm using various subsets of assignment spectra. The assignment of a chemical shift value was classified as confident by the program, when for the residue the algorithm converged to the same value in 80% of the cases. A residue was counted as confidently assigned when it had at least two of its backbone atoms confidently assigned, and was part of a stretch of at least two such residues. The set of basic 6 experiments comprises hCANH, hcoCAcoNH, hCONH, hCOcaNH, hcaCBcaNH and hcaCBcacoNH assignment spectra.

Peak lists used for automated assignments	Confidently assigned residues	% of protein confidently assigned
Basic 6	74	26.1%
Basic 6, hCOCANH, hcoCAcoNH	150	53.0%
Basic 6, hNcocaNH, hNcacoNH	137	48.4%
Basic 6, hCOCANH, hcoCAcoNH, hNcocaNH, hNcacoNH	162	57.2%
Basic 6, hNCAH, hNcoCAH	102	36.0%
hCOCANH, hcoCAcoNH, hcaCBcaNH, hcaCBcacoNH	57	20.1%
Basic 6, hNCAH, hNcoCAH, hCOCANH, hcoCAcoNH, hNcocaNH, hNcacoNH	163	57.6%
Basic 6, hCOCANH, hcoCAcoNH, hNcocaNH, hNcacoNH, hCOCAhH	161	56.9%
Basic 6, hNCAH, hNcoCAH, hCOCANH, hcoCAcoNH, hNcocaNH, hNcacoNH, hCOCAhH	163	57.6%

Interestingly, although in theory a combination of the 4D C-C-N-H experiments with 3D C β linking spectra contains more unambiguous information than the “basic six” set of spectra, only 57 assignments were obtained from these four spectra. Taken together, this shows that to achieve high levels of assignment, it is crucial to use spectra with high sensitivity (“basic six” 3Ds), as well as spectra that might have lower sensitivity but suffer less from peak overlap (C-C-N-H or N-N-H datasets). The 163 confident assignments achieved with the algorithm are shown in **Figure 1-9**. The automated assignments were then confirmed and extended manually using a backbone walk (**Figure 1-7**, **Figure 1-9**). The assigned hNH spectrum is shown on **Figure 1-10**.

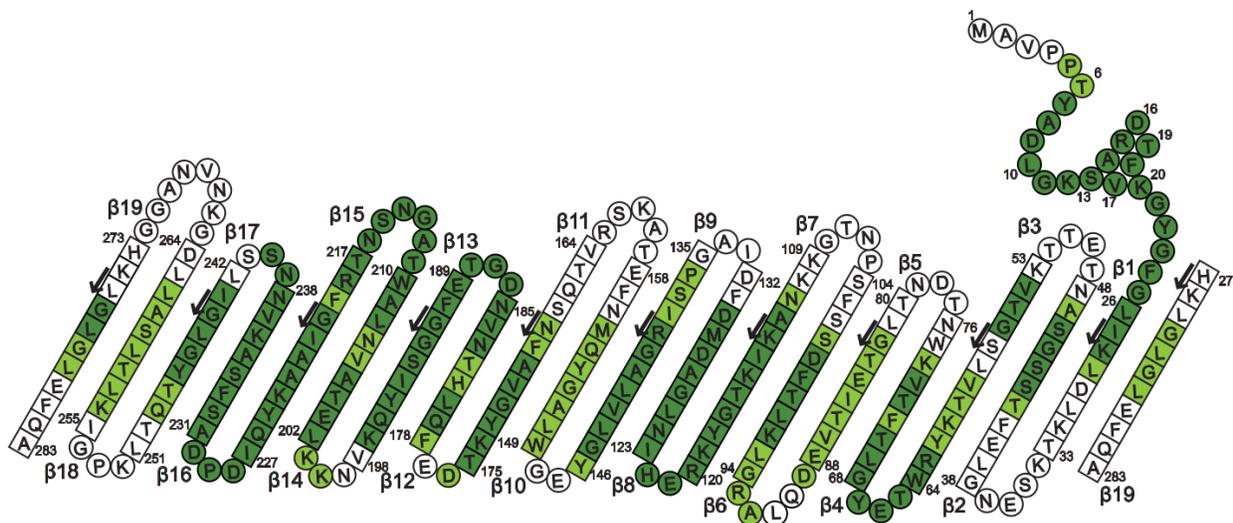


Figure 1-9: Comparison of automated and manual assignments for E73VC127SC232A-hVDAC1. Assignments obtained with the FLYA algorithm and confirmed manually are shown in dark green. Residues for which assignments were not marked as “confident” by the FLYA algorithm but assignments could be obtained manually are colored in light green. Residues K53, Q90, L91, K274, L275 had confident automated assignments, but these were not confirmed manually, thus they are not considered assigned.

Visualizing the assigned residues on the topology map of the protein (**Figure 1-9**), it is evident that the assigned residues are not statistically dispersed in the protein, rather they are clustered into regions more readily assigned than others. The very beginning of the N-terminus remains unassigned, which can be explained by the two prolines at residues 4 and 5. These break the backbone walk applied in the assignment strategy, thus leaving a very short assignable stretch. Furthermore, terminal residues can be fairly mobile due to a lack of structure, reducing the efficiency of magnetization transfer by cross-polarization in the assignment sequences. From P5, however, all of the N-terminal α -helix could be assigned, and peaks were well-resolved with high intensities, indicating a structured region. The only exception was G11, with an intensity at least

50% lower than of all other peaks in the N-terminus, a possible due to the residue being more dynamic. This result confirms the findings of previous studies, where with the exception of the disordered N-terminus of the structure published by Hiller et al. (Hiller et al., 2008), all subsequent studies both in micelles (Bayrhuber et al., 2008; Jaremko et al., 2016), and in a lipid bilayer (Schneider et al., 2010) found the N-terminus of the protein to be α -helical (or consisting of two α -helical segments, with a kink at L10 and G11 (Ujwal et al., 2008)).

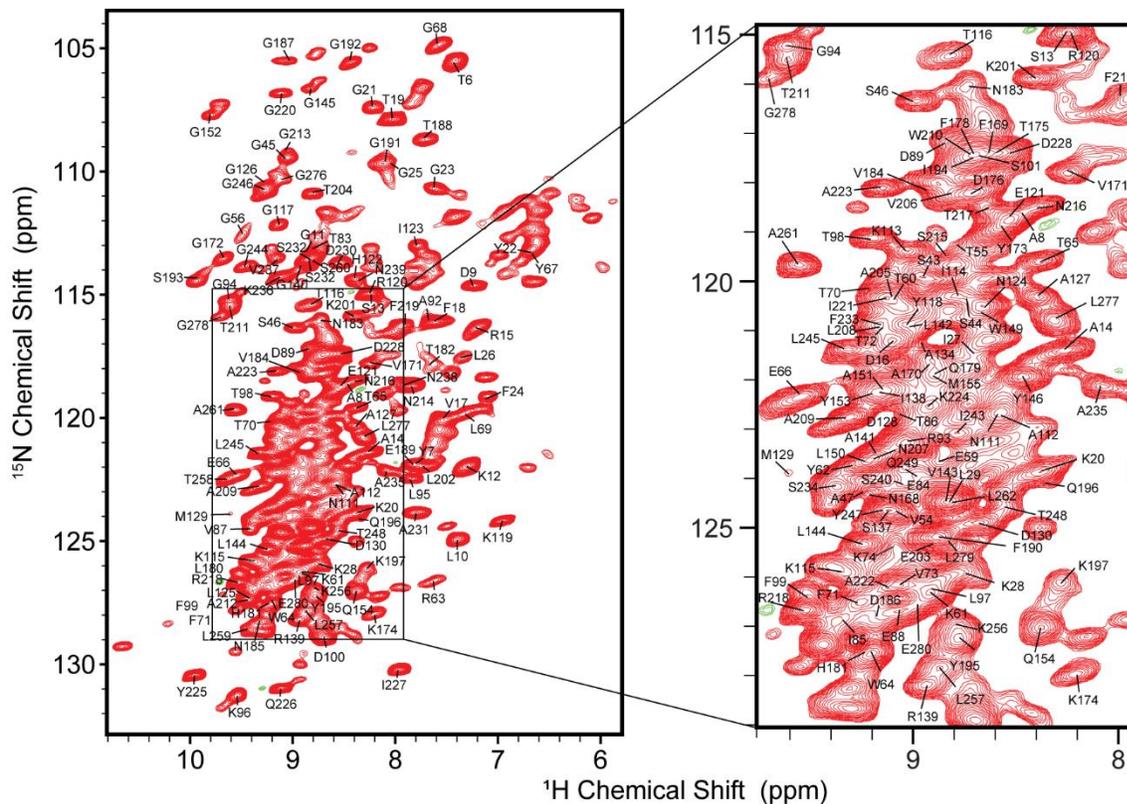


Figure 1-10: De novo assignment of E73VC127SC232A-hVDAC1 The hNH spectrum with assignments based on FLYA results obtained by including peak lists from all recorded assignment spectra, which were then inspected and extended manually. The separation of peaks belonging to residues in the α -helix or the β -strands can be observed: most helical ^1H shifts appear between 7 and 8 ppm, whereas residues in the β -strands tend to appear at higher ^1H chemical shifts.

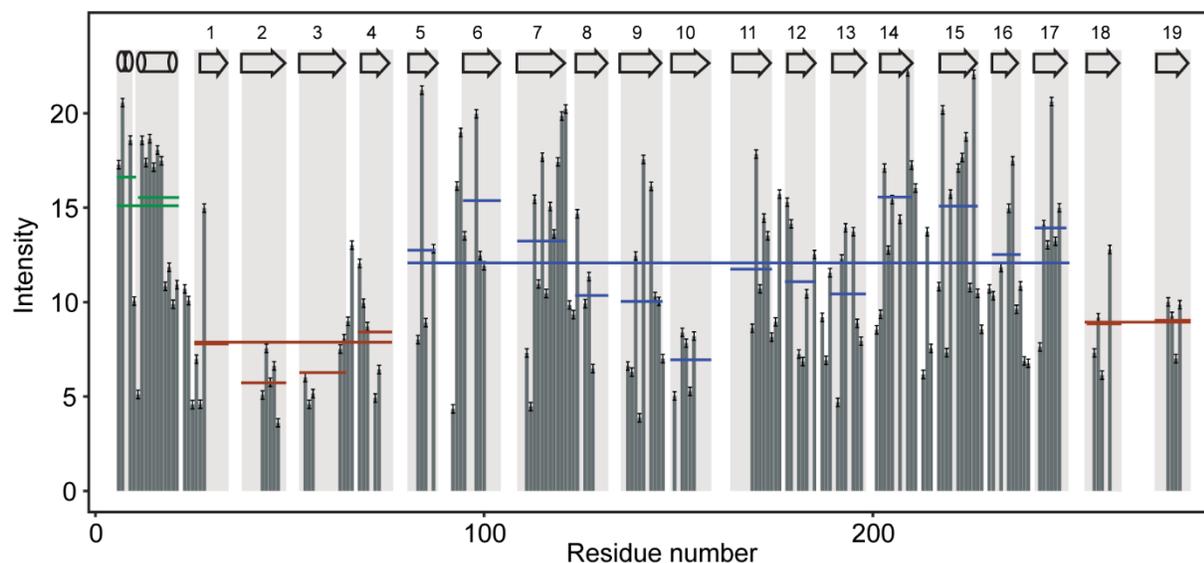


Figure 1-11: Signal intensities of E73V/C127A/C232S-hVDAC1 in a lipid bilayer as determined from the hCANH spectrum. Gray rectangles indicate secondary structural elements as determined previously in micelles (Jaremko et al., 2016). For each secondary structural element (α -helical segment or β -strand), the average of peak intensities in that motif are shown as colored lines, and average peak intensities are also indicated for the whole region (including loops). Green shows average intensities in the helical region, red in the 4 N-terminal β -strands and the 2 C-terminal strands connecting to them, blue shows average intensities in the rest of the β -barrel.

Although the assignment of several residues in the N-terminal β -strands was possible, signals in this region suffer from low intensity and severe overlap. (**Figure 1-11**) Motion on the μ s timescale has been shown to interfere with magic-angle spinning (Saito et al., 2002), resulting in signal loss (Chevelkov et al., 2007). In the presence of dynamics, hydrogen bonds may also be weaker, increasing H_{H_2O}/H_{HN} chemical exchange rates, which would again result in the broadening of signals. Since residues undergoing fast motion experience a variety of chemical environments, their chemical shifts will appear as an average of these environments, decreasing their chemical shift dispersion (the extreme case of this effect can be observed in intrinsically disordered proteins). Low signal intensity and significant overlap in the N-terminal barrel thus might stem from increased mobility in this region, as seen previously in micelles. In the wild-type sequence Villinger et al. observed low signal intensity and elevated μ s-ms timescale motion N-terminal 6 β -strands of the wild-type sequence. In their study, the E73V point mutation reduced dynamics in the whole N-terminal part of the barrel (β 1- β 9), as well as in the last strand (β 19); this effect was ascribed to removal of the charged Glu-sidechain pointing into the apolar membrane (Villinger et al., 2010). Our data show however, that in a lipid bilayer even in the presence of the E73V mutation, this part of the barrel is intrinsically more mobile. A logical explanation for this

phenomenon could be that the helix that runs along the C-terminal part of the barrel has a stabilizing effect, reducing the motion in its proximity.

1.3.3 Determination of hVDAC1's topology in a lipid bilayer

Sequential assignments are a valuable source of structural information, as chemical shifts of backbone atoms and C β -s are indicative of the ϕ and ψ torsion angles of the backbone. Using the TALOS-N web server (Shen and Bax, 2013) secondary structure of VDAC in lipid bilayers was predicted using the protein's sequence and chemical shift assignments. (**Figure 1-12**)

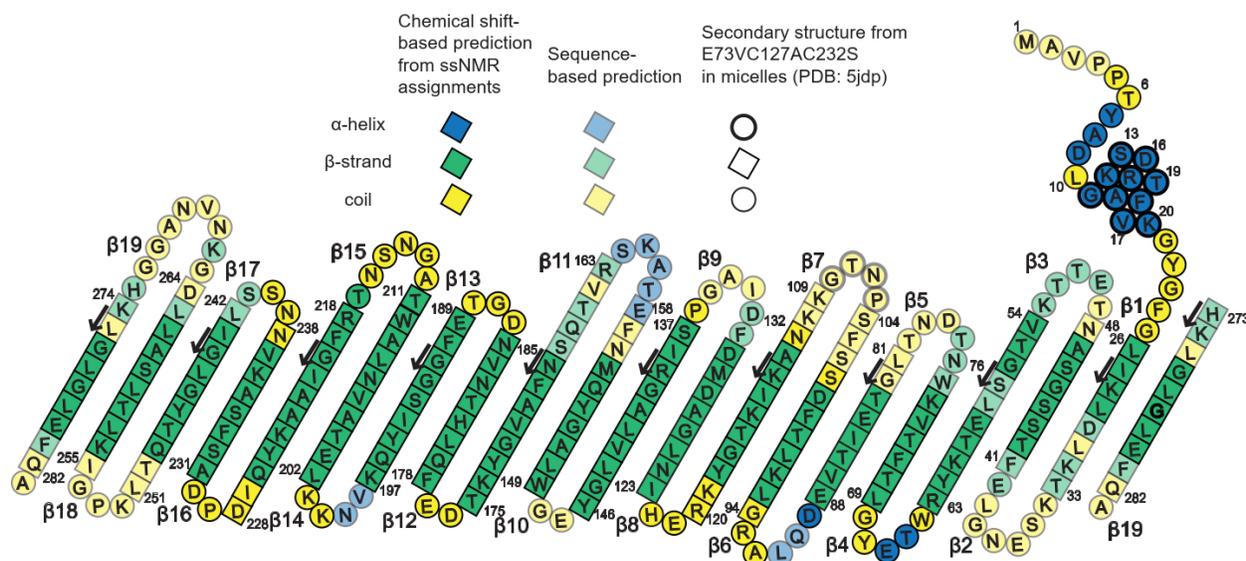


Figure 1-12: Secondary structure prediction of E73VC127AC232S-hVDAC1 Secondary structure of VDAC based on ssNMR chemical shift assignments was predicted using the TALOS-N webserver. All predictions based on chemical shifts are mapped in dark, saturated colors: α -helix – blue, β -sheet – green, loops – yellow. Talos N predictions based only on the primary sequence are colored using the same color code, but faded, lighter. The secondary structural elements from the E73V-hVDAC1 structure in micelles (PDB: 5jdp, lowest energy structure) is shown as follows: α -helix – thick circles, β -sheet – squares, loops – thin circles.

The secondary structure predicted by TALOS-N based on chemical shift assignment in lipid bilayers matches very well with the published structure of the E73V mutant in micelles (PDB: 5jdp). According to the prediction from ssNMR, the helix appears to be somewhat longer than in the solution structure, comprising the 14 residues between Y7 and K20, instead of starting only at G11, although NOE contacts in E73V-hVDAC1 indicate a helical structure even in micelles

between T6 and L10 (Schmidt, 2017). The β -strands predicted from the chemical shifts also overlap very well with those in the solution structure. Some slight deviations occur in the extent of the strands, close to the loops, where according to the prediction, β -strands might be somewhat shorter than in the published structure (e.g. I225 and D226 in β 15), or might extend further by a few residues (e.g. I133-P136 in β 9). The latter might be an effect stemming from the fact that 2D crystals tend to have lower mobility in loop regions due to close stacking of the channels, and this might give rise to a weak H-bonding in facing loop residues. Interestingly, based on chemical shifts, some isolated residues (D89, H122), as well as the linker between β 3 and β 4 are predicted to adopt a helical conformation. Due to the fact, however, that even the linker consists of only 3 residues, a proper helical turn is impossible in these cases, as an α -helix requires formation of $i \rightarrow i+4$ bonds.

VDAC's topology (strand alignment) in lipid bilayers was determined by recording the 4D HNhhNH spectrum (**Figure 1-13**), in which a radio-frequency driven recoupling (RFDR) element (Bennett et al., 1992) is used to recouple distance dependent dipolar couplings, correlating $^1\text{H}^{\text{N}}$ - ^{15}N pairs in close spatial proximity.

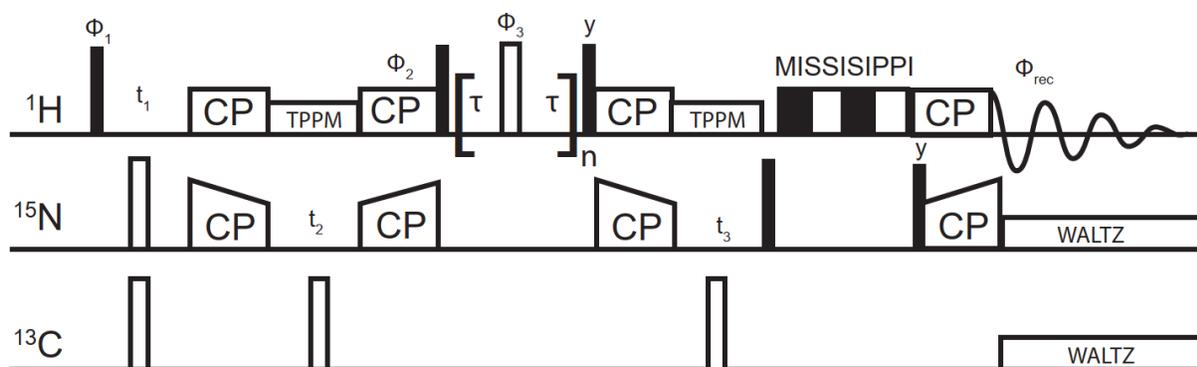


Figure 1-13: HNhhNH pulse sequence for acquiring spatial restraints. Full and empty rectangles denote hard excitation and inversion pulses respectively. ^1H - ^1H dipolar mixing is achieved by repeating the RFDR block enclosed by brackets n times. The refocusing pulses in the RFDR sequence are applied in a rotor synchronized manner, and the length of one block is set to exactly one rotor period ($1 / \nu_{\text{rot}}$). All phases were x unless noted otherwise. Phase cycling is as follows: $\phi_1 = y - y$, $\phi_2 = (y)_2(-y)_2$, $\phi_3 = (x y)_2(y x)_2$, and $\phi_{\text{rec}} = x - x - x x$.

The assigned contacts in the HNhhNH spectrum, based on which the hydrogen-bonding pattern in the protein can be inferred, are shown on **Figure 1-14**. Due to the difficulties in the assignment of β 2 and residues in the neighboring strands stemming from low peak intensities and severe signal

overlap, the contacts of $\beta 2$ to neighboring strands were somewhat ambiguous. It is clearly visible though that the contacts between residues 26-283 define a 19-stranded β -barrel structure, with all strands adopting an antiparallel arrangement except for strands 1 and 19, which have a parallel orientation, also indicated by the characteristic H^N-H^N contacts.

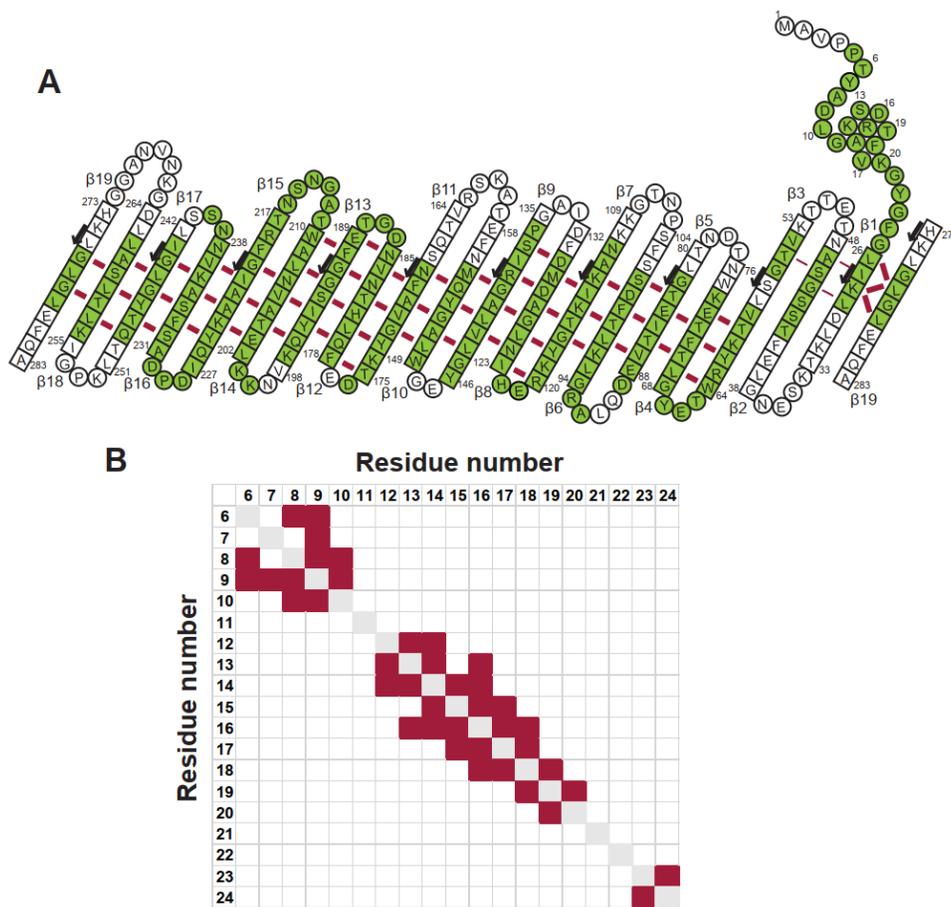


Figure 1-14: Topology of VDAC in a lipid bilayer as determined by ssNMR A) Long-range, inter β -strand contacts as measured in the HNhhNH experiment. The hydrogen bonds shown are based on the measured contacts, and are indicated by thick burgundy lines. Uncertain contacts (cross-peaks appeared close to the noise level for residues of strand $\beta 2$) are shown by thin burgundy lines. Residues assigned by ssNMR are colored in green. Contacts observed within the barrel are mapped onto the topology model of E73V VDAC determined in micelles (PDB: 5jdp). B) Contacts in the N-terminal α -helix, as measured in the HNhhNH experiment, shown in burgundy. Contacts were mapped symmetrically (i.e. if residue i is contacting residue $i+2$, it is indicated on both sides of the diagonal).

Upon inspection of the N-terminal contacts, a helical topology is clearly defined by the $i-i+2$ and $i-i+3$ contacts (**Figure 1-14B**). Residues T6-L10, as well as K12-F18 all show medium-range contacts (i to $i+2$ or $i+3$), while T19 and K20 show contacts to the neighboring residues only. Taken together with the TALOS results, this indicates the helical region extending between T6

and K20. No residues were however observed to contact G11, a residue already suspected to be more dynamic due to its relatively low peak intensity. This observation is in line with the kink around G11 observed in both a detergent (Ujwal et al., 2008; Jaremko et al., 2016) and a lipid environment (Schneider et al., 2010). We can thus conclude that the N-terminal residues form two helical segments, one between T6-L10, and the other between K12-K20. K20 is followed by another kink, evidenced by the lack of contacts for G21 and Y22, connecting the helix to the barrel. One more *i-i+1* contact was observed between G23 and F24, however, this does not necessarily indicate the presence of a helix, considering that no contacts were observed for any neighboring residues, and since from L26 onwards, characteristic inter- β -strand contacts are present. This contact possibly arises from the helix interacting with the beta barrel, somewhat restricting the conformational space of the linker, and bringing the amide protons of G23 and F24 into close proximity. When compared to the hVDAC1 structures obtained by solution NMR (Jaremko et al., 2016; Bohm et al., 2020), the observed long-range $^1\text{H-N-}^1\text{H-N}$ contacts both in the N-terminal helical region as well as in the β -barrel are in accordance with the previously published structures.

In solution NMR studies, due to the method's higher sensitivity, the overall number of $^1\text{H-N-}^1\text{H-N}$ contacts was much higher than the 59 contacts identified in this study by ssNMR. Villinger describes 267 contacts (counting two peaks symmetric to the diagonal as one contact), while Schmidt further improved NOE assignments with an additional 450 contacts (Villinger, 2012; Schmidt, 2017). Many of these contacts are redundant in terms of defining the protein's topology, e.g. 198 contacts in the barrel alone were observed between neighboring residues, and ~400 contacts were observed to residues in the opposite β -strands neighboring the hydrogen-bonding partner, or to close by residues in the loop regions. The 6 helix-barrel contacts observed by solution NMR (T6-A141, Y7-L142, Y7-G126, A14-Y195, V17-A206, and F24-L262), however proved crucial for positioning the N-terminus relative to the barrel. Although we could not detect any helix-barrel contacts by ssNMR on the perdeuterated sample used in our study, solid state NMR studies on fully protonated samples had previously shown that A14 C β is in close proximity (~4-6 Å) with S193 C β (Eddy et al., 2015a) and the proximity of residues L10 and V143 (Schneider et al., 2010).

In order to gain more information in the location of the helix within the barrel, we used paramagnetic relaxation enhancement, a method often used for determining inter-domain distances

in proteins. Proximities of up to 35 Å (Clare and Iwahara, 2009) can be identified with the approach, significantly longer than the ~6 Å limit of NOE- or dipolar recoupling-based experiments. An A2C point mutation was introduced at the N-terminus of E73VC127AC232S-hVDAC1 (A2CE73VC127AC232S-hVDAC1), and the cysteine was then coupled to an MTSL spin label. Changes in signal intensity throughout the protein sequence were monitored by recording an hCANH spectrum. The MTSL-labeled protein's fold was the same as that of the unlabeled protein, as was shown by the nearly identical fingerprint (hNH) spectra (**Figure 1-15B**). It was clearly visible even on the relatively overlapped hNH spectra, that some peaks were bleached out completely (e.g. K119), while others' intensity decreased significantly (e.g. R15, G192).

To quantify the peak intensity changes induced by the MTSL label, relative intensities were calculated (I/I_0), and the quotients were then scaled to one. (Scaling was necessary due to differences in rotor packing). **Figure 1-15** shows that the intensities measured in the N-terminal α -helix are overall lower than the intensities in the barrel, which is to be expected from the N-terminal location of the MTSL label. Complete bleaching is observed at the ends of β 7 and β 8, as well as the loop connecting the two strands. This is in perfect accordance with the location of the helix in the previously published structures (Ujwal et al., 2008; Jaremko et al., 2016), where the carbonyl oxygens of A2 and P4 form a hydrogen bond with the H^N of H122 and $N\delta 2$ of N124. Indeed, residues K119-N124 are bleached completely, as well as the nearest residues from the nearby β -strands (G145, K174), and the closest residue in the helix (G11). In this region, all residues bleached completely are closer than 12 Å (distance between the $C\beta$ of A2, which, when mutated to cysteine carries the MTSL label and the H^N of the bleached residue).

Several residues are also bleached close to the ends of strands β 14-16, on the opposite edge of the barrel. These residues are much farther from the first landing site than the ~13 Å limit of bleaching, e.g. R218 is 30.1 Å from the MTSL-labeled A2C residue. Although this could be the indication of a second landing site of the helix, it is most probably stemming from the stacking of the channels in the lamellar 2D crystals. To clarify the origin of this effect, further experiments are needed, where u -[2H , ^{13}C , ^{15}N] and MTSL labeled protein is diluted with unlabeled VDAC to minimize the interaction between the protein molecules.

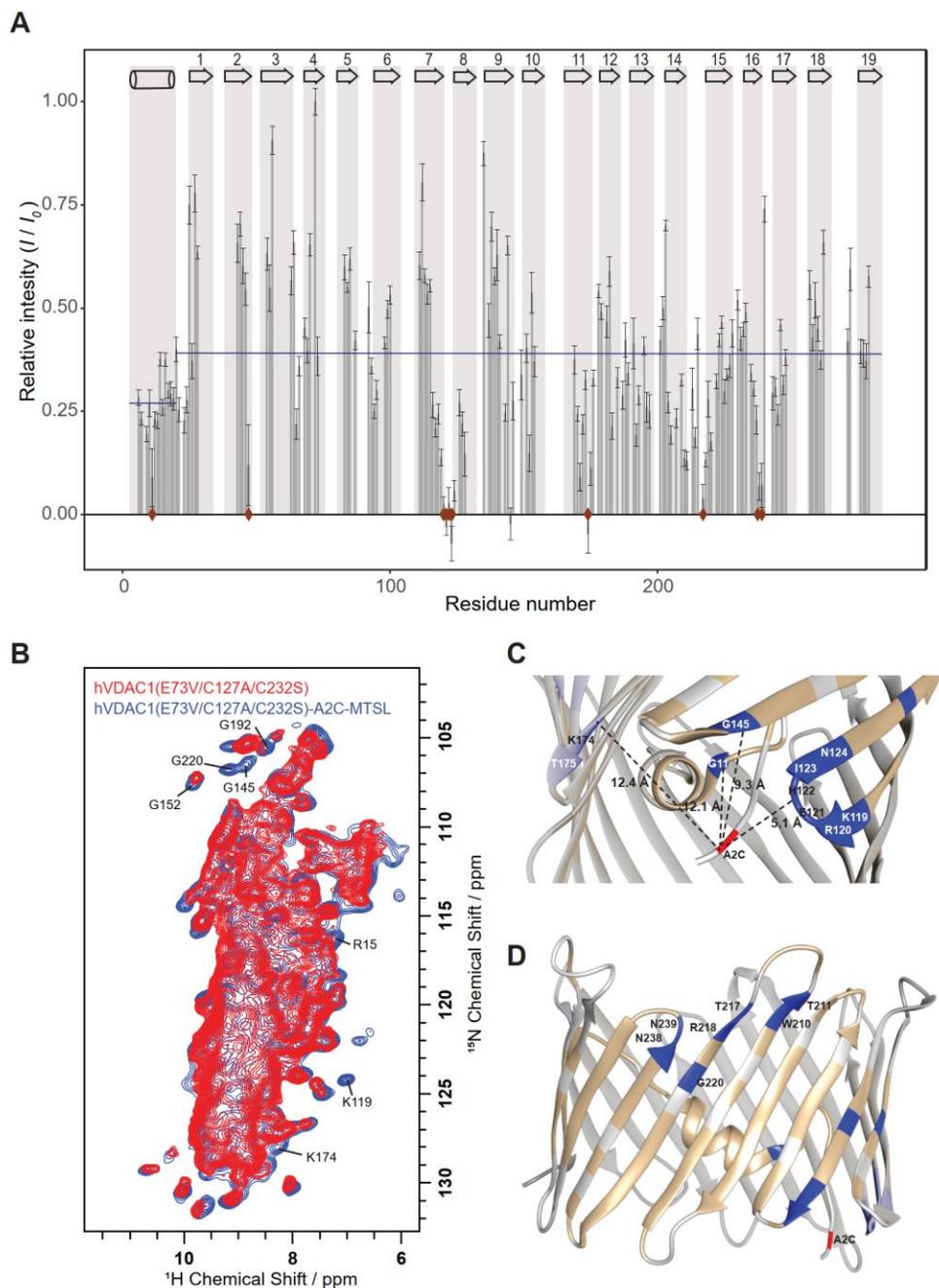


Figure 1-15: Locating VDAC's N-terminal α -helix with MTSL labeling. A) Intensity changes of E73VC127AC232S-hVDAC1 upon MTSL labeling at the A2C position. Relative intensities were scaled to 1. Red diamonds show residues where intensities were below twice the noise level (no peak was observable). In these cases, the intensities at the expected peak positions are displayed. Average relative intensities in the helix and the barrel are shown by blue lines. B) hNH correlation spectrum of hVDAC1 with (red) and without the MTSL label. Resolved peaks broadened beyond detection upon MTSL labeling are assigned. C) and D) Peaks with relative intensities less than $\overline{I_{rel}} - \sigma_{I_{rel}}$ (0.178) are mapped in blue onto PDB 5JDP. The A2C mutation labeled with MTSL is colored in red. Residues that were not assigned or were too overlapped to read off accurate intensity values are colored light grey. C) Landing site of the helix in micelles (Jaremko et al., 2016). Distances to each neighboring loops or β -strand are shown. D) Bleached residues in strands β 14-16.

In order to learn about the water-accessibility of VDAC in a lipid bilayer, and thus gain more information about its tertiary structure, H/D exchange was performed. Several hours are required for the sample preparation (buffer exchange and rotor packing), and since the resolution in a lipid bilayer is too low to gain extensive site-specific information from a 2D hNH spectrum, recording a 3D hCANH spectrum is necessary. Because of this, the time resolution of H/D exchange in ssNMR is significantly lower than in solution, and instead of hours, it is close to a day. Thus, in H/D exchange by ssNMR, only residues that are in particularly slow exchange with the surrounding buffer can be observed.

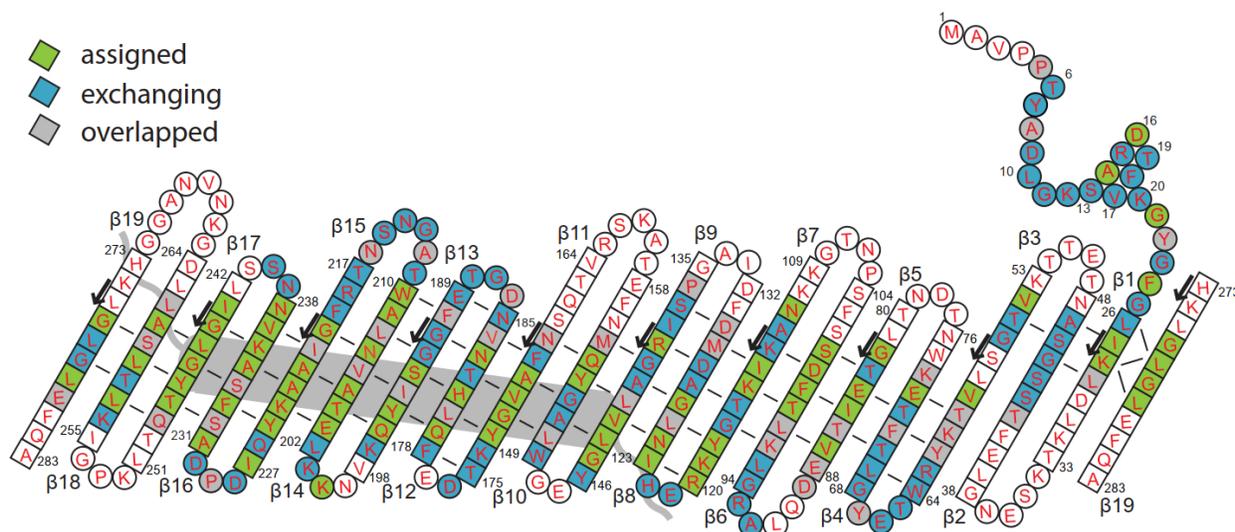


Figure 1-16: H/D exchange of E73VC127AC232S-hVDAC1. Residues undergoing complete exchange ($I < 3\sigma$, where σ is the noise level) in the 1 day timeframe of the experiment are shown in blue, other assigned residues are shown in green. Assigned residues where signal overlap interferes with determining exact signal intensities are shown in grey. The helix running along the barrel wall is shown in grey.

Figure 1-16 shows residues prone to H/D exchange in VDAC. When mapping the exchanging residues onto the lipid-bound topology model, most of the loops show a higher tendency for exchange, which is unsurprising, as they are fully exposed to water outside the lipid membrane, and their amide protons do not participate in strong hydrogen bonds which could slow down the exchange process. On the timescale of the experiment, however, the helical amide protons also exchange almost completely, as well as most of the N-terminal β -barrel ($\beta 1$ - $\beta 4$). This agrees well with previous observations (Bayrhuber et al., 2008; Hiller et al., 2008), where increased exchange rates were found for the N-terminus, and for most residues in $\beta 1$ - $\beta 3$ in wild-type hVDAC1, though

exchange rates in the N-terminal barrel could be somewhat quenched with the E73V mutation (Bayrhuber et al., 2008). The faster H/D exchange of the N-terminal barrel can most probably be explained by less stable hydrogen bonds, as indicated by the elevated dynamics of wild-type hVDAC1 in micelles (Villinger et al., 2010), and the lower signal intensities and higher signal overlap observed in ssNMR spectra of E73VC127AC232S-hVDAC1 in a lipid bilayer (as described in Chapter 1.3.2.).

Apart from a few exceptions (A151, G152), residues shielded by the helix in the barrel show no exchange in strands β 9- β 17. This could be due to the fact that the barrel is more rigid in this region (more stable hydrogen-bonds), and the helix running along the barrel could also hinder the access of water to the barrel wall. To understand even more about water and lipid accessible regions of the protein, we developed new pulse sequences utilizing z-mixing of magnetization (NOE) of mobile residues to the protein, allowing a residue-by-residue mapping of a membrane protein's environment. The method and the results are described in Chapter 2.

1.3.4 Influence of the membrane environment on VDAC

It had been previously shown that VDAC1's structure remains unchanged upon 2D crystal formation in various lipids (Eddy et al., 2012). It had however been previously suggested that some membrane components, such as cholesterol might also be essential for the channels to be fully functional (Pfaller et al., 1985; Hiller et al., 2008). We thus investigated the influence of the membrane environment on VDAC's structure.

Using a total lipid extract from porcine brain, VDAC was reconstituted in a membrane with a near native lipid composition. The hNH spectrum was used as a fingerprint to follow any structural changes related with the membrane environment. The VDAC spectra in single-component phosphocholine lipid DMPC and the total lipid extract from porcine brain extract are shown in **Figure 1-17**. Only minor perturbations can be observed between the two sample conditions, with the overall shape of the spectrum remaining unchanged, indicating identical protein folds. In particular, the peak positions for helical residues labeled in **Figure 1-17** are unchanged, but most resolved peaks from different parts of the barrel (strands β 4, β 6, β 7, β 9, β 10, β 13, β 15, β 16, β 18) also remain unperturbed. The minor chemical shift changes (~ 0.1 ppm, calculated from ^1H and ^{15}N

chemical shifts) observed for β -strand residues are therefore most likely caused by local interaction with the different lipid components of the brain extract. This could also be the reason for the broadening of some peaks, or changes to their shape, though the linewidths of most isolated peaks appear very similar.

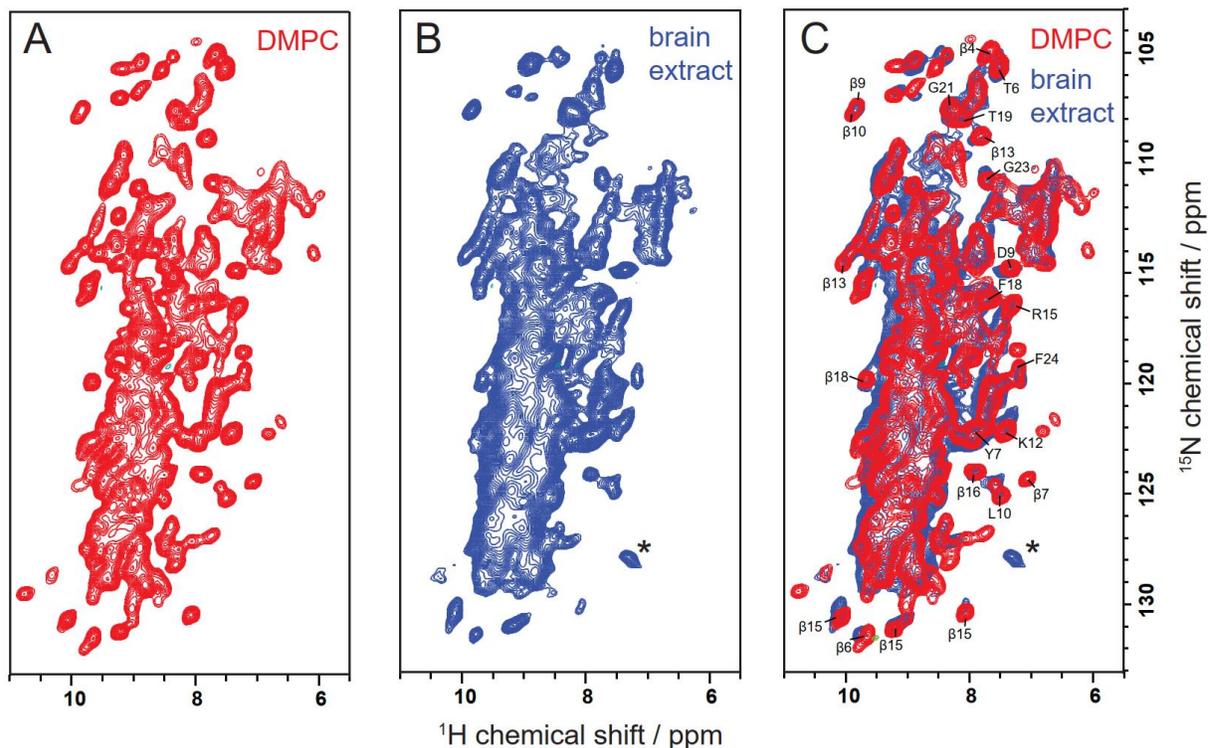


Figure 1-17: hNH correlation spectra of E73V hVDAC1 reconstituted in A) DMPC and B) brain extract. C) shows the overlay of spectra recorded in DMPC and brain extract, as well as assignments for isolated peaks from the α -helix, the kink between the helix and the barrel, and from several β -sheets. Due to the lower spectral width of the spectrum recorded in brain extract lipids, an aliased sidechain appears, marked with an asterisk.

1.3.4.1 Cholesterol binding to VDAC

Since cholesterol was in some cases necessary to obtain normal function of recombinant VDAC (Hiller et al., 2008), and has been shown to bind strongly to the channel (De Pinto et al., 1989), we investigated the interaction of cholesterol and VDAC in a lipid bilayer. Chemical shift perturbations, however, do not always give direct information of proximity between the protein and the binding partner, as structural rearrangements can also have an effect on the chemical shifts further from the binding site (allosteric effect), as was shown for the cholesterol binding of the

mitochondrial translocator protein TSPO, where the structure of the transmembrane regions was perturbed as far as 35 Å from the cholesterol's binding site (Jaipuria et al., 2017). It is possible to observe direct contacts between protein and ligand by employing z - z mixing of magnetization, which correlates nuclei in close spatial proximity. We have developed a method for the residue-specific investigation of a membrane protein's environment by proton-detected ssNMR (Najbauer et al., 2019), which is described in detail in Chapter 2.

The same principle as described in the publication was applied for observing direct contacts between the mobile aliphatic cholesterol sidechains and the protein backbone, and both protein and DMPC lipid components were deuterated to suppress aliphatic signals not stemming from cholesterol (Najbauer et al., 2019). Since cholesterol was fourfold less abundant than DMPC in the original study, the 4D HhnCANH pulse sequence was modified to HhCANH, compromising some spatial resolution (of accessibility) for sensitivity. In the latter, one less CP transfer step is applied: the two shorter distance, thus more selective H-N then N-CA magnetization transfer steps are substituted with a longer-range, but direct H-CA magnetization transfer. This slight loss in resolution, however, does not substantially affect the structural information gained from the experiment.

We could identify three of the binding sites predicted through docking (Weiser et al., 2014). The approximate placement of cholesterol in these binding modes is shown in **Figure 1-18A-B**, with the cholesterol molecules binding to hydrophobic grooves on the barrel wall formed by the side chains of Leu, Ile, Phe, Tyr, Trp, Val, and Ala. Contacts with cholesterol are indicated as circled residues in **Figure 1-18C-E**. Good agreement is seen with the predicted contacts based on molecular docking (shaded in blue) in **Figure 1-18C-E** for 3 cholesterol binding sites. There are in total 6 contacts in accordance with sites shown on **Figure 1-18 C** and **D**, whereas the third possible binding site shows only 1 contact, possibly explained by a lower population.

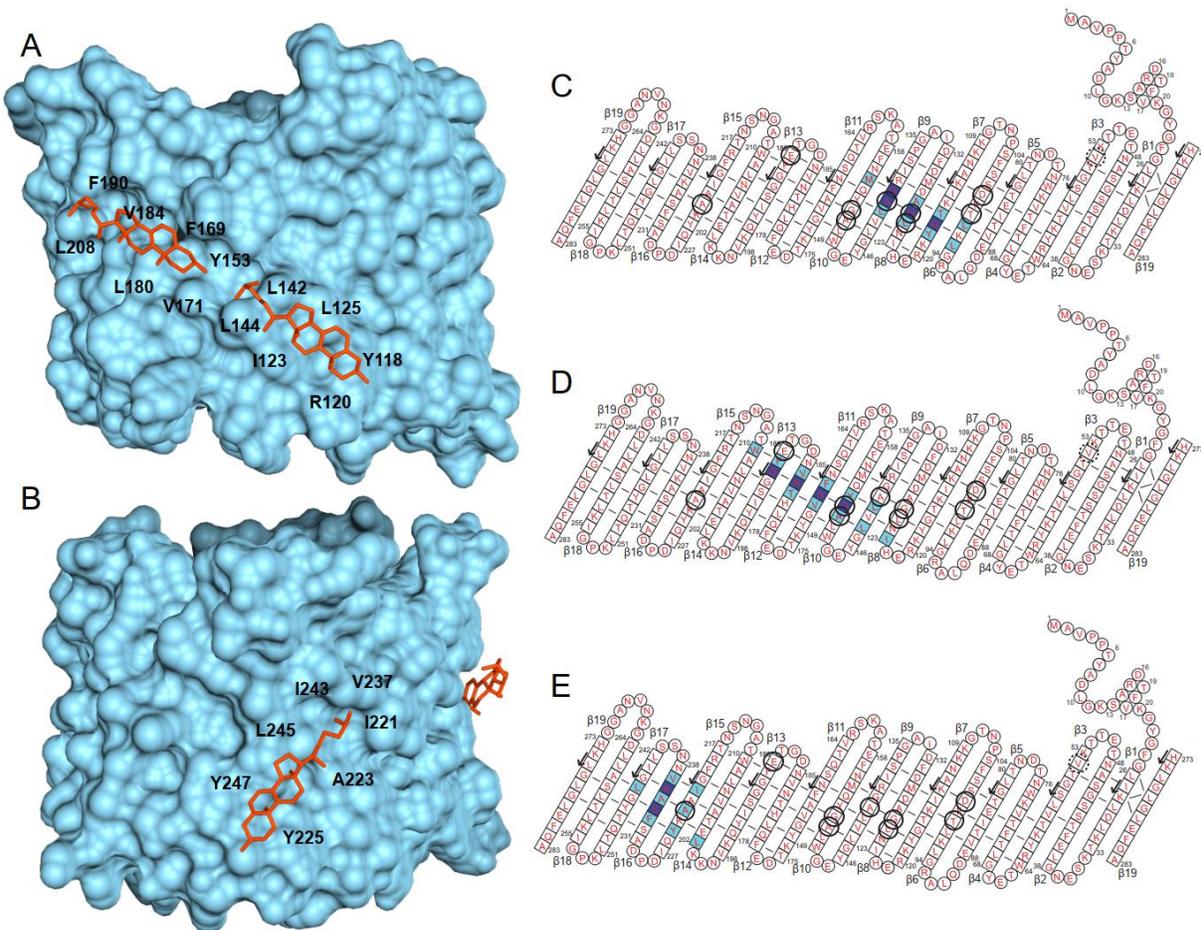


Figure 1-18. Interaction of VDAC with cholesterol. **A and B** show the three cholesterol binding sites of VDAC identified by solid-state NMR spectroscopy. These are a subset of the five binding sites predicted by docking (Weiser et al., 2014). The hydrophobic sidechains forming the grooves on the side of the barrel are labeled. **C, D,** and **E** show the cholesterol aliphatic chain– protein amide contacts observed by solid-state NMR circled in black. Backbone and sidechain contacts predicted by docking are colored in dark and light blue, respectively.

Signal intensities are an excellent indicator of mobility in solid-state NMR. We thus plotted the relative intensities ($I_{\text{cholesterol-bound}} / I_{\text{unbound}}$) measured in the hCANH spectrum of the cholesterol-bound sample and in the unbound sample (**Figure 1-19**). We found that relative signal intensities of residues in the barrel averaged ~ 1.5 higher compared to the N-terminal helical region, a clear indication of cholesterol’s stabilizing effect on barrel dynamics. Furthermore, the relative intensities of $\beta 7$ – $\beta 13$, strands with 6 of the 7 identified cholesterol contacts, were all somewhat higher than average. This is in accordance with our observation that most probably sites shown on **Figure 1-18C** and **D** are more populated than the one shown on **Figure 1-18E**, and the cholesterol bound to these two sites stabilizes the contacting β -strands.

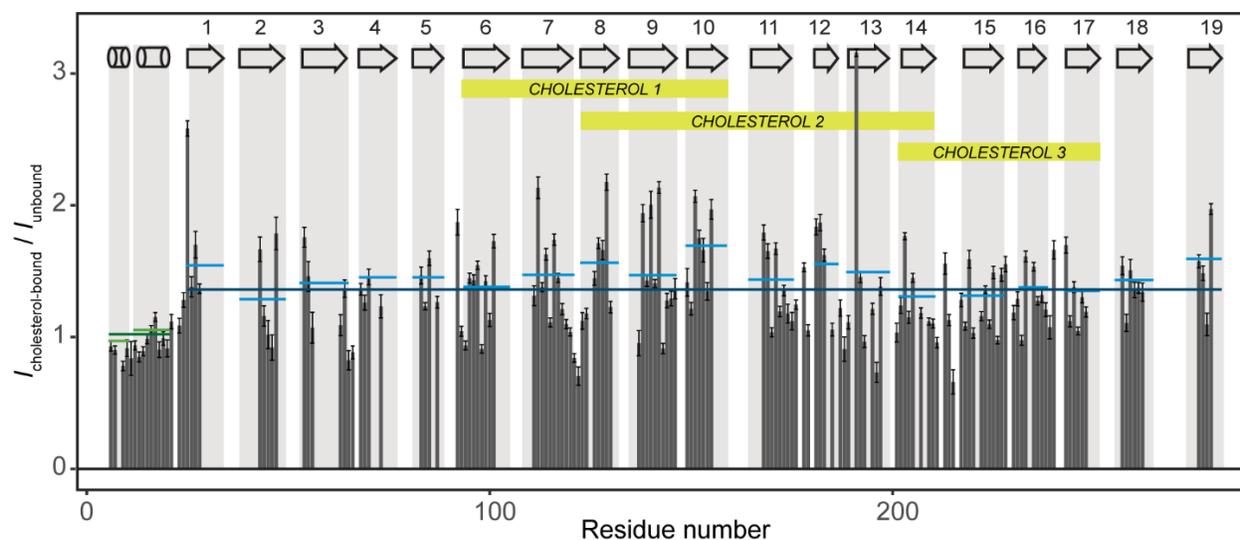


Figure 1-19: Change of peak intensities upon binding of cholesterol to E73V/C127A/C232S-hVDAC1 at a 1:5 protein:cholesterol ratio. Relative intensities ($I_{\text{cholesterol-bound}} / I_{\text{unbound}}$, no scaling applied) are shown in function of residue number. The average relative peak intensity of each secondary structural element is shown in colored lines: green – α -helix, blue – β -strand. The average relative peak intensity for the whole of the N-terminal helical region, as well as the barrel has been calculated, and is shown as a longer line. Secondary structural elements are indicated with cylinders (α -helix) and arrows (β -strand) in the grey rectangles showing the extent of each structural element. The extent of the cholesterol binding sites, as shown by z - z mixing, is marked in yellow. Site 1 corresponds to C, site 2 to D, and site 3 to E of **Figure 1-18**.

1.3.5 VDAC's interaction with G3139

The 18-mer phosphorothioate oligonucleotide G3139 (TCTCCCAGCGTGCGCCAT, the individual nucleotides linked by thiophosphate instead of phosphate bonds (PubChem, 2019)) has been shown to selectively block the VDAC channel, and although the mechanism is unclear, a partial entry into the pore was suggested. 2D crystals of E73VC127AC232S-hVDAC1 were incubated at a 1:1 molar ratio of G3139 solution, and chemical shift perturbations, as well as intensity changes were determined by measuring an hCANH spectrum, as well as the linking hcoCAcoNH experiment to confirm assignments. No chemical shift perturbations >0.2 ppm (calculated from H, ^1H N and $\text{C}\alpha$ shifts, see Methods) were observed for residues that could be unambiguously identified, however chemical shift perturbations larger than 0.1 ppm were identified in strands β 13 (G191, G192) and β 14 (T211). Even more striking was the fact that several residues disappeared from the hCANH spectrum, shown on **Figure 1-20**. Residues, whose intensities were most affected were G11, T19, and K20 in the α -helix, G23 and G25 in the kink between the helix and the barrel, and residues N238, N239 and I243, close to/in the loop between

strands β 16 and β 17, on the cytosolic side of the channel (Tomasello et al., 2013). A significant decrease in peak intensities was also observed in strands β 12 (F178, Q179, T182), β 13 (G192), and β 14 (N207). With the exception of T70, and S43, a residue whose assignment was ambiguous in the G3139-bound state, residues most affected are on the side of the barrel contacting the helix, suggesting a binding site at the helix, and the neighboring, C-terminal strands of the barrel. As discussed in more detail in Chapter 1.4.6, this binding site overlaps very well with the interaction sites identified for other nucleotides (ATP (Villinger, 2012), GTP, UTP (Villinger et al., 2014), β -NADH (Bohm et al., 2020)).

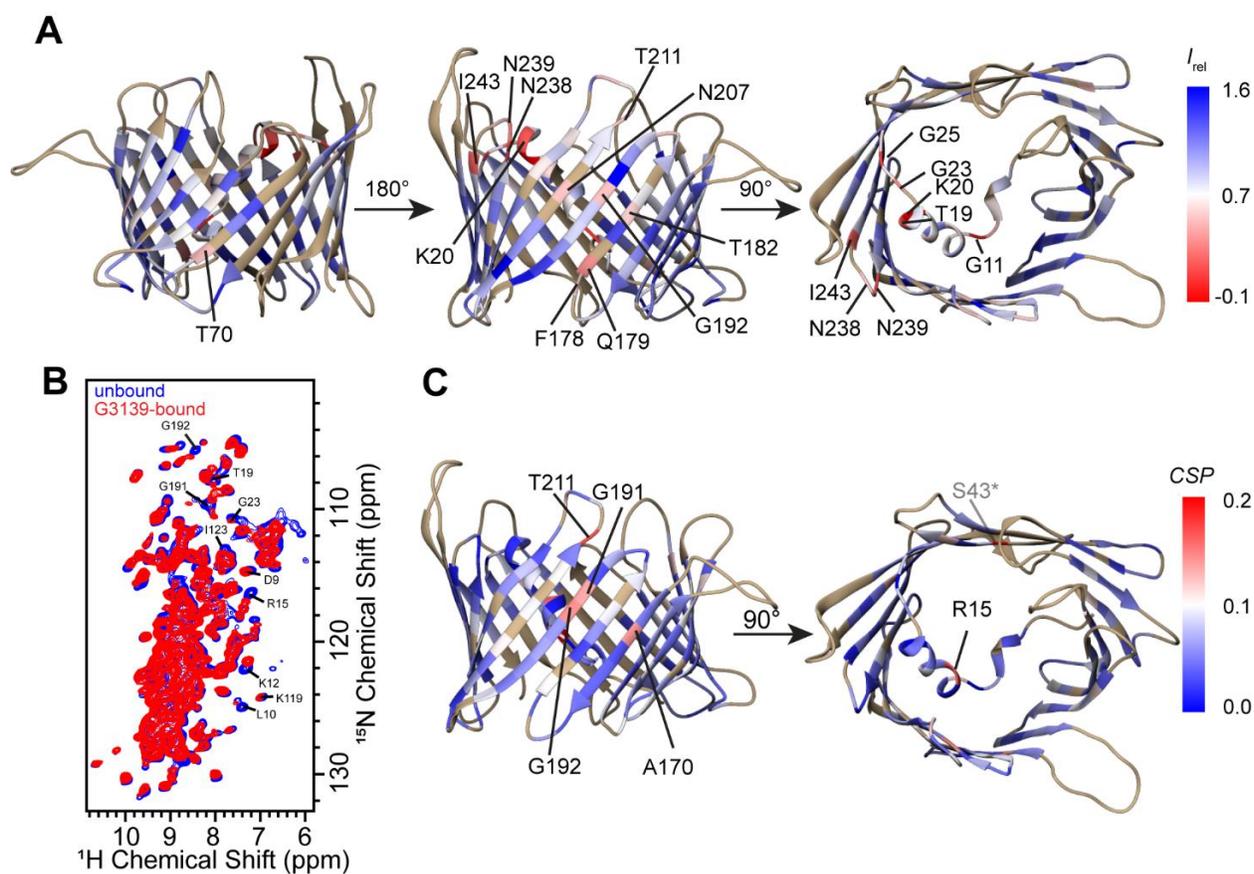


Figure 1-20: Binding of G3139 to E73V/C127A/C232S-hVDAC1. A) Intensity changes occurring upon binding are plotted onto the structure of E73V-hVDAC1 (PDB: 5jdp) with the Chimera software’s “Render by attribute” option. Relative intensities were calculated from the intensity of peaks in the hCANH spectra ($I_{\text{G3139-bound}} / I_{\text{unbound}}$). If no peak could be identified, a peak was placed at the expected peak position. In the color scheme, blue corresponds to the highest relative intensity (1.5), and red to minimal relative intensity (-0.1). B) Overlay of hNH spectra of the unbound state (blue) and G3139-bound state (red). C) Chemical shift perturbations mapped onto the E73V-hVDAC1 structure (PDB: 5jdp). Blue corresponds to no perturbation, red shows perturbed residues (maximal observed perturbation was 0.2 ppm). Residue S43 (marked with an asterisk) shows a large perturbation, however the assignment for this residue in the bound state was uncertain.

1.3.6 Investigation of VDAC's closed state through point mutations

Investigation of VDAC's gating mechanism and determining its closed state poses a problem for all structural biology methods, as the measurement would have to take place under applied voltage. This can be circumvented however with the investigation of VDAC mutants permanently exhibiting the channel's low-conductance state. Such a mutant was initially suggested by Geula et al. (Geula et al., 2012), where the G21V/G23V mutations were shown to partially lock the channel into a closed-state conformation. These mutations served as a basis for developing the closed-state G21V/G23V/E73V/C127A/C232S quintuple mutant by introducing mutations necessary for sample stability. The electrophysiology of this mutant was shown to be identical to that of the G21V/G23V mutant, and its structure in micelles was investigated in detail by Schmidt (Schmidt, 2017), using solution NMR spectroscopy.

Schmidt suggested that the helix moves away from its original position along the barrel wall, and observed an alternative landing site using NOE measurements at the hydrophobic patch around L242 and L262. To further force the dissociation of the helix from its open state position on the barrel wall, mutation of the hydrophobic contact sites V143 and L150 to alanine was suggested, resulting in the G21V/G23V/E73V/C127A/V143A/L150A/C232S septuple mutant, although its electrophysiological properties had not previously been tested.

To assess sample quality, uniformly [^{13}C , ^{15}N]-labeled samples of the closed-state quintuple and septuple mutants were prepared and compared to the open-state triple mutant (E73VC127AC232S-hVDAC1) under identical experimental conditions. Unlike the sample of the triple mutant, no 2D crystalline patches were observed with negative stain EM for the quintuple or the septuple VDAC mutants.

Interestingly, despite the lack of 2D crystals, the fully protonated quintuple mutant showed some resolved peaks with narrow, ~ 0.5 ppm ^{15}N linewidths in the hNH spectrum. (**Figure 1-21**) These peaks were only slightly broader than the ~ 0.4 ppm lines of the 2D crystalline triple mutant. This broadening could be a result of lower sample quality due to the absence of 2D crystals, or could stem from an increase in mobility. The septuple mutant on the other hand showed no isolated peaks, and very low overall spectral resolution. Thus the G21V/G23V/E73V/C127A/C232S quintuple mutant was chosen for investigation of the closed state.

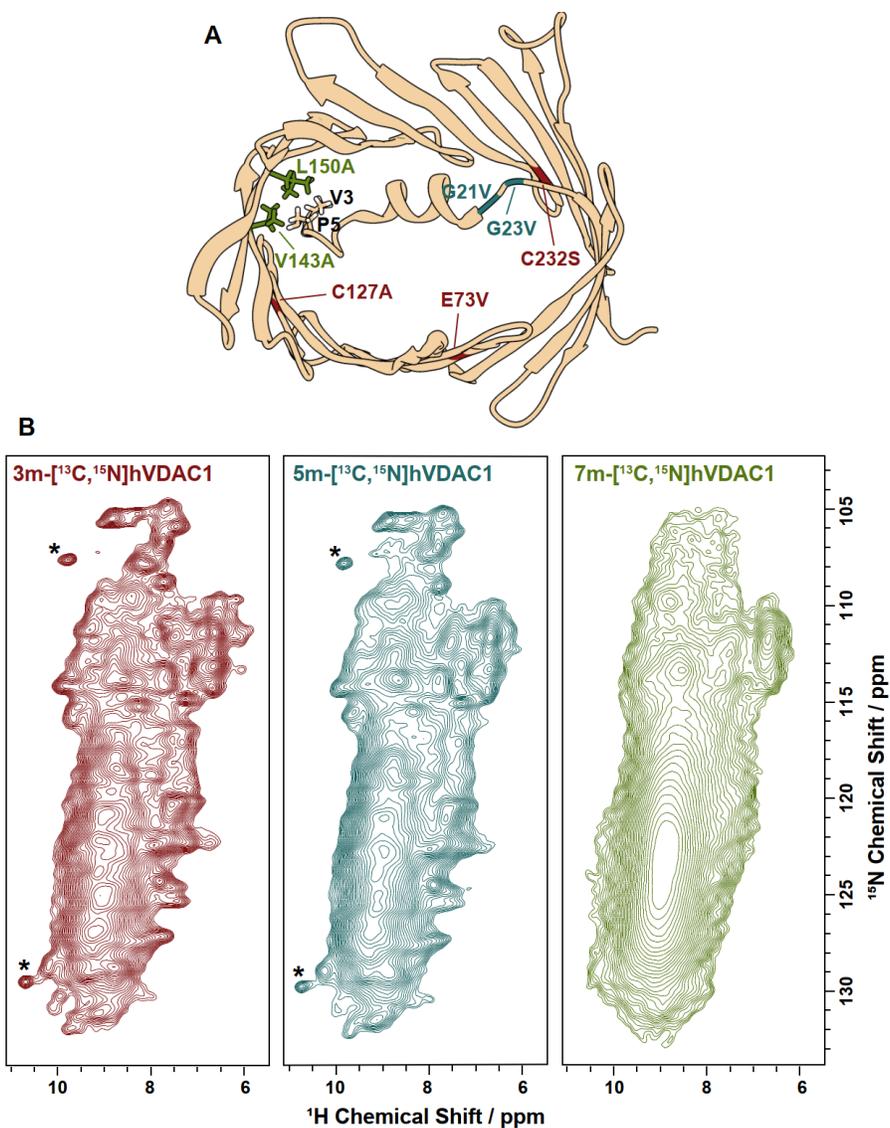


Figure 1-21: Mutants of hVDAC1 and their sample quality. A) Mutations introduced to hVDAC1. E73V/C127A/C232S mutations in the barrel yielding the triple mutant (red); G21V/G23V were further introduced in the kink between the helix and the barrel in the G21V/G23V/E73V/C127A/C232S-hVDAC1 quintuple mutant (blue); V143A/L150A mutations introduced to the quintuple mutant, yielding the septuple mutant (7m) G21V/G23V/E73V/C127A/V143A/L150A/C232S-hVDAC1 (green), destroying the hydrophobic landing site of the helix on the barrel wall. Hydrophobic sidechains partaking in the interaction are shown. B) hNH correlation spectra of the VDAC mutants. All spectra were recorded at 800 MHz, 55 kHz MAS on uniformly [^{13}C , ^{15}N]-labeled samples. hNH correlation spectra of open-state, triple mutant (3m) E73V/C127A/C232S-hVDAC1 (red); closed-state quintuple mutant (5m) G21V/G23V/E73V/C127A/C232S-hVDAC1 (blue); closed-state septuple mutant (7m) G21V/G23V/E73V/C127A/V143A/L150A/C232S-hVDAC1 (green) ^{15}N linewidths were determined from the peaks marked with an asterisk.

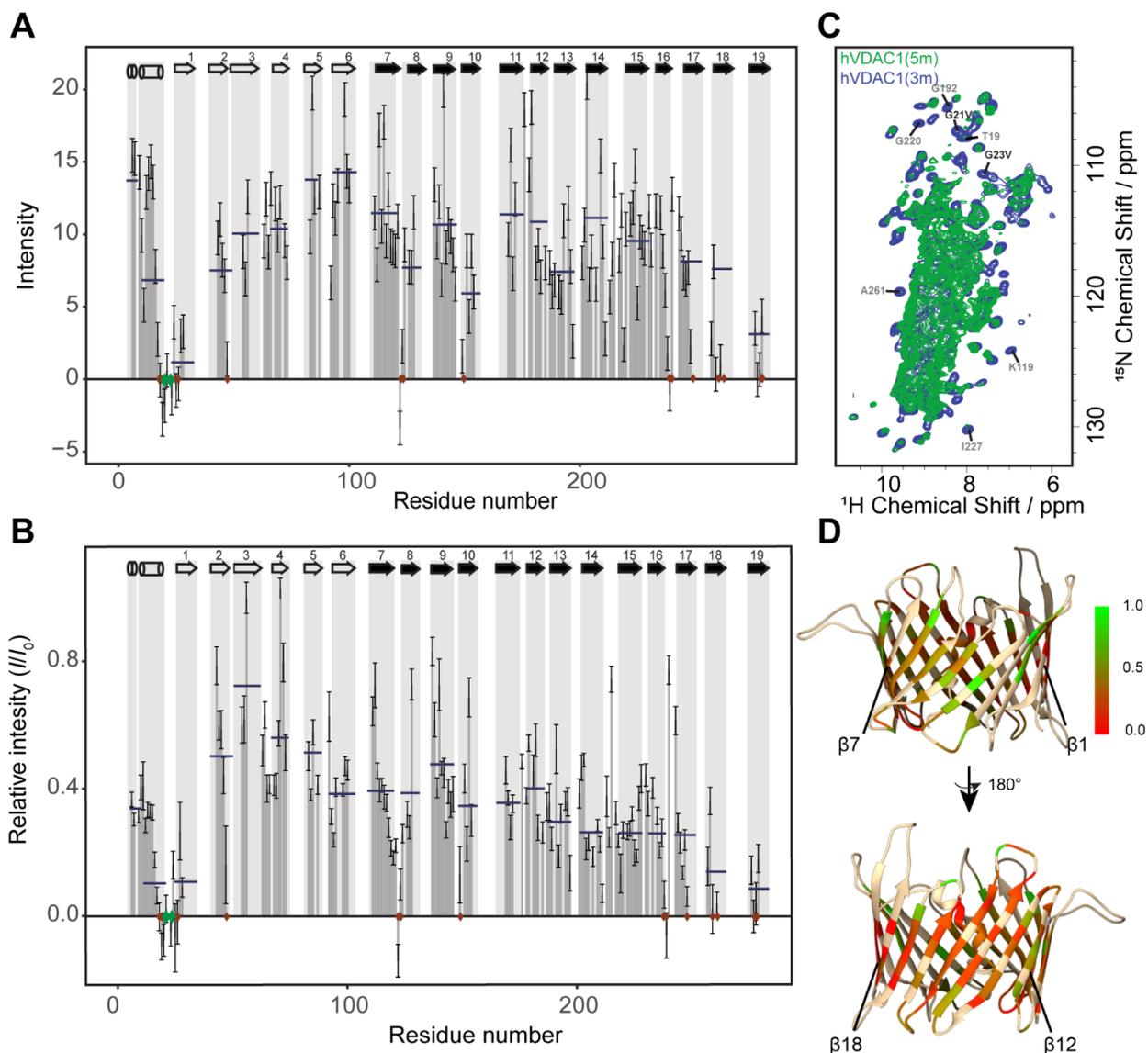


Figure 1-22: Closure of the VDAC channel - effect of the G21V/G23V mutations. A) Peak intensities in the hCANH spectrum of the the G21V/G23V/E73V/C127A/C232S-hVDAC1 quintuple mutant. B) Relative intensity changes in the hCANH spectra upon introduction of the G21V/G23V mutations (I denotes intensities of the quintuple G21V/G23V/E73V/C127A/C232S-hVDAC1, I_0 the intensities of the E73V/C127A/C232S-hVDAC1 triple mutant). Relative intensities are scaled such that the greatest value is 1 (scaling factor: 0.41). Overlapping residues are not plotted in A) or B) due to the inaccuracy of intensities. Green diamonds show the position of the two mutations, red diamonds mark residues where the peak's intensity in the quintuple mutant dropped below 2σ , where σ is the noise level. Blue lines show the average relative intensities in the N- and C-terminal part of the α -helical region (separated by the kink at G11), as well as in the β -strands. β -strands with the helix running in front are marked by black arrows. C) Overlay of the hNH spectra of the closed-state quintuple (green) and the open-state triple mutant (blue). The peaks of G21 and G23 mutated to V are labeled in black, and as an example some residues disappearing from the spectrum have been assigned (grey). D) Relative intensities plotted onto the triple mutant structure (PDB: 5jdp). Red corresponds to a relative intensity of 0, green to a relative intensity of one. The side of the barrel in contact with the helix is shown on the bottom.

Further investigations on the quintuple mutant were conducted on a uniformly [^2H , ^{13}C , ^{15}N]-labeled sample. The quintuple mutant's hNH spectrum was found to overlay excellently with that of the triple mutant, with no significant changes in chemical shifts, and as expected, signals belonging to G21 and G23 could not be observed. It was however quite surprising that in the quintuple mutant's hNH spectrum, many peaks were of low relative intensity or had disappeared entirely from the spectrum (**Figure 1-22C**). An hCANH spectrum was recorded to quantify any changes in intensity. Interestingly, peaks from the $\alpha 2$ helix and strands $\beta 7$ -17 were no longer higher in intensity than the rest of the barrel (**Figure 1-22A**, cf. **Figure 1-11**). The region with the highest relative intensities included strands $\beta 2$ -5 (located on the opposite side of the barrel where the N-terminal helix runs), while $\beta 18$ and 19, as well as the C-terminal part of the helix ($\alpha 2$), and the neighboring $\beta 1$ were almost completely broadened beyond detection. $\beta 7$ -17, the strands neighboring the helix in the open-state structure were also significantly reduced in intensity, as well as the N-terminal part of the helix ($\alpha 1$) (**Figure 1-22D**).

In an attempt to identify the helix's landing site in the closed state, the quintuple mutant was prepared with an MTSL label at the A2C position. The protein, however, was different from other samples upon macroscopic observation: the density of the sample appeared higher, and a film was formed at the surface, in contrast to the previous puffy, loose lipid crystalline samples. The MTSL-labeled quintuple mutant gave spectra of low quality, showing broad lines and perturbed peak positions (**Figure 1-23**).

The broader lines and the overall low spectral quality could well be a result of the helix sampling a large number of conformations inside the barrel, bringing the MTSL label into close proximity with a large number of residues. The perturbed peak positions could, however, possibly be indicators of an altered protein fold. This could be caused by the introduction of the hydrophobic MTSL label to an already dynamic structure, where the label could easily interact with ("stick to") various hydrophobic patches in the barrel. Upon reducing the MTSL-labeled quintuple mutant with ascorbic acid, the original spectrum was not fully recovered, a further indication of perturbation of the protein by the label, and even more surprisingly several peaks disappeared. As expected though, some previously bleached peaks became once more observable (D230, G244), providing some support that the MTSL-labeled helix indeed has a binding site at the hydrophobic patch of $\beta 17$.

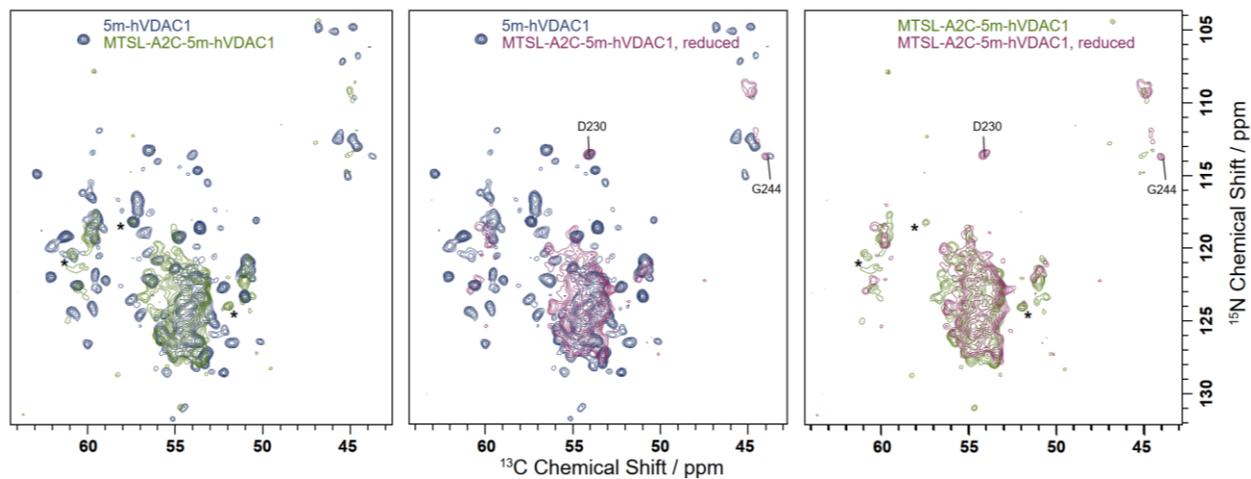


Figure 1-23: MTSL-labeling of (A2C)/G21V/G23V/E73V/C127A/C232S-hVDAC1 (5m-hVDAC1) The ^{13}C - ^{15}N projections of hCANH spectra are shown for the unlabeled quintuple mutant (blue), the quintuple mutant coupled to an MTSL label at the A2C position (green), and the latter reduced with ascorbic acid (purple). Peaks disappearing upon reduction of the sample are labeled with an asterisk.

1.4 Discussion

1.4.1 Comparison of detergent and membrane structures

The relevance of refolded mammalian VDAC's 19-stranded β -barrel structure as determined by X-ray crystallography (Ujwal et al., 2008; Hosaka et al., 2017) and NMR spectroscopy (Bayrhuber et al., 2008; Hiller et al., 2008; Jaremko et al., 2016; Bohm et al., 2020) has long been a matter of debate. Even as recently as 2016, a barrel incorporating 13 β -strands as well as an α -helix has been proposed as the channel's functional structure based on biochemical studies (Colombini, 2016). Several lines of evidence have indicated however that the structures determined in bicelles and micelles were fully functional. Ujwal et al. had shown that mVDAC1's gating profile after membrane insertion was in good agreement with that of endogenous VDAC1 channels isolated from rats (Ujwal et al., 2008). In her thesis, Villinger also described fast and functional insertion of refolded hVDAC1, suggesting that the spontaneous insertion of the channels is incompatible with a large-scale rearrangement required to form the structure determined from various biochemical studies by Colombini (Villinger, 2012). Hiller et al. however showed that hVDAC1 did not acquire full function upon membrane insertion in the absence of Triton X-100 and cholesterol (Hiller et al., 2008). Other studies have also demonstrated that a membrane environment is required for VDAC to achieve full functionality (Shanmugavadivu et al., 2007), underlining the necessity of investigating VDAC's structure in a native-like lipid bilayer environment.

The present study yields a detailed insight into VDAC topology in a lipid bilayer. Our results show good agreement with previous structural information obtained by X-ray crystallography, as well as solution- and ssNMR spectroscopy.

1.4.2 VDAC assignments

In previous solution NMR studies, relatively complete assignments of VDAC have been achieved for both wild-type (wt) hVDAC1 (Villinger, 2012) and the E73V mutant (Schmidt, 2017; Bohm et al., 2020). Recently, Bohm et al. increased assignments to 88% for wild-type VDAC, and to

91% for the E73V mutant in LDAO detergent micelles. They found that with the exception of some loop residues (broadened out due to exchange), all residues in the barrel could be confidently assigned (**Figure 1-24**). The assignment of the N-terminal region still remains somewhat problematic, however, here only 12 out of the 15 helical residues (T6-K20) could be assigned, an indication of elevated exchange in this region in micelles.

Membrane proteins of ~30 kDa and above are challenging systems to investigate by solid-state NMR. To date, using only ssNMR data, the assignment and structure of only two beta barrel membrane proteins in lipids is available (Retel et al., 2017; Schubeis et al., 2020a). The bottleneck in the assignment process is the usually insufficient sample quality resulting in severe overlap for large proteins. This was previously also the case for VDAC, where DMPC liposome samples only allowed assignment of most of the α -helix (up to V17), with the exception of a few barrel residues that could clearly be identified with selective labeling (Schneider et al., 2010). The introduction of a 2D crystalline sample preparation (Dolder et al., 1999) in DMPC lipids for the solid state NMR study of VDAC was however a real breakthrough, as these samples gave spectra of exceptional resolution (Eddy et al., 2012). With a ^{13}C -detected assignment strategy on at least four selectively labeled 2D crystalline samples, Eddy et al. could assign not only the entire N-terminus, but also stretches from 7 different β -strands, a total of 88 residues (31% of all residues) (Eddy et al., 2015b) (**Figure 1-24**).

Using the same 2D crystalline sample preparation on $u\text{-}[^2\text{H},^{13}\text{C},^{15}\text{N}]\text{-hVDAC1}$ (the E73V/C127A/C232S mutations were introduced to improve sample stability and spectral quality), and following a proton-detected strategy for assignment (Barbet-Massin et al., 2014; Fraga et al., 2017; Fricke et al., 2017), we assigned a total of 194 residues (69% of the protein) from a single sample (the $\text{H}\alpha$ -labeled sample was used only to confirm assignments). This is comparable to the assignment completeness recently achieved in membrane proteins of similar size in lipid bilayers: 60% in the 34 kDa outer membrane protein G (OmpG) using a total of 10 amino-acid-type selective labeling schemes (Retel et al., 2017), and 84% in the 24 kDa membrane protein AlkL (Schubeis et al., 2020a). Using only automated assignments 58% of the protein could be assigned. Although this is lower than the 82% achieved on the microcrystalline 42.5 kDa maltose-binding protein (MBP) (Stanek et al., 2020), it is reasonable considering that VDAC is in a membrane environment.

Similar to previous solid state NMR studies, the N-terminal helix could be assigned confidently, although the first two residues assigned by Eddy et al are missing, due to prolines breaking the backbone walk, and separating only a stretch of three residues from the rest. More significantly, residues from all 19-beta strands could be assigned, giving a comprehensive picture of the β -barrel in a lipid bilayer. As discussed previously by Eddy et al. (Eddy et al., 2015b), although loops often disappear from ssNMR spectra due to their mobility, 2D lipid crystalline samples tend to introduce rigidity to these regions through the tight packing of the molecules. Indeed, we were able to assign residues from a total of 8 loops in the barrel (**Figure 1-24**).

Comparison of ^{13}C -detected solid state NMR assignments given by Eddy et al. for wild-type hVDAC1 (Eddy et al., 2015b) and our proton-detected assignments of E73V/C127A/C232S-hVDAC1, both in a DMPC lipid bilayer, show large discrepancies for 26 residues (chemical shift perturbation >2 ppm on ^1H N or >1 ppm on $\text{C}\alpha$) (**Figure 1-24**). These residues do not cluster in any particular region, but are scattered evenly across the barrel. This is quite surprising, as with the exception of strands around the E73V mutation site (residues ~ 40 -100), no major chemical shift perturbations were observed in micelles when comparing the wild type and the E73V mutant of the protein (Bohm et al., 2020). Large heteronuclear (especially ^{13}C) chemical shift perturbations are also not expected upon slight changes in experimental conditions such as temperature or buffer composition. These, by perturbing hydrogen bonding, would mostly influence only ^1H chemical shifts, and, to a lesser extent ^1H N shifts, as they are directly bound to the amide protons. A possible explanation is that in a 2D crystal, where the molecules are tightly packed against each other, the E73V mutation could also affect other regions of the neighboring β -barrels. Of course another possible reason for chemical shift differences could be occasional misassignment. Since all perturbed residues could be assigned in the highly unambiguous 4D hCACONH and hcoCACONH spectra, it is possible that the increased overlap in the 3D spectra used in the ^{13}C -detected assignment strategy led to ambiguities.

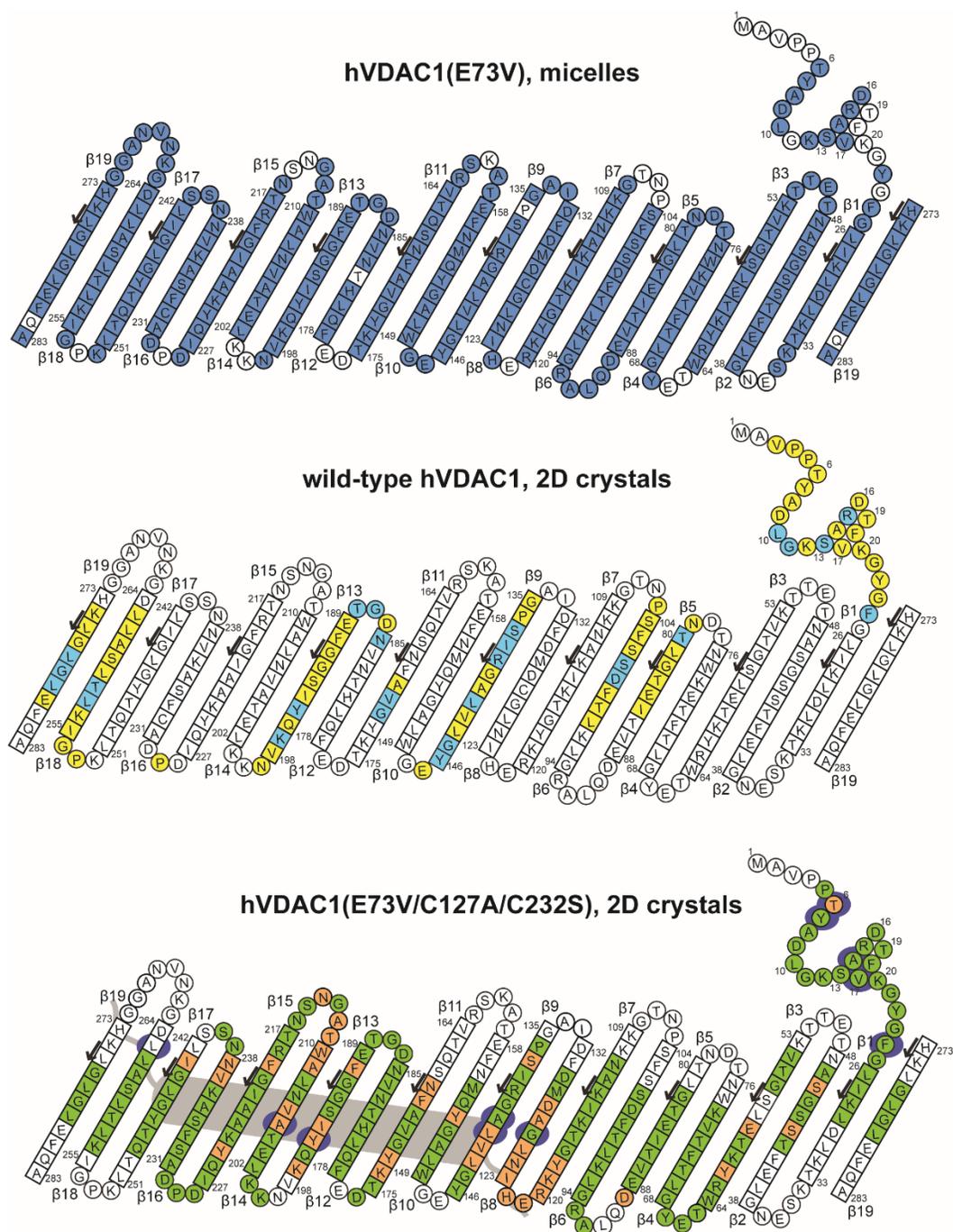


Figure 1-24: Extent of VDAC assignments. Top, blue: E73V-hVDAC1 assignments in LDAO micelles, solution NMR (Bohm et al., 2020). Middle, yellow: assignments of wt-VDAC in 2D crystals, DMPC, ssNMR (Eddy et al., 2015b). Residues colored in light blue show significant chemical shift perturbations compared to our assignments of E73V/C127A/C232S-hVDAC1 in DMPC 2D crystals. Bottom, green: E73V/C127A/C232S-hVDAC1 assignments, in 2D crystals, DMPC, ssNMR. Peach color indicates chemical shift perturbations larger than 0.2 ppm (calculated from H, ^1H N, C α , and C β shifts, weighted by 1, 0.17, 0.3, and 0.3, respectively) (Long and Yang, 2009; Williamson, 2013). These were calculated from our assignments in 2D crystals, and assignments in LDAO micelles (Schmidt, 2017). Squares depict β -sheet, circles helical and loop residues at the N-terminus, and between strands, respectively. Colored fill indicates assigned residue. Blue circles indicate helix-barrel ^1H N- ^1H N contact sites, as determined by solution NMR (Schmidt, 2017). The location of the helix, as determined in solution, is shown in grey.

Our solid state NMR assignments, though not as complete as those achieved in solution, present a significant step forward towards the residue-specific characterization and structure determination of the protein. The extent of the assignments allows an all-round view of the protein, as assignments from all secondary structural elements are available, and allow determination of the protein topology in a lipid bilayer.

To pinpoint any differences between VDAC structure in solution and solid state, we calculated chemical shift perturbations between assignments of E73V/C127A/C232S-hVDAC1 in 2D DMPC lipid crystals and in LDAO micelles. Chemical shifts matched well overall, with 41 residues showing perturbations larger than 0.2 ppm, the largest values being 1.15 ppm (F169), 0.70 ppm (Y62), and 0.65 ppm (R120 and A134). These 41 residues are located almost exclusively in strands which have the helix running next to them, close to the helix-barrel contact sites determined by solution NMR (**Figure 1-24**). The loop between strands $\beta 7$ and $\beta 8$, where two hydrogen bonds between the carbonyl oxygens of A2 and P4 and the ^HN of H122 and the ND2 of N124 were identified (Ujwal et al., 2008), was particularly affected by the change of environment. The chemical shift perturbations could be a result of helix-barrel contacts being influenced by the binding of detergent molecules, which are known to bind to hydrophobic regions (Yang et al., 2014). The particularly large ^HN and ^HN chemical shift changes of residues I123 and N124 (^HN : 0.32 ppm, and 0.53 ppm, respectively, ^HN : 4.9 ppm in both cases), could be an indication of not only an altered hydrogen bonding between the protein's N-terminus to the loop (Williamson, 2013), but also a conformational change in this protein region (significant $\text{C}\alpha$ and $\text{C}\beta$ perturbations (R120 $\text{C}\beta$: 3.2 ppm, I123 $\text{C}\alpha$: 1.3 ppm, N124 $\text{C}\beta$: 1.5 ppm)).

1.4.3 VDAC's structure in a lipid bilayer

Even though the N-terminus was postulated to play a role in voltage gating (Hiller and Wagner, 2009), its structure was controversial even after publication of the first structures of VDAC. This was mostly due to difficulties in assignment of the helical residues in solution: the initial NMR structure had only residues T6-L10 assigned (Hiller et al., 2008) and even the most recent assignments are missing G11, and F18-K20 (Schmidt, 2017; Bohm et al., 2020). The first three

structures of VDAC had three different conformations for the N-terminus: not determined (Hiller et al., 2008), an α -helix formed by residues Y7-V17 (Bayrhuber et al., 2008), and two α -helical segments extending from T6-K20, broken by L10 and G11 (Ujwal et al., 2008) - this kinked helical structure was seen in other solution NMR studies as well (Jaremko et al., 2016; Bohm et al., 2020). As the necessity of a native-like environment on VDAC function was realized early on (Shanmugavadivu et al., 2007), solid-state NMR studies on VDAC in a lipid bilayer have been conducted in parallel to studies in a detergent environment. Schneider et al. recorded assignment spectra on samples in liposomes to determine the native conformation of the N-terminus. Dihedral angle predictions based on chemical shift assignments revealed two helical segments at residues 7-8 as well as 12-17, separated by a kink formed by D9-G11 (Schneider et al., 2010). In 2D crystals, with the exception of G11, Eddy et al. found good agreement with torsion angles predicted from liposome assignments, as well as with the assignments of Hiller et al. in LDAO micelles (Eddy et al., 2015a).

To compare secondary structure in our preparation of E73VC127AC232S-hVDAC1, the 2D crystalline preparation of wild-type hVDAC1 (Eddy et al., 2015a), and the micelle preparation of E73V hVDAC1 (Jaremko et al., 2016), phi and psi torsion angles were predicted from chemical shift assignments using TALOS-N (Shen and Bax, 2013). Torsion angles predicted from our assignments and from the assignments for the E73V mutant in micelles matched well overall in the N-terminus. The only residues showing a large difference in torsion angles ($>100^\circ$) are G11 (both phi and psi) and S13 (**Figure 1-25**, **Figure 1-26**). For both these residues, however, the only nucleus assigned in solution was $C\alpha$, and (especially) the lack of $C\beta$ information makes them poor predictors of secondary structure. The torsion angles predicted from our assignments also overlap well with those obtained from 2D crystalline samples of wild-type VDAC. The only significant differences are in the torsion angle psi for residues T6 (at the edge of the helical region), and for L10 and G11, the residues expected to be at the kink between the two helices (Ujwal et al., 2008; Schneider et al., 2010). Although this could be an indication in a different relative orientation of the two helical segments, it is most probably due to the uncertainty of torsion angle prediction, since $C\beta$ assignments for wild-type VDAC are missing for both L10 and – naturally – G11.

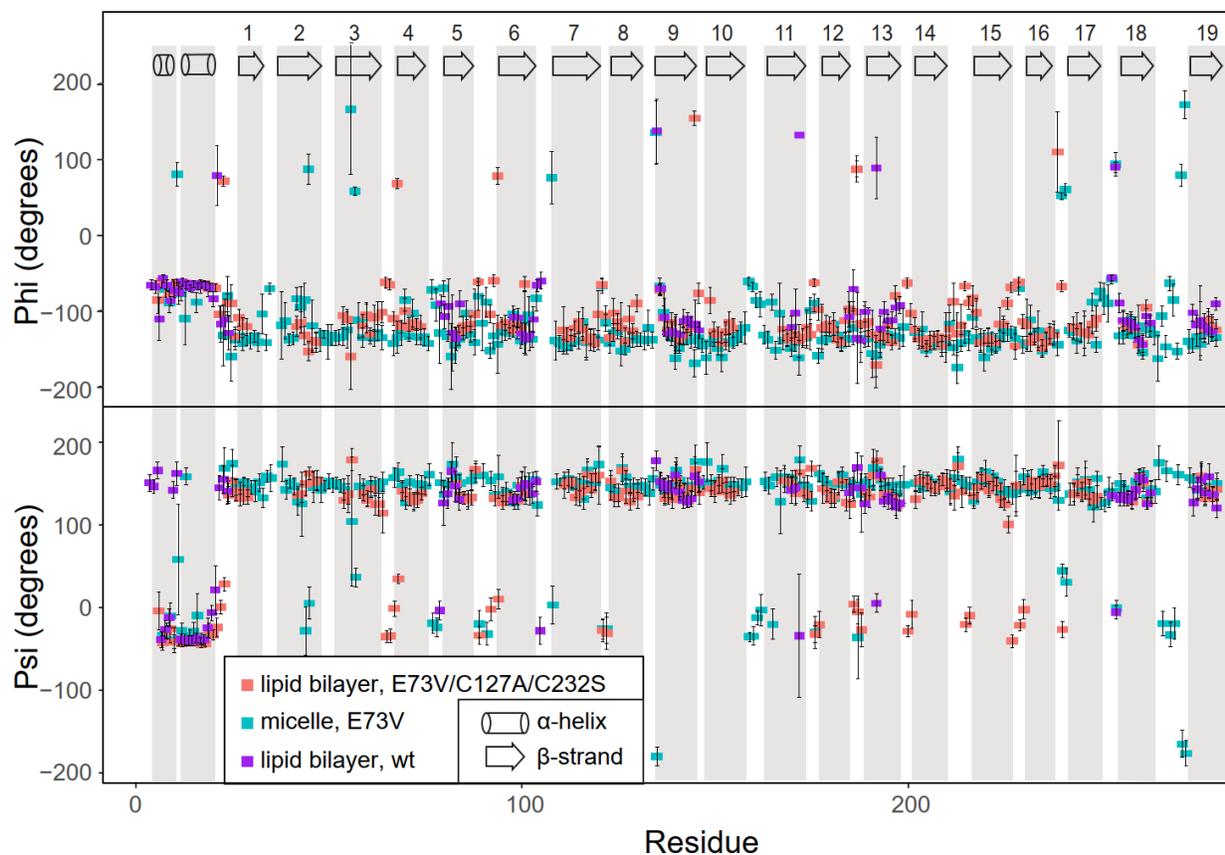


Figure 1-25: Plot of predicted torsion angles for E73V/C127A/C232S-hVDAC1 in micelles and in a lipid bilayer. Torsion angle predictions were carried out with TALOS-N. Angles based on assignments in a DMPC lipid bilayer, 2D crystals (present study) are shown in blue, while red depicts angle predictions from assignments in LDAO micelle (Jaremko et al., 2016). Purple shows torsion angle predictions of wild-type VDAC in DMPC 2D crystals, based on assignments by Eddy et al. (Eddy et al., 2015b). Secondary structural elements are shown on top and in the figure legend, and are depicted based on the solution NMR structure of the E73V mutant (Jaremko et al., 2016).

Although the exact structure of the N-terminus was unclear, all structural studies in detergents (micelles and bicelles) (Bayrhuber et al., 2008; Hiller et al., 2008; Ujwal et al., 2008; Jaremko et al., 2016; Bohm et al., 2020) showed a 19-stranded β -barrel forming the bulk of the protein, and the partial assignments obtained in lipids, as well as the long-range L10-V143 (Schneider et al., 2010) and A14-S193 (Eddy et al., 2015a) contacts, were in accordance with the detergent structures. These, however were in contrast with the structure determined in functional studies by Colombini and coworkers (Colombini, 2004, 2012): a barrel consisting of 13 β -strands and an α -helix. Our extensive assignments gave us a chance to predict torsion angles for 194 residues (out of 283) throughout the protein sequence. We found that these agree well with previous structures

obtained in detergents (a comparison for E73V hVDAC1 in micelles, wild-type hVDAC1 in liposomes and our sample can be seen on **Figure 1-25**). Major differences ($\Delta\phi$ or $\Delta\psi > 100^\circ$) were found almost exclusively only in loops (**Figure 1-26**). The only exceptions to this were residues S44, G45, and G56 in one cluster. Two of these residues do not have a C β , making the predictions less reliable. These residues are however in the more mobile N-terminal β -strands, thus it is also possible that they sample a variety of conformations.

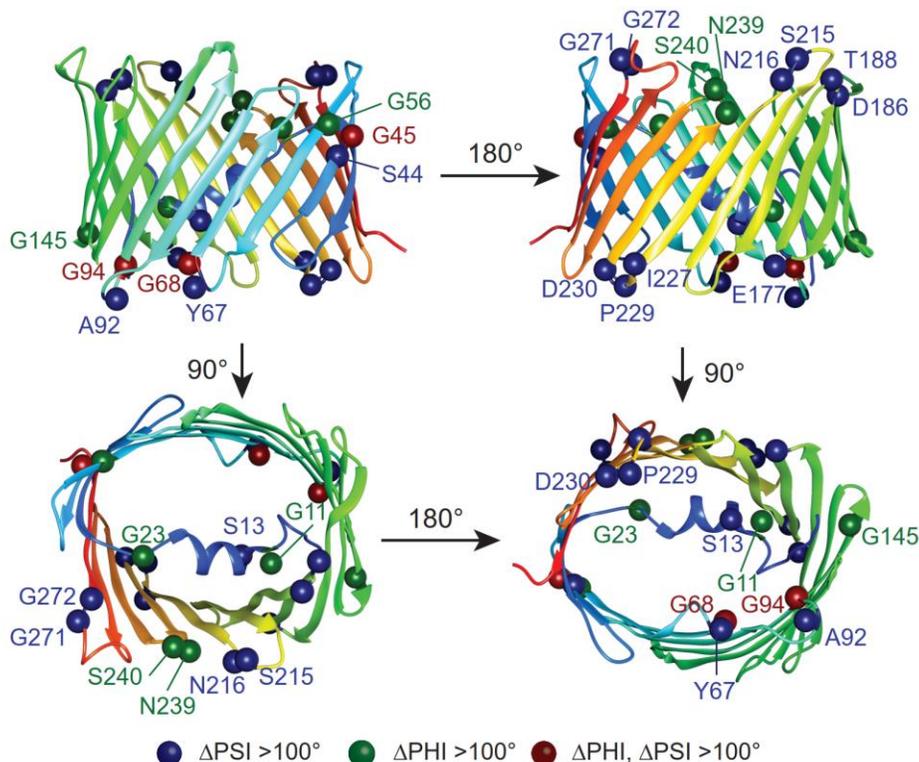


Figure 1-26: Large discrepancies between predicted torsion angles in micelles and a lipid bilayer for E73V/C127A/C232S-hVDAC1. Differences larger than 100° in phi and psi angles are shown as spheres: green for phi angles, blue for psi angles, red for both phi and psi angles. The protein is colored according to a rainbow scheme, starting with blue at the N-terminus, and finishing with red at the C-terminus. The figure was created using the lowest energy structure of the protein's structure in micelles (PDB: 5jdp) (Jaremko et al., 2016) with the Chimera software.

We could obtain more in-depth structural information in 2D crystals through the measurement of dipolar recoupling experiments, from which long-range contacts and the protein's hydrogen-bonding pattern can be deduced. Furthermore, the measurement of an MTSL-labeled mutant also allowed us to locate the N-terminus of the protein. We complemented these measurements with experiments relying on z - z -mixing (Najbauer et al., 2019) (described in detail in Chapter 2) to

give direct, residue-specific information on the protein's environment. All this information allowed confident determination of the protein's topology in lipid bilayers.

The N-terminus clearly showed the characteristic $i-i+1$, $i-i+2$ and $i-i+3$ contact pattern of α -helices for residues between T6 and K20 (**Figure 1-14**), with a breaking point at G11, a residue showing no contacts to neighboring residues. This allows us to conclude that the native structure of the N-terminus is in accordance with the N-terminal helix observed in most previous studies (including functional studies by Colombini), including the kink between the two helical segments, also observed around G11 in several detergent structures (Ujwal et al., 2008; Schneider et al., 2010; Jaremko et al., 2016; Bohm et al., 2020).

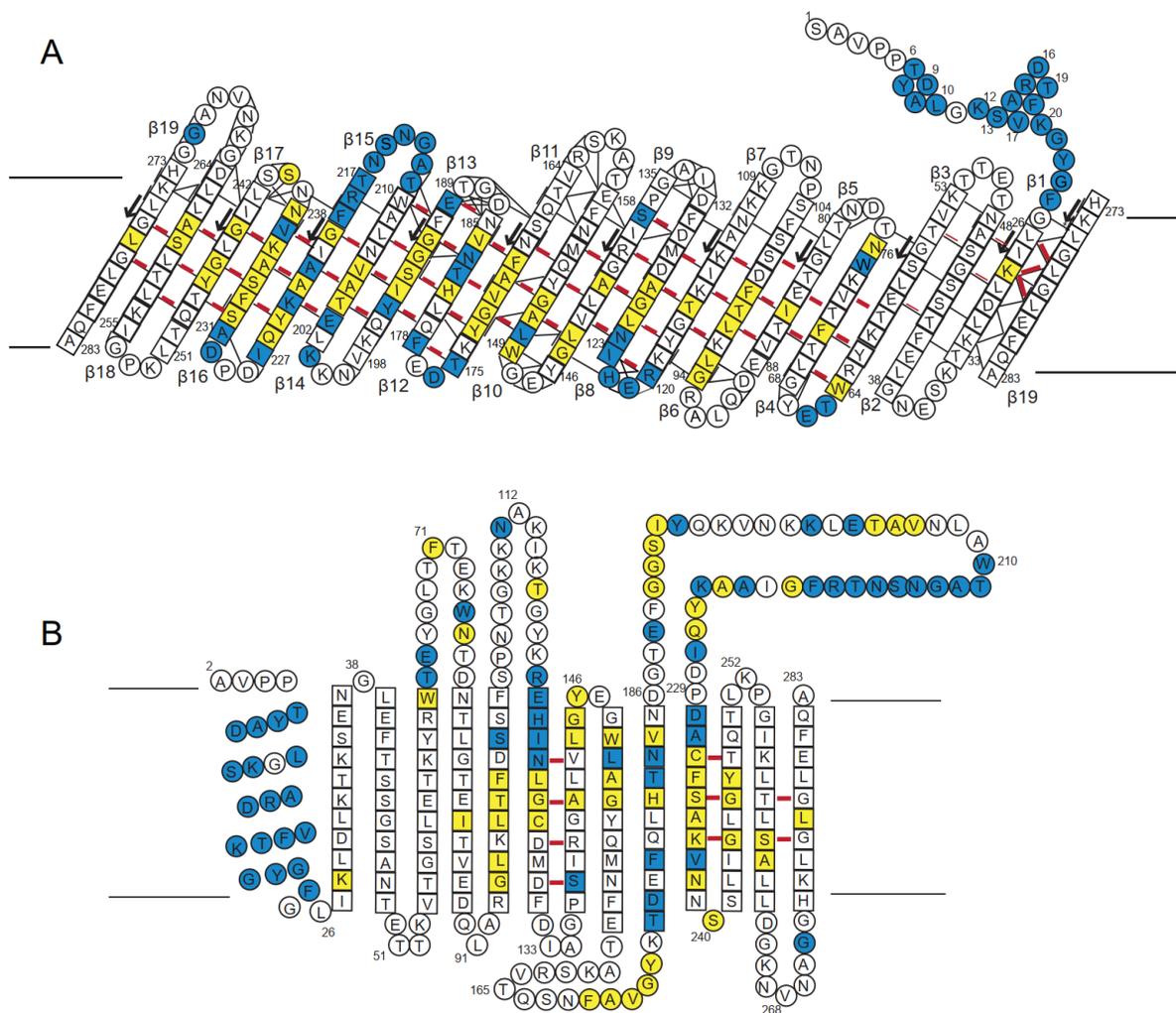


Figure 1-27: Possible topologies of VDAC. A) Topology model of E73-hVDAC1 in LDAO micelles from Villinger's thesis (Villinger, 2012). Contacts observed by solution NMR are shown in black lines. (An updated contact map showing more NOE contacts is available in Schmidt's thesis (Schmidt, 2017), however Villinger's map was chosen

for visibility reasons). Contacts observed in E73V/C127A/C232S-hVDAC1 in DMPC 2D crystals are shown in red lines. B) Topology map of the functional structure suggested by Colombini (Colombini, 2004). Contacts observed in E73V/C127A/C232S-hVDAC1 in DMPC 2D crystals agreeing with the topology proposed by Colombini are shown in red lines. Contacts from z - z mixing experiments (Najbauer et al., 2019) are shown in yellow for lipid, and blue for water contacts.

The spatial proximities we obtained from dipolar recoupling experiments overlapped perfectly with the NOE contacts observed in micelles (**Figure 1-27A**). The protein-lipid- and protein-water contacts we identified in 2D crystals also confirmed the structure determined in micelles, and contrasted starkly with the structure suggested by Colombini (**Figure 1-27B**). In the latter, many residues shown to be in contact with aliphatic lipid chains would be located in loops (especially between β 9 and β 10). The helix, though shown to be completely exposed to water, and with no lipid contacts were observed, is embedded in the membrane in Colombini's model.

Although the chemical shift perturbations in the strands along the helix-barrel contact sites (**Figure 1-24**) indicate some minor changes in the interaction of the helix and the barrel when comparing assignments from micelles and 2D crystals, the MTSL labeling experiments confirmed the overall orientation of the helix shown in detergent structures. Ujwal et al showed hydrogen bonds forming between the N-terminus of the helix and residues in the loop between β 7 and β 8 (Ujwal et al., 2008), and indeed residues K119-N124 were bleached completely. Bleaching of residues on the opposite side of the channel in β 14- β 16 raises the possibility of a second landing site of the helix. This would be in accordance with the second landing site suggested by Schmidt at the hydrophobic patch at around L242 (Schmidt, 2017). However, a likely explanation for this is the lamellar structure of the 2D crystals, where the MTSL label of one protein molecule is close enough to the channel below to bleach these residues. To completely clarify this question, however, further experiments on an MTS-labeled sample diluted with unlabeled VDAC will need to be performed.

While the overall topology of the protein could be clearly defined from solid-state NMR experiments, the barrel's shape remains unclear. A slight ellipticity of the channel had been previously observed in hVDAC1 and mVDAC ($\epsilon = 0.1$) (Bayrhuber et al., 2008; Ujwal et al., 2008), and the high-resolution structure of E73V-hVDAC1 showed an even more pronounced elliptical deformation ($\epsilon = 0.2$) (Jaremko et al., 2016; Schmidt, 2017). The possibility was raised that this stems from the pressure exerted by the micelles onto the channel (Laplace pressure (Mukerjee, 1970)), and in case of the E73V mutant, the mutation stabilizes a certain barrel conformation. However, since an AFM study on yeast revealed highly elliptical channels ($\epsilon = 0.3$)

under native conditions in the mitochondrial outer membrane (Goncalves et al., 2007), it is possible that the channel is naturally non-circular (while the Laplace pressure could deform the channel in micelles, in a lipid bilayer the pressure could be coming from the tight packing of the neighboring molecules). Determining the exact ellipticity of VDAC in a lipid bilayer would, however, be extremely difficult, if at all possible by solid state NMR due to the lack of highly accurate long-distance restraints (Koehler Leman et al., 2018).

1.4.4 VDAC's mobility and gating behavior

VDAC's voltage gating behavior is one of the channel's most characteristic properties (Colombini, 1989). A multitude of electrophysiology studies have shown the channel to be in a high-conductance, anion-selective open state in the absence of applied voltage, while under voltage a drop in conductance and a change of selectivity can be observed (Schein et al., 1976).

There is increasing evidence that channel gating is a highly dynamic process, with multiple open- and closed substates. Both the barrel and the helix, located inside the pore in an ideal position to regulate conductance, have been suggested to partake in voltage gating (Ujwal et al., 2008; Hiller and Wagner, 2009; Villinger et al., 2010; Zachariae et al., 2012; Bohm et al., 2020).

Although we did not measure relaxation rates on the 2D crystalline sample, peak intensities in MAS-NMR are excellent indicators of dynamics, as molecular motion can manifest as weaker dipolar coupling strength, resulting in a decrease of signal intensity (Matlahov and van der Wel, 2018). In E73V/C127A/C232S-hVDAC1 (implementing the channel's open state) we observed increased mobility in the N-terminal β -strands β 1- β 4, as well in the adjoining C-terminal strands β 18 and β 19 (peaks in β 10 were also of lower intensity, however the surrounding strands all showed strong signals). This observation agrees well with the significant broadening observed by Villinger et al. in the 4 N-terminal β -strands, as well as β 16-19 in wild-type hVDAC1, and to some extent in E73V-hVDAC1 (Villinger et al., 2010). The higher intensity of peaks in β 5- β 17, the strands in front of which the α -helix is running indicates a stabilizing effect of the helix on the barrel wall. Indeed Schneider et al. had observed destabilization of the barrel upon truncation of the helix in Δ N(1-20)-hVDAC1 (Schneider et al., 2010). Compared to the rest of the barrel, the average intensity of residues in strands β 1-4, 18, 19 was roughly 65%, while the N-terminal α -

helix was 20% more intense than β 5- β 17. The latter is in stark contrast with the increased mobility observed recently by Bohm et al. in the N-terminal segments in the ground (open) state (Bohm et al., 2020). This difference in the mobility of the N-terminal helices could possibly (at least partially) stem from an altered interaction between the helices and the barrel wall in a detergent environment, as suggested by the chemical shift perturbations between lipid and detergent assignments (**Figure 1-24**) for residues close to helix-barrel contacts.

Upon investigating the closed-state quintuple (G21V/G23V/E73V/C127A/C232S) mutant, Schmidt had observed far-reaching effects of the G21V/G23V mutations in the linker between the helix and the barrel in a micelle sample, observing a drop of signal intensities for residues not only in the linker and β 1, but also β 15- β 19 (Schmidt, 2017), and had noted that based on preliminary ssNMR studies, these effects are even more pronounced in a membrane environment, the disappearance of the L10-V143 contact indicating a dissociation of the helix from the barrel.

Indeed, for this mutant in 2D lipid crystals we found a strong increase of dynamics especially in the helix (most apparent in the α 2 segment), as well as in a large part of the barrel: β 1, and in the adjoining C-terminal part, β 6- β 19 (**Figure 1-22**). This fits well with the previous observation in liposomes that the barrel is mobilized upon truncation of the N-terminal helix in Δ (1-20)-hVDAC1 (Schneider et al., 2010), and the MD results that large-scale deformations are needed to reproduce channel selectivity (Zachariae et al., 2012). The region β 6- β 19 overlaps remarkably well with strands β 5- β 19 shown in MD simulations in a DMPC bilayer to be stabilized by interaction with the helix (Zachariae et al., 2012), raising the possibility of an at least partial dissociation of the helix (especially the α 2 segment) from the barrel wall in the quintuple mutant (and thus the channel's closed state).

The position of the N-terminal α 1 segment in the closed state (T6-L10) is unclear. The disappearance of the L10-V143 contacts in a preliminary ssNMR study on the quintuple mutant in DPhPC (Schmidt, 2017), as well as the significant mobilization of the loop between β 7- β 8, known to form hydrogen bonds with A2 and P4, suggest a possible loss of barrel contact even for the α 1 helical segment. Moreover, though very low in intensity, Schmidt observed possible new contacts (e.g. T6-L242, T6-A270, Y7-S240) supporting a possible landing site for the helix in the closed state at the hydrophobic patch (L242 and L262) in β 17- β 18 (Schmidt, 2017). The less pronounced increase in dynamics in the α 1 segment and in the strands with the α 1 segment running in front of

them (β 7- β 12) (**Figure 1-22**), as well as an electrophysiology study showing the channel to gate normally even in the presence of an L10C-A170C cross link (Teijido et al., 2012) however suggest a partial attachment of the N-terminal α 1 segment to the barrel wall. It is possible that there are multiple conformations of the quintuple mutant, one with the helix entirely detached, moving to a new position, and one where only α 2 moves away from the barrel wall. Indeed, Geula et al. observed multiple conductance states (and multiple closed states) for the G21V/G23V mutant, and suggested that these two mutations in the linker between the N-terminal helix and the β -barrel could lock the helix into multiple conformations (Geula et al., 2012).

The conservation of the G11 residue (Amodeo et al., 2014) could also be related to voltage gating. Since a hinge-like role (Rosenhouse-Dantsker and Logothetis, 2006; Hogel et al., 2018) has frequently been attributed to glycines due to their conformational flexibility, G11 could impart the possibility for the α 1 and α 2 helical segments to move independently.

1.4.5 Cholesterol binding to VDAC

Electrophysiology experiments have shown the binding of cholesterol not only to facilitate VDAC's membrane insertion, but also to slightly reduce the % closure ($(1-G_{50mV}/G_0)100$) of the channel (Popp et al., 1995). Our data showing a relative stabilization of the barrel compared to the helix in the presence of cholesterol (**Figure 1-19**) supports this nicely, considering that NMR and MD simulations have shown barrel deformations to be crucial for voltage gating (Schneider et al., 2010; Villinger et al., 2010; Zachariae et al., 2012).

Using z - z mixing experiments to identify cholesterol binding sites, we found the two binding sites previously identified by solution NMR in LDAO micelles (Hiller et al., 2008) to be populated, as well as indication of a third binding site suggested by MD simulations (Weiser et al., 2014). The binding site at E73 identified only recently using photo-affinity labeling and mass spectrometry (Cheng et al., 2019) was not observed in our sample. A possible explanation might be that the E73V mutation reduces the binding affinity, as observed for the E73Q mutant, while a contributing factor might also be the lower starting signal intensity observed in strands β 1- β 4.

1.4.6 Comparison of blockage mechanisms

Although an important player in apoptosis, cancer, and neurodegeneration, VDAC's druggability is problematic due to a lack of specific binding partners (Magri et al., 2018). Since the Bcl-2 antisense oligonucleotide G3139 is a promising drug candidate shown to greatly reduce channel conductance by directly binding VDAC1 (Lai et al., 2006), we investigated the mechanism through which it inhibits VDAC's normal function. We did not find changes of peak intensities in complete regions within the protein, which indicates that the mechanism of channel closure is different than the observation in the closed state mutant (G21V/G23V/E73V/C127A/C232S), where channel closure is accompanied by mobilization of a large portion of the barrel, as well as part of the helix. Rather a decrease in intensity for individual residues in β 16 and β 17, as well as β 12- β 14, in the α 2 helix, as well as in the linker between the helix and the barrel was observed, a sign of intermediate exchange occurring upon ligand binding (Ziarek et al., 2011). This data, taken together with the fact that the effect of the oligonucleotide is length-dependent, and upon investigating both the full-length and truncated versions of G3139, only the longest ones (14-16, and 18-mers) were found to cause significant losses of channel conductance (Lai et al., 2006), suggests that G3139 acts by sterically blocking the channel.

The dinucleotide β -NADH was recently shown by solution NMR and MD simulations to block the channel by steric occlusion by binding to VDAC in the exact same pocket at β 16 and β 17 (Bohm et al., 2020) that we identified as the binding site of G3139. Moreover, the nucleoside triphosphate ATP has been shown to possess a binding site encompassing the α -helix, the linker between helix and barrel, and strands β 12- β 19 (Villinger, 2012), which overlaps perfectly with G3139's binding site, and other nucleotides, such as GTP and UTP have also been shown to share this site (Villinger et al., 2014). These results indicate a common mechanism of action for all nucleotides binding to VDAC. Similar steric occlusion as a blockage mechanism for the channel was also speculated for hexokinase-I and the inorganic polycationic dye ruthenium red (RuR) (Zaid et al., 2005).

1.5 Conclusion and outlook

We have performed a comprehensive study of the human voltage-dependent anion channel's (hVDAC1's) structure and interactions in a lipid bilayer. Using a 2D crystalline sample giving spectra of exceptional resolution, we obtained extensive assignments for the membrane-bound channel, allowing us to determine the protein's topology in a lipid bilayer with high certainty. Apart from indications of a somewhat altered helix-barrel contact, we found the topology of the protein to be the same as previously identified in micelles: an N-terminal α -helix broken into two segments by G11, and a 19-stranded β -barrel, showing a parallel orientation of strands at β 1 and β 19. We could identify elevated dynamics in the N-terminal β -barrel (β 1- β 4), as well as in the C-terminus adjacent to it (β 18- β 19), as was also observed in micelles. We, however, found the α -helix to be exceptionally rigid in the open state, a stark contrast to the dynamics of the helix found in micelles.

The closed state of VDAC, implemented by introducing the G21V/G23V mutations to the linker between the helix and the barrel, showed an increase of dynamics in the N-terminal helix, particularly the α 2 segment, as well as in the β -strands previously stabilized by the helix. This provides further evidence for the closed state of VDAC being highly dynamic, with the participation of both helix and barrel in the gating.

We showed cholesterol to further increase barrel stability in the open state, and identified three of the binding sites predicted previously by docking studies. Furthermore, we showed the binding site of the specific binding partner and cancer drug-candidate G3139 to coincide with nucleotide binding sites previously seen in micelles.

Future perspectives for the study of VDAC include calculation of the solid-state NMR structure, combining solid-state NMR chemical shifts and long-range contacts with modelling approaches, such as in CS-ROSETTA. To deepen the understanding of the dynamics accompanying gating, dynamics measurements in a lipid bilayer (e.g. $R_{1\rho}$) on both the open- and the closed state of the channel would be beneficial.

At the moment, little is known about the molecular basis of VDAC's interaction with other proteins and small molecules. Identifying some of VDAC's further binding sites for the myriad of known small molecular binding partners (Magri et al., 2018) would be highly beneficial to the design of

selective binding partners, improving the protein's druggability. Moreover, VDAC interacts with α -synuclein (Rostovtseva et al., 2015), a key player in neurodegenerative diseases, and the translocator protein (TSPO) (Shoshan-Barmatz et al., 2019) in the outer mitochondrial membrane, with important functions in steroid synthesis, immune response, and apoptosis. Since TSPO is an integral membrane protein, and α -synuclein has been shown to partially fold upon membrane binding, solid-state NMR would be the method of choice to study these interactions in a lipid bilayer.

1.6 References

Akbey, U., Lange, S., Trent Franks, W., Linser, R., Rehbein, K., Diehl, A., van Rossum, B.J., Reif, B., and Oschkinat, H. (2010). Optimum levels of exchangeable protons in perdeuterated proteins for proton detection in MAS solid-state NMR spectroscopy. *J. Biomol. NMR* *46*, 67-73.

Al Bitar, F., Roosens, N., Smeyers, M., Vauterin, M., Van Boxtel, J., Jacobs, M., and Homble, F. (2003). Sequence analysis, transcriptional and posttranscriptional regulation of the rice vdac family. *Bba-Gene Struct Expr* *1625*, 43-51.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*, 4 edn (New York: Garland Science).

Amodeo, G.F., Scorciapino, M.A., Messina, A., De Pinto, V., and Ceccarelli, M. (2014). Charged residues distribution modulates selectivity of the open state of human isoforms of the voltage dependent anion-selective channel. *PLOS ONE* *9*, e103879.

Andreas, L.B., Jaudzems, K., Stanek, J., Lalli, D., Bertarello, A., Le Marchand, T., Cala-De Paepe, D., Kotelovica, S., Akopjana, I., Knott, B., *et al.* (2016). Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 9187-9192.

Andreas, L.B., Le Marchand, T., Jaudzems, K., and Pintacuda, G. (2015a). High-resolution proton-detected NMR of proteins at very fast MAS. *J. Magn. Reson.* *253*, 36-49.

Andreas, L.B., Stanek, J., Le Marchand, T., Bertarello, A., Cala-De Paepe, D., Lalli, D., Krejcikova, M., Doyen, C., Oster, C., Knott, B., *et al.* (2015b). Protein residue linking in a single spectrum for magic-angle spinning NMR assignment. *J. Biomol. NMR* *62*, 253-261.

Araiso, Y., Tsutsumi, A., Qiu, J., Imai, K., Shiota, T., Song, J., Lindau, C., Wenz, L.S., Sakaue, H., Yunoki, K., *et al.* (2019). Structure of the mitochondrial import gate reveals distinct preprotein paths. *Nature* *575*, 395-+.

Aram, L., Geula, S., Arbel, N., and Shoshan-Barmatz, V. (2010). VDAC1 cysteine residues: topology and function in channel activity and apoptosis. *Biochem. J.* *427*, 445-454.

Barbet-Massin, E., Pell, A.J., Jaudzems, K., Franks, W.T., Retel, J.S., Kotelovica, S., Akopjana, I., Tars, K., Emsley, L., Oschkinat, H., *et al.* (2013). Out-and-back ¹³C-¹³C scalar transfers in protein resonance assignment by proton-detected solid-state NMR under ultra-fast MAS. *J. Biomol. NMR* *56*, 379-386.

Barbet-Massin, E., Pell, A.J., Retel, J.S., Andreas, L.B., Jaudzems, K., Franks, W.T., Nieuwkoop, A.J., Hiller, M., Higman, V., Guerry, P., *et al.* (2014). Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J. Am. Chem. Soc.* *136*, 12489-12497.

- Bayrhuber, M., Meins, T., Habeck, M., Becker, S., Giller, K., Villinger, S., Vorrhein, C., Griesinger, C., Zweckstetter, M., and Zeth, K. (2008). Structure of the human voltage-dependent anion channel. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 15370-15375.
- Ben-Hail, D., and Shoshan-Barmatz, V. (2016). VDAC1-interacting anion transport inhibitors inhibit VDAC1 oligomerization and apoptosis. *Biochim. Biophys. Acta* *1863*, 1612-1623.
- Bennett, A.E., Ok, J.H., Griffin, R.G., and Vega, S. (1992). Chemical-Shift Correlation Spectroscopy in Rotating Solids - Radio Frequency-Driven Dipolar Recoupling and Longitudinal Exchange. *J. Chem. Phys.* *96*, 8624-8627.
- Bennett, A.E., Rienstra, C.M., Auger, M., Lakshmi, K.V., and Griffin, R.G. (1995). Heteronuclear Decoupling in Rotating Solids. *J. Chem. Phys.* *103*, 6951-6958.
- Benz, R. (1994). Permeation of Hydrophilic Solutes through Mitochondrial Outer Membranes - Review on Mitochondrial Porins. *Bba-Rev. Biomembranes.* *1197*, 167-196.
- Benz, R., and Brdiczka, D. (1992). The Cation-Selective Substate of the Mitochondrial Outer-Membrane Pore - Single-Channel Conductance and Influence on Intermembrane and Peripheral Kinases. *J. Bioenerg. Biomembr.* *24*, 33-39.
- Blachly-Dyson, E., and Forte, M. (2001). VDAC channels. *IUBMB Life* *52*, 113-118.
- Blachly-Dyson, E., Peng, S., Colombini, M., and Forte, M. (1990). Selectivity changes in site-directed mutants of the VDAC ion channel: structural implications. *Science* *247*, 1233-1236.
- Blachly-Dyson, E., Peng, S.Z., Colombini, M., and Forte, M. (1989). Probing the structure of the mitochondrial channel, VDAC, by site-directed mutagenesis: a progress report. *J. Bioenerg. Biomembr.* *21*, 471-483.
- Bohm, R., Amodeo, G.F., Murlidaran, S., Chavali, S., Wagner, G., Winterhalter, M., Brannigan, G., and Hiller, S. (2020). The Structural Basis for Low Conductance in the Membrane Protein VDAC upon beta-NADH Binding and Voltage Gating. *Structure* *28*, 206-214.
- Briones, R., Weichbrodt, C., Paltrinieri, L., Mey, I., Villinger, S., Giller, K., Lange, A., Zweckstetter, M., Griesinger, C., Becker, S., *et al.* (2016). Voltage Dependence of Conformational Dynamics and Subconducting States of VDAC-1. *Biophys. J.* *111*, 1223-1234.
- Casadio, R., Jacoboni, I., Messina, A., and De Pinto, V. (2002). A 3D model of the voltage-dependent anion channel (VDAQ). *FEBS letters* *520*, 1-7.
- Caterino, M., Ruoppolo, M., Mandola, A., Costanzo, M., Orru, S., and Imperlini, E. (2017). Protein-protein interaction networks as a new perspective to evaluate distinct functional roles of voltage-dependent anion channel isoforms. *Mol. Biosyst.* *13*, 2466-2476.
- Chacinska, A., Rehling, P., Guiard, B., Frazier, A.E., Schulze-Specking, A., Pfanner, N., Voos, W., and Meisinger, C. (2003). Mitochondrial translocation contact sites: separation of dynamic

and stabilizing elements in formation of a TOM-TIM-preprotein supercomplex. *EMBO J.* *22*, 5370-5381.

Cheng, W.W.L., Budelier, M.M., Sugawara, Y., Bergdoll, L., Queralt-Martin, M., Rosencrans, W., Rostovtseva, T.K., Chen, Z.W., Abramson, J., Krishnan, K., *et al.* (2019). Multiple neurosteroid and cholesterol binding sites in voltage-dependent anion channel-1 determined by photo-affinity labeling. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids.* *1864*, 1269-1279.

Chevelkov, V., Faelber, K., Schrey, A., Rehbein, K., Diehl, A., and Reif, B. (2007). Differential line broadening in MAS solid-state NMR due to dynamic interference. *J. Am. Chem. Soc.* *129*, 10195-10200.

Chevelkov, V., Rehbein, K., Diehl, A., and Reif, B. (2006). Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew. Chem. Int. Ed. Engl.* *45*, 3878-3881.

Choudhary, O.P., Paz, A., Adelman, J.L., Colletier, J.P., Abramson, J., and Grabe, M. (2014). Structure-guided simulations illuminate the mechanism of ATP transport through VDAC1. *Nat. Struct. Mol. Biol.* *21*, 626-632.

Clore, G.M., and Iwahara, J. (2009). Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* *109*, 4108-4139.

Colombini, M. (1989). Voltage gating in the mitochondrial channel, VDAC. *J. Membr. Biol.* *111*, 103-111.

Colombini, M. (2004). VDAC: the channel at the interface between mitochondria and the cytosol. *Mol. Cell. Biochem.* *256-257*, 107-115.

Colombini, M. (2012). VDAC structure, selectivity, and dynamics. *Bba-Biomembranes* *1818*, 1457-1465.

Colombini, M. (2016). The VDAC channel: Molecular basis for selectivity. *Biochim. Biophys. Acta* *1863*, 2498-2502.

Colombini, M., Yeung, C.L., Tung, J., and Konig, T. (1987). The Mitochondrial Outer-Membrane Channel, Vdac, Is Regulated by a Synthetic Polyanion. *Biochim. Biophys. Acta* *905*, 279-286.

De Pinto, V., Benz, R., and Palmieri, F. (1989). Interaction of non-classical detergents with the mitochondrial porin. A new purification procedure and characterization of the pore-forming unit. *Eur. J. Biochem.* *183*, 179-187.

De Pinto, V., Tomasello, F., Messina, A., Guarino, F., Benz, R., La Mendola, D., Magri, A., Milardi, D., and Pappalardo, G. (2007). Determination of the conformation of the human VDAC1 N-terminal peptide, a protein moiety essential for the functional properties of the pore. *ChemBioChem* *8*, 744-756.

- De Pinto, V.D., and Palmieri, F. (1992). Transmembrane arrangement of mitochondrial porin or voltage-dependent anion channel (VDAC). *J. Bioenerg. Biomembr.* *24*, 21-26.
- Depinto, V., Prezioso, G., Thinnes, F., Link, T.A., and Palmieri, F. (1991). Peptide-Specific Antibodies and Proteases as Probes of the Transmembrane Topology of the Bovine Heart Mitochondrial Porin. *Biochemistry* *30*, 10191-10200.
- Dolder, M., Zeth, K., Tittmann, P., Gross, H., Welte, W., and Wallimann, T. (1999). Crystallization of the human, mitochondrial voltage-dependent anion-selective channel in the presence of phospholipids. *J. Struct. Biol.* *127*, 64-71.
- Dubey, A.K., Godbole, A., and Mathew, M.K. (2016). Regulation of VDAC trafficking modulates cell death. *Cell Death Discov.* *2*, 16085.
- Duee, E.D., and Vignais, P.V. (1965). Echange Entre Adenine-Nucleotides Extra-Et Intramitochondriaux. *Biochim. Biophys. Acta* *107*, 184-188.
- Eddy, M.T., Andreas, L., Tejjido, O., Su, Y., Clark, L., Noskov, S.Y., Wagner, G., Rostovtseva, T.K., and Griffin, R.G. (2015a). Magic angle spinning nuclear magnetic resonance characterization of voltage-dependent anion channel gating in two-dimensional lipid crystalline bilayers. *Biochemistry* *54*, 994-1005.
- Eddy, M.T., Ong, T.C., Clark, L., Tejjido, O., van der Wel, P.C., Garces, R., Wagner, G., Rostovtseva, T.K., and Griffin, R.G. (2012). 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.
- Eddy, M.T., Su, Y., Silvers, R., Andreas, L., Clark, L., Wagner, G., Pintacuda, G., Emsley, L., and Griffin, R.G. (2015b). Lipid bilayer-bound conformation of an integral membrane beta barrel protein by multidimensional MAS NMR. *J. Biomol. NMR* *61*, 299-310.
- Eddy, M.T., Yu, T.Y., Wagner, G., and Griffin, R.G. (2019). Structural characterization of the human membrane protein VDAC2 in lipid bilayers by MAS NMR. *J. Biomol. NMR* *73*, 451-460.
- Florke, H., Thinnes, F.P., Winkelbach, H., Stadtmuller, U., Paetzold, G., Morys-Wortmann, C., Hesse, D., Sternbach, H., Zimmermann, B., Kaufmann-Kolle, P., *et al.* (1994). Channel active mammalian porin, purified from crude membrane fractions of human B lymphocytes and bovine skeletal muscle, reversibly binds adenosine triphosphate (ATP). *Biol. Chem. Hoppe Seyler* *375*, 513-520.
- Forte, M., Guy, H.R., and Mannella, C.A. (1987). Molecular genetics of the VDAC ion channel: structural model and sequence analysis. *J. Bioenerg. Biomembr.* *19*, 341-350.
- Fraga, H., Arnaud, C.A., Gauto, D.F., Audin, M., Kurauskas, V., Macek, P., Krichel, C., Guan, J.Y., Boisbouvier, J., Sprangers, R., *et al.* (2017). Solid-State NMR H-N-(C)-H and H-N-C-C 3D/4D Correlation Experiments for Resonance Assignment of Large Proteins. *ChemPhysChem* *18*, 2697-2703.

- Fricke, P., Chevelkov, V., Zinke, M., Giller, K., Becker, S., and Lange, A. (2017). Backbone assignment of perdeuterated proteins by solid-state NMR using proton detection and ultrafast magic-angle spinning. *Nat. Protoc.* *12*, 764-782.
- Gattin, Z., Schneider, R., Laukat, Y., Giller, K., Maier, E., Zweckstetter, M., Griesinger, C., Benz, R., Becker, S., and Lange, A. (2015). Solid-state NMR, electrophysiology and molecular dynamics characterization of human VDAC2. *J. Biomol. NMR* *61*, 311-320.
- Ge, L., Villinger, S., Mari, S.A., Giller, K., Griesinger, C., Becker, S., Muller, D.J., and Zweckstetter, M. (2016). Molecular Plasticity of the Human Voltage-Dependent Anion Channel Embedded Into a Membrane. *Structure* *24*, 585-594.
- Geula, S., Ben-Hail, D., and Shoshan-Barmatz, V. (2012). Structure-based analysis of VDAC1: N-terminus location, translocation, channel gating and association with anti-apoptotic proteins. *Biochem. J.* *444*, 475-485.
- Gincel, D., Zaid, H., and Shoshan-Barmatz, V. (2001). Calcium binding and translocation by the voltage-dependent anion channel: a possible regulatory mechanism in mitochondrial function. *Biochem. J.* *358*, 147-155.
- Goncalves, R.P., Buzhynskyy, N., Prima, V., Sturgis, J.N., and Scheuring, S. (2007). Supramolecular assembly of VDAC in native mitochondrial outer membranes. *J. Mol. Biol.* *369*, 413-418.
- Gunter, T.E., and Sheu, S.S. (2009). Characteristics and possible functions of mitochondrial Ca(2+) transport mechanisms. *Biochim. Biophys. Acta* *1787*, 1291-1308.
- Guo, X.W., Smith, P.R., Cognon, B., D'Arcangelis, D., Dolginova, E., and Mannella, C.A. (1995). Molecular design of the voltage-dependent, anion-selective channel in the mitochondrial outer membrane. *J. Struct. Biol.* *114*, 41-59.
- Hiller, M., Krabben, L., Vinothkumar, K.R., Castellani, F., van Rossum, B.J., Kuhlbrandt, W., and Oschkinat, H. (2005). Solid-state magic-angle spinning NMR of outer-membrane protein G from *Escherichia coli*. *ChemBioChem : A European Journal of Chemical Biology* *6*, 1679-1684.
- Hiller, S., Garces, R.G., Malia, T.J., Orekhov, V.Y., Colombini, M., and Wagner, G. (2008). Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* *321*, 1206-1210.
- Hiller, S., and Wagner, G. (2009). The role of solution NMR in the structure determinations of VDAC-1 and other membrane proteins. *Curr. Opin. Struct. Biol.* *19*, 396-401.
- Hinsch, K.D., De Pinto, V., Aires, V.A., Schneider, X., Messina, A., and Hinsch, E. (2004). Voltage-dependent anion-selective channels VDAC2 and VDAC3 are abundant proteins in bovine outer dense fibers, a cytoskeletal component of the sperm flagellum. *J. Biol. Chem.* *279*, 15281-15288.

- Hodge, T., and Colombini, M. (1997). Regulation of metabolite flux through voltage-gating of VDAC channels. *J. Membr. Biol.* *157*, 271-279.
- Hogel, P., Gotz, A., Kuhne, F., Ebert, M., Stelzer, W., Rand, K.D., Scharnagl, C., and Langosch, D. (2018). Glycine Perturbs Local and Global Conformational Flexibility of a Transmembrane Helix. *Biochemistry* *57*, 1326-1337.
- Hoogenboom, B.W., Suda, K., Engel, A., and Fotiadis, D. (2007). The supramolecular assemblies of voltage-dependent anion channels in the native membrane. *J. Mol. Biol.* *370*, 246-255.
- Hosaka, T., Okazaki, M., Kimura-Someya, T., Ishizuka-Katsura, Y., Ito, K., Yokoyama, S., Dodo, K., Sodeoka, M., and Shirouzu, M. (2017). Crystal structural characterization reveals novel oligomeric interactions of human voltage-dependent anion channel 1. *Protein Sci.* *26*, 1749-1758.
- Isaacson, R.L., Simpson, P.J., Liu, M., Cota, E., Zhang, X., Freemont, P., and Matthews, S. (2007). A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J. Am. Chem. Soc.* *129*, 15428-15429.
- Jain, M.G., Lalli, D., Stanek, J., Gowda, C., Prakash, S., Schwarzer, T.S., Schubeis, T., Castiglione, K., Andreas, L.B., Madhu, P.K., *et al.* (2017). Selective H-1-H-1 Distance Restraints in Fully Protonated Proteins by Very Fast Magic-Angle Spinning Solid-State NMR. *J. Phys. Chem. Lett.* *8*, 2399-2405.
- Jaipuria, G., Leonov, A., Giller, K., Vasa, S.K., Jaremko, L., Jaremko, M., Linser, R., Becker, S., and Zweckstetter, M. (2017). Cholesterol-mediated allosteric regulation of the mitochondrial translocator protein structure. *Nat. Commun.* *8*, 14893.
- Jap, B.K., Downing, K.H., and Walian, P.J. (1990). Structure of PhoE porin in projection at 3.5 Å resolution. *J. Struct. Biol.* *103*, 57-63.
- Jaremko, L., Jaremko, M., Giller, K., Becker, S., and Zweckstetter, M. (2014). Structure of the Mitochondrial Translocator Protein in Complex with a Diagnostic Ligand. *Science* *343*, 1363-1366.
- Jaremko, M., Jaremko, L., Villinger, S., Schmidt, C.D., Griesinger, C., Becker, S., and Zweckstetter, M. (2016). High-Resolution NMR Determination of the Dynamic Structure of Membrane Proteins. *Angew. Chem. Int. Ed. Engl.* *55*, 10518-10521.
- Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A., and Guntert, P. (2006). Optimal isotope labelling for NMR protein structure determinations. *Nature* *440*, 52-57.
- Koehler-Leman, J., D'Avino, A.R., Bhatnagar, Y., and Gray, J.J. (2018). Comparison of NMR and crystal structures of membrane proteins and computational refinement to improve model quality. *Proteins* *86*, 57-74.
- Kozuch, J., Weichbrodt, C., Millo, D., Giller, K., Becker, S., Hildebrandt, P., and Steinem, C. (2014). Voltage-dependent structural changes of the membrane-bound anion channel hVDAC1

probed by SEIRA and electrochemical impedance spectroscopy. *Phys. Chem. Chem. Phys.* *16*, 9546-9555.

Kuhlbrandt, W., and Wang, D.N. (1991). Three-dimensional structure of plant light-harvesting complex determined by electron crystallography. *Nature* *350*, 130-134.

Lai, J.C., Tan, W.Z., Benimetskaya, L., Miller, P., Colombini, M., and Stein, C.A. (2006). A pharmacologic target of G3139 in melanoma cells may be the mitochondrial VDAC. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 7494-7499.

Lalli, D., Idso, M.N., Andreas, L.B., Hussain, S., Baxter, N., Han, S., Chmelka, B.F., and Pintacuda, G. (2017). Proton-Based Structural Analysis of a Heptahelical Transmembrane Protein in Lipid Bilayers. *Journal of the American Society* *139*, 13006-13012.

Lewandowski, J.R., Dumez, J.N., Akbey, U., Lange, S., Emsley, L., and Oschkinat, H. (2011). Enhanced Resolution and Coherence Lifetimes in the Solid-State NMR Spectroscopy of Perdeuterated Proteins under Ultrafast Magic-Angle Spinning. *J. Phys. Chem. Lett.* *2*, 2205-2211.

Lewis, B.A., Harbison, G.S., Herzfeld, J., and Griffin, R.G. (1985). NMR structural analysis of a membrane protein: bacteriorhodopsin peptide backbone orientation and motion. *Biochemistry* *24*, 4671-4679.

Lin, Y.L., Cheng, Y.S., Ho, C.I., Guo, Z.H., Huang, S.J., Org, M.L., Oss, A., Samoson, A., and Chan, J.C.C. (2018). Preparation of fibril nuclei of beta-amyloid peptides in reverse micelles. *Chem. Commun.* *54*, 10459-10462.

Linser, R., Bardiaux, B., Higman, V., Fink, U., and Reif, B. (2011). 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.

Linser, R., Gelev, V., Hagn, F., Arthanari, H., Hyberts, S.G., and Wagner, G. (2014). Selective methyl labeling of eukaryotic membrane proteins using cell-free expression. *J. Am. Chem. Soc.* *136*, 11308-11310.

Lipper, C.H., Stofleth, J.T., Bai, F., Sohn, Y.S., Roy, S., Mittler, R., Nechushtai, R., Onuchic, J.N., and Jennings, P.A. (2019). Redox-dependent gating of VDAC by mitoNEET. *Proc. Natl. Acad. Sci. U. S. A.* *116*, 19924-19929.

Long, D., and Yang, D.W. (2009). Buffer Interference with Protein Dynamics: A Case Study on Human Liver Fatty Acid Binding Protein. *Biophys. J.* *96*, 1482-1488.

Magri, A., Reina, S., and De Pinto, V. (2018). VDAC1 as Pharmacological Target in Cancer and Neurodegeneration: Focus on Its Role in Apoptosis. *Front. Chem.* *6*, 108.

Maldonado, E.N., Sheldon, K.L., DeHart, D.N., Patnaik, J., Manevich, Y., Townsend, D.M., Bezrukov, S.M., Rostovtseva, T.K., and Lemasters, J.J. (2013). Voltage-dependent Anion Channels Modulate Mitochondrial Metabolism in Cancer Cells REGULATION BY FREE TUBULIN AND ERASTIN. *J. Biol. Chem.* *288*, 11920-11929.

- Malia, T.J., and Wagner, G. (2007). NMR structural investigation of the mitochondrial outer membrane protein VDAC and its interaction with antiapoptotic Bcl-xL. *Biochemistry* 46, 514-525.
- Mannella, C.A. (1982). Structure of the Outer Mitochondrial-Membrane - Ordered Arrays of Pore-Like Subunits in Outer-Membrane Fractions from Neurospora-Crassa Mitochondria. *J. Cell Biol.* 94, 680-687.
- Mannella, C.A., and Guo, X.W. (1990). Interaction between the Vdac Channel and a Polyanionic Effector - an Electron-Microscopic Study. *Biophys. J.* 57, 23-31.
- Martin, W.F., and Mentel, M. (2010). The origin of mitochondria. *Nature Education* 3, 58.
- Matlahov, I., and van der Wel, P.C.A. (2018). Hidden motions and motion-induced invisibility: Dynamics-based spectral editing in solid-state NMR. *Methods* 148, 123-135.
- Menzel, V.A., Cassara, M.C., Benz, R., de Pinto, V., Messina, A., Cunsolo, V., Saletti, R., Hinsch, K.D., and Hinsch, E. (2009). Molecular and functional characterization of VDAC2 purified from mammal spermatozoa. *Bioscience. Rep.* 29, 351-362.
- Mertins, B., Psakis, G., Grosse, W., Back, K.C., Salisowski, A., Reiss, P., Koert, U., and Essen, L.O. (2012). Flexibility of the N-terminal mVDAC1 segment controls the channel's gating behavior. *PLOS ONE* 7, e47938.
- Messina, A., Reina, S., Guarino, F., and De Pinto, V. (2012). VDAC isoforms in mammals. *Bba-Biomembranes* 1818, 1466-1476.
- Moulder, S.L., Symmans, W.F., Booser, D.J., Madden, T.L., Lipsanen, C., Yuan, L., Brewster, A.M., Cristofanilli, M., Hunt, K.K., Buchholz, T.A., *et al.* (2008). Phase I/II study of G3139 (Bcl-2 antisense oligonucleotide) in combination with doxorubicin and docetaxel in breast cancer. *Clin. Cancer Res.* 14, 7909-7916.
- Movellan, K.T., Najbauer, E.E., Pratihari, S., Salvi, M., Giller, K., Becker, S., and Andreas, L.B. (2019). Alpha protons as NMR probes in deuterated proteins. *J. Biomol. NMR* 73, 81-91.
- Mukerjee, P. (1970). Odd-Even Alternation in Chain Length Variation of Micellar Properties - Evidence of Some Solid-Like Character of Micelle Core. *Kolloid Z. Z. Polym.* 236, 76-79.
- Najbauer, E.E., and Andreas, L.B. (2019). Correcting for magnetic field drift in magic-angle spinning NMR datasets. *J. Magn. Reson.* 305, 1-4.
- Najbauer, E.E., Movellan, K.T., Schubeis, T., Schwarzer, T., Castiglione, K., Giller, K., Pintacuda, G., Becker, S., and Andreas, L.B. (2019). Probing Membrane Protein Insertion into Lipid Bilayers by Solid-State NMR. *ChemPhysChem* 20, 302-310.
- Noskov, S.Y., Rostovtseva, T.K., Chamberlin, A.C., Tejjido, O., Jiang, W., and Bezrukov, S.M. (2016). Current state of theoretical and experimental studies of the voltage-dependent anion channel (VDAC). *Biochim Biophys Acta* 1858, 1778-1790.

O'Brien, S., Moore, J.O., Boyd, T.E., Larratt, L.M., Skotnicki, A., Koziner, B., Chanan-Khan, A.A., Seymour, J.F., Bociek, R.G., Pavletic, S., *et al.* (2007). Randomized phase III trial of fludarabine plus cyclophosphamide with or without oblimersen sodium (Bcl-2 antisense) in patients with relapsed or refractory chronic lymphocytic leukemia. *J. Clin. Oncol.* *25*, 1114-1120.

Oakley, H., Cole, S.L., Logan, S., Maus, E., Shao, P., Craft, J., Guillozet-Bongaarts, A., Ohno, M., Disterhoft, J., Van Eldik, L., *et al.* (2006). Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: Potential factors in amyloid plaque formation. *J. Neurosci.* *26*, 10129-10140.

Paulson, E.K., Morcombe, C.R., Gaponenko, V., Dancheck, B., Byrd, R.A., and Zilm, K.W. (2003). Sensitive high resolution inverse detection NMR spectroscopy of proteins in the solid state. *J. Am. Chem. Soc.* *125*, 15831-15836.

Pavlov, E., Grigoriev, S.M., Dejean, L.M., Zweihorn, C.L., Mannella, C.A., and Kinnally, K.W. (2005). The mitochondrial channel VDAC has a cation-selective open state. *Biochim. Biophys. Acta* *1710*, 96-102.

Penzel, S., Oss, A., Org, M.L., Samoson, A., Bockmann, A., Ernst, M., and Meier, B.H. (2019). Spinning faster: protein NMR at MAS frequencies up to 126kHz. *J. Biomol. NMR* *73*, 19-29.

Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997). 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.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* *25*, 1605-1612.

Pfaff, E., Klingenberg, M., and Heldt, H.W. (1965). Unspecific Permeation and Specific Exchange of Adenine Nucleotides in Liver Mitochondria. *Biochim. Biophys. Acta* *104*, 312-315.

Pfaller, R., Freitag, H., Harmey, M.A., Benz, R., and Neupert, W. (1985). A Water-Soluble Form of Porin from the Mitochondrial Outer-Membrane of *Neurospora-Crassa* - Properties and Relationship to the Biosynthetic Precursor Form. *J. Biol. Chem.* *260*, 8188-8193.

Popp, B., Schmid, A., and Benz, R. (1995). Role of Sterols in the Functional Reconstitution of Water-Soluble Mitochondrial Porins from Different Organisms. *Biochemistry* *34*, 3352-3361.

PubChem (2019). Oblimersen (National Library of Medicine).

Queralt-Martin, M., Bergdoll, L., Jacobs, D., Bezrukov, S.M., Abramson, J., and Rostovtseva, T.K. (2019). Assessing the role of residue E73 and lipid headgroup charge in VDAC1 voltage gating. *Biochimica et Biophysica Acta - Bioenergetics* *1860*, 22-29.

- Raschle, T., Hiller, S., Yu, T.Y., Rice, A.J., Walz, T., and Wagner, G. (2009). Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *J. Am. Chem. Soc.* *131*, 17777-17779.
- Rauch, G., and Moran, O. (1994). On the Structure of Mitochondrial Porins and Its Homologies with Bacterial Porins. *Biochem Biophys Res Commun* *200*, 908-915.
- Reif, B., Jaroniec, C.P., Rienstra, C.M., Hohwy, M., and Griffin, R.G. (2001). 1H-1H MAS correlation spectroscopy and distance measurements in a deuterated peptide. *J. Magn. Reson.* *151*, 320-327.
- Repalli, J. (2014). Translocator protein (TSPO) role in aging and Alzheimer's disease. *Curr. Aging Sci.* *7*, 168-175.
- Retel, J.S., Nieuwkoop, A.J., Hiller, M., Higman, V.A., Barbet-Massin, E., Stanek, J., Andreas, L.B., Franks, W.T., van Rossum, B.J., Vinothkumar, K.R., *et al.* (2017). Structure of outer membrane protein G in lipid bilayers. *Nat. Commun.* *8*.
- Roos, N., Benz, R., and Brdiczka, D. (1982). Identification and Characterization of the Pore-Forming Protein in the Outer-Membrane of Rat-Liver Mitochondria. *Biochim. Biophys. Acta* *686*, 204-214.
- Rosenhouse-Dantsker, A., and Logothetis, D.E. (2006). New roles for a key glycine and its neighboring residue in potassium channel gating. *Biophys. J.* *91*, 2860-2873.
- Rostovtseva, T., and Colombini, M. (1996). ATP flux is controlled by a voltage-gated channel from the mitochondrial outer membrane. *J. Biol. Chem.* *271*, 28006-28008.
- Rostovtseva, T., and Colombini, M. (1997). VDAC channels mediate and gate the flow of ATP: Implications for the regulation of mitochondrial function. *Biophys. J.* *72*, 1954-1962.
- Rostovtseva, T.K., Gurnev, P.A., Protchenko, O., Hoogerheide, D.P., Yap, T.L., Philpott, C.C., Lee, J.C., and Bezrukov, S.M. (2015). alpha-Synuclein Shows High Affinity Interaction with Voltage-dependent Anion Channel, Suggesting Mechanisms of Mitochondrial Regulation and Toxicity in Parkinson Disease. *J. Biol. Chem.* *290*, 18467-18477.
- Rostovtseva, T.K., Kazemi, N., Weinrich, M., and Bezrukov, S.M. (2006). Voltage gating of VDAC is regulated by nonlamellar lipids of mitochondrial membranes. *J. Biol. Chem.* *281*, 37496-37506.
- Saccone, C., Caggese, C., D'Erchia, A.M., Lanave, C., Oliva, M., and Pesole, G. (2003). Molecular clock and gene function. *J. Mol. Evol.* *57*, S277-S285.
- Saito, H., Tsuchida, T., Ogawa, K., Arakawa, T., Yamaguchi, S., and Tuzi, S. (2002). Residue-specific millisecond to microsecond fluctuations in bacteriorhodopsin induced by disrupted or disorganized two-dimensional crystalline lattice, through modified lipid-helix and helix-helix interactions, as revealed by 13C NMR. *Biochim. Biophys. Acta* *1565*, 97-106.

- Saraste, M. (1999). Oxidative phosphorylation at the fin de siecle. *Science* 283, 1488-1493.
- Schein, S.J., Colombini, M., and Finkelstein, A. (1976). Reconstitution in planar lipid bilayers of a voltage-dependent anion-selective channel obtained from paramecium mitochondria. *J. Membr. Biol.* 30, 99-120.
- Schleiff, E., Silvius, J.R., and Shore, G.C. (1999). Direct membrane insertion of voltage-dependent anion-selective channel protein catalyzed by mitochondrial Tom20. *J. Cell Biol.* 145, 973-978.
- Schmidt, C.D. (2017). NMR spectroscopic investigations on VDAC. In Max Planck Institute of Biophysical Chemistry (Göttingen: University of Göttingen).
- Schmidt, E., and Guntert, P. (2012). A new algorithm for reliable and general NMR resonance assignment. *J. Am. Chem. Soc.* 134, 12817-12829.
- Schneider, R., Etkorn, M., Giller, K., Daebel, V., Eisfeld, J., Zweckstetter, M., Griesinger, C., Becker, S., and Lange, A. (2010). The native conformation of the human VDAC1 N terminus. *Angew. Chem. Int. Ed. Engl.* 49, 1882-1885.
- Schredelseker, J., Paz, A., Lopez, C.J., Altenbach, C., Leung, C.S., Drexler, M.K., Chen, J.N., Hubbell, W.L., and Abramson, J. (2014). High resolution structure and double electron-electron resonance of the zebrafish voltage-dependent anion channel 2 reveal an oligomeric population. *J. Biol. Chem.* 289, 12566-12577.
- Schubeis, T., Le Marchand, T., Daday, C., Kopec, W., Tekwani Movellan, K., Stanek, J., Schwarzer, T.S., Castiglione, K., de Groot, B.L., Pintacuda, G., *et al.* (2020a). A beta-barrel for oil transport through lipid membranes: Dynamic NMR structures of AlkL. *Proc. Natl. Acad. Sci. U. S. A.* 117, 21014-21021.
- Schubeis, T., Schwarzer, T.S., Le Marchand, T., Stanek, J., Movellan, K.T., Castiglione, K., Pintacuda, G., and Andreas, L.B. (2020b). Resonance assignment of the outer membrane protein AlkL in lipid bilayers by proton-detected solid-state NMR. *Biomol. NMR Assign.* 14, 295-300.
- Shaka, A.J., Keeler, J., and Freeman, R. (1983a). Evaluation of a New Broad-Band Decoupling Sequence - Waltz-16. *J. Magn. Reson.* 53, 313-340.
- Shaka, A.J., Keeler, J., Frenkiel, T., and Freeman, R. (1983b). An Improved Sequence for Broad-Band Decoupling - Waltz-16. *J. Magn. Reson.* 52, 335-338.
- Shanmugavadivu, B., Apell, H.J., Meins, T., Zeth, K., and Kleinschmidt, J.H. (2007). Correct folding of the beta-barrel of the human membrane protein VDAC requires a lipid bilayer. *J. Mol. Biol.* 368, 66-78.
- Shastri, S., Vonck, J., Pflieger, N., Haase, W., Kuehlbrandt, W., and Glaubit, C. (2007). Proteorhodopsin: characterisation of 2D crystals by electron microscopy and solid state NMR. *Biochim. Biophys. Acta* 1768, 3012-3019.

Shen, Y., and Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* *56*, 227-241.

Shimizu, S., Ide, T., Yanagida, T., and Tsujimoto, Y. (2000). Electrophysiological study of a novel large pore formed by Bax and the voltage-dependent anion channel that is permeable to cytochrome c. *J. Biol. Chem.* *275*, 12321-12325.

Shimizu, S., Narita, M., and Tsujimoto, Y. (1999). Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. *Nature* *399*, 483-487.

Shoshan-Barmatz, V., Krelin, Y., and Shteinifer-Kuzmine, A. (2018a). VDAC1 functions in Ca(2+) homeostasis and cell life and death in health and disease. *Cell Calcium* *69*, 81-100.

Shoshan-Barmatz, V., Nahon-Crystal, E., Shteinifer-Kuzmine, A., and Gupta, R. (2018b). VDAC1, mitochondrial dysfunction, and Alzheimer's disease. *Pharmacol. Res.* *131*, 87-101.

Shoshan-Barmatz, V., Pittala, S., and Mizrachi, D. (2019). VDAC1 and the TSPO: Expression, Interactions, and Associated Functions in Health and Disease States. *Int. J. Mol. Sci.* *20*, 3348.

Song, J., Midson, C., Blachly-Dyson, E., Forte, M., and Colombini, M. (1998a). The sensor regions of VDAC are translocated from within the membrane to the surface during the gating processes. *Biophys. J.* *74*, 2926-2944.

Song, J.M., Midson, C., Blachly-Dyson, E., Forte, M., and Colombini, M. (1998b). The topology of VDAC as probed by biotin modification. *J. Biol. Chem.* *273*, 24406-24413.

Stahlberg, H., Fotiadis, D., Scheuring, S., Remigy, H., Braun, T., Mitsuoka, K., Fujiyoshi, Y., and Engel, A. (2001). Two-dimensional crystals: a powerful approach to assess structure, function and dynamics of membrane proteins. *FEBS letters* *504*, 166-172.

Stanek, J., Andreas, L.B., Jaudzems, K., Cala, D., Lalli, D., Bertarello, A., Schubeis, T., Akopjana, I., Kotelovica, S., Tars, K., *et al.* (2016). 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.

Stanek, J., Schubeis, T., Paluch, P., Güntert, P., Andreas, L.B., and Pintacuda, G. (2020). Automated Backbone NMR Resonance Assignment of Large Proteins Using Redundant Linking from a Single Simultaneous Acquisition. *J. Am. Chem. Soc.* *142*, 5793-5799.

Tan, W., and Colombini, M. (2007). VDAC closure increases calcium ion flux. *Biochim. Biophys. Acta* *1768*, 2510-2515.

Tan, W., Loke, Y.H., Stein, C.A., Miller, P., and Colombini, M. (2007). Phosphorothioate oligonucleotides block the VDAC channel. *Biophys. J.* *93*, 1184-1191.

Tejjido, O., Ujwal, R., Hillerdal, C.O., Kullman, L., Rostovtseva, T.K., and Abramson, J. (2012). Affixing N-terminal alpha-helix to the wall of the voltage-dependent anion channel does not prevent its voltage gating. *J. Biol. Chem.* *287*, 11437-11445.

- Thomas, L., Blachly-Dyson, E., Colombini, M., and Forte, M. (1993). Mapping of residues forming the voltage sensor of the voltage-dependent anion-selective channel. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 5446-5449.
- Tomasello, M.F., Guarino, F., Reina, S., Messina, A., and De Pinto, V. (2013). The voltage-dependent anion selective channel 1 (VDAC1) topography in the mitochondrial outer membrane as detected in intact cell. *PLOS ONE* *8*, e81522.
- Tsujimoto, Y., and Shimizu, S. (2000). VDAC regulation by the Bcl-2 family of proteins. *Cell Death Differ.* *7*, 1174-1181.
- Tucker, K., and Park, E. (2019). Cryo-EM structure of the mitochondrial protein-import channel TOM complex at near-atomic resolution. *Nat. Struct. Mol. Biol.* *26*, 1158-1166.
- Tugarinov, V., and Kay, L.E. (2003). 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.
- Ujwal, R., Cascio, D., Colletier, J.P., Faham, S., Zhang, J., Toro, L., Ping, P., and Abramson, J. (2008). 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.
- Unger, V.M., Kumar, N.M., Gilula, N.B., and Yeager, M. (1997). Projection structure of a gap junction membrane channel at 7 Å resolution. *Nat. Struct. Biol.* *4*, 39-43.
- Vasa, S.K., Singh, H., Grohe, K., and Linser, R. (2019). Assessment of a Large Enzyme-Drug Complex by Proton-Detected Solid-State NMR Spectroscopy without Deuteration. *Angew. Chem. Int. Ed. Engl.* *58*, 5758-5762.
- Veenman, L., and Gavish, M. (2006). The peripheral-type benzodiazepine receptor and the cardiovascular system. Implications for drug development. *Pharmacol. Therapeut.* *110*, 503-524.
- Veenman, L., Papadopoulos, V., and Gavish, M. (2007). Channel-like functions of the 18-kDa translocator protein (TSPO): Regulation of apoptosis and steroidogenesis as part of the host-defense response. *Curr. Pharm. Des.* *13*, 2385-2405.
- Veenman, L., Vainshtein, A., Yasin, N., Azrad, M., and Gavish, M. (2016). Tetrapyrroles as Endogenous TSPO Ligands in Eukaryotes and Prokaryotes: Comparisons with Synthetic Ligands. *Int. J. Mol. Sci.* *17*.
- Villinger, S. (2012). Dynamics and interactions of the voltage-dependent anion channel 1 studied by NMR spectroscopy. In Max Planck Institute for Biophysical Chemistry (Göttingen: University of Göttingen).
- Villinger, S., Briones, R., Giller, K., Zachariae, U., Lange, A., de Groot, B.L., Griesinger, C., Becker, S., and Zweckstetter, M. (2010). Functional dynamics in the voltage-dependent anion channel. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 22546-22551.

- Villinger, S., Giller, K., Bayrhuber, M., Lange, A., Griesinger, C., Becker, S., and Zweckstetter, M. (2014). Nucleotide interactions of the human voltage-dependent anion channel. *J. Biol. Chem.* *289*, 13397-13406.
- Wagner, G., Nasr, M., Ziarek, J., Baptista, D., Arthanari, H., Sun, Z.Y., Hagn, F., and Pluckthun, A. (2017). Engineering nanodiscs for membrane protein studies. *Abstr. Pap. Am. Chem. S.* *254*.
- Walz, T., and Grigorieff, N. (1998). Electron Crystallography of Two-Dimensional Crystals of Membrane Proteins. *J. Struct. Biol.* *121*, 142-161.
- Weiser, B.P., Salari, R., Eckenhoff, R.G., and Brannigan, G. (2014). Computational Investigation of Cholesterol Binding Sites on Mitochondrial VDAC. *J. Phys. Chem. B* *118*, 9852-9860.
- Wiedemann, N., Kozjak, V., Chacinska, A., Schonfisch, B., Rospert, S., Ryan, M.T., Pfanner, N., and Meisinger, C. (2003). Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature* *424*, 565-571.
- Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* *73*, 1-16.
- Xiang, S., Grohe, K., Rovo, P., Vasa, S.K., Giller, K., Becker, S., and Linser, R. (2015). Sequential backbone assignment based on dipolar amide-to-amide correlation experiments. *J. Biomol. NMR* *62*, 303-311.
- Xu, X., and Colombini, M. (1996). Self-catalyzed insertion of proteins into phospholipid membranes. *J. Biol. Chem.* *271*, 23675-23682.
- Xu, X., Decker, W., Sampson, M.J., Craigen, W.J., and Colombini, M. (1999). Mouse VDAC isoforms expressed in yeast: Channel properties and their roles in mitochondrial outer membrane permeability. *J. Membr. Biol.* *170*, 89-102.
- Yang, Z.R., Wang, C., Zhou, Q.X., An, J.L., Hildebrandt, E., Aleksandrov, L.A., Kappes, J.C., DeLucas, L.J., Riordan, J.R., Urbatsch, I.L., *et al.* (2014). Membrane protein stability can be compromised by detergent interactions with the extramembranous soluble domains. *Protein Sci.* *23*, 769-789.
- Yehezkel, G., Abu-Hamad, S., and Shoshan-Barmatz, V. (2007). An N-terminal nucleotide-binding site in VDAC1: involvement in regulating mitochondrial function. *J. Cell. Physiol.* *212*, 551-561.
- Zachariae, U., Schneider, R., Briones, R., Gattin, Z., Demers, J.P., Giller, K., Maier, E., Zweckstetter, M., Griesinger, C., Becker, S., *et al.* (2012). beta-Barrel mobility underlies closure of the voltage-dependent anion channel. *Structure* *20*, 1540-1549.
- Zaid, H., Abu-Hamad, S., Israelson, A., Nathan, I., and Shoshan-Barmatz, V. (2005). The voltage-dependent anion channel-1 modulates apoptotic cell death. *Cell Death Differ.* *12*, 751-760.

- Zalk, R., Israelson, A., Garty, E.S., Azoulay-Zohar, H., and Shoshan-Barmatz, V. (2005). Oligomeric states of the voltage-dependent anion channel and cytochrome c release from mitochondria. *Biochem. J.* *386*, 73-83.
- Zalman, L.S., Nikaido, H., and Kagawa, Y. (1980). Mitochondrial Outer-Membrane Contains a Protein Producing Nonspecific Diffusion Channels. *J. Biol. Chem.* *255*, 1771-1774.
- Zheng, L., Fishbein, K.W., Griffin, R.G., and Herzfeld, J. (1993). Two-dimensional solid-state proton NMR and proton exchange. *J. Am. Chem. Soc.* *115*, 6254-6261.
- Zhou, D.H., and Rienstra, C.M. (2008). High-performance solvent suppression for proton detected solid-state NMR. *J. Magn. Reson.* *192*, 167-172.
- Ziarek, J.J., Peterson, F.C., Lytle, B.L., and Volkman, B.F. (2011). Binding site identification and structure determination of protein-ligand complexes by NMR a semiautomated approach. *Methods Enzymol.* *493*, 241-275.
- Zimmerberg, J., and Parsegian, V.A. (1986). Polymer inaccessible volume changes during opening and closing of a voltage-dependent ionic channel. *Nature* *323*, 36-39.

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

This chapter is based entirely on the following publication, and contains identical text passages:

Eszter E. Najbauer*, Kumar Tekwani Movellan*, Tobias Schubeis, Tom Schwarzer, Kathrin Castiglione, Karin Giller, Guido Pintacuda, Stefan Becker, Loren B. Andreas. Probing Membrane Protein Insertion into Lipid Bilayers by Solid-State NMR, *ChemPhysChem*, 2019, 20:302-310.

*denotes equal contributions

2.1 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 bilayer may be necessary for maintaining membrane protein structure and function, it is crucial to develop methods for the study of membrane proteins in such near-native environments (Cross et al., 2011; Frey et al., 2017; Kurauskas et al., 2018). 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 (Eddy et al., 2015; Retel et al., 2017; Saurel et al., 2017). 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 (Lesage and Bockmann, 2003; Lesage et al., 2008; Bockmann et al., 2009) or membrane association (Comellas et al., 2012).

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 (Huster et al., 1999), which is still at least an order of magnitude longer than for the protein under moderate spinning of ~ 10 kHz. This allows the water or lipid signals to be selected based on differential relaxation times (Huster et al., 2002; Gallagher et al., 2004; Ader et al., 2009) or to distinguish multiple water pools with distinct relaxation properties within the same sample (Bockmann et al., 2009; Wang et al., 2017).

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 (Paulson et al., 2003). 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 (Lesage et al., 2008; Van Melckebeke et al., 2011). 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 (Fernandez et al., 2002). 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 2-1**.

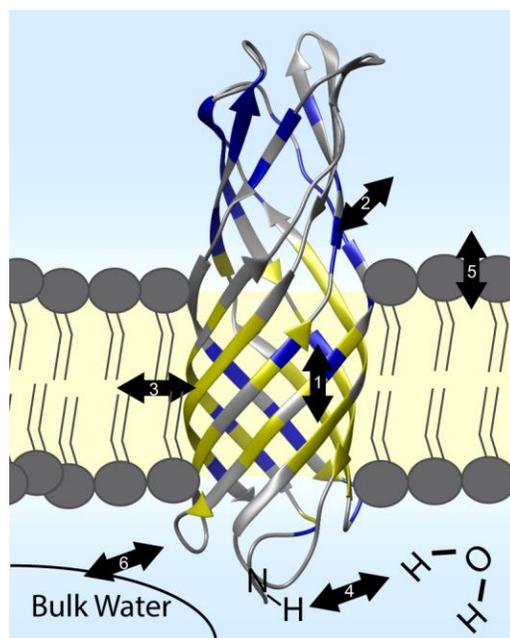


Figure 2-1. Schematic representation of longitudinal magnetization transfer in a lipid bilayer sample. 1-6 shows transfer between 1, protein and protein, 2, water and protein, 3, lipid and protein, 4, water and protein by chemical exchange, 5, lipid and water, and 6, bulk and associated water.

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 (Lange et al., 2002). 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 (Huster et al., 2002). 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 (Luo and Hong, 2010). 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 (Jia et al., 2008), and the contribution of spin diffusion to magnetization transfer is reduced dramatically, allowing

site resolved measurement of water proximity (Paulson et al., 2003; Lesage et al., 2008; Van Melckebeke et al., 2011). 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 (Lesage et al., 2008; Van Melckebeke et al., 2011). 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 (Zhou et al., 2007; Wickramasinghe et al., 2009; Huber et al., 2011; Knight et al., 2011; Linser et al., 2011; Kobayashi et al., 2013; Nishiyama et al., 2014; Andreas et al., 2016). 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 deuteration²⁹ 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 (Andreas et al., 2015). 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 (Kumashiro et al., 1998; Luo and Hong, 2006) or a selective excitation pulse (Ader et al., 2009). 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 (Huster et al., 2002). Alternatively, a spectral dimension can be used to distinguish mobile components from resonances of the protein (Paulson et al., 2003; Lesage et al.,

2008; Van Melckebeke et al., 2011; Comellas et al., 2012). 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.

2.2 Materials and methods

^2H , ^{13}C , ^{15}N -labeled AlkL was prepared as previously described (Jain et al., 2017; Schwarzer et al., 2017). 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 (Malia and Wagner, 2007; Bayrhuber et al., 2008; Hiller et al., 2008). Preparation of 2D crystals was carried out using the protocol modified by Eddy et al. (Eddy et al., 2012) from the original protocol published by Dolder et al. (Dolder et al., 1999).

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 (Shaka et al., 1983) at 10 kHz on both ^{13}C and ^{15}N . Protons were decoupled using 7 kHz TPPM decoupling (Bennett et al., 1995) during indirect acquisition times. The NOE mixing time was 50 ms. The recycle delay was 0.8 s. 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 (Goddard and Kneller, 2008).

2.3 Results and discussion

Figure 2-2 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 2-2A** shows the 3-dimensional HhNH pulse sequence, which is only slightly modified from the hNHH sequence of Zilm and coworkers (Paulson et al., 2003) 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 (**Figure 2-2B**) 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 (**Figure 2-2B**) or transferred echo double resonance (TEDOR, **Figure 2-2C**) (Hing et al., 1992). Similar to recent observations, the TEDOR transfer was found to be more efficient (Lends et al., 2018). 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.

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 CH_3 , acyl chain CH_2 and choline CH_3 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.

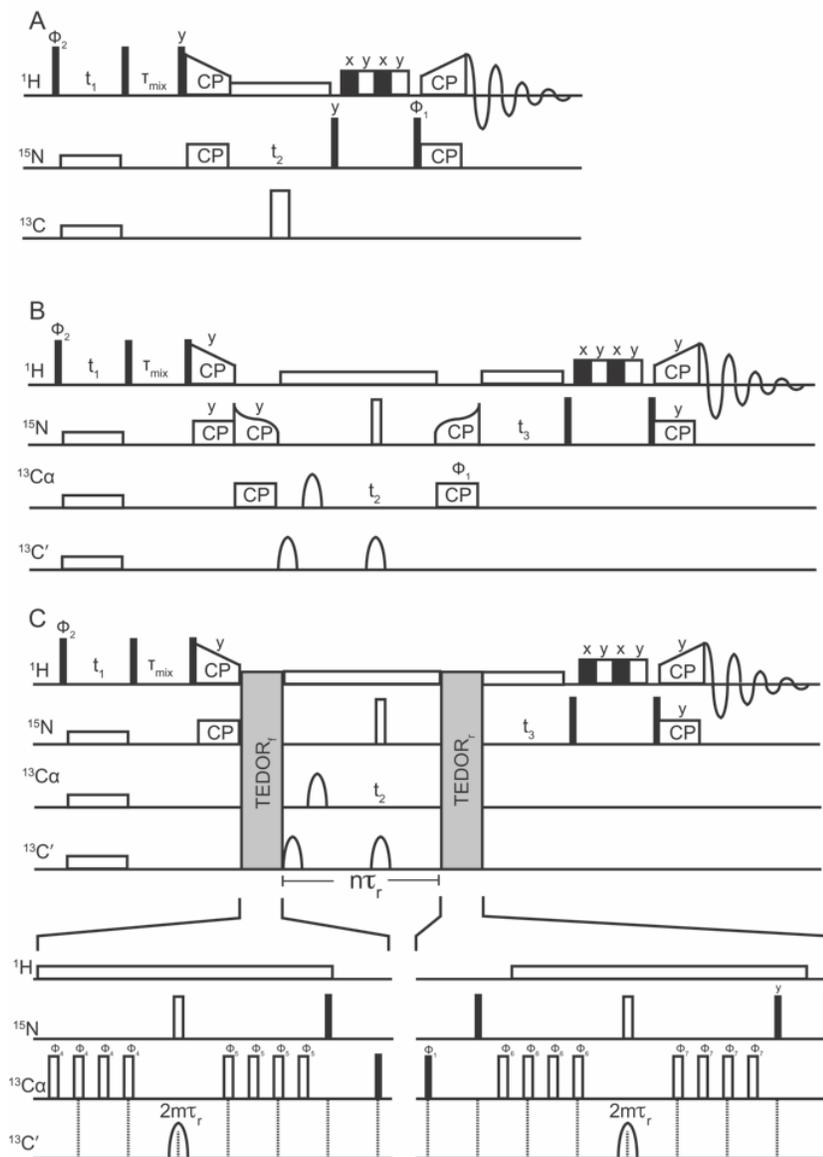


Figure 2-2: Pulse sequences used to record site resolved water and lipid proximity in membrane proteins. In a, the 3D HhNH sequence. In b, the CP based HhCANH sequence. In c, the TEDOR based HhCANH sequence. Solid and empty rectangles represent 90 and 180 degree pulses, respectively. Empty half-ovals represent selective 180 degree pulses, Q3 pulses (Emsley and Bodenhausen, 1992) in our implementation. Tangent ramped pulses were used on the ^{15}N channel for $^{13}\text{C}^{15}\text{N}$ 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 $\phi_1 = y - y$, $\phi_2 = x x - x - x$, $\phi_{\text{rec}} = y - y - y y$, in B, $\phi_1 = x - x$, $\phi_2 = x x - x - x$, $\phi_3 = y y - y - y - y - y y y$, $\phi_{\text{rec}} = y - y y - y - y y - y - y y$, and in C, $\phi_1 = x - x$, $\phi_2 = x x - x - x$, $\phi_3 = x x x x - x - x - x - x$, $\phi_{\text{rec}} = x - x - x x - x x x - x$. REDOR pulses on the carbon channel in C followed the xy-16 scheme. All other pulses were applied with phase x.

Figure 2-3 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 2-3A**, 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 (Veshtort and Griffin, 2011; Wittmann et al., 2016). 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.

An initial rate approximation shows transfer rates of 3-15 s⁻¹ within the protein (green dashed lines in **Figure 2-3**). 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 order of magnitude reduction in the deuterated sample at 55 kHz as compared with protonated samples at 4 - 5 kHz MAS (Kumashiro et al., 1998; Huster et al., 2002; Luo and Hong, 2010). 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 (Kumashiro et al., 1998; Huster et al., 2002). A value of D of 0.8 nm²/ms was measured in rigid organic polymers (Clauss et al., 1993). 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 (Kumashiro et al., 1998; Huster et al., 2002; Luo and Hong, 2010). 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 (Fick, 1995), 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 (Macura and Ernst, 1980; Huster et al., 1999). 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 (Ader et al., 2009; Luo and Hong, 2010). 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,

$$\frac{\partial M_p(t)}{\partial t} = (-R_{1,p} - R_{mp} - R_{pp'})M_p + (R_{mp})M_m + (R_{pp'})M_{p'} \quad (2 - 1)$$

$$\frac{\partial M_m(t)}{\partial t} = \left(-R_{1,m} - \frac{1}{N}R_{mp}\right)M_m + \frac{1}{N}(R_{mp})M_p \quad (2 - 2)$$

Where the subscript p is protein, and m is for the mobile species, either lipid or water. $R_{l,p}$ is the longitudinal relaxation rate of protein, and the cross relaxation between mobile species and protein is R_{wp} . 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, $R_{1,m_{eff}}$

$$\frac{\partial M_m(t)}{\partial t} = \left(-R_{1,m_{eff}}\right) M_m \quad (2 - 3)$$

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 (Bockmann et al., 2009). 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,

$$M_p(t) = M_{0,m} \frac{R_{mp}}{R_{1,p} + R_{mp} - R_{1,m_{eff}}} \left(e^{-R_{1,m_{eff}}t} - e^{-(R_{1,p}+R_{mp})t} \right) \quad (2 - 4)$$

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 2-3**.

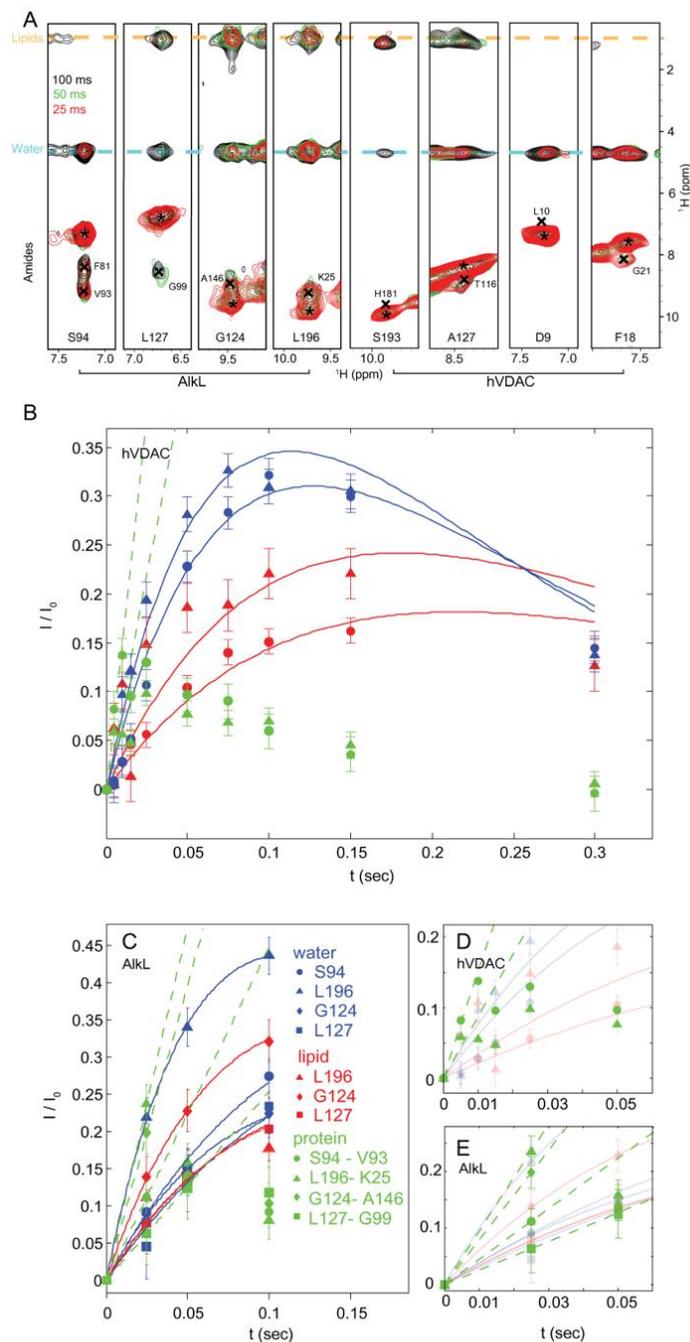


Figure 2-3: Evolution of magnetization in the HhNH spectrum of AlkL and hVDAC for selected residues. In A, the proton-proton planes from isolated residues are shown for 25, 50, and 100 ms in red, green, and black, respectively. The evolution of cross-peak intensities in hVDAC in B shows the water-protein 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 S2 in the Appendix. Error bars correspond to the root mean squared noise of the spectra.

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 T_1 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 ($T_1 \sim 185$ ms), we can safely assume that the protein T_1 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 (Paulson et al., 2003). 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 (Macura and Ernst, 1980; Huster et al., 1999). 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} (Figure S2 in the Appendix 5.2.2). 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 (Paulson et al., 2003).

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 $^1\text{H } T_1$ 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 (Oster et al., 2017). 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 s_1 even with protein deuteration and very fast MAS.

Figure 2-4 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 (Eddy et al., 2015) 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 (Hiller et al., 2008), as well as the placement of the helix within the large water-filled pore in crystals and in micelles (Bayrhuber et al., 2008; Ujwal et al., 2008; Jaremko et al., 2016). 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.

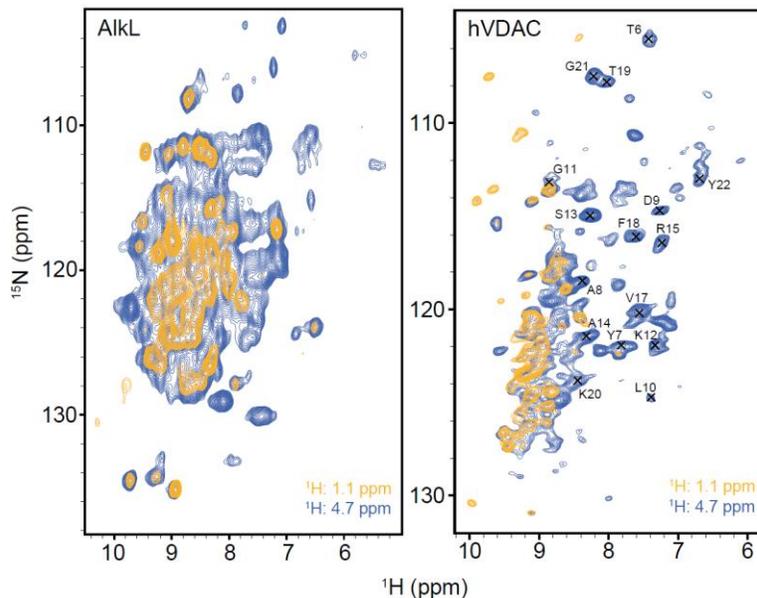


Figure 2-4: Lipid and water signal detected at the protein amides for AlkL and hVDAC. The 2D planes of the 3D HhNH spectrum are shown corresponding to the lipids (yellow, 1.1 ppm) and water (blue, 4.7 ppm). Assignments of the N-terminal helix are indicated for hVDAC. Proton-proton mixing proceeded for 50 ms.

Site-resolved correlations between the protein and the surrounding small molecules can be obtained in the HhnCANH spectrum. **Figure 2-5** 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 CH₂ and CH₃ 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 (Figure S1 in the Appendix 5.2.2). 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, ¹³CA, ¹⁵N, and ¹H.

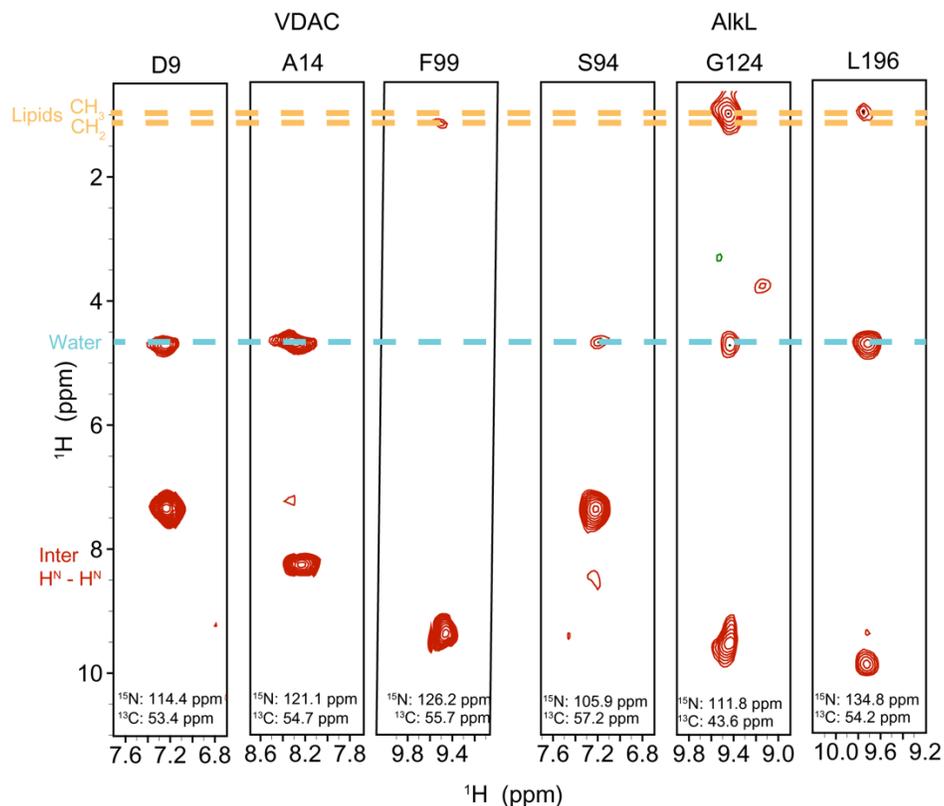


Figure 2-5: Selected strips from the 4D spectra of hVDAC and AlkL using 50 ms proton-proton mixing. The position of water and lipids are indicated by blue and yellow dashed lines, respectively. In most slices, only a single intense autocorrelation peak is observed in the amide region, indicating minimal magnetization transfer within the protein.

Figure 2-6 shows water- and lipid-protein correlations mapped onto the OmpW homology model of AlkL, generated with swissmodel server (Waterhouse et al., 2018) and the solution NMR structure of hVDAC (Jaremko et al., 2016). 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 (Hong et al., 2006; Touw et al., 2010; Horst et al., 2014; Kucharska et al., 2015) 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.

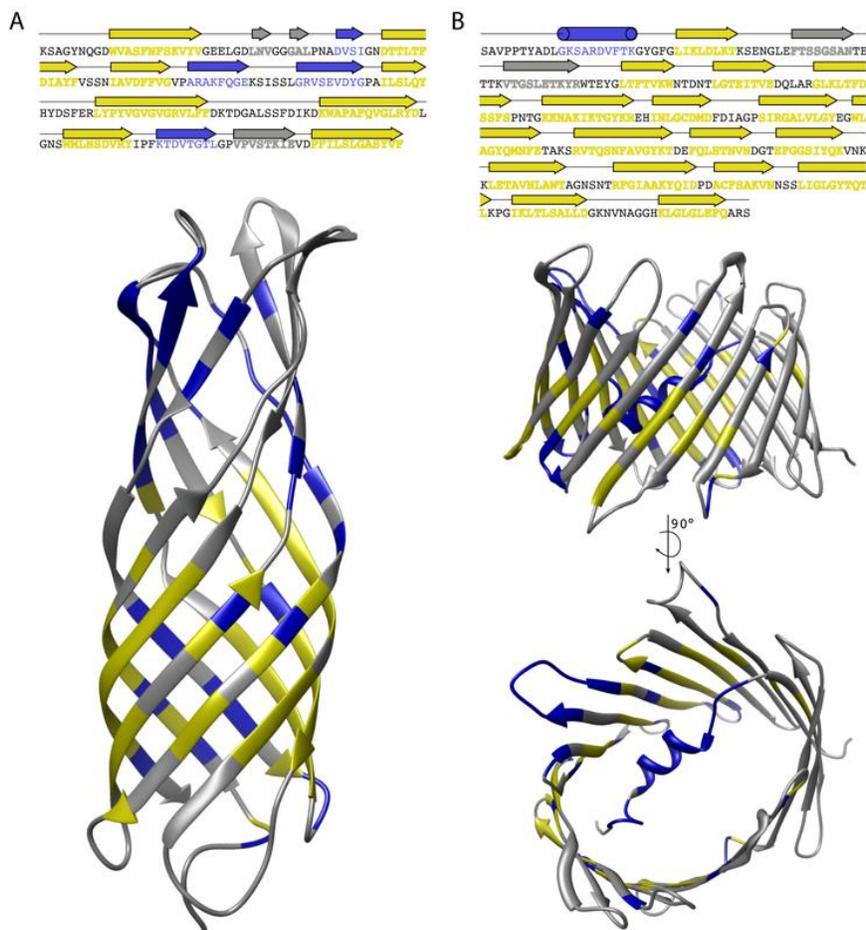


Figure 2-6: Water and lipid contacts shown on the homology model of AlkL using OmpW (PDB 2F1T) as a template (A) and the solution NMR structure (PDB 5JDP) of hVDAC (B). All residues that show a lipid contact are colored yellow, while residues that show only water contact are colored blue, and residues 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.

For hVDAC, combination of both 3D and 4D data (**Figure 2-6**) 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 (Dev et al., 2016). Whereas PRE has also been used for defining protein-protein (Oster et al., 2017) or water-protein (Linser et al., 2009) 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.

2.4 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.

2.5 References

- Ader, C., Schneider, R., Seidel, K., Etzkorn, M., Becker, S., and Baldus, M. (2009). Structural rearrangements of membrane proteins probed by water-edited solid-state NMR spectroscopy. *J. Am. Chem. Soc.* *131*, 170-176.
- Andreas, L.B., Jaudzems, K., Stanek, J., Lalli, D., Bertarello, A., Le Marchand, T., Cala-De Paepe, D., Kotelovica, S., Akopjana, I., Knott, B., *et al.* (2016). Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc. Natl. Acad. Sci. U S A* *113*, 9187-9192.
- Andreas, L.B., Le Marchand, T., Jaudzems, K., and Pintacuda, G. (2015). High-resolution proton-detected NMR of proteins at very fast MAS. *Journal of magnetic resonance* *253*, 36-49.
- Bayrhuber, M., Meins, T., Habeck, M., Becker, S., Giller, K., Villinger, S., Vonrhein, C., Griesinger, C., Zweckstetter, M., and Zeth, K. (2008). Structure of the human voltage-dependent anion channel. *Proc. Natl. Acad. Sci. U S A* *105*, 15370-15375.
- Bennett, A.E., Rienstra, C.M., Auger, M., Lakshmi, K.V., and Griffin, R.G. (1995). Heteronuclear Decoupling in Rotating Solids. *J. Chem. Phys.* *103*, 6951-6958.
- Bockmann, A., Gardiennet, C., Verel, R., Hunkeler, A., Loquet, A., Pintacuda, G., Emsley, L., Meier, B.H., and Lesage, A. (2009). Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. NMR.* *45*, 319-327.
- Clauss, J., Schmidt-Rohr, K., and Spiess, H.W. (1993). Determination of Domain Sizes in Heterogeneous Polymers by Solid-State Nmr. *Acta Polym.* *44*, 1-17.
- Comellas, G., Lemkau, L.R., Zhou, D.H., George, J.M., and Rienstra, C.M. (2012). Structural intermediates during alpha-synuclein fibrillogenesis on phospholipid vesicles. *J. Am. Chem. Soc.* *134*, 5090-5099.
- Cross, T.A., Sharma, M., Yi, M., and Zhou, H.X. (2011). Influence of solubilizing environments on membrane protein structures. *Trends Biochem. Sci.* *36*, 117-125.
- Dev, J., Park, D., Fu, Q., Chen, J., Ha, H.J., Ghantous, F., Herrmann, T., Chang, W., Liu, Z., Frey, G., *et al.* (2016). Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* *353*, 172-175.
- Dolder, M., Zeth, K., Tittmann, P., Gross, H., Welte, W., and Wallimann, T. (1999). Crystallization of the human, mitochondrial voltage-dependent anion-selective channel in the presence of phospholipids. *J. Struct. Biol.* *127*, 64-71.
- Eddy, M.T., Ong, T.C., Clark, L., Tejjido, O., van der Wel, P.C., Garces, R., Wagner, G., Rostovtseva, T.K., and Griffin, R.G. (2012). 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.

Eddy, M.T., Su, Y., Silvers, R., Andreas, L., Clark, L., Wagner, G., Pintacuda, G., Emsley, L., and Griffin, R.G. (2015). Lipid bilayer-bound conformation of an integral membrane beta barrel protein by multidimensional MAS NMR. *J. Biomol. NMR* *61*, 299-310.

Emsley, L., and Bodenhausen, G. (1992). Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. *Journal of magnetic resonance* *97*, 135-148.

Fernandez, C., Hilty, C., Wider, G., and Wuthrich, K. (2002). Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR spectroscopy. *Proc. Natl. Acad. Sci. U S A* *99*, 13533-13537.

Fick, A. (1995). On Liquid Diffusion (Reprinted from the London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Vol 10, Pg 30, 1855). *J. Membrane Sci.* *100*, 33-38.

Frey, L., Lakomek, N.A., Riek, R., and Bibow, S. (2017). Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics. *Angew. Chem. Int. Ed. Engl.* *56*, 380-383.

Gallagher, G.J., Hong, M., and Thompson, L.K. (2004). Solid-state NMR spin diffusion for measurement of membrane-bound peptide structure: gramicidin A. *Biochemistry* *43*, 7899-7906.

Goddard, T.D., and Kneller, D.G. (2008). SPARKY 3 (University of California, San Francisco).

Hiller, S., Garces, R.G., Malia, T.J., Orekhov, V.Y., Colombini, M., and Wagner, G. (2008). Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* *321*, 1206-1210.

Hing, A.W., Vega, S., and Schaefer, J. (1992). Transferred-echo double-resonance NMR. *Journal of magnetic resonance* *96*, 205-209.

Hong, H., Patel, D.R., Tamm, L.K., and van den Berg, B. (2006). The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. *J. Biol. Chem.* *281*, 7568-7577.

Horst, R., Stanczak, P., and Wuthrich, K. (2014). NMR polypeptide backbone conformation of the *E. coli* outer membrane protein W. *Structure* *22*, 1204-1209.

Huber, M., Hiller, S., Schanda, P., Ernst, M., Bockmann, A., Verel, R., and Meier, B.H. (2011). A proton-detected 4D solid-state NMR experiment for protein structure determination. *ChemPhysChem* *12*, 915-918.

Huster, D., Arnold, K., and Gawrisch, K. (1999). Investigation of lipid organization in biological membranes by two-dimensional nuclear overhauser enhancement spectroscopy. *J. Phys. Chem. B* *103*, 243-251.

Huster, D., Yao, X.L., and Hong, M. (2002). Membrane protein topology probed by H-1 spin diffusion from lipids using solid-state NMR spectroscopy. *J. Am. Chem. Soc.* *124*, 874-883.

Jain, M.G., Lalli, D., Stanek, J., Gowda, C., Prakash, S., Schwarzer, T.S., Schubeis, T., Castiglione, K., Andreas, L.B., Madhu, P.K., *et al.* (2017). 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.

Jaremko, M., Jaremko, L., Villinger, S., Schmidt, C.D., Griesinger, C., Becker, S., and Zweckstetter, M. (2016). High-Resolution NMR Determination of the Dynamic Structure of Membrane Proteins. *Angew. Chem. Int. Ed. Engl.* *55*, 10518-10521.

Jia, Z.L., Zhang, L.L., Chen, Q., and Hansen, E.W. (2008). Proton spin diffusion in polyethylene as a function of magic-angle spinning rate. A phenomenological approach. *J. Phys. Chem. A* *112*, 1228-1233.

Knight, M.J., Webber, A.L., Pell, A.J., Guerry, P., Barbet-Massin, E., Bertini, I., Felli, I.C., Gonnelli, L., Pierattelli, R., Emsley, L., *et al.* (2011). 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.

Kobayashi, T., Mao, K., Paluch, P., Nowak-Krol, A., Sniechowska, J., Nishiyama, Y., Gryko, D.T., Potrzebowski, M.J., and Pruski, M. (2013). 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.

Kucharska, I., Seelheim, P., Edrington, T., Liang, B., and Tamm, L.K. (2015). OprG Harnesses the Dynamics of its Extracellular Loops to Transport Small Amino Acids across the Outer Membrane of *Pseudomonas aeruginosa*. *Structure* *23*, 2234-2245.

Kumashiro, K.K., Schmidt-Rohr, K., Murphy, O.J., Ouellette, K.L., Cramer, W.A., and Thompson, L.K. (1998). A novel tool for probing membrane protein structure: Solid-state NMR with proton spin diffusion and X-nucleus detection. *J. Am. Chem. Soc.* *120*, 5043-5051.

Kurauskas, V., Hessel, A., Ma, P.X., Lunetti, P., Weinhaupl, K., Imbert, L., Brutscher, B., King, M.S., Sounier, R., Dolce, V., *et al.* (2018). How Detergent Impacts Membrane Proteins: Atomic-Level Views of Mitochondrial Carriers in Dodecylphosphocholine. *J. Chem. Phys. Lett.* *9*, 933-+.

Lange, A., Luca, S., and Baldus, M. (2002). Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *J. Am. Chem. Soc.* *124*, 9704-9705.

Lends, A., Ravotti, F., Zandomenighi, G., Bockmann, A., Ernst, M., and Meier, B.H. (2018). Direct amide (15)N to (13)C transfers for solid-state assignment experiments in deuterated proteins. *J. Biomol. NMR*.

Lesage, A., and Bockmann, A. (2003). Water-protein interactions in microcrystalline crh measured by 1H-13C solid-state NMR spectroscopy. *J. Am. Chem. Soc.* *125*, 13336-13337.

Lesage, A., Gardiennet, C., Loquet, A., Verel, R., Pintacuda, G., Emsley, L., Meier, B.H., and Bockmann, A. (2008). Polarization transfer over the water-protein interface in solids. *Angew. Chem. Int. Ed. Engl.* *47*, 5851-5854.

Linser, R., Dasari, M., Hiller, M., Higman, V., Fink, U., Lopez del Amo, J.M., Markovic, S., Handel, L., Kessler, B., Schmieder, P., *et al.* (2011). Proton-detected solid-state NMR spectroscopy of fibrillar and membrane proteins. *Angew. Chem. Int. Ed. Engl.* *50*, 4508-4512.

Linser, R., Fink, U., and Reif, B. (2009). Probing surface accessibility of proteins using paramagnetic relaxation in solid-state NMR spectroscopy. *J. Am. Chem. Soc.* *131*, 13703-13708.

Luo, W., and Hong, M. (2006). A 1D sensitivity-enhanced ¹H spin diffusion experiment for determining membrane protein topology. *Solid State Nucl. Magn. Reson.* *29*, 163-169.

Luo, W., and Hong, M. (2010). Conformational changes of an ion channel detected through water-protein interactions using solid-state NMR spectroscopy. *J. Am. Chem. Soc.* *132*, 2378-2384.

Macura, S., and Ernst, R.R. (1980). Elucidation of Cross Relaxation in Liquids by Two-Dimensional Nmr-Spectroscopy. *Mol. Phys.* *41*, 95-117.

Malia, T.J., and Wagner, G. (2007). NMR structural investigation of the mitochondrial outer membrane protein VDAC and its interaction with antiapoptotic Bcl-xL. *Biochemistry* *46*, 514-525.

Nishiyama, Y., Malon, M., Ishii, Y., and Ramamoorthy, A. (2014). 3D ¹⁵N/¹⁵N/¹H chemical shift correlation experiment utilizing an RFDR-based ¹H/¹H mixing period at 100 kHz MAS. *Journal of magnetic resonance* *244*, 1-5.

Oster, C., Kosol, S., Hartmuller, C., Lamley, J.M., Iuga, D., Oss, A., Org, M.L., Vanatalu, K., Samoson, A., Madl, T., *et al.* (2017). Characterization of Protein-Protein Interfaces in Large Complexes by Solid-State NMR Solvent Paramagnetic Relaxation Enhancements. *J. Am. Chem. Soc.* *139*, 12165-12174.

Paulson, E.K., Morcombe, C.R., Gaponenko, V., Dancheck, B., Byrd, R.A., and Zilm, K.W. (2003). 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.

Retel, J.S., Nieuwkoop, A.J., Hiller, M., Higman, V.A., Barbet-Massin, E., Stanek, J., Andreas, L.B., Franks, W.T., van Rossum, B.J., Vinothkumar, K.R., *et al.* (2017). Structure of outer membrane protein G in lipid bilayers. *Nat. Commun.* *8*, 2073.

Saurel, O., Iordanov, I., Nars, G., Demange, P., Le Marchand, T., Andreas, L.B., Pintacuda, G., and Milon, A. (2017). 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.

Schwarzer, T.S., Hermann, M., Krishnan, S., Simmel, F.C., and Castiglione, K. (2017). Preparative refolding of small monomeric outer membrane proteins. *Protein Expr. Purif.* *132*, 171-181.

Shaka, A.J., Keeler, J., Frenkiel, T., and Freeman, R. (1983). An Improved Sequence for Broad-Band Decoupling - Waltz-16. *J. Magn. Reson.* *52*, 335-338.

Touw, D.S., Patel, D.R., and van den Berg, B. (2010). The crystal structure of OprG from *Pseudomonas aeruginosa*, a potential channel for transport of hydrophobic molecules across the outer membrane. *PLoS One* *5*, e15016.

Ujwal, R., Cascio, D., Colletier, J.P., Faham, S., Zhang, J., Toro, L., Ping, P., and Abramson, J. (2008). 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.

Van Melckebeke, H., Schanda, P., Gath, J., Wasmer, C., Verel, R., Lange, A., Meier, B.H., and Bockmann, A. (2011). Probing water accessibility in HET-s(218-289) amyloid fibrils by solid-state NMR. *J. Mol. Biol.* *405*, 765-772.

Veshtort, M., and Griffin, R.G. (2011). Proton-driven spin diffusion in rotating solids via reversible and irreversible quantum dynamics. *J. Chem. Phys.* *135*.

Wang, T., Jo, H., DeGrado, W.F., and Hong, M. (2017). Water Distribution, Dynamics, and Interactions with Alzheimer's beta-Amyloid Fibrils Investigated by Solid-State NMR. *J. Am. Chem. Soc.* *139*, 6242-6252.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* *46*, W296-W303.

Wickramasinghe, N.P., Parthasarathy, S., Jones, C.R., Bhardwaj, C., Long, F., Kotecha, M., Mehboob, S., Fung, L.W.M., Past, J., Samoson, A., *et al.* (2009). Nanomole-scale protein solid-state NMR by breaking intrinsic H-1 T-1 boundaries. *Nat. Meth.* *6*, 215-218.

Wittmann, J.J., Agarwal, V., Hellwagner, J., Lends, A., Cadalbert, R., Meier, B.H., and Ernst, M. (2016). Accelerating proton spin diffusion in perdeuterated proteins at 100 kHz MAS. *J. Biomol. NMR* *66*, 233-242.

Zhou, D.H., Shea, J.J., Nieuwkoop, A.J., Franks, W.T., Wylie, B.J., Mullen, C., Sandoz, D., and Rienstra, C.M. (2007). Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* *46*, 8380-8383.

3 Correcting for magnetic field drift in MAS NMR

This chapter is based entirely on the following publication, and contains identical text passages:

Eszter E. Najbauer, Loren B. Andreas. Correcting for magnetic field drift in magic-angle spinning NMR datasets. *J Magn Reson*, 2019, 305:1-4

3.1 Introduction

Drift in the main magnetic field in magnetic resonance techniques is a well-known issue. If not corrected for, the drift of the B_0 magnetic field results in broadening of spectral peaks and distortion of lineshapes. In magnetic resonance imaging (MRI), using spectral registration, each spectral average is fit to a reference scan in the time domain by adjusting both frequency and phase terms (Near et al., 2015). In solution NMR spectroscopy, the issue is usually solved by correcting the magnetic field while the measurement is recorded. This field lock is typically implemented by tracking the field drift by detecting deuterium and constantly adjusting the main field with the use of room temperature electromagnets (Baker and Burd, 1957; Hofer et al., 1978).

In magic angle spinning (MAS) NMR measurements, this solution can be impractical, due to the small sample volume, which results in low sensitivity of deuterium and an unstable lock. Additionally, many probes built for proton detected MAS NMR do not have a deuterium channel. An alternative approach for carbon detected spectra is to use the proton channel for a lock (Maly et al., 2006). External locks can also be used to track the main field. In this case, a small sample (e.g. containing deuterium oxide in solution) is placed close to the MAS rotor in its own dedicated detection coil (Paulson and Zilm, 2005). Such external lock systems are available from instrument

manufacturers, but they do not entirely remove drift, since they are not detecting the field at the sample, and in addition, they are often not temperature controlled, resulting in a long equilibration time before the temperature sensitive D₂O sample can be used. The result is that MAS NMR data is often acquired while the main field drifts. Even adjusting the linear drift compensation of the spectrometer, we often observe drift of up to 0.075 ppm after a 24 hour measurement. For ¹³C-detected data, this is usually insignificant, since linewidths are generally greater than 0.3 ppm. However, proton-detected spectra of deuterated microcrystals can have linewidths below 0.05 ppm (Lewandowski et al., 2011; Agarwal et al., 2014), and even the less ideal preparations of membrane proteins can have linewidths of around 0.1 ppm (Lakomek et al., 2017).

In solution NMR, the lock is also important in order to maintain good water suppression. This is because the typically used water suppression schemes, such as presaturation methods (Neuhaus et al., 1996; Simpson and Brown, 2005; Mo and Raftery, 2008) and WATERGATE (Piotto et al., 1992; Sklenar et al., 1993) are highly dependent on the carrier frequency. With a small drift, the water suppression can be severely compromised. This is not the case in cross polarization-based proton detected MAS NMR, where the water is suppressed by relatively strong saturation pulses (Zhou and Rienstra, 2008; Barbet-Massin et al., 2014; Fricke et al., 2017). It is therefore possible to correct for large drift in the field, provided the drift is known. Alternatively, the drift may be assumed to be linear if the acquisition time is made sufficiently short, such that a linear correction is a good approximation of the actual drift.

3.2 Results

3.2.1 Linear phase correction of the FID

We apply two principles to acquire the best resolution possible under conditions of a slow field drift. First, we acquire data in short blocks of about 24 hours, or even shorter, such that the drift over this time will be approximately linear. Longer acquisitions can be divided into 24-36 hour acquisitions by use of reduced phase cycles and non-uniform sampling. Next, we correct the data in the time domain assuming linear drift occurred during the measurement. This drift is determined from a 1D spectrum with a sensitive and narrow line. Correction of a signal with frequency ν by a frequency offset ν_0 in the time domain is simply a linear ramp in the phase, according to the well-known Fourier relation (Fourier, 1822, 2009). Applying a linear phase ramp of $2\pi\nu_0$ to an FID of $S(t) = S_0 \cdot \exp(i2\pi\nu t_{aq})$, we obtain the expected frequency shift of ν_0 , which can be seen in either the time domain, or in the frequency domain after Fourier transformation:

$$\begin{aligned} S(t) \cdot \exp(i2\pi\nu_0 t_{aq}) &= S_0 \cdot \exp(i2\pi\nu t_{aq}) \cdot \exp(i2\pi\nu_0 t_{aq}) \\ &= S_0 \cdot \exp(i2\pi(\nu + \nu_0)t_{aq}) \xrightarrow{FT} S(\nu + \nu_0) \end{aligned} \quad (3 - 1)$$

For multidimensional data, the correction is made across each dimension, keeping track of multiple time values: the actual point in time when the data was acquired, as well as the evolution time in each indirect dimension. The actual time determines the frequency correction needed, while the indirect evolution determines the location on the linear ramp. For best results, all the phases for a complex point should be acquired before moving to the next point, such that all phases are acquired at nearly the same magnetic field. In practice, we have found that the drift is small enough that the data can be acquired as planes, and it can be assumed that the full plane was acquired at a single magnetic field.

For correction of a complex point for 2D spectra, the correction is extended as follows. In a 2D spectrum, one complex point comprises 4 FIDs: rr (real in both indirect and direct dimensions), ri

(real in indirect, imaginary in direct dimension), *ir* (imaginary in indirect, real in direct dimension), and *ii* (imaginary in both dimensions).

The phase correction to be applied to any complex point consisting of 4 FIDs in the direct dimension is:

$$\phi_{\text{dir}} = \left(f_{\text{start}} + i \frac{f_{\text{end}} - f_{\text{start}}}{n} \right) t_{\text{dir}} * 2\pi \quad (3 - 2)$$

where ϕ_{dir} is the phase correction needed for correction the direct dimension, f_{start} and f_{end} are the required frequency shifts (measured on reference 1D spectra) at the beginning and end of recording the given spectrum, n is the total number of complex points in the spectrum, i is the index of the complex point, and $t_{\text{dir}} = j * dw_{\text{dir}}$ is the acquisition time up to that point in the direct dimension, where j is the corrected point's index within the FID, and dw_{dir} is the dwell time in the direct dimension. The phase correction for the direct dimension is done looping through the 4 FIDs belonging to the complex point.

After correction of drift in the direct dimension, the FIDs are: rr_c , ri_c , ir_c , ii_c :

$$rr_c = \cos \phi_{\text{dir}} * rr - \sin \phi_{\text{dir}} * ri \quad (3 - 3)$$

$$ri_c = \sin \phi_{\text{dir}} * rr + \cos \phi_{\text{dir}} * ri \quad (3 - 4)$$

$$ir_c = \cos \phi_{\text{dir}} * ir - \sin \phi_{\text{dir}} * ii \quad (3 - 5)$$

$$ii_c = \sin \phi_{\text{dir}} * ir + \cos \phi_{\text{dir}} * ii \quad (3 - 6)$$

The correction of the indirect dimension is done similarly. The phase is calculated:

$$\phi_{\text{id}} = \left(f_{\text{start}} + i \frac{f_{\text{end}} - f_{\text{start}}}{n} \right) t_{\text{id}} * 2\pi * \frac{\gamma_{\text{id}}}{\gamma_{\text{dir}}} \quad (3 - 7)$$

where all notations previously defined remain the same, γ_{dir} and γ_{id} are the gyromagnetic ratios of the detected nucleus and the nucleus in the indirect dimension, respectively, and $t_{\text{id}} = dw_{\text{id}} * i$, with dw_{id} being the dwell time in the indirect dimension.

The FIDs also corrected for drift in the indirect dimension thus are:

$$rr_{cc} = \cos \phi_{id} * rr_c - \sin \phi_{id} * ir_c \quad (3 - 8)$$

$$ir_{cc} = \sin \phi_{id} * rr_c + \cos \phi_{id} * ir_c \quad (3 - 9)$$

$$ri_{cc} = \cos \phi_{id} * ri_c - \sin \phi_{id} * ii_c \quad (3 - 10)$$

$$ii_{cc} = \sin \phi_{id} * ri_c + \cos \phi_{id} * ii_c \quad (3 - 11)$$

This principle can be extended to an arbitrary number of dimensions. The script used for drift correction of phase-sensitive data up to 4 dimensions is available in the Supplemental Information in the Appendix.

3.2.2 Application of drift correction

While datasets may contain no drift at all, we have found that occasionally, data would need to be thrown away if correction were not used. For example, the rate of drift is different and unpredictable for several days after filling helium. Also, when recording data over the course of several weeks, the linear drift compensation of the magnet may not be sufficient, and drifts of up to hundreds of proton Hz are not uncommon. The linear drift compensation might also be switched off automatically if the current in the Z0 shim coil reaches its maximal value, resulting in up to hundreds of proton Hz drift within one day. The linear approximation of the field drift during measurements can be determined by recording a 1D spectrum directly before and after the experiment. For biological samples, the proton 1D spectrum typically contains sufficient water, lipids, or other narrow signals with high intensity that can be used to precisely determine the drift.

We demonstrate drift correction using 2D, 3D and 4D spectra recorded on a sample of ^2H , ^{13}C , ^{15}N -labeled VDAC (voltage-dependent anion channel) prepared as previously described (Dolder et al., 1999; Eddy et al., 2012; Zachariae et al., 2012). All spectra were measured using a 1.3 mm HCN probe on a Bruker Avance III 800 MHz narrow bore spectrometer at 55 kHz MAS.

A 2D spectrum was recorded in 96 scans per point during an 8 hour period during which the linear drift compensation (31 proton Hz/hour) was turned off (**Figure 3-1**). The hCONH spectrum was measured over 12 hours following a reset of the Z0 shim coil (**Figure 3-2, Figure 3-3**), which is

needed on our 800 MHz instrument every ~6 months. As an example of a 4D spectrum, we recorded the HhCANH sequence, modified from the out-and-back variants in (Najbauer et al., 2019). The spectrum was recorded in one day blocks of 2 scans per point by applying 3.55% non-uniform sampling. In total, 19 blocks were recorded over 3 weeks and averaged either before or after drift correction (**Figure 3-4**). Although the linear drift compensation was active during the whole measurement, the magnet drifted within a range of ~350 Hz. The multi-dimensional decomposition (MDD) algorithm (Orekhov and Jaravine, 2011) in Topspin was used for reconstruction.

The drift correction for each spectrum was performed in Topspin using the script provided in the SI. Since the script is based on the C programming language, it can be easily modified for other data acquisition and processing environments by adapting the Bruker-specific AU commands used to read acquisition parameters. The runtime of the script is less than a second on typical personal computers, even for 4D data with > 5000 complex points.

The 2D hNH spectrum is characterized by distorted triangular (“butterfly”) peak shapes. Essentially, one side of the line is artificially narrowed, while the other side is artificially broadened during the drift (**Figure 3-1A**). Correction of the drift eliminates peak shape distortions and restores the expected round peak shapes (**Figure 3-1B**).

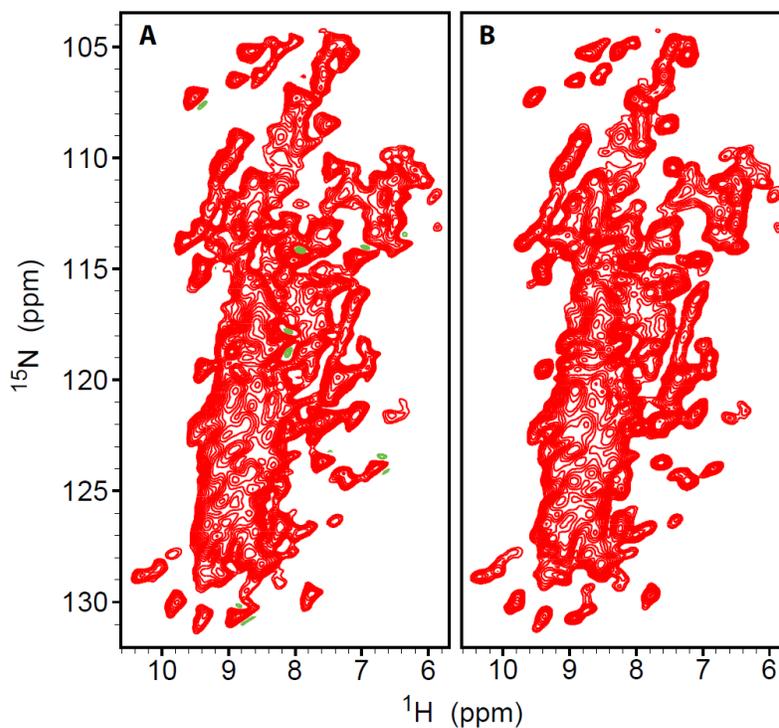


Figure 3-1: Effect of linear drift compensation on peak shapes. 2D hNH spectrum acquired without linear drift compensation (A). Several peaks are distorted and show “butterfly” peak shapes characteristic of a large drift (294 Hz). After applying linear drift correction to the recorded data (B), the peak distortion disappears and the expected round peak shape is restored.

In **Figure 3-2**, the drift during the 3D dataset is shown after transformation of the direct dimension. The drift over the duration of the acquisition is seen on the residual t_1 noise from water (top) and the improvement is seen after correction (bottom). In the 3D and 4D spectra, the effect of the drift in the processed spectra is most apparent in the reduced peak intensities (**Figure 3-3** and **Figure 3-4**), as well as in elongated peak shapes in the proton dimension (**Figure 3-4**). In the 3D spectrum drift correction resulted in a 20 percent improvement in peak height.

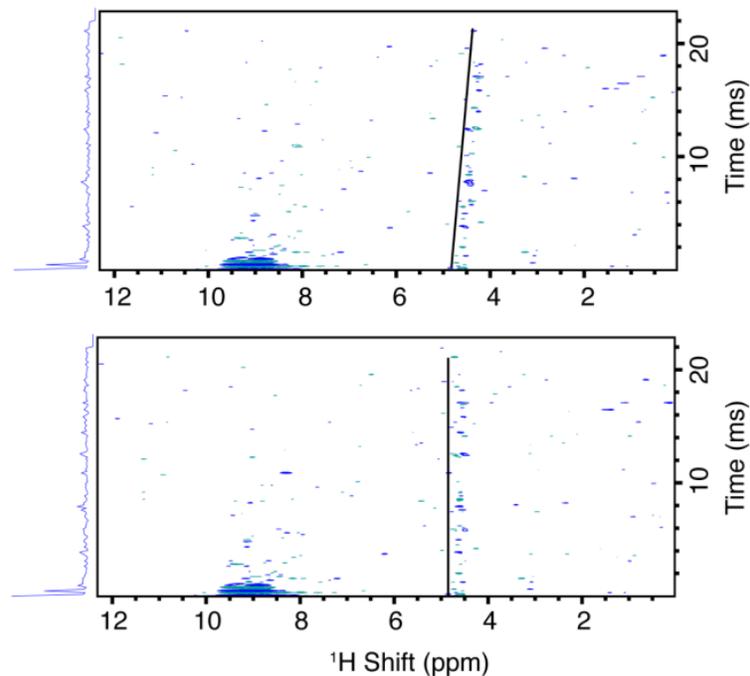


Figure 3-2. Drift during a 12 hour acquisition of the 3D spectrum hCONH. The spectrum was Fourier transformed in the direct dimension, and shows the drift over the total acquisition on the time axis. The acquisition followed a reset of the room temperature Z0 shim coil when the drift rate is unpredictable. The residual water shows that the drift is linear to a good approximation. The total drift during the acquisition was 304 Hz (0.38 ppm) as determined from proton 1D spectra acquired directly before and after the measurement.

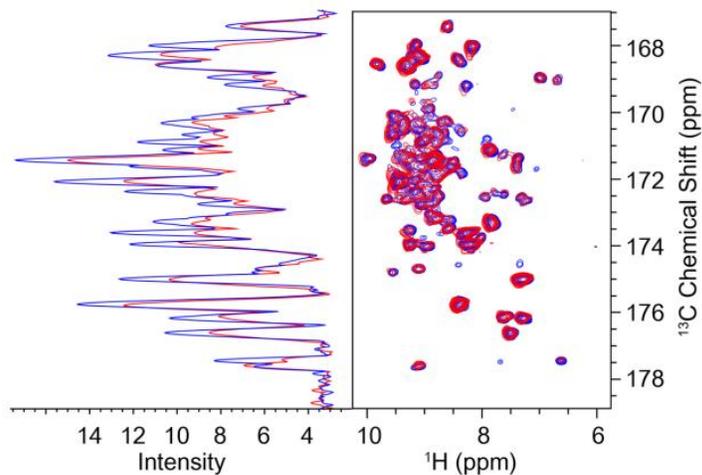


Figure 3-3. Comparison of the corrected (blue) and uncorrected (red) 3D hCONH spectrum of Figure 3-2. The CO-H projection of the full 3D dataset is shown, together with a double projection onto the ^{13}C axis to show the improvement in sensitivity after correction.

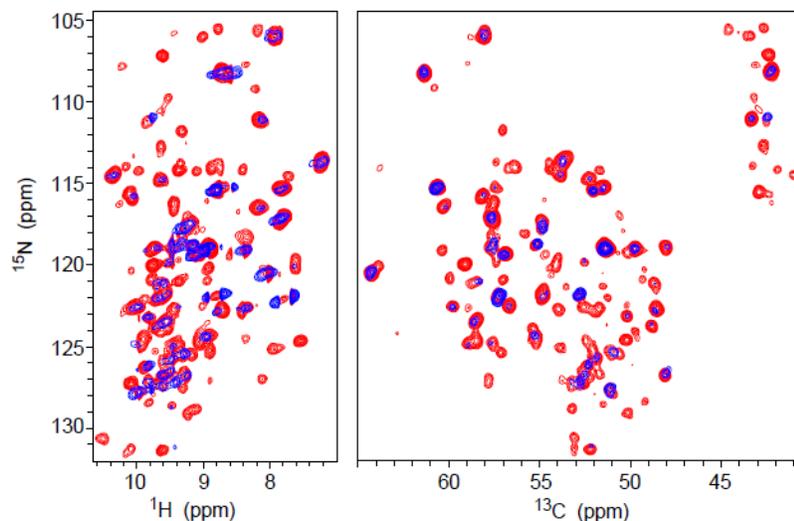


Figure 3-4: Overlay of ^1H - ^{15}N and the ^{13}C - ^{15}N projections of a 4D HhCANH spectrum recorded over 3 weeks with (red) and without (blue) correcting the field drift (maximum 350 Hz during the time of acquisition). The two spectra are contoured at the same intensity levels. The reduced intensity is clearly visible in the projections, where the drift of the magnetic field was not corrected. The HN projection also displays elongated peak shapes in the ^1H dimension, as proton resonances are narrow in ppm, thus the field drift also has a greater effect on the peak shapes of this dimension. 75 ms longitudinal ^1H - ^1H mixing was used.

3.3 Conclusions

We present a method to correct for drift occurring while recording MAS NMR spectra in the absence of a lock. We demonstrate that the detrimental effect of a significant field drift can be minimized if the data is recorded in short blocks and, after acquisition, a linear phase ramp calculated from the drift during each experiment is applied to each spectrum across all dimensions. We show that drift correction leads to a significant increase in peak intensity and improved peak shapes. The script provided in the Appendix is capable of correcting linear drift in all dimensions of a 2D, a 3D or a 4D spectrum recorded in planes, points or using non-uniform sampling. A second script is provided for the case that a dataset with minimum phase cycle exceeds the time over which linear compensation can be used. In this case, the dataset can be subdivided and later concatenated. Details are provided in the header of the script.

3.4 References

Agarwal, V., Penzel, S., Szekely, K., Cadalbert, R., Testori, E., Oss, A., Past, J., Samoson, A., Ernst, M., Bockmann, A., *et al.* (2014). 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.

Baker, E.B., and Burd, L.W. (1957). High Stability Nuclear Magnetic Resonance Spectrograph. *Rev. Sci. Instrum.* *28*, 313-321.

Barbet-Massin, E., Pell, A.J., Retel, J.S., Andreas, L.B., Jaudzems, K., Franks, W.T., Nieuwkoop, A.J., Hiller, M., Higman, V., Guerry, P., *et al.* (2014). Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J. Am. Chem. Soc.* *136*, 12489-12497.

Dolder, M., Zeth, K., Tittmann, P., Gross, H., Welte, W., and Wallimann, T. (1999). Crystallization of the human, mitochondrial voltage-dependent anion-selective channel in the presence of phospholipids. *J. Struct. Biol.* *127*, 64-71.

Eddy, M.T., Ong, T.C., Clark, L., Tejjido, O., van der Wel, P.C., Garces, R., Wagner, G., Rostovtseva, T.K., and Griffin, R.G. (2012). 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.

Fourier, J.B.J. (1822). *Théorie analytique de la chaleur* (Paris: Chez Firmin Didot, Père et Fils).

Fourier, J.B.J. (2009). *The Analytical Theory of Heat* (Cambridge: Cambridge University Press).

Fricke, P., Chevelkov, V., Zinke, M., Giller, K., Becker, S., and Lange, A. (2017). Backbone assignment of perdeuterated proteins by solid-state NMR using proton detection and ultrafast magic-angle spinning. *Nat. Protoc.* *12*, 764-782.

Hofer, D.C., Kahwaty, V.N., and Valentino, C.R. (1978). NMR field frequency lock system (IBM (International Business Machines Corporation)).

Lakomek, N.A., Frey, L., Bibow, S., Bockmann, A., Riek, R., and Meier, B.H. (2017). Proton-Detected NMR Spectroscopy of Nanodisc-Embedded Membrane Proteins: MAS Solid-State vs Solution-State Methods. *J. Phys. Chem. B* *121*, 7671-7680.

Lewandowski, J.R., Dumez, J.N., Akbey, U., Lange, S., Emsley, L., and Oschkinat, H. (2011). Enhanced Resolution and Coherence Lifetimes in the Solid-State NMR Spectroscopy of Perdeuterated Proteins under Ultrafast Magic-Angle Spinning. *J. Phys. Chem. Lett.* *2*, 2205-2211.

Maly, T., Bryant, J., Ruben, D., and Griffin, R.G. (2006). A field-sweep/field-lock system for superconducting magnets--Application to high-field EPR. *J. Magn. Reson.* *183*, 303-307.

Mo, H.P., and Raftery, D. (2008). Pre-SAT180, a simple and effective method for residual water suppression. *J. Magn. Reson.* *190*, 1-6.

Najbauer, E.E., Movellan, K.T., Schubeis, T., Schwarzer, T., Castiglione, K., Giller, K., Pintacuda, G., Becker, S., and Andreas, L.B. (2019). Probing Membrane Protein Insertion into Lipid Bilayers by Solid-State NMR. *ChemPhysChem* 20, 302-310.

Near, J., Edden, R., Evans, C.J., Paquin, R., Harris, A., and Jezzard, P. (2015). Frequency and phase drift correction of magnetic resonance spectroscopy data by spectral registration in the time domain. *Magn. Reson. Med.* 73, 44-50.

Neuhaus, D., Ismail, I.M., and Chung, C.W. (1996). "FLIPSY" - A new solvent-suppression sequence for nonexchanging solutes offering improved integral accuracy relative to 1D NOESY. *J Magn Reson Ser A* 118, 256-263.

Orekhov, V.Y., and Jaravine, V.A. (2011). Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Progress in nuclear magnetic resonance spectroscopy* 59, 271-292.

Paulson, E.K., and Zilm, K.W. (2005). External field-frequency lock probe for high resolution solid state NMR. *Rev. Sci. Instrum.* 76, 026104.

Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661-665.

Simpson, A.J., and Brown, S.A. (2005). Purge NMR: Effective and easy solvent suppression. *J. Magn. Reson.* 175, 340-346.

Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993). Gradient-Tailored Water Suppression for H-1-N-15 Hsqc Experiments Optimized to Retain Full Sensitivity. *J Magn Reson Ser A* 102, 241-245.

Zachariae, U., Schneider, R., Briones, R., Gattin, Z., Demers, J.P., Giller, K., Maier, E., Zweckstetter, M., Griesinger, C., Becker, S., *et al.* (2012). beta-Barrel mobility underlies closure of the voltage-dependent anion channel. *Structure* 20, 1540-1549.

Zhou, D.H., and Rienstra, C.M. (2008). High-performance solvent suppression for proton detected solid-state NMR. *J. Magn. Reson.* 192, 167-172.

4 Investigating the phase separation and dynamics of Nup98-derived FG-domains

This chapter is based on the following publication, and contains identical text passages: Eszter E. Najbauer, Sheung Chun Ng, Christian Griesinger, Dirk Görlich, Loren B. Andreas. Atomic resolution dynamics of cohesive interactions in phase-separated Nup98 FG domains, *Nat Commun*, 2022, *in press*

4.1 Introduction

One of the hallmarks of a living organism is its ability to encode and transmit the information required for its replication. In prokaryotes and eukaryotes, the genetic information is stored as DNA, and transcribed into RNA, from which proteins with an abundance of functions (structural, enzymatic, etc.) are synthesized. The realization of this process varies widely, however, between prokaryotes and eukaryotes. While in prokaryotes the steps of transcription and translation are coupled (meaning synthesis of protein from mRNA, and the transcription of DNA into mRNA happen in parallel), in eukaryotes these processes are separated in both time and space. One of the main features of a eukaryotic cell is its nucleus, comprising most importantly the cell's genome. The nucleus is enclosed by the nuclear envelope (Kite, 1913), a highly specialized and regulated double membrane continuous with the endoplasmic reticulum (Watson, 1955), separating nuclear transcription and cytoplasmic translation. This separation allowed a large increase in genome size, as a separated and mechanically supported nucleus in eukaryotes can stably contain DNA 1000 times the size as that of prokaryotes, leading to the evolution of more complex organisms. Separation of translation and transcription also lead to evolution of RNA splicing (Gilbert, 1978), which, combined with exon shuffling lead to the appearance of multidomain proteins. These, depending on the arrangement and combination of domains are capable of performing a wide

variety of functions. Lastly, confining chemical processes (e.g. transcription) to a smaller space results in higher concentrations of the reactants, leading to increased reaction rates.

4.1.1 Nuclear transport

Protein synthesis takes place in the cytoplasm, while messenger RNA is produced in the nucleus, making the existence of a transport system between the two spaces necessary. A bidirectional transport of various macromolecules is ensured by the nuclear transport machinery consisting of the nuclear pore complexes (NPC), nuclear transport receptors (NTRs) and the RanGTPase system. Mature RNA and ribonucleoprotein complexes are exported to the cytoplasm, while nuclear proteins are imported into the nucleus.

Nuclear pore complexes (NPC) are large protein assemblies embedded in the nuclear envelope, acting as a gate for the regulation of material flow, and also (indirectly) as an important point of regulation for gene expression (**Figure 4-1**). The size of NPCs in vertebrates can be as large as 124 MDa for the NPCs of *Xenopus laevis* oocytes (Reichelt et al., 1990). Vertebrate NPCs have an outer diameter of ~120 nm (Rout and Wentz, 1994), an inner diameter of about ~40 nm (Davis, 1995), and a depth of ~70 nm (Rout and Wentz, 1994).

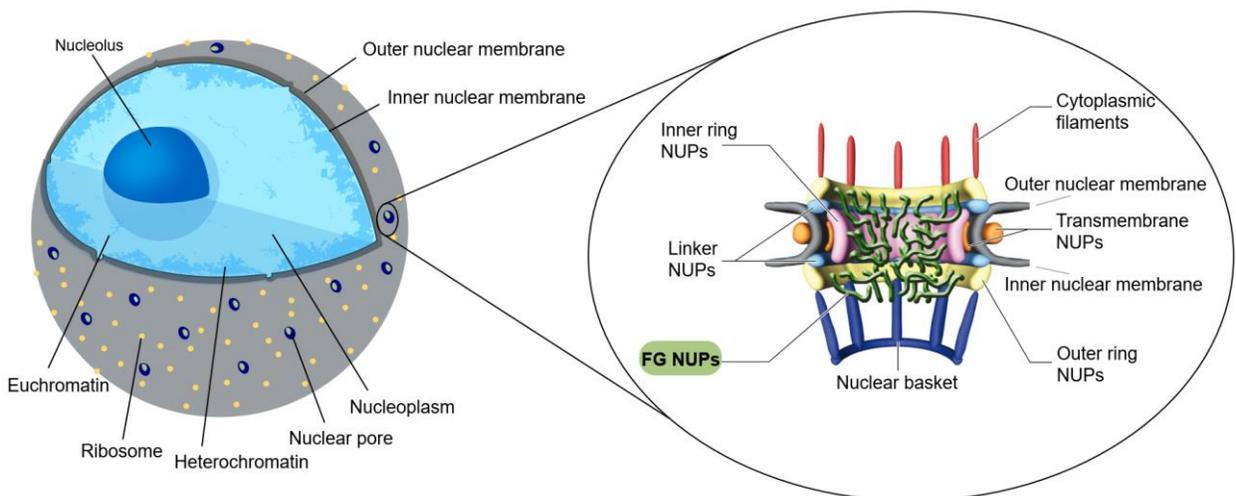


Figure 4-1: Schematic structure of the cell nucleus and the nuclear pore complex (NPC) The NPC (right) is embedded in the nuclear envelope formed by the inner and outer membranes, surrounding the cell nucleus (left). The transmembrane, linker, outer and inner ring nucleoporins (NUPs) form the scaffold of the NPC; FG-NUPs are anchored to the latter, reaching into and filling the central pore, forming a selective barrier. The NPC's symmetry to the plane of the membrane is broken by proteinaceous filaments extending into the cytoplasm (cytoplasmic filaments), and converging to form a basket-like structure (the nuclear basket) inside the nucleus. The figure showing the NPC (right) was adapted from Grossman et al. (Grossman et al., 2012).

The complex possesses an eightfold rotational symmetry perpendicular to the membrane, and a twofold rotational symmetry in the plane of the membrane (Wischnitzer, 1958; Hinshaw et al., 1992; Akey and Radermacher, 1993), which could be indicative of modular assembly of the complex. The ~30 different, evolutionarily conserved proteins (Rout et al., 2000; Cronshaw et al., 2002; DeGrasse et al., 2009) forming the NPC are called nucleoporins (NUPs), and are either embedded in the nuclear envelope, forming the scaffold of the NPC, or are attached to the scaffold, and some of these form a selective barrier. Altogether there are ~1000 individual nucleoporin molecules in an NPC, since due to the NPC's symmetry, nucleoporins are present in several, in human NPCs mostly 32 copies (Lin et al., 2016).

The nuclear pore complex conducts a large flux of molecules (up to 100 MDa/NPC/s, corresponding to ~1000 translocation events) (Ribbeck and Görlich, 2001), including the exchange of all proteins between the cytoplasm and the nucleus. Molecules smaller than ~30 kDa can diffuse through the pore unhindered, while molecules above this size experience increasingly restricted diffusion, requiring active transport (Bonner, 1975; Paine et al., 1975; Görlich and Kutay, 1999).

This large protein flux is made possible by **nuclear transport receptors** (NTRs), which, through forming a complex with the cargo, increase the speed of diffusion through the barrier up to 600-fold (Frey et al., 2018). Most NTRs are large proteins (90-150 kDa) with an overall negative charge (Freitas and Cunha, 2009), belonging to the importin- β superfamily of proteins (Kim et al., 2017). Structurally, they contain multiple HEAT (Huntingtin, elongation factor 3, protein phosphatase 2A, and the yeast kinase TOR1) repeats (Andrade and Bork, 1995), each of which consists of two antiparallel α -helices connected by a short loop. 18-20 HEAT repeats then stack into a highly flexible superhelical solenoid structure (Kim et al., 2017). Other nuclear carriers also exist, such as nuclear transport factor 2 (NTF2) belonging to the NTF2 family. This protein forms ~30 kDa homodimers (Wagner et al., 2015), and is responsible for binding and shuttling Ran through the selective barrier (see below) (Ribbeck et al., 1998). Depending on the direction of nuclear transport, NTRs can be grouped as importins or exportins. While importins transport molecules into the nucleus, exportins ensure transport into the cytoplasm. In a few cases, a protein may act both as an exportin and an importin (e.g . Exportin 4 (Lipowsky et al., 2000; Gontan et al., 2009), Importin 13 (Mingot et al., 2001)).

Directionality of nuclear transport is ensured by the RanGTPase system at the expense of chemical energy (Cautain et al., 2015). **Ran** (Ras-related nuclear protein) is a small, 25 kDa G protein capable of nucleotide binding, acting as a molecular switch, as it changes conformation depending on whether it binds GTP or GDP (Moore, 1998). In this way, it acts as a molecular switch, and its on/off state influences the binding of interacting NTRs. Importins and exportins respond differently to RanGTP binding: while for importins, RanGTP binding induces dissociation of the cargo complex (Görllich et al., 1996), exportins only bind cargo in the presence of RanGTP, forming a trimeric complex (Kutay et al., 1997). The dissociation of this complex happens with the transformation of RanGTP to RanGDP (Cautain et al., 2015).

Importins bind cargo in the cytosol, at low RanGTP concentrations, however once the importin-cargo complex diffuses across the membrane, the high RanGTP concentration in the nucleus causes dissociation of the complex. Exportins, on the other hand, bind cargo only as a trimeric (exportin·RanGTP·cargo) complex, which can then diffuse across the selective barrier. Once in the cytosol, Ran's intrinsic GTPase activity is enhanced through its interaction with RanGAP (Ran GTPase activating protein) (Hopper et al., 1990), which in turn occurs through complex formation with RanBP1 (Ran binding protein 1) (Bischoff et al., 1995; Cautain et al., 2015), causing the hydrolysis of RanGTP to RanGDP, and the dissociation of the trimeric complex. To avoid accumulation of RanGDP in the cytoplasm, the nuclear transport factor 2 (NTF2) shuttles it back to the nucleus. Here, RanGTP is regenerated by the action of RCC1, a nucleotide exchange factor also known as RanGEF (Ran guanine nucleotide exchange factor). Since RanGAP is only present in the cytoplasm (Hopper et al., 1990), while RCC1 is associated with the chromatin located in the nucleus (Makde et al., 2010), a steep RanGTP gradient arises through the nuclear envelope, responsible for establishing the chemical potential necessary for the directionality of nuclear transport. The RanGTPase system is shown on **Figure 4-2**.

Translocation of the cargo through the selectivity barrier is a multi-step process. The cargo is often recognized by a member of the the importin- β family of transport factors at the **nuclear localization signal** (NLS) (consisting of one or two short, 5-7 residue long, basic sequences in loop regions of the cargo protein containing lysine or arginine (Marfori et al., 2011; Cautain et al., 2015)) or at the **nuclear export signals** (NES) (short peptide sequences rich in leucines (Cautain et al., 2015)), however recognition may also take place by the cargo's three-dimensional fold or through

an adaptor protein, such as importin- α (Kau et al., 2004; Riddick and Macara, 2007; Lin and Hoelz, 2019). The NTR·cargo complex then docks to the NPC, translocates through the permeability barrier, then exits the permeability barrier. This is followed by the disassembly of the NTR·cargo complex. Finally, the NTR is returned to its starting compartment (Kim et al., 2017).

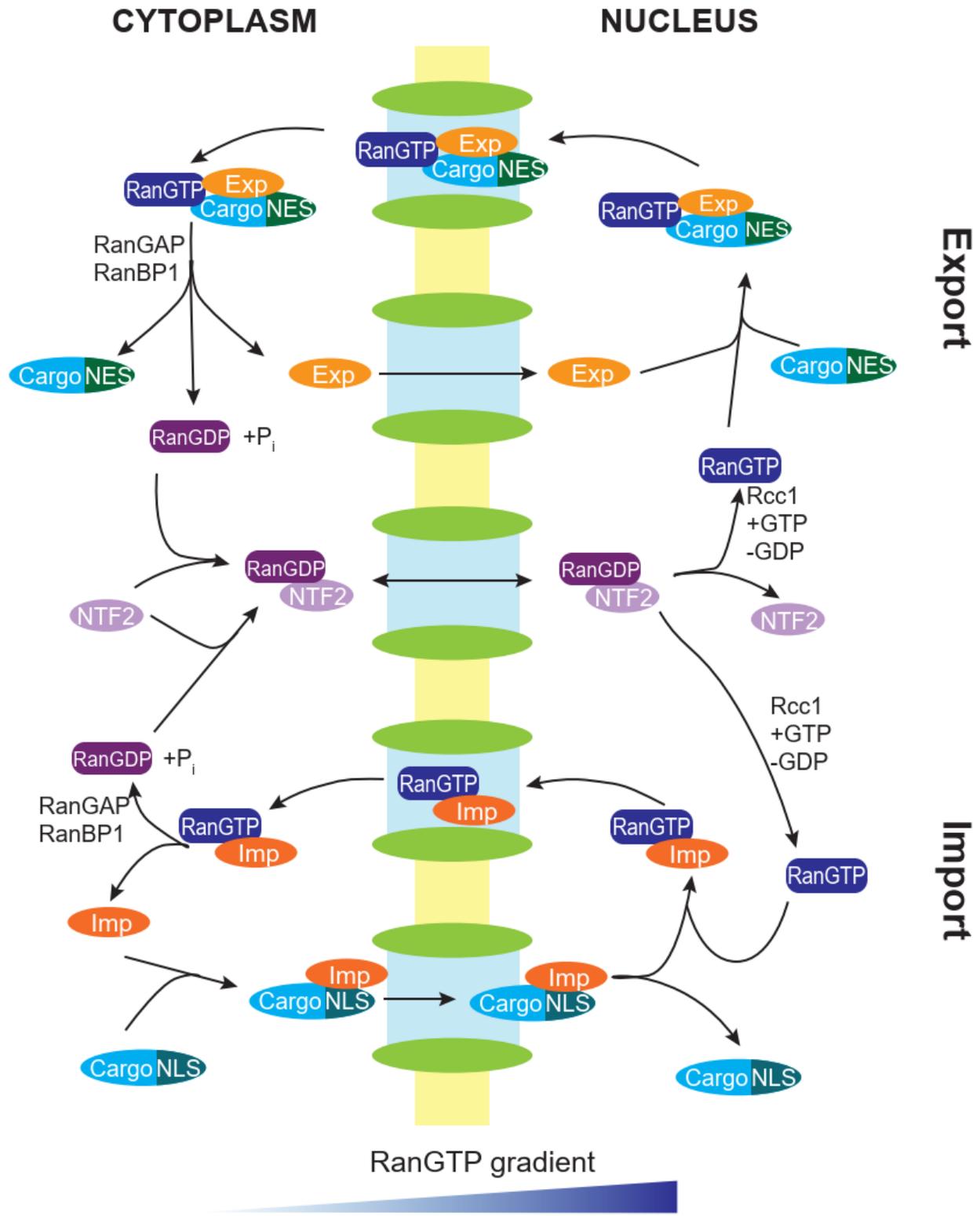


Figure 4-2: The RanGTPase system. Exp and Imp denote exportin and importin, respectively, NLS and NES stand for the nuclear localizations signal and the nuclear export signals in the cargo sequences.

4.1.2 FG-Nucleoporins

FG nucleoporins (FG-NUPs) form the selective barrier of the nuclear pore. They derive their name from the phenylalanines occurring frequently in their sequences, often part of a phenylalanine-glycine (FG) dipeptide motif. FG-NUPs are intrinsically disordered proteins also containing at least one folded domain (often with a coiled-coil structure), which anchors them to the NPC, and the FG motifs are located in the intrinsically disordered low sequence complexity regions called FG repeat domains or FG domains (Terry and Went, 2009; Chug et al., 2015).

Phenylalanines are often part of a variety of FG-motifs (FG, GLFG, FxFG, PxFG, Fx, FxFx, SLFG, SPFG, NxFG, FxxFG as well as GF) (Cushman et al., 2006; Denning and Rexach, 2007; Milles et al., 2015), occurring at a high, up to 260 mM concentration in the central channel of the NPC (Aramburu and Lemke, 2017). Bioinformatics studies have shown that as characteristic for most IDPs, FG-domains do not show a high level of sequence conservation per se, and have high evolution rates (Denning and Rexach, 2007). Their propensity for disorder, however, has been conserved, as well as the hydrophilic nature and length of the spacers between the hydrophobic FG motifs (~10-20 amino acids) (Denning and Rexach, 2007).

Various approaches have demonstrated that FG-domains are the main constituents of the NPC's selective barrier. Upon cleavage (Flather and Semler, 2015), genetic deletion (Strawn et al., 2004) or biochemical depletion (Finlay and Forbes, 1990; Hulsmann et al., 2012) of individual FG-domains, the barrier deteriorates or may even collapse. On the other hand, the selective behavior of the NPC can be recapitulated in hydrogels of purified FG-domains formed *in vitro*. This hydrogel mirrors the selectivity of the NPC by excluding inert species, but allowing the entry of cargo-NTR complexes (Frey and Görlich, 2007). About 12 different proteins make up the FG-Nups found in NPCs, among which the FG-domain of Nup98 was found to be among the most critical for barrier formation (Hulsmann et al., 2012; Lin and Hoelz, 2019). In contrast to the majority of intrinsically disordered proteins (IDPs), which are abundant in charged and polar residues allowing them to maintain an unfolded state under physiological solvent conditions (Uversky et al., 2000), Nup98 FG-domains are 400-700 residues long, and have a low net charge combined with a high mean hydrophobicity, comparable to levels found in folded proteins (Schmidt and Görlich, 2015). They are poorly soluble in water, and phase separate from dilute (sub- μ M) solutions to form ~250 mg/ml hydrogels, which is conserved across a variety of species: mammals, insects, lancelets, plants,

fungi, nematodes, amoebas, and ciliates (Schmidt and Görlich, 2015). Of 10 *Xenopus* nucleoporins, the FG-domain of Nup98 formed the hydrogel with the strictest sieving effect, allowing passage of the small NTR protein NTF2, excluding inert proteins >27 kDa in size, and retaining importin β -type NTRs at the surface of the hydrogel “particles” (rim staining) (Labokha et al., 2013).

The selective barrier formed by FG-domains is a challenging system to study *in situ*. It is below the resolution limit of confocal microscopy, and although several high-resolution structures of the NPC have been obtained by cryogenic electron tomography (Bui et al., 2013; Lin et al., 2016; Kim et al., 2018), where folded proteins could be fitted into the images at secondary structure resolution, the dynamic FG-domains at the center of the pore appear averaged out, and thus are invisible to the approach. This makes NMR, often used in the study of IDPs, an ideal technique to learn more about this system and its dynamics at an atomic scale.

The physicochemical properties of the selective barrier, and the apparent paradox that transport by NTRs can be fast and selective at the same time has been the focus of intense research in recent years. Ader et al. characterized the hydrogel obtained from full-length Nsp1 FG/FxFG repeat domain using solid-state NMR spectroscopy, and observed both mobile and fibril-like structures, with Phe residues occurring in both. The C terminus of the protein rich in Asp, Glu, and Lys-rich charged spacers appeared mobile, while the more rigid N-terminal part containing NTSQ-rich spacers was shown to form interchain β -sheets. NOE contacts were observed between the aromatic Phe ring and aliphatic methyl groups, suggesting the presence of transient hydrophobic interactions within the hydrogel. Based on these lines of evidence, a possible model for selective transport was suggested, where the fibril structures provide stability for a firm barrier, while dynamic phenylalanines bind NTRs (Ader et al., 2010). Also using ssNMR, Labokha et al. found that rigid regions of the Nup98 gel are not necessarily fibril structures, but still associated cohesiveness with rigidity (observed through a high intensity cross-polarization signal) (Labokha et al., 2013).

Using a combined approach of stopped-flow kinetic measurements, molecular dynamics simulations and NMR spectroscopy on the PxFG domain of Nup153 to formulate a model of a more dynamic system, Milles et al. showed that FG-domains populate a number of rapidly fluctuating conformations that bind NTRs at near-diffusion limited on-rates in a polyvalent complex (Milles et al., 2015).

FG-domains show a reduced affinity towards NTR proteins in a cellular milieu (Tetenbaum-Novatt et al., 2012), which is most probably due to the sequestering of the FG groups from each other and the competition of NTRs with other weak binding partners found in the cellular environment (Tetenbaum-Novatt et al., 2012; Wall and Hough, 2018). A recent in-cell NMR study on the FSFG-K fragment of the FG-NUP Nsp1 revealed a difference in interaction patterns within the protein sequence even between the cellular environments of *E. coli* bacteria and *S. cerevisiae* (yeast) (Wall and Hough, 2018). Regardless of the organism, however, the sequence was found to be disordered and to undergo fast (microsecond-millisecond) timescale dynamics (Wall and Hough, 2018). Elevated R_2 and R_2/R_1 ^{15}N relaxation rates were measured at or near the FSFG repeats, suggesting interaction between these motifs. A comparison of relaxation measurements in *E. coli* bacteria and *S. cerevisiae* revealed a different interaction pattern in the cellular environments, and indicated that even the hydrophilic spacers participate in weak interactions.

A protein family found in insects, closely related to FG-domains in sequence, shows several similar properties, despite its differing functions. Resilins are highly elastic proteins that provide mechanical flexibility to insect wings (Donoughe et al., 2011), store energy upon folding and unfolding of the wings (Haas et al., 2000), and are also present in the jumping mechanism of fleas (Bennet-Clark and Lucey, 1967), cicadas (Gorb, 2004), and other insects. Resilin gains its elastomeric properties not only from a high percentage of glycine and proline residues in the sequences (Li and Kiick, 2013) but also from the tyrosines (appearing as YG motifs), which (opposed to phenylalanines in the FG motifs of FG-domains) can cross link with each other into a three dimensional structure containing di- and tri-tyrosine links between the YG motifs (Tjin et al., 2014). Resilins, similarly to FG-domains, are present as random coils (Tjin et al., 2014), while showing a slight propensity towards β -sheet conformations upon mechanical stress (stretching) (Reddin, 2008).

4.1.3 Models for selectivity

An adequate model for the selective barrier of the NPC should be able to explain several aspects of its behavior. The most obvious one is how the barrier lets small cargoes pass unhindered, also facilitating transport of large NTR-bound cargoes, while inert molecules >30 kDa are excluded. Since these cargoes are still more than tenfold smaller in diameter than the inner diameter of the

NPC, it is also crucial to selective behavior that the barrier seals perfectly, excluding the possibility of the formation of a solvent channel allowing the passage of inert cargoes.

Several models emphasizing various aspects of nuclear transport have been proposed to explain the selective behavior of the nuclear pore (Schmidt and Görlich, 2016).

In the **virtual gate model** (Rout et al., 2003) (**Figure 4-3A**), the main consideration is the entropy decrease of macromolecules upon confinement into a narrow space (the channel of the NPC). FG-domains are regarded as entropic brushes (Hoh, 1998), serving to further limit the space and thus the number of conformations available to the macromolecule traversing the nuclear envelope. Without any enthalpic contributions from binding of the macromolecule to the NPC, the decrease in entropy upon entering the channel will always result in a positive ΔG and an activation barrier. However, if the macromolecule (or the cargo·NTR complex) interacts with a large number of low-affinity FG motifs, the favorable change in enthalpy can compensate for the decrease in entropy. In order to speed up the passage and exit of the cargo molecules, it is much more favorable to have many weak interactions with high off rates, rather than a few strong ones, which would pose a kinetic hindrance for nuclear transport. The experimental evidence however, that replacing the cohesive Nup98 FG repeats by non-cohesive ones makes the barrier more uniformly permeable (Hulsmann et al., 2012) contradicts this model, where the elimination of FG domains should lead to the exclusion of all large cargoes.

The **reduction of dimensionality model** (Peters, 2005) (**Figure 4-3B**) relies on the observation that for most molecules, the NPC acts as a pore of ~10 nm in diameter (Paine et al., 1975; Keminer et al., 1999), however if the cargoes can bind to FG domains, molecules up to 39 nm in size can be transported through the pore (Pante and Kann, 2002). The assumption in this model is that the FG-domains provide a hydrophobic lining to the pore, while hydrophilic spacers extend into the pore lumen, leaving only a 10 nm channel free, through which small molecules can diffuse. Upon the transport of cargo·NTR complexes, the NTRs bind to the FG motifs. The firm attachment of NTRs to the dense array of FG motifs on the channel surface grants them the ability to perform a 2D random walk on the pore's wall, in the meantime pushing the loose hydrophobic spacers out of the way. At the exit point, once the cargo·NTR complexes are exposed to high concentrations of Ran-GTP (nuclear side) or Ran-GAP and RanBP (cytoplasmic side), they disassemble, and are released

from the FG binding (Peters, 2009). In this case, however it is unclear how NTRs dissociate from the strong interaction with the FG domains.

The **forest model** (Yamada et al., 2010) (**Figure 4-3C**) assumes the presence of both cohesive and non-cohesive FG domains in the NPC. The first adopts globular, collapsed coil conformations (shrubs), while the latter more dynamic, extended coil conformations (trees). In the model these domains do not intermix, and form two transport zones: one in the middle of the pore, and one through the extended chains between the collapsed coils. Depending on their surface charge, hydrophobicity, and size, nuclear transport takes place in one of these zones (Yang, 2011), a possible explanation for the existence of multiple NTR translocation pathways, as observed in yeast (Strawn et al., 2004). The model is, however, incompatible with the finding that hydrogel particles self-assembled only from the cohesive Nup98 domains show permselectivity properties similar to those of intact NPCs (Schmidt and Görlich, 2015).

The **selective phase model** (Ribbeck and Görlich, 2001; Frey and Görlich, 2007) (**Figure 4-3D**) is based on cohesive interactions, possibly hydrophobic interactions between the FG motifs, which cross link the FG nucleoporin protein chains into a mesh, whose size determines the diameter of molecules that can pass through freely. The cross links between the phenylalanine rich regions are responsible for preventing the plug from leaking, while the hydrophilic spacers between the FG motifs act against the collapse of the structure. When a molecule larger than the mesh size is transported, the NTR binding the cargo is capable of interacting with the FG motifs, and locally disrupting the cross links, “melting” its way through the gel. Once the NTR has passed, the mesh seals itself on a timescale of μs or faster (Moussavi-Baygi and Mofrad, 2016). This model is supported by the observation that when cohesive FG motifs are replaced by non-cohesive ones, the NPC loses its selectivity, even if the non-cohesive motifs can bind NTRs as well (Hulsmann et al., 2012).

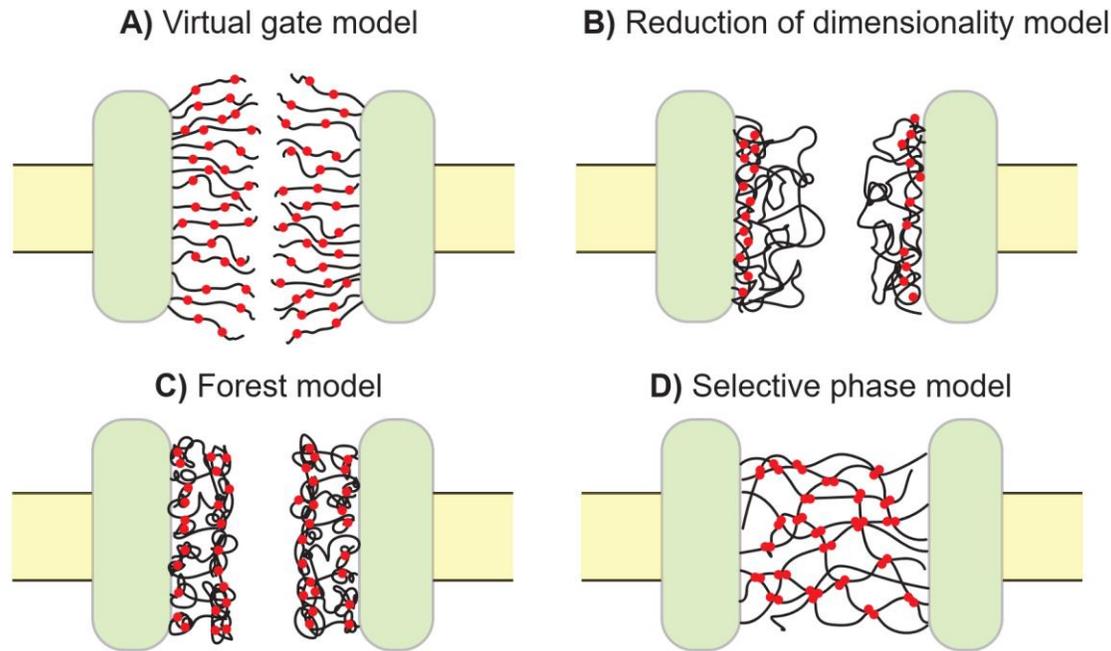


Figure 4-3: Models for selective transport through the nuclear pore complex (NPC). In the cross section of the NPC, the lipid bilayer is depicted in yellow, the wall of the NPC in green. The disordered protein chains are shown as black lines, the FG motifs as red dots. The models shown are the A) Virtual gate model, B) Reduction of dimensionality model, C) Forest model, D) Selective phase model.

Some models for selective transport are based solely on an interaction between NTRs and FG-domains. The **reversible collapse model** was based on the observation that gold grafted Nup153 undergoes collapse when incubated with concentrations as low as femtomolar of Importin β (Lim et al., 2007). The model is relatively similar to the virtual gate model relying on entropic bristles in repelling non-specific cargos, differing in the assumption that upon NTR binding, the FG domains collapse into compact conformations, thus drawing the cargo·NTR complex through the NPC (Lim et al., 2007). However, since NPCs are located in a cellular environment where the equilibrium concentration of NTRs is in the $\sim\mu\text{M}$ range (Hahn and Schlenstedt, 2011), nucleoporins would be found at all time in their collapsed states, which would eliminate the effect of the cargo·NTR complex being drawn into the NPC.

The **Kap-centric model** assumes the presence of two populations of NTRs: one tightly bound to the FG-domains, an integral part of the NPC (termed “slow phase”), and another one weakly bound to the remaining binding sites (“fast phase”). It is the latter one that ensures fast transport through the NPC (Kapinos et al., 2014). In this model, FG-domains would alone not be sufficient to produce

a selective hydrogel. The observation that the <30 kDa mCherry is efficiently excluded from FG-domain-derived hydrogels, even in the absence of NTRs contradicts this model (Labokha et al., 2013).

Further models for selective transport include the oily spaghetti model (Macara, 2001), the Brownian affinity gating model (Rout et al., 2000), or the di-block copolymer brush gate model (Ando et al., 2014; Ando and Gopinathan, 2017), differing in mostly only minutiae from the models described above. It has also been suggested that the mechanism of transport is a hybrid of the above models (Strawn et al., 2004; Patel et al., 2007; Wentz and Rout, 2010). As of now, no consensus for the mechanism of selective transport in NPCs has yet been reached, as almost all these models have some extent of contradictory experimental evidence.

4.1.4 Phase separation in biological systems

A well-known phenomenon in polymer chemistry, the diverse and important roles of phase separation in biological systems were discovered only relatively recently. Although liquid-liquid separation of hemoglobin at high concentrations *in vitro* (Galkin et al., 2002), as well as liquid droplet formation of proteins in crystallization trials had been previously observed (Dumetz et al., 2008), it was roughly only in the last decade that phase separation was described as the organizing principle behind membraneless organelles (e.g. stress granules (Brangwynne et al., 2009), the nucleolus (Frottin et al., 2019), and synaptic vesicles (Chen et al., 2020)), as well as a mode of regulation in cellular processes (Li et al., 2012). Since many biological stress factors (changes in temperature, ion concentration, osmotic pressure, pH) can influence phase separation behavior, this phenomenon could help explain changes in compartmentalization and formation of molecular condensates upon cellular stress (Martin and Mittag, 2018). Protein aggregation via phase separation has been implicated in age-related diseases (Elbaum-Garfinkle, 2019), such as amyotrophic lateral sclerosis (ALS) (Pakravan et al., 2020), Alzheimer's disease (Ambadipudi et al., 2017; Kanaan et al., 2020), and Parkinson's disease (Ray et al., 2020), emphasizing the importance of understanding and harnessing this phenomenon.

Phase separation is a physical transition, where above a certain (critical) concentration, molecular species of the initially homogeneously mixed solution eventually occupy defined spatial

compartments at specific concentrations (Boeynaems et al., 2018). The resulting newly formed phase may have properties ranging from liquid-, through gel- to solid-like, depending on organization at the molecular level. Gels are held together by a network of covalent (chemical gel) or non-covalent interactions (physical gel) spanning the whole of the droplet. In biological systems hydrogels consisting of water-soluble polymer molecules are the most common.

The driving forces behind phase separation can range from hydrogen bonding, dipole-dipole, electrostatic, cation- π , π - π , to hydrophobic interactions. These interactions (and the physical nature of the resulting phase) are encoded in the protein sequence, mostly in low-complexity, intrinsically disordered regions. These regions do not have a defined structure, but undergo continuous rapid chemical exchange, and compared to a statistical distribution are often rich in just a few types of amino acids, which are organized into repetitive short interaction motifs or charge blocks. Characteristic polar amino acids in low-complexity domains include Asn, Gln, Ser, Pro, as well as Gly, where the residue's characteristics are dominated by its polar backbone, while charged residues are usually Arg, Lys, Asp, or Glu. Residues responsible for aromatic interactions may be Phe or Tyr. Phase separation behavior can be influenced both by the composition as well as the pattern of the protein sequence, although the influence of the latter has not yet been studied in detail (Martin and Mittag, 2018).

Derivation of the Flory-Huggins equation is discussed in detail in the review by (Martin and Mittag, 2018), but is summarized briefly as follows. The Gibbs free energy of phase separation is given by the well-known equation:

$$\Delta G_{\text{ph.sep.}} = \Delta H_{\text{ph.sep.}} - T\Delta S_{\text{ph.sep.}} \quad (4 - 1)$$

where T is the temperature, and $\Delta H_{\text{ph.sep.}}$ and $\Delta S_{\text{ph.sep.}}$ are the enthalpy and entropy changes associated with phase separation, respectively. The enthalpy term contains pairwise self and cross interaction potentials between protein and solvent molecules, and in the simplest case can be approximated by the mean of these interactions:

$$\Delta H_{\text{ph.sep.}} = -\phi_s\phi_p\chi \frac{kTV}{v_m}, \text{ where } \chi = \frac{z}{kT} \left[\frac{1}{2}(\varepsilon_{ss} + \varepsilon_{pp}) - \varepsilon_{ps} \right] \quad (4 - 2)$$

ϕ_s and ϕ_p are the volume fractions of the solvent and the polymer, respectively, V is the total volume, k is the Boltzmann constant, ϵ_{ss} , ϵ_{pp} , and ϵ_{ps} are the interaction energies between solvent, protein and protein and solvent molecules. v_m is the average volume of a monomer, often approximated as the square root of the product of the volume of one monomer of the polymer and one molecule of solvent. z is the total of potential interactions in the lattice.

In the simplest case the entropy term can be described by the entropic cost of confining the polymer and the solvent into a limited space. Using this approximation, the decrease of entropy upon phase separation can be given as follows:

$$\Delta S_{\text{ph.sep.}} = -kV \left(\frac{\phi_s}{v_s} \ln \phi_s + \frac{\phi_p}{v_p} \ln \phi_p \right) \quad (4-3)$$

where all notations are as previously defined, v_s and v_p correct for the volume of a single solvent or protein molecule. Usually v_s is defined as 1, while v_p is proportional to the sequence length. Increasing the polymer's length increases $\Delta S_{\text{ph.sep.}}$, making phase separation more favorable.

By combining the enthalpy (Equation 4-2) and entropy (Equation 4-3) terms, we obtain the Flory-Huggins equation:

$$\Delta G_{\text{ph.sep.}} = -\phi_s \phi_p \chi \frac{kTV}{v_m} - TkV \left(\frac{\phi_s}{v_s} \ln \phi_s + \frac{\phi_p}{v_p} \ln \phi_p \right) \quad (4-4)$$

Spontaneous phase separation occurs within the boundary given by:

$$\frac{\partial^2 \Delta G_{\text{ph.sep.}}}{\partial \phi_p^2} = 0 \quad (4-5)$$

The Flory-Huggins theory (Flory, 1942; Huggins, 1942) describes the behavior of polar polymers relatively accurately, where mixing of the dense and the light phase is entropically favorable. For these systems increasing the temperature makes mixing more favorable, and these polymers have an upper critical solution temperature (UCST behavior, **Figure 4-4A**), above which only one phase can exist, regardless of the polymer concentration. Flory-Huggins theory has been used to understand the behavior of the germ granule protein Ddx4 (Nott et al., 2015), as well as the *C. elegans* protein LAF-1, forming P granule-like structures *in vitro* (Wei et al., 2017).

Although useful in describing UCST systems, the Flory-Huggins theory comes short when applied to polymers rich in hydrophobic groups, due to its inaccurate description of the entropy term. The assumptions, from which entropy change is calculated in the Flory-Huggins theory do not take into account the number of conformations available to the molecules. For example, if the molecule forms a globular ensemble in solution (e.g. due to abundant hydrophobic groups), the increased number of available (extended) conformations in the dense phase also has to be taken into account, resulting in a positive $\Delta S_{\text{ph.sep.}}$. This is of particular importance for FG-domains, where the highly repetitive hydrophobic FG motifs linked by flexible spacers have many more energetically equivalent states in the dense phase, which lowers critical concentration by an order of magnitude in comparison to a fully hydrophilic sequence (Boeynaems et al., 2018). Entropically driven solvent release also leads to an increase in entropy upon phase separation, as water excluded from the protein becomes more entropic (greater rotational, translational entropy) in the dilute phase (hydrophobic effect). For these systems decreasing the temperature leads to mixing, and these polymers have a lower critical solution temperature (LCST) (**Figure 4-4B**).

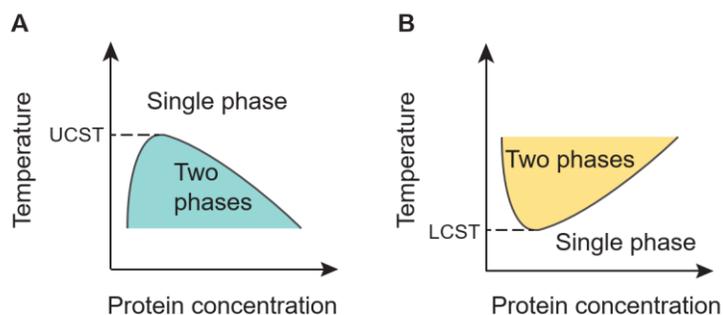


Figure 4-4: Schematic phase diagrams of UCST (A) and LCST (B) behavior. In the colored areas, the system exists as two phases. Above (A) or below (B) the critical point, the system can only exist as a single phase. The critical point is characterized by the temperature (upper critical solution temperature – UCST, lower critical solution temperature – LCST) (Holehouse and Pappu, 2015).

Although surface plasmon resonance has indicated a behavior of FG-domains consistent with the LCST phase separation shown by polymers (Schoch, 2017), the temperature dependence of their phase separation behavior has not yet been explicitly determined. For resilin-mimetic proteins displaying many sequence similarities with FG-domains, an LCST behavior has been observed, in some cases also combined with a UCST-type phase transition (Balu et al., 2014). While the LCST transition was suggested to be driven by the interaction of exposed hydrophobic groups, UCST

behavior was attributed to intermolecular assembly, and driven by hydrogen bonding between neighboring segments/molecules. The LCST behavior of a synthetic block-polymer sequence inspired by resilin was found to be tunable to UCST by replacing the hydrophobic residues in the sequence with pairs of zwitterionic residues, arginine being the preferred cationic residue (Quiroz and Chilkoti, 2015). Rich in valine, proline, and glycine, and hydrophobic in nature, elastin-like polypeptides (ELPs) also show LCST behavior, and are model systems often used to study the LCST transition (Li et al., 2014; Zhao et al., 2016).

Mutation or post-translational modification of certain residues (such as phosphorylation of serine/threonine/tyrosine or arginine methylation/citrullination) in the interaction motifs can alter phase separation behavior (Owen and Shewmaker, 2019), and this can be a mechanism in the cell to control the concentration and intracellular distribution of the protein component. For example, phosphorylation of the tau protein precedes its aggregation, and thus might be pivotal in the formation of the characteristic fibrils observed in Alzheimer's disease (Noble et al., 2013). *In vitro*, mutations can be excellent tools to uncover the driving forces upon phase separation of a protein. Understanding and harnessing phase separation for therapy in medicine is an exciting area of research currently undergoing rapid development (Wheeler, 2020).

4.1.5 Aims and outline

In this chapter we investigate the atomistic properties of *Tetrahymena thermophila* MacNup98A-derived hydrogels. In Sections 4.3.1 and 4.3.2 we compare hydrogels of wild-type MacNup98A with two perfectly repetitive sequences focusing on permselectivity and spectral properties. We then assign the perfectly repetitive sequence recapitulating the properties of wild-type MacNup98A (Section 4.3.3), and characterize the dynamics of the protein chain in the hydrogel phase by TRACT measurements (Section 4.3.4).

In Section 4.3.5 we investigate the influence of hydrophobic leucine on the phase separation and on the dynamics of the protein chain. We use FRAP to measure the diffusion of various mutants containing or lacking leucine within hydrogel particles, as well as recording NOESY spectra to identify spatial proximities of residues within the hydrogel.

Next, we study the impact of salt concentration and temperature on phase separation and on properties of the hydrogel. We not only characterize dynamics, but also measure critical concentrations under various conditions in order to construct a phase diagram (Sections 4.3.6 and 4.3.7), which allows us to obtain quantitative information on the thermodynamics of phase separation of Nup98-like FG-domains (Section 4.3.8).

4.2 Materials and methods

4.2.1 Recombinant protein expression and purification

TtMacNup98A (Mac98A) FG domain: Procedures were as published previously (Ng et al., 2021) with modifications here for $^{15}\text{N}/^{13}\text{C}$ isotope labelling: *E. coli* NEB Express cells transformed with a plasmid, encoding the protein fused to a N-terminal histidine tag (Table 5.4.2), were allowed to grow in a modified M9 medium (supplemented with trace elements) containing the isotopes (Table 5.4.3) at 37°C until OD₆₀₀ reached ≈0.8. Expression of the target protein was induced by 0.4 mM IPTG, 30°C for 16 h. Then the protein was extracted, purified by Ni-chromatography and lyophilized as previously described. Lyophilized powder was weighed for quantification and stored at -20°C until use.

prf.GLFG_{52x12}: Since a histidine tag may introduce unwanted signals in high-resolution NMR measurements, we designed a construct with a SUMO protease cleavage site (Frey and Görlich, 2014a, b) such that the histidine tag was cleaved after Ni chromatography (as described below), namely His₁₄-ZZ-*sc*SUMO-*prf.GLFG_{52x12}* (plasmid number: 'pSNG064', Table 5.4.2). The construct was expressed in *E. coli* cells in the M9 medium as described above. The cells were resuspended in cold 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 300 mM (NH₄)₂SO₄, 1 mg/ml lysozyme, and lysed by a freeze-thaw cycle followed by a mild sonication. Inclusion bodies containing the construct were recovered by centrifugation (k-factor: 798; 45 min; 25°C) and washed. The construct was extracted with 4 M guanidine hydrochloride, 50 mM Tris/HCl pH 7.5, 10 mM DTT. The extract was cleared by ultracentrifugation (k-factor: 135; 90 min) and applied overnight at room temperature to a Ni(II) chelate column. The column was washed in extraction buffer containing 20 mM imidazole, then extensively in 0 M guanidine hydrochloride, 50 mM Tris/HCl pH 7.5, 10 mM DTT and the proteins were eluted with 500 mM imidazole, 50 mM Tris/HCl pH 7.5, 10 mM DTT. The eluate was digested with 30 nM *bd*SEN1 protease (Frey and Görlich, 2014a) for 18 hours, at 21 °C. The removal of the His₁₄-ZZ-*sc*SUMO tag from *prf.GLFG_{52x12}* decreased its solubility and induced its phase separation. Thus, the untagged *prf.GLFG_{52x12}*, in form of insoluble material, was recovered by centrifugation (k-factor: 798, 25 °C, 1 h) and then washed twice in 300 mM ammonium acetate pH 7.5. The protein was solubilized with 30% acetonitrile. Residual uncut construct and His₁₄-ZZ-*sc*SUMO tag in the extract were

cleared by passing through a Ni(II) chelate column. The flow-through was lyophilized. The lyophilized powder was weighed for quantification and stored at -20°C until use. Note that the product is completely free of cysteine residues, and thus free of disulfide cross-links, which could change local mobility of the peptide chain in the proximity of the cross-links.

For FRAP experiments, the FG domain was expressed with an additional C-terminal Cysteine (plasmid number: pSNG102; Table 5.4.2) for labelling with Atto488 maleimide. Production of the domain was as above, except that the bacteria were cultured in Terrific Broth (TB). The lyophilized, purified protein was dissolved in 2 M guanidine hydrochloride and the protein was allowed to react with Atto488 maleimide (ATTO-TEC, Germany) at 1:1 molar ratio at pH 6.8. The labelled protein was further purified by gel filtration on a PD10 Sephadex column and quantified by the absorbance of Atto488.

Pro-free prf.GLFG_{52x12}: Procedures were as described (Ng et al., 2021), except that M9 medium containing isotopes was used for cell culture.

GLFG_{52x12}, GAFG_{52x12}, GXFG//L_{52x12}, GXFG//V_{52x12} and GLLG//L_{52x12}: Procedures were the same as that for the Mac98A FG domain (for constructs containing a His₁₈-tag, Table 5.4.2)/ *prf.GLFG_{52x12}* (for constructs containing a His₁₄-ZZ-*sc*SUMO tag), except that TB medium was used for cell cultures.

prf.GLFG_{7x12} fragment (7x perfect repeats), GLLG_{52x12} and GLAG_{52x12}: His₁₄-ZZ-*sc*SUMO constructs similar to the above were constructed and expressed as described (the *prf.GLFG_{7x12}* fragment was expressed in the M9 medium containing ¹⁵N isotope; GLLG_{52x12} and GLAG_{52x12} were expressed in TB medium). These constructs did not form inclusion bodies in the *E.coli* host and thus the soluble cell lysates cleared by ultracentrifugation were applied directly to Ni(II) chelate columns. The columns were washed extensively in 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 20 mM DTT and then in protease buffer: 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 5 mM DTT. 50 nM *bd*SEN1 in protease buffer was applied for overnight on-column cleavage (Frey and Görlich, 2014a, b). The cleaved target proteins were soluble and eluted with the protease buffer.

Protein probes for FG particles permeation assays: NTF2 and mCherry were expressed as His-tagged-fusions (Table 5.4.2) and purified by native Ni(II) chelate chromatography, as described previously (Schmidt and Görlich, 2015; Frey et al., 2018).

4.2.2 FG particles preparation for permeation assays and confocal laser scanning microscopy

A previously described procedure (Schmidt and Görlich, 2015) was applied with minor modifications: 2 μl of a 1 mM FG domain stock solution (lyophilized powder dissolved in 2 M guanidinium-HCl) was rapidly diluted with 100 μl assay buffer (20 mM NaPi pH 6.8, 150 mM NaCl, 5 mM DTT) and 15 μl of the mixture was mixed with 15 μl permeation probes (2 μM Alexa488-labeled rat NTF2 (rNTF2) or 10 μM mCherry in assay buffer). The resulting mixture ([FG protein] = 10 μM) was placed on collagen-coated μ -slides 18-well (IBIDI, Germany). FG particles were allowed to sediment under gravity for 30 min (and for reaching an equilibrium state). Then Alexa488 or mCherry signals were acquired with 488 or 561 nm excitation respectively with a Leica SP5 confocal scanning microscope equipped with a 63x oil immersion objective and hybrid detectors (standard mode, in which non-linear response of the detector was auto-corrected). The method for staining FG particles with Thioflavin-T (ThT) was similar, except that 15 μl of 4 μM ThT was mixed with 15 μl of the FG particles suspension. ThT was excited at 405 nm and detected in a 460-500 nm window. ThT signals were measured with the same settings and the image processing was the same for all figures.

4.2.3 Analysis of phase-separation by centrifugation (centrifugation assay)

FG-domain stock solutions (lyophilized powder dissolved in 2 M guanidinium-HCl, except for GLLG_{52x12} and GLAG_{52x12}, which were buffered in the protease buffer described above) were rapidly diluted with assay buffer (typically 20 mM sodium phosphate (NaPi) buffer pH 6.8, 5 mM DTT, 150 mM NaCl or [NaCl] specified in the figures), to the final [FG protein] stated in the figures. After incubation of 1 min, the FG phase (insoluble content) was separated from the soluble content by ultracentrifugation (14680 rpm, 30 mins, by Eppendorf 5424 Centrifuge) at the temperature specified in the figures. The corresponding cut-off sedimentation coefficient $\approx 240\text{S}$ according to the following formula for fixed-angle rotor suggested by Livshits et al., 2015:

$$S = \frac{10^5}{4\pi^2} \cdot \frac{1}{(RPM/10000)^2} \cdot \frac{1}{t} \cdot \frac{L}{R_{av}} \quad (4 - 6)$$

where S is the sediment coefficient in Svedberg; t is the centrifugation time in hour; L is the sedimentation path length (in cm) of the rotor and R_{av} is the average radius of rotation.

A fraction of pellet (FG phase) and supernatant (soluble content), each 3 or 6% of the total was loaded for SDS-PAGE, followed by Coomassie blue staining. Critical concentration was defined by the concentration of the supernatant, which was estimated with standard loading controls. Some experiments were repeated three times to derive the standard deviations (S.D.), which were shown as error bars, and the representative gel images were shown. In case if the critical concentration is above 20 μM , the experiment was repeated by dissolving the lyophilized powder of FG domain directly in assay buffer to a concentration of 200 μM , and the loaded amount on the gel was adjusted accordingly.

4.2.4 Analysis of phase-separation by Dynamic Light Scattering (DLS)

Lyophilized powder of prf.GLFG_{52x12} was dissolved and diluted in NaCl solutions containing 20 mM NaPi pH 6.8. Final protein concentration and NaCl concentration were as stated in the main figure. The solutions were analyzed in closed cuvettes using a DynaPro NanoStar DLS instrument (Wyatt Technologies). To acquire temperature-dependent phase separation curves, the temperature was automatically raised by 1°C per min, typically from 2 to 40°C (or as indicated in **Figure 4-21B**, also cooling from 40 to 2°C). DLS signals were acquired continuously. The Dynamics 7.1.5 software was used for autocorrelation analysis and computation of radii. Three datasets for each condition were averaged, and standard deviations (S.D.) were shown as error bars in **Figure 4-23**.

4.2.5 Fluorescence recovery after photobleaching (FRAP) and estimation of diffusion coefficients

2 μl of unlabeled FG domain stock (1 mM prf.GLFG_{52x12} in 2 M guanidinium-HCl) was rapidly diluted with 100 μl pre-cooled assay buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT) on ice. At this point phase separation was suppressed by the ice temperature. Then 30 μl of the FG solution was quickly mixed with 1 μl of each of the Atto488-labelled FG domains, which have

been adjusted to typically a concentration of 15 μM (in 2 M guanidinium-HCl), except for GAFG_{52x12}: 90 μM . Therefore, final protein concentration of unlabeled FG protein was 20 μM ; Atto488-labeled species was 0.5 μM (Atto488-labeled GAFG_{52x12}: 3 μM). The mixtures were incubated at room temperature to allow phase-separation and then transferred to micro-slides for confocal laser scanning microscopy. FG particles (n=6 typically) of diameters \sim 8-10 μm were photo-bleached near the centers at 488 nm. Bleached area: circular with a diameter of 3 μm . After bleaching, Atto488 signals were acquired at 2 s intervals. Raw data were corrected for photo-bleaching during acquisition and normalized to the initial signal. The recovery curves were fitted to a simple exponential model (Taylor et al., 2019) as follows:

$$C(t) = A(1 - \exp - (t/\tau)) \quad (4 - 7)$$

where $C(t)$ is the normalized signal at time t . A is a preexponential scaling factor close to 1, which was used to account for incomplete recovery (due to a small immobile fraction of protein in the sample). τ is the characteristic diffusion time, during which $1/e$ of the fluorescence recovers. The diffusion coefficient is calculated from the characteristic diffusion time using: $D = R^2 / \tau$, where R is the radius of the bleached spot.

4.2.6 NMR measurements

4.2.6.1 Sample packing

Formation of hydrogels has been observed both by dilution of concentrated guanidinium-HCl stocks (Schmidt and Görlich, 2015; Konishi and Yoshimura, 2020) and by hydration of lyophilized protein powder (Milles and Lemke, 2011). For solid-state NMR measurements we used the latter method: roughly 1.1 mg of lyophilized powder was packed into 1.3 mm rotors. Using excess buffer, samples were hydrated in a tabletop centrifuge at 16700 rpm for 1 h at 25°C. Buffers used for NMR spectroscopy contained 20 mM NaPi, 0.01% NaN₃, 5 mM EDTA and 150 mM NaCl. For tracking the effect of salt concentration, a series of buffers containing 75 mM, 200 mM, 300 mM, or 600 mM NaCl were applied. The pH of all buffers was set to 6.80 with 6 M NaOH or 3 M HCl.

4.2.6.2 General spectral parameters for solid-state measurements

All spectra, except when noted otherwise were recorded in a 1.3 mm MAS probe at 20 kHz MAS spinning frequency in a 600 MHz wide bore Bruker Avance III HD spectrometer. Unless indicated otherwise, the temperature was set to 310 K at 500 l/h VT gas flow. Typical pulse lengths were 2.5 μ s on ^1H , 4.2 μ s on ^{13}C , and 5.0 μ s on ^{15}N . In the all experiments, where applicable, water suppression was achieved by a 100 ms MISSISSIPPI pulse train (Zhou and Rienstra, 2008). Decoupling during acquisition periods was set to 5 kHz TPPM (Bennett et al., 1995) on protons, and to 5 kHz WALTZ (Shaka et al., 1983a; Shaka et al., 1983b) on ^{15}N , and for double-labeled samples, also on ^{13}C . Recycle delays were set to 1.0 s. Unless noted otherwise, all spectra were processed using a cosine squared window function using the Bruker Topspin 4.0.8. software package.

4.2.6.3 ^{13}C spectra with direct excitation and ^1H - ^{13}C cross-polarization

^{13}C -detected experiments were recorded on u- $[^{13}\text{C}$ - $^{15}\text{N}]$ labeled samples, averaging 1024 scans. Cross-polarization (CP) spectra on TtMacNup98A and prf.GLFG $_{52\times 12}$ were recorded at an optimal CP condition of $\omega_{^1\text{H}} = \omega_{^{13}\text{C}}$, (average fields of 30 kHz on both nuclei), using an 80-100% ramp on ^{13}C with an optimal length of 13 ms. On the Pro-free-prf.GLFG $_{52\times 12}$, the optimal CP condition was found to be 50 kHz on ^{13}C , using a 80-100% ramp, and 68 kHz on ^1H (average fields), length was 4 ms. Relaxation delays were optimized to allow for full signal recovery between scans. Spectra were processed with an exponential window function, applying a line broadening corresponding to half the linewidth (60 Hz, 60 Hz, and 80 Hz, in **Figure 4-5A**, **Figure 4-6** left and right, respectively).

4.2.6.4 ^1H - ^{15}N and ^1H - ^{13}C correlation spectra

INEPT-based hNH (using refocused INEPT blocks to allow for MISSISSIPPI water suppression) and TROSY (Pervushin et al., 1997) spectra were recorded at 20 kHz MAS spinning frequency at 304.1 K in a 600 MHz wide bore Bruker Avance III HD spectrometer. ^1H - ^{15}N INEPT delays were set to 1.8 ms. Each spectrum averaged 4 scans per point. Spectra were processed with 100 ms acquisition time in the indirect ^{15}N , and 400 ms in the direct ^1H dimension.

INEPT-based hCH spectra were recorded with a pulse sequence analogous to the INEPT-based hNH experiment. The hCH experiment in **Figure 4-7C** was recorded on a u-[¹⁵N]-labeled sample to gain resolution. The ¹³C dimension was sampled to 40 ms.

The spectrum used for overlay with a solution spectrum (**Figure 4-13**) was measured on a u-[¹³C, ¹⁵N]-labeled sample to make it comparable to the solution spectrum that had to be recorded on an isotopically labeled sample for sensitivity reasons. The cooling gas was set to a temperature of 296 K at 500 l/h gas flow, which was found to be 304.4 K at the sample after calibration. The spectrum was sampled to 50 ms in the ¹³C dimension, but only 10 ms were used for processing.

For the HhCH NOESY experiment the NOE mixing block was added to the INEPT-based hCH block analogously as described in Chapter 2 (Najbauer et al., 2019). The spectrum was recorded on an 800 MHz Bruker Avance III HD spectrometer in a narrow-bore 1.3 mm probe. All experimental conditions were identical as described above. The NOE mixing time was set to 100 ms. Assignment of the side chains was accomplished with a similar HhCH pulse sequence on the 600 MHz Avance III HD spectrometer, the ¹H-¹H mixing correlating nuclei within the same spin system was achieved with a WALTZ-16 (Shaka et al., 1983b) scheme with 12.5 μs pulse lengths applied for 9.6 ms.

4.2.6.5 Assignment experiments

Assignment experiments were recorded on a Bruker Avance III HD 850 MHz spectrometer using a wide bore H/X/Y/D 3.2 mm probe. MAS spinning frequency was set to 12.5 kHz, temperature of the cooling gas was set to 310 K. The sample was u-[¹³C, ¹⁵N]-labeled. INEPT delays were set to 2.7 ms for the H/N transfer, 10 ms for the N/C, 4.5 ms for the CO-CA transfer, and 11.2 ms for the CB/CA and CA/N simultaneous transfer. The HNCA spectrum (Cavanagh, 2007) was acquired in 20 hours, and was sampled to 18.6 ms in ¹⁵N, 24.8 ms in ¹³C, and 27.0 ms in the ¹H dimension. The HNcoCA experiment (Cavanagh, 2007) was acquired using the same sampling as the HNCA experiment. For the HNCACB experiment (Cavanagh, 2007), acquired in 32 hours, sampling was identical to the previous experiments in the ¹H and ¹⁵N dimensions, and 6.0 ms in the ¹³C dimension. Although these acquisition times in the direct and indirect dimensions are not sufficient to take full advantage of the narrow linewidths offered by the sample, they provide sufficient resolution to accomplish sequential assignment of the repeating units.

4.2.6.6 Determining rotational correlation times with TRACT

TRACT experiments (Lee et al., 2006) were recorded using buffers with varying NaCl concentrations (75 mM, 200 mM, 300 mM, or 600 mM NaCl). Residue-specific correlation times were measured for the peptide chain using pseudo-3D experiments (Lee et al., 2006), where a series of 2D ^1H - ^{15}N planes were recorded with varying increments for both the α and the β states. The ^1H dimension was sampled to 100 ms, the ^{15}N dimension to 40 ms (sweep widths were 41.7 ppm on ^1H and 25.0 ppm on ^{15}N). Each spectrum averaged 32 scans per point. For both states, 7 points were sampled in the pseudo dimension such that signal in the last increment was ~30% of the original. Each peak in a pseudo 3D spectrum was fitted with an exponential decay using the Bruker Dynamics Center. Errors of the relaxation times were estimated from the goodness of the fit. τ_c values were calculated as described by Lee et al. (Lee et al., 2006) using Wolfram Alpha (Wolfram|Alpha, 2018). Errors of τ_c were determined using the Monte Carlo method (10000 runs).

4.2.6.7 Temperature calibration

Temperature calibration in solid-state probes was achieved using a solution consisting of 80% ethylene glycol and 20% DMSO- d_6 filled into a 1.3 mm rotor, which was then spun at 20 kHz and cooled/heated with the variable temperature gas flow. The exact temperature was determined from the chemical shift difference of the ethylene and the hydroxyl protons using an NMR thermometer application (NMR Thermometer, 2020).

4.2.6.8 Liquid state experiments

Solution samples of the prf.GLFG_{52x12} were prepared by dissolving the lyophilized sample (u- ^{15}N]-labeled for [^{15}N , ^1H]-TRACT experiments, u- ^{13}C , ^{15}N]-labeled for ^1H - ^{15}N and ^1H - ^{13}C HSQC spectra) in ice-cold buffer (20 mM NaPi, 150 mM NaCl, pH 6.8) such that its concentration was 15 μM . This solution was then centrifuged at 25°C (for the [^{15}N - ^1H]-TRACT measurement) or 35°C (for the HSQC measurements) in a tabletop centrifuge at 16700 rpm for 1 h. The supernatant (containing ~1-2 μM prf.GLFG_{52x12}) was then quickly transferred to a 3 mm NMR tube. This procedure ensured the absence of phase separation in solution samples, since the critical

concentration for the prf.GLFG_{52x12} at 35°C is lower than at any of the lower temperatures applied during the measurements (**Figure 4-23**). Since D₂O is necessary for locking, however it could influence phase separation behavior of FG-domains, the 3 mm NMR tube was placed in a 5 mm tube containing pure D₂O. Due to the concentric placement of the tubes, the samples could still be shimmed adequately.

All spectra were recorded on a Bruker Avance III HD spectrometer equipped with a 5 mm CPPTCI Z-GRD cryoprobe. HSQC spectra were sampled to 1 s on ¹H (for processing, only 400 ms were used) and 50 ms on the heteronucleus. The ¹H-¹⁵N HSQC spectra were recorded with the fhsqcf3gpqh pulse sequence from the Bruker library, using the standard parameter set, the ¹H-¹³C HSQC spectra were measured with the hsqcetgpsi standard Bruker pulse sequence, also using the standard parameter set.

[¹⁵N, ¹H]-TRACT experiments were sampled to 20 ms on ¹⁵N (sweep width was 36 ppm), and 200 ms on ¹H (sweep width was 13.0 ppm) averaging 8 scans per point. 16 increments were measured for both states such that the signal in the last increment was ~30% of that in the first increment. τ_c values were obtained as described above in Chapter 4.2.6.6.

The concentration of the prf.GLFG_{7x12} sample was 100 μ M. All spectra ([¹⁵N, ¹H]-TRACT, ¹H-¹⁵N HSQC) were recorded as described above on a Bruker Avance Neo spectrometer equipped with a 5 mm CPTCI Z-GRD cryoprobe.

All spectra were processed using a cosine squared window function using the Bruker Topspin 4.0.8. software package.

4.3 Results

The results shown here were obtained in collaboration with Dr. Sheung-Chun Ng and Prof. Dr. Dirk Görlich. All NMR spectroscopy studies were performed by the author of this thesis, design of perfectly repetitive sequences, confocal microscopy experiments, dynamic light scattering (DLS) measurements, centrifugation assays, and fluorescence recovery after photobleaching (FRAP) experiments were carried out by Dr. Sheung-Chun Ng. The design and selectivity study of the perfectly repetitive sequences is described in detail in a manuscript by Ng. et al. with the title “Recapitulation of selective nuclear transport with a perfectly repeated 12mer GLFG peptide” (Ng et al., 2021), which was still under preparation at the time this thesis was written. For this reason, as necessary for coherence, some findings from that manuscript are also summarized briefly in the Results section below.

4.3.1 Selectivity and spectroscopy of the *Tetrahymena thermophila* MacNup98A (TtMacNup98A) hydrogel

Following recombinant expression and purification of TtMacNup98A, guanidinium hydrochloride stocks prepared from the protein was diluted to a concentration above the critical concentration for phase separation as described in the Methods section, and the resulting phase observed with a confocal microscope. As previously observed (Schmidt and Görlich, 2015), upon dilution of the guanidinium stock, the protein immediately formed round gel-like particles of about 4-8 μm in diameter (**Figure 4-5B**). Upon incubation of the droplets with Thioflavin T (ThT), no fluorescent signal was detected upon excitation at 405 nm, indicating the absence of amyloid fibrils (**Figure 4-5C**). The selectivity of the hydrogel TtMacNup98A droplets against Alexa488-NTF2, a 29 kDa nuclear transport receptor, and mCherry, a 28 kDa inert, fluorescent protein was tested using fluorescence microscopy. We found that NTF2 had entered the droplets and after an incubation of 30 min, we found that the partition coefficient at equilibrium was ~ 2000 , while mCherry remained excluded (partition coefficient ~ 0.05) (**Figure 4-5D**). These observations perfectly reproduce the permselectivity observed previously by Schmidt and Görlich (Schmidt and Görlich, 2015).

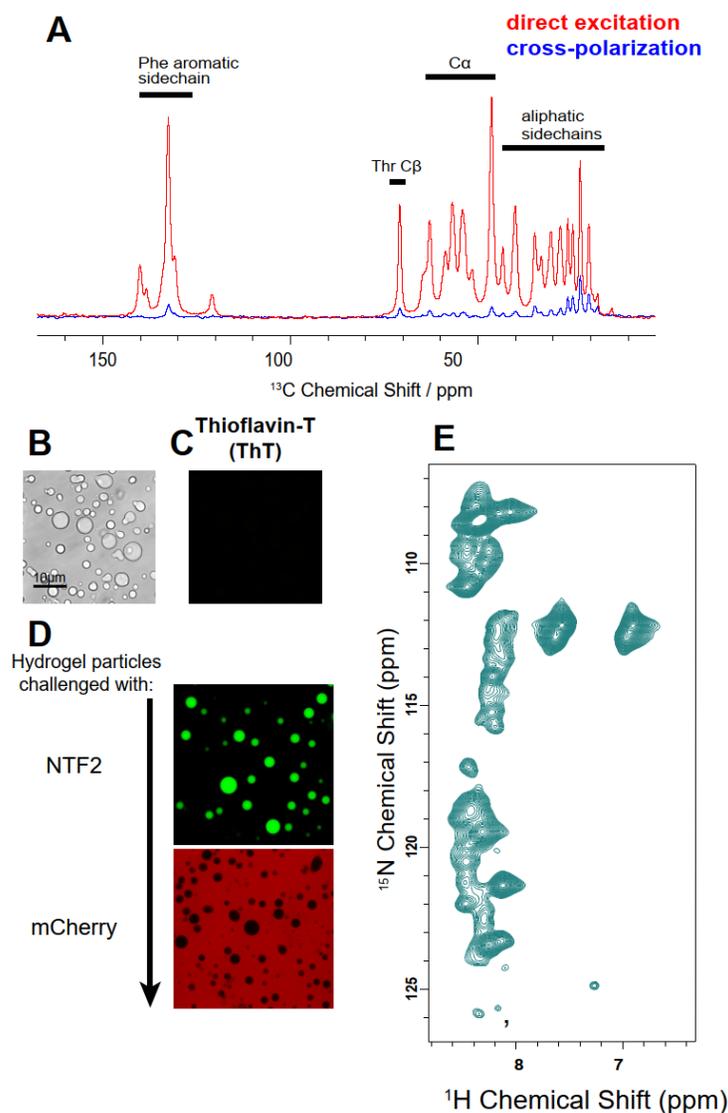


Figure 4-5: Droplet formation and NMR spectroscopy of the TtMacNup98A hydrogel. A) ^{13}C spectra of the TtMacNup98A hydrogel. The direct excitation spectrum is shown in red, the cross-polarization (CP) spectrum in blue, both recorded in 1024 scans. B) Phase-contrast microscopy image of TtMacNup98A droplets, phase separated into spherical particles following dilution from 2M GuHCl (Materials and Methods). C) Thioflavin T staining of TtMacNup98A droplets. D) Selectivity of the hydrogel particles when challenged with NTF2 (labeled with Alexa Fluor 488, green) and mCherry protein (red). Images of the equilibrium state recorded with a confocal laser scanning microscope. E) ^1H - ^{15}N INEPT-based correlation spectrum of the TtMacNup98A hydrogel recorded at 20 kHz MAS.

To assess the spectral quality of the phase-separated droplets, the hydrogel was formed inside a 1.3 mm rotor, and, as described in the Methods section, low-frequency magic angle spinning was used to eliminate any line broadening stemming from inhomogeneity in magnetic susceptibility. A ^{13}C direct excitation spectrum was recorded using low-power proton decoupling (**Figure 4-5A**). The observed lines were narrow under these conditions (~ 70 -80 Hz), pointing to a large degree of

inherent mobility in the gel, which averages out most of the chemical shift anisotropy and heteronuclear ^1H - ^{13}C dipolar couplings within the gel. Although the characteristic regions of the individual functional groups (CO, aromatic groups, Thr C β , C α , Gly C α , alkyl sidechains) are well resolved and can be identified, significant spectral overlap makes sequence specific assignment impossible (e.g. all Thr C β groups appear as one resonance). This makes multidimensional experiments necessary for residue-specific characterization of the gel, as can be expected from a protein sequence of close to 600 residues.

Potentially rigid regions in the hydrogel were probed through ^1H - ^{13}C cross polarization (CP) (**Figure 4-5A**). This experiment revealed an optimal CP condition of $\omega_{^1\text{H}} = \omega_{^{13}\text{C}}$, as well as very low efficiency of cross polarization for both TtMacNup98A, as low as <0.1 for C α resonances (defining CP efficiency as the maximal signal achievable by ^1H - ^{13}C CP vs. ^{13}C direct excitation in an equal number of scans, the maximal theoretical value being $\gamma_{^1\text{H}}/\gamma_{^{13}\text{C}} \approx 4$). The low level of CP signal further confirms the absence of rigid protein regions. Since the CP condition matches that of a static sample, it is clear that the interaction through which the magnetization transfer takes place is not dependent on spinning frequency. Since it has been previously shown that CP can take place even in solution, through scalar couplings (Eykn et al., 2003), the low CP signal can be explained by magnetization transfer through weak scalar couplings, independent of MAS spinning frequency, rather than the much stronger, MAS-dependent dipolar couplings.

To better resolve individual peaks, a 2D ^1H - ^{15}N correlation spectrum of TtMacNup98A was recorded. Due to the low efficiency of cross-polarization, an INEPT-based sequence was used, with a full transfer of in-phase ^1H magnetization to in-phase ^{15}N magnetization to enable MISSISSIPPI water suppression (**Figure 4-5E**). Again, lines were found to be fairly narrow (~ 0.08 ppm on ^1H and ~ 0.3 ppm on ^{15}N on the narrowest peaks), even under slow spinning and low-power proton decoupling. The chemical shift dispersion in the ^1H dimension is very low, 0.4 ppm, including sidechain amide resonances, which is characteristic of intrinsically disordered proteins. Despite the narrow resonances, due to the intrinsically disordered nature and the highly repetitive, degenerate sequence of the protein, there is significant overlap among the peaks, which makes a residue-specific assignment of TtMacNup98A intractable in the hydrogel form.

4.3.2 Simplification of spectra – design of perfectly repetitive sequences

To enable residue-specific characterization of the protein, perfectly repetitive sequences were designed, and their selectivity was tested. Both the design of the sequences and the selectivity of the droplets are described in detail by Ng et al{Ng, 2021 #598}, but summarized here briefly. After identifying the average distance between the FG motifs to be 8 residues, a “GLFG_{52x12}” sequence was designed with spacers of uniform lengths between the GLFG motifs, their overall composition reflecting the prevalence of each amino acid in full-length TtMacNup98A. In a further step, the sequence of the spacers was also made uniform. Two different sequences were designed, with spacers of different composition, but both very close to the wild-type amino acid frequency. In the perfectly repetitive GLFG_{52x12} (prf.GLFG_{52x12}) the proline content was slightly higher than in TtMacNup98A, so another perfectly repetitive, proline-free variant (Pro-free-prf.GLFG_{52x12}) was designed to investigate the influence of prolines on the protein’s behavior. These sequences are shown on **Figure 4-6**.

Both the perfectly repetitive GLFG_{52x12} (prf.GLFG_{52x12}) and the proline-free GLFG_{52x12} (Pro-free-prf.GLFG_{52x12}) were recombinantly expressed and purified. The prf.GLFG_{52x12} formed droplets very similar to those of TtMacNup98A, while phase separation of the Pro-free-prf.GLFG_{52x12} resulted mostly in irregular clumps. The selectivity of the prf.GLFG_{52x12} against Alexa488-NTF2, and mCherry agreed perfectly with that of TtMacNup98A. Although the aggregates of Pro-free-prf.GLFG_{52x12}, were somewhat selective, they did not fully recapitulate the selectivity observed for the droplets of TtMacNup98A and the prf.GLFG_{52x12}. When incubated with ThT, the droplets of the prf.GLFG_{52x12} were not stained, however the irregular clumps formed by the Pro-free-prf.GLFG_{52x12} showed a strong fluorescent signal, indicating the presence of a fibril-like structure (**Figure 4-6**).

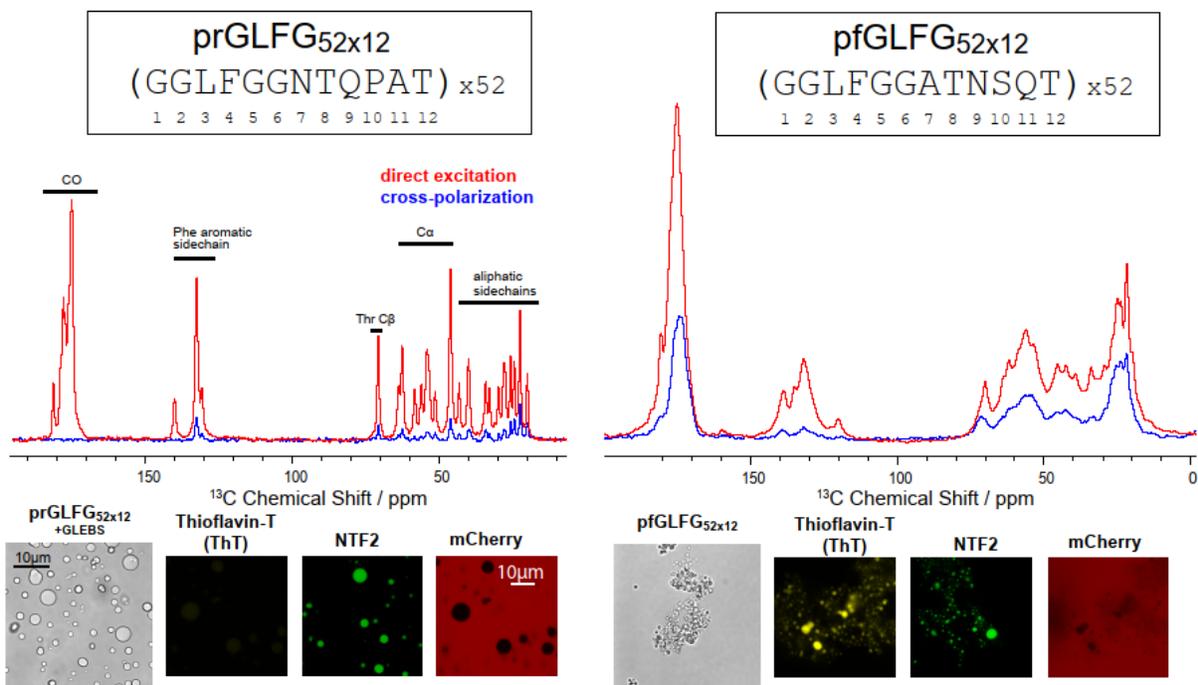


Figure 4-6: Sequence, selectivity and ^{13}C spectra of particles prepared from perfectly repetitive sequences mimicking the wild-type TtMacNup98A composition. Direct excitation spectra of the prf.GLFG $_{52 \times 12}$ and the proline-free prf.GLFG $_{52 \times 12}$ (Pro-free-prf.GLFG $_{52 \times 12}$) are shown in red, cross-polarization spectra in blue. All spectra were recorded by averaging 1024 FIDs, and processed with exponential apodization using half of the linewidth as line broadening. Brightfield microscopy images of the droplets, thioflavin T (ThT) staining, and selectivity of the sequences against NTF2 (labeled with Alexa Fluor 488, green) and mCherry (red) obtained with a confocal laser scanning microscope are shown on the bottom. For the prf.GLFG $_{52 \times 12}$, microscopy images are of the sequence containing the GLEBS domain.

The lack of fibril formation in the prf.GLFG $_{52 \times 12}$, and fibrilization in the Pro-free-prf.GLFG $_{52 \times 12}$ is in line with the observation that proline tends to break secondary structural elements. Since the residue lacks an amide proton, it can only form one of the two hydrogen bonds necessary for the formation of helices or sheets, and tends to appear in irregular structures, such as turns (Levitt, 1978). Although there are other amino acids in the sequence that favor turns (e.g. Gly, Asn – previously the possibility was raised that the particles were ThT negative due to the high glycine content (Schmidt and Görlich, 2015)), Ng et al (Ng and Görlich, 2021) (unpublished) showed that the absence of proline is enough to change the sequence’s structural preference.

The resonances in the ^{13}C direct excitation spectrum of the prf.GLFG $_{52 \times 12}$ were similarly narrow to those in the spectrum of TtMacNup98A (~70-80 Hz), pointing to the same inherent mobility. The

spectrum of the Pro-free-prf.GLFG_{52x12} showed much broader lines of >200 Hz however, an indication of a less dynamic system.

In the ¹H-¹³C cross polarization experiment (**Figure 4-6**), similarly to TtMacNup98A, the optimal CP condition for the prf.GLFG_{52x12} was found to be $\omega_{1H} = \omega_{13C}$, and a similarly low efficiency of cross polarization was observed (< 4% for C α -s). The Pro-free-prf.GLFG_{52x12} however shows much higher CP efficiencies of ~40% in the C α -region. The optimal condition for cross polarization in this case matches the Hartmann-Hahn condition of $\omega_{13C} - \omega_{1H} = \omega_{MAS}$. Although the theoretical efficiency of ¹H-¹³C CP is $\gamma_{1H}/\gamma_{13C} \approx 4$, this is rarely observed experimentally. In practice, completely rigid samples typically show a CP efficiency of 2-2.5. The value of 0.4 obtained for the Pro-free-prf.GLFG_{52x12} is much lower than this, however still 10-fold higher than the efficiencies of TtMacNup98A and the prf.GLFG_{52x12}. This shows that while the Pro-free-prf.GLFG_{52x12} maintains some inherent mobility, it is significantly more rigid than both the TtMacNup98A and the prf.GLFG_{52x12} proteins.

4.3.3 High resolution ¹H-¹⁵N correlation spectra and backbone assignment of the prf.GLFG_{52x12}

The resonances in the 2D INEPT-based ¹H-¹⁵N correlation spectrum were again narrow, ~0.051 ppm on ¹H (~31 Hz) and ~0.19 ppm on ¹⁵N (~11 Hz) on average. Excluding sidechain resonances, 11 peaks were observed in the spectrum. Since each repeat contains 12 residues, one of which is proline, which lacks an amide proton, these 11 peaks correspond to the 11 amide groups in one repeat of the sequence, proving that each repeat within the sequence is equivalent. (Some low intensity peaks were also observed in the hNH spectrum, showing even narrower linewidths than the major peaks. These resonances most probably belong to the first few residues remaining after cleavage of the SUMO tag (see Methods)).

When overlaid with the ¹H-¹⁵N correlation spectrum of TtMacNup98A, the spectrum of the prf.GLFG_{52x12} shows very similar linewidths and chemical shift dispersion (**Figure 4-7A**). This confirms the presence of similar dynamics and the same intrinsic disorder as in the wild type sequence, and shows that the prf.GLFG_{52x12} is indeed a suitable model sequence, even when investigating atomic-level properties.

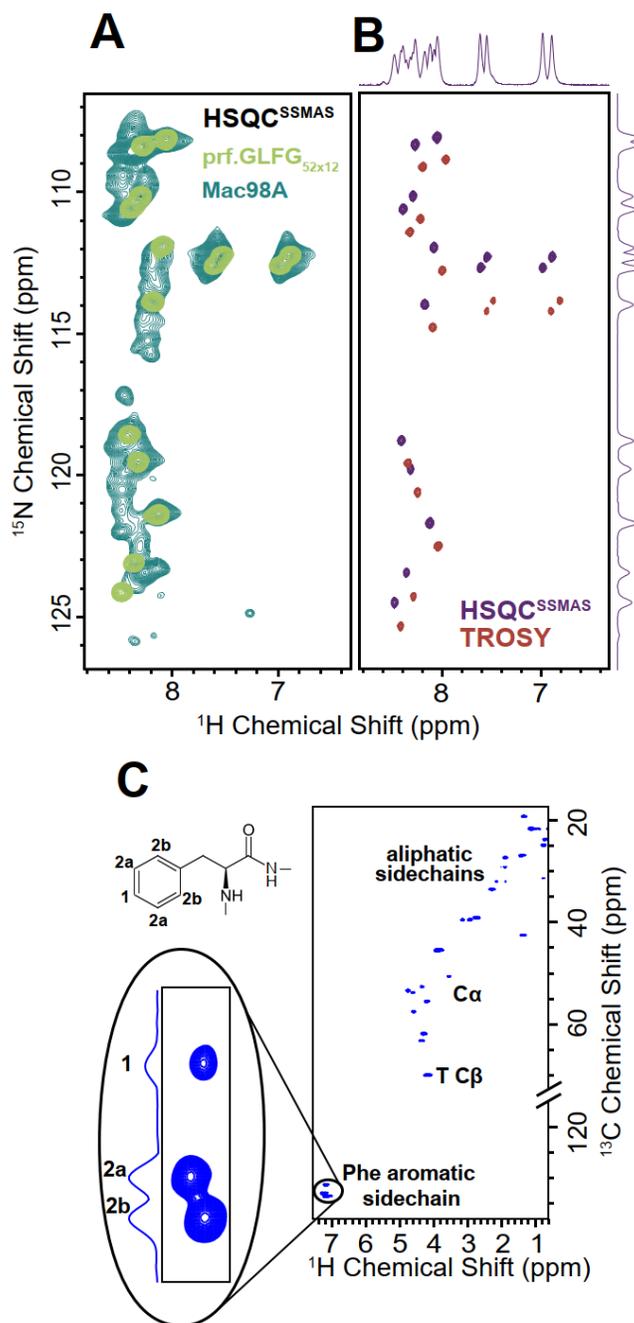


Figure 4-7: The engineered perfectly repetitive sequence, prf.GLFG_{52x12} improves the resolution of NMR measurements (a) Overlay of ¹⁵N-¹H correlation spectra (¹⁵N-¹H HSQC^{SSMAS}) of phase-separated Mac98A FG domain (forest green) and prf.GLFG_{52x12} (light green), where each peak corresponds to an amide group with a distinct chemical environment. (b) ¹⁵N-¹H HSQC^{SSMAS} of prf.GLFG_{52x12} (purple) and the TROSY spectrum (red). The peak positions are shifted by half the ¹J(¹⁵N,¹H) coupling in both dimensions as expected. Proline residues lack HN, and thus are not detected. (c) ¹³C-¹H HSQC^{SSMAS} spectrum of prf.GLFG_{52x12}. The cross-peaks corresponding to CH bonds of the Phe aromatic sidechain are enlarged and labelled with their relative intensities (2:2:1).

Interestingly, the TROSY spectrum (Pervushin et al., 1997; Cavanagh, 2007) (**Figure 4-7B**) showed linewidths only very slightly narrower: 0.046 ppm (28 Hz) on ^1H and 0.14 ppm (8.6 Hz) on ^{15}N , a sign that in the protein short segments are tumbling independently. The hCH spectrum (**Figure 4-7C**) also shows well-resolved, narrow peaks. Remarkably, only three peaks are visible in the aromatic region, at 1:2:2 intensities, corresponding to the 3 C-H groups in phenylalanine's aromatic ring. Several possible stacking arrangements of phenylalanine rings would break the equivalency of the aromatic groups (Martinez and Iverson, 2012), however there are several other arrangements that would conserve equivalency of the aromatic (e.g. C_n symmetry for an n-mer). Thus, while we cannot exclude the possibility, our data does not provide evidence for the formation of a long-lasting structure formed by phenylalanine rings on the NMR timescale.

Due to the previously observed low CP efficiency, the assignment strategy for the prf.GLFG_{52x12} followed the INEPT-transfer based approach used in solution NMR spectroscopy. Since solid-state NMR probes are not equipped with gradients, gradient-free sequences were used for recording HNCA (**Figure 4-8**, modified based on the sequence in (Cavanagh, 2007)) and HNcoCA (Cavanagh, 2007) spectra for residue linking, as well as the HNcaCB (Cavanagh, 2007) spectrum for residue typing.

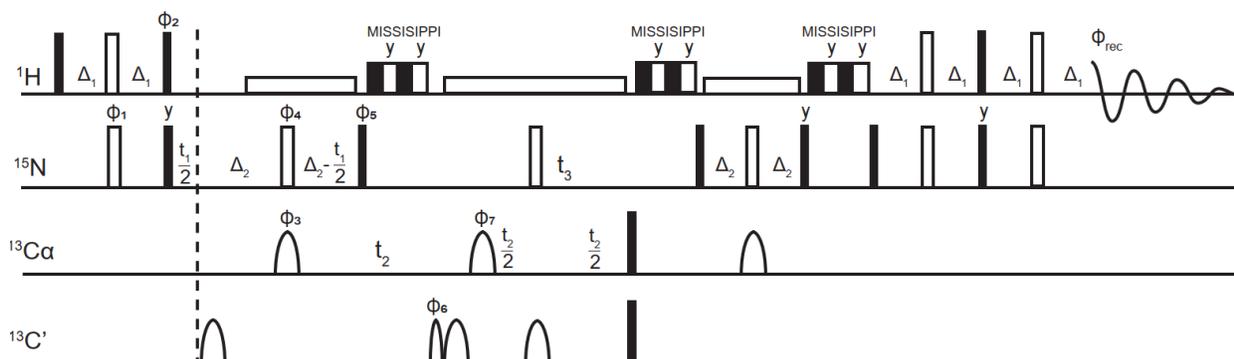


Figure 4-8: INEPT-based HNCA pulse sequence used for residue linking Black rectangles denote $\pi/2$, white (hollow) rectangles π pulses. Shaped pulses are denoted by half-ellipses (narrow: $\pi/2$, wide: π pulses). Phases are x unless shown otherwise. $\phi_1 = x -x$, $\phi_2 = y -y$, $\phi_3 = x*8 y*8$, $\phi_4 = x x x x y y y y$, $\phi_5 = y*8 -y*8$, $\phi_6 = x x x x -x -x -x -x$, $\phi_7 = x -x y -y$, $\phi_{rec} = (y x x -y)*2 (-y -x -x y)*2$. The pulse program can be found in the Appendix.

The high spectral resolution of the assignment spectra allowed a straightforward assignment procedure. The only difficulty in the otherwise unambiguous residue linking was distinguishing

Gly2 and Gly6 from each other. Since the prf.GLFG_{52x12} is intrinsically disordered, the chemical shifts of identical residue types are very similar – in this case the C α chemical shifts of glycines are nearly identical. Since both Gly2 and Gly6 are preceded by another glycine, their peaks in the ¹³C dimension are broadened due to overlap between the *i* (G2, G6) and the *i-1* (G1, G5, respectively) resonances. The exact C α resonances of G2 and G6 could be read off as *i-1* peaks of L3 and N7 in the HNCA spectrum. By determining the C α resonances of the preceding G1 and G5 as *i* resonances in the HNCA spectrum, the H^N and N resonances of G2 and G6 could then be identified unambiguously from the HNcoCA spectrum. The residue linking process and the assigned ¹H-¹⁵N correlation spectrum is shown on **Figure 4-9A** and **B**, respectively, the chemical shifts of each residue, are shown in **Table 4-1**. To identify any secondary structure propensities, $\Delta\delta C\alpha - \Delta\delta C\beta$ values were also determined for all residues except glycine. Positive values are expected for α -helical, negative values β -sheet structures. A secondary structural element is considered fully formed if consecutive residues show values of about ± 2 ppm. This common method for calculating secondary chemical shifts has the advantage that any errors stemming from mis-referencing of the spectra are canceled (Marsh et al., 2006).

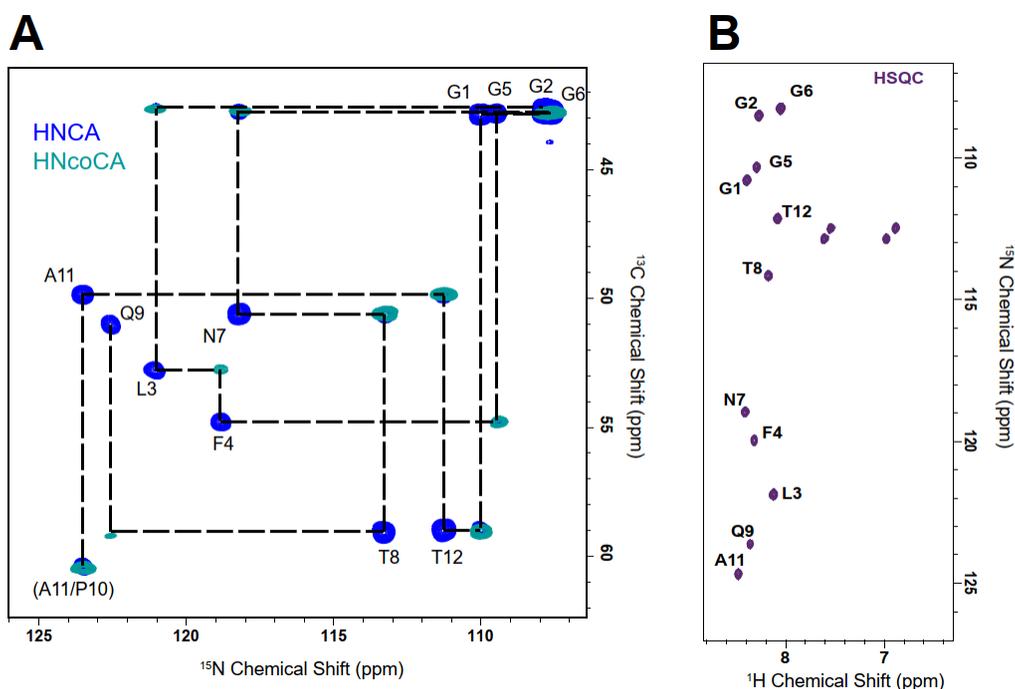


Figure 4-9: Assignment of the prf.GLFG_{52x12} hydrogel. A) Residue linking shown on the ¹³C-¹⁵N projections of the HNCA (dark blue) and the HNcoCA spectra (turquoise). B) Assignment of the ¹H-¹⁵N correlation spectra.

Table 4-1: Chemical shifts and secondary chemical shift of the prf.GLFG_{52x12} Random coil chemical shifts were taken from the BMRB database (Ulrich et al., 2008)

Residue	Chemical shifts (ppm)				$\Delta\delta C\alpha - \Delta\delta C\beta$
	¹ H	¹⁵ N	C α	C β	
G1	8.13	110.1	45.6	-	-
G2	8.03	107.9	45.4	-	-
L3	7.86	121.0	55.5	42.635	-0.903
F4	8.03	118.9	57.5	39.527	-0.578
G5	8.04	109.5	45.5	-	-
G6	7.84	107.7	45.5	-	-
N7	8.18	118.2	53.4	39.146	-0.39
T8	7.91	113.3	61.9	69.913	-3.015
Q9	8.10	122.6	53.8	29.307	-1.649
P10	-	-	63.3	32.081	-0.005
A11	8.20	123.527	52.6	19.353	-0.217
T12	7.82	111.316	61.7	69.939	-3.241

All $\Delta\delta C\alpha - \Delta\delta C\beta$ are slightly negative, indicating a slight propensity towards an extended conformation, however except for T8 and T12, all secondary chemical shift values are smaller than 2 ppm, and no two consecutive residues have values of 2 ppm or above, meaning there is no fully formed β sheet in the protein. The propensity towards an extended conformation is not surprising, considering the relatively high CP efficiency of the proline free variant (Pro-free-prf.GLFG_{52x12}) indicating fibril formation in the absence of the secondary structure-breaking prolines.

Since all peaks are well resolved and could be unambiguously assigned in the ¹H-¹⁵N correlation spectrum, this opened the possibility of residue-specific investigations into the dynamic properties of the hydrogel.

4.3.4 Nanosecond-timescale dynamics of the prf.GLFG_{52x12} hydrogel

Observing the narrow linewidths measured at low MAS frequencies in both the ¹³C direct excitation and the ¹H-¹⁵N correlation spectrum, the low ¹H-¹³C CP efficiency, and the spinning frequency-independent Hartmann-Hahn condition it was already clear that the hydrogel possesses large degree inherent mobility. In order to characterize these dynamics in a residue-specific manner, a [¹⁵N,¹H]-TRACT (TROSY for rotational correlation times) (Lee et al., 2006) spectrum was recorded.

TRACT was originally developed to determine the rotational correlation times (τ_c) of proteins by measuring relaxation rates of both the α and the β spin states of the $^{15}\text{N}^{\text{H}}$ nuclei. Based on the TROSY principle, it eliminates contributions of dipolar couplings to remote protons as well as chemical exchange, factors that would otherwise be detrimental to the accuracy of the measured τ_c -s.

The basis for this is that in expressions R_α and R_β , all contribution to relaxation (λ – auto relaxation rate, R_H – contribution from dipole-dipole coupling with remote protons, R_{CS} – contribution from chemical exchange) appear with identical signs, except for the transverse cross-correlated relaxation rate (η_{xy}):

$$R_\alpha = \lambda - \eta_{xy} + R_{CS} + R_H \quad (4 - 8)$$

$$R_\beta = \lambda + \eta_{xy} + R_{CS} + R_H \quad (4 - 9)$$

These contributions can be eliminated by taking the difference of R_α and R_β :

$$R_\beta - R_\alpha = 2\eta_{xy} \quad (4 - 10)$$

Rate of transverse cross-correlated relaxation can be expressed as (Lee et al., 2006; Cavanagh, 2007):

$$\eta_{xy} = 2p\delta_N(3\cos^2\theta - 1)(4J(0) + 3J(\omega_N)) \quad (4 - 11)$$

where $p = \frac{\mu_0\gamma_H\gamma_N\hbar}{16\pi^2\sqrt{2}r_{NH}^3}$, differing only by a constant multiplier from the spatial function for dipolar relaxation, $\delta_N = \gamma_N B_0 \Delta\delta_N / (3\sqrt{2})$, and is the spatial function for CSA relaxation, and the spectral density function at ω frequency, $J(\omega_N) = 0.4\tau_c / (1 + \omega^2\tau_c^2)$. γ_H and γ_N are the gyromagnetic ratios of protons, and ^{15}N , respectively, \hbar is the Planck constant, r_{NH} is the distance between the amide ^{15}N and ^1H , B_0 is the field of the NMR spectrometer, and τ_c is the rotational correlation time of the molecule or a selected residue.

To determine the residue-specific τ_c of each residue, a pseudo-3D version of the pulse sequence was used (**Figure 4-10A**), where the H-N correlations were resolved in two dimensions, and the relaxation rate of each peak in both α and β states of ^{15}N was determined by recording a series of 2D spectra. **Figure 4-10B** shows for residue F4 the fit of R_α and R_β .

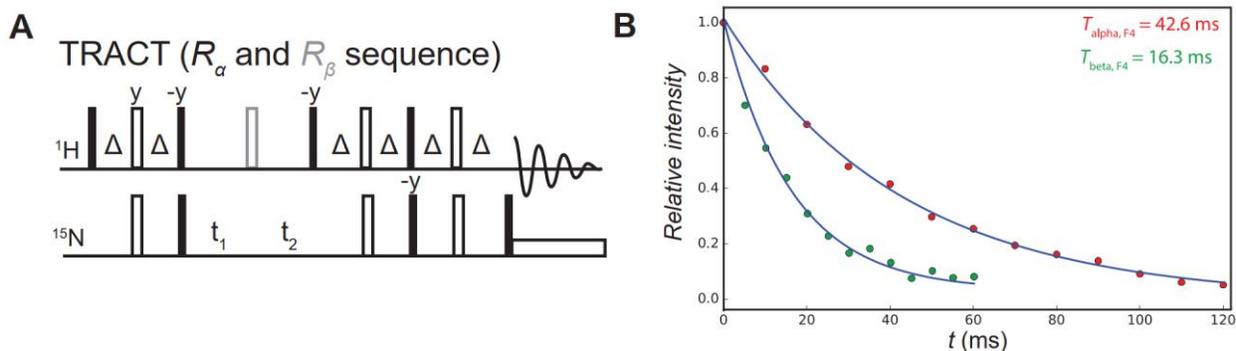


Figure 4-10: A) Gradient-free pseudo-3D TRACT pulse sequence and B) example of fit for R_α and R_β . A) All phases are x, except when indicated otherwise. The Δ delays were set to 1.8 ms. The π proton pulse shown in gray is only executed in the R_β sequence. B) The R_α and R_β fits are shown for the residue F4 in the hydrogel formed by prf.GLFG_{52x12} at 308.7 K. Relative intensities are calculated dividing all intensity values with that of the first time point (0 delay) for both curves.

With the exception of proline (which does not have an amide proton, and thus does not appear in the ^1H - ^{15}N spectrum), τ_c values could be determined for each residue of a repeat (**Figure 4-11**).

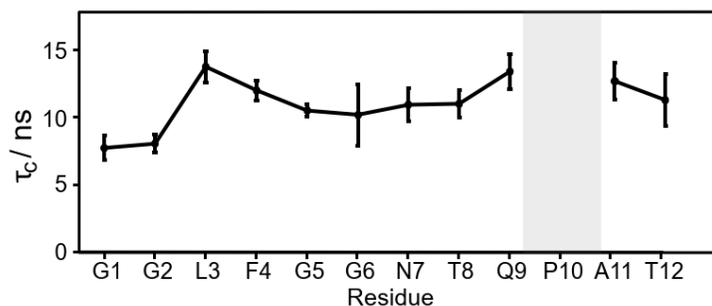


Figure 4-11: Rotational correlation times for each residue in a repeat of the hydrogel formed by prf.GLFG_{52x12}, as determined by [^{15}N , ^1H]-TRACT on the amide ^1H spins. Due to the absence of amide protons, the rotational correlation time of proline could not be determined.

The inherent mobility evident in CP and INEPT spectra is confirmed by the nanosecond timescale motion determined in the TRACT experiment. When plotting τ_c against the residue number, two maxima are observed: one at residues L3 and F4, and another one around P10 (Q9 and A11). The latter can most probably be explained with the bulky five-membered pyrrolidine ring in proline. Proline's fixed ϕ angle ($\phi \approx -60^\circ$) (www.cryst.bbk.ac.uk, 1996) limits the number of the amino acid's possible conformations, and introduces rigidity into the protein backbone (Jacob et al., 1999). Proline also restricts the conformations available to the neighboring residues, especially the

preceding one (MacArthur and Thornton, 1991), and this also appears in their increased rotational correlation times. This effect is clearly visible through the higher τ_c values of Q9 and A11, where both are increased compared to neighboring residues, but especially the τ_c of Q9.

The maximum at L3 and F4 is similarly high as the other one at Q9, with τ_c values ~ 1.5 -times higher than for the spacer residues. The motion of these two residues is most probably slowed down by interactions of their sidechains. This phenomenon is not unexpected for phenylalanine, since phenylalanines were previously shown to be crucial for phase separation behavior, and it was the hydrophobic and the π - π interactions of the aromatic ring that were thought to contribute to cohesion (Frey et al., 2006; Krishnan et al., 2008; Dolker et al., 2010). Strikingly, leucine appears to be even more rigid than phenylalanine. Since leucine residues lack aromaticity, but have a long, purely hydrophobic alkyl sidechain, this result highlights the role of hydrophobic interactions in the gel. Although the interacting motifs in the sequence were previously referred to as FG-motifs, glycine residues do not show elevated rotational correlation times, and probably do not partake in the network of interactions in the gel. Glycine has been shown to have a preference for appearing in turns, rather than structured elements (helices, sheets), due to its flexibility, its large range of available ϕ and ψ angles (SwissModel, 2020). Although we have shown that proline is pivotal in preventing fibrilization of the prf.GFLG_{52x12}, glycine could also have a role in lending flexibility to the sequence and thus further contributing to its disorder by inhibiting aggregation (Parrini et al., 2005).

4.3.5 Leucine's role in phase separation

Since leucine showed a similar degree of rigidity as phenylalanine in the relaxation experiments, we hypothesized that leucine might also take part in interactions crucial to gel formation. Indeed, spatial vicinity between aliphatic sidechain protons and aromatic ring protons had previously been observed in NOE spectra of Nsp1, a yeast nucleoporin that exhibits both gelation and fibrilization propensity (Ader et al., 2010). First we assigned all resonances in the ^1H - ^{13}C correlation spectrum using an HhCH WALTZ-mixing based scheme (**Figure 4-12**).

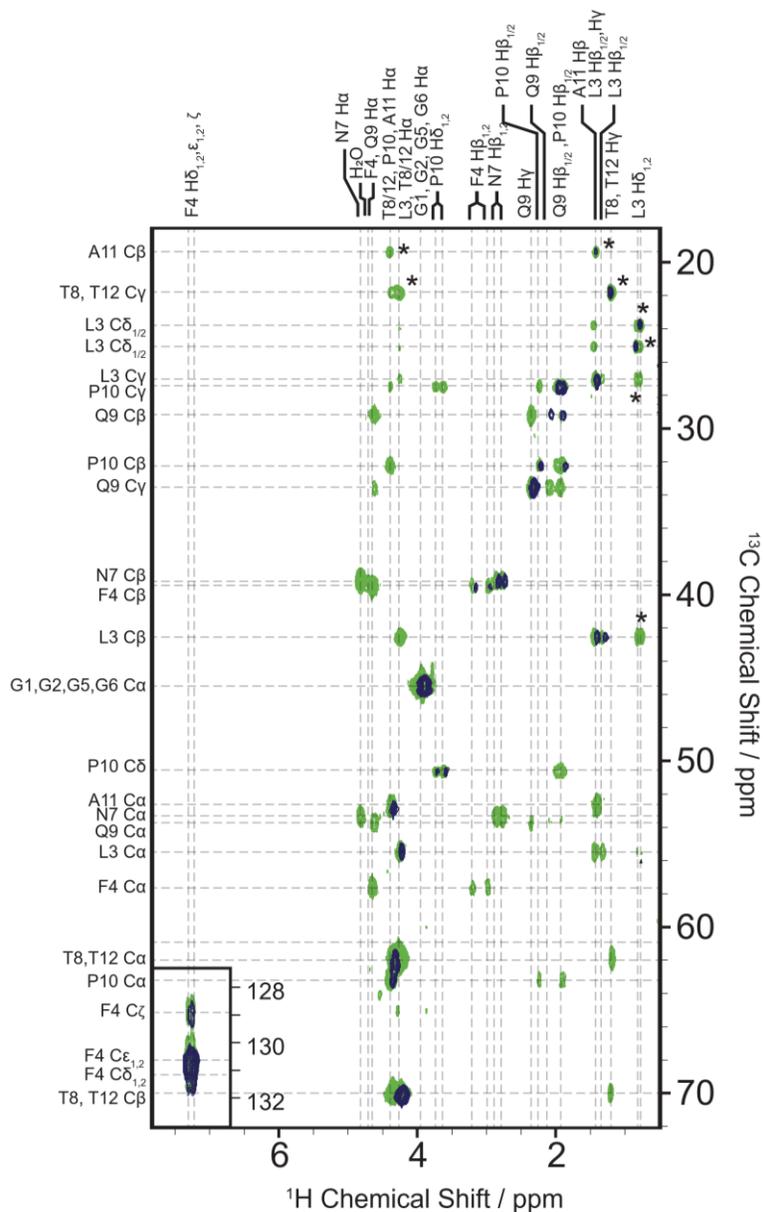


Figure 4-12: Assignment of prf.GLFG_{52x12} sidechains. Mixing of magnetization between the carbons of the same spin system was achieved using the WALTZ (Shaka et al., 1983a; Shaka et al., 1983b) scheme. The resulting hCCH spectrum is shown in light green, in dark blue, the hCH one-bond correlation spectrum. Both spectra were recorded at 308.7K. Dashed lines are used to guide the eye to the chemical shift values of the assigned ¹³C resonances (left) as well as ¹H resonances (top). In case of overlapping chemical shifts, all possibilities are listed. Peaks marked with an asterisk (*) are shown at 5x contour levels compared to the rest of the spectrum. In the insert showing the aromatic F4 sidechain, the hCCH spectrum is contoured at 2x. The strong water signal has been removed from the hCH correlation spectrum to allow observation of the H α resonances clearly visible in the hCCH spectrum, but its position has been marked among the ¹H chemical shift values.

We then compared ¹H-¹⁵N and ¹H-¹³C correlation spectra of prf.GLFG_{52x12} in solution, where the protein concentration is below the critical concentration for phase separation and in a hydrogel

The ^1H - ^{13}C correlation spectra also overlay very well, with practically identical peak positions for all groups. A slight upfield shift can be observed, however, on the CH-groups of the leucine sidechain in the ^1H dimension, on the order of ~ 0.03 ppm. Also considering the evidence from the elevated rotational correlation time of leucine alongside that of phenylalanine, it is likely that this shift is induced by the ring currents in the phenylalanine's aromatic group, and signals an interaction between the hydrophobic leucine sidechain with phenylalanine. Considering that upon tight binding, ring currents can cause chemical shift perturbations of several ppm (Platzer et al., 2020), the small chemical shift perturbation observed for the leucine sidechain would indicate a weak, transient interaction.

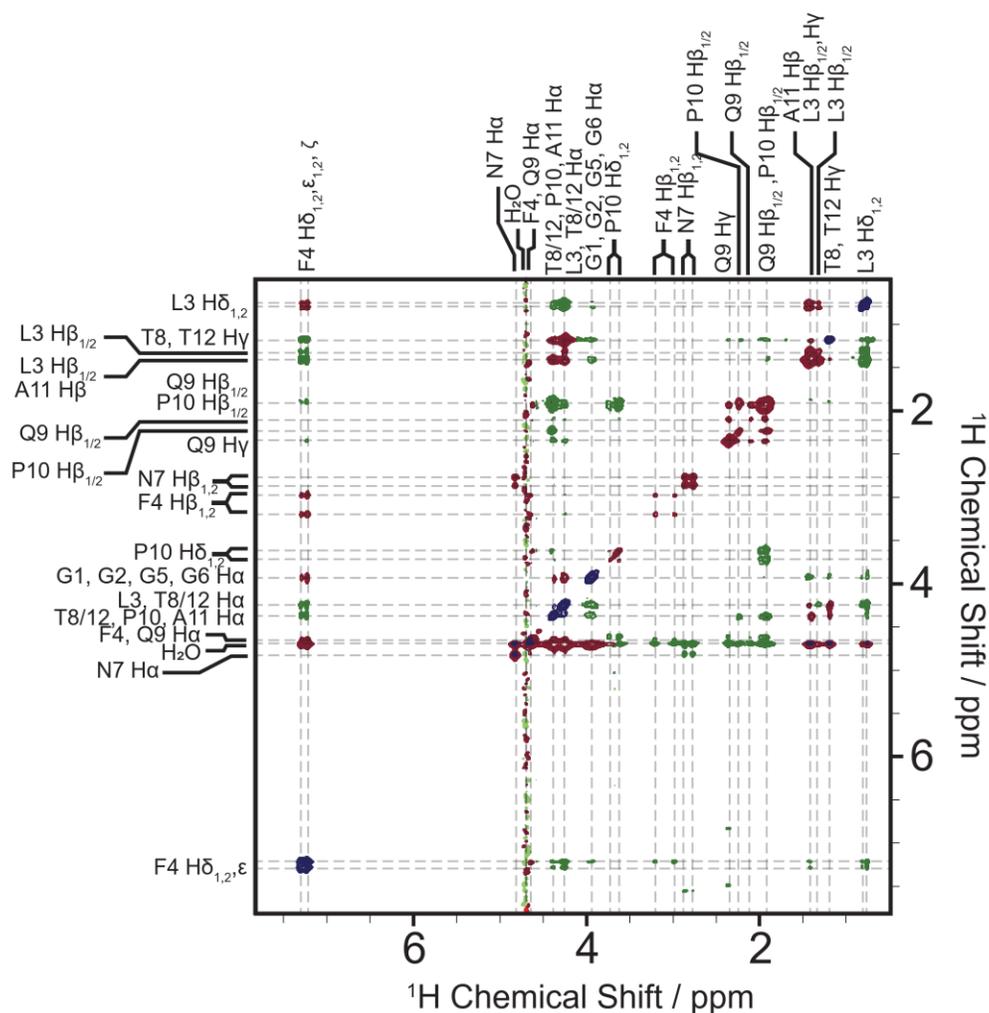


Figure 4-14: ^{13}C -filtered ^1H - ^1H NOESY spectrum of the prf.GLFG $_{52 \times 12}$. Due to the large dynamic range of peaks, the contours are 1x (dark green), 2x (red), 9.5x (dark blue). The applied mixing time was 100 ms, the measurement took place at 304.3 K. Dashed lines guide the eye to the assignments of chemical shifts, in case of overlap all possibilities are listed.

In the NOESY spectrum (**Figure 4-14**) relatively few inter-residue cross peaks could be identified, a further indication of an unstructured peptide chain (Ader et al., 2010). Cross peaks between phenylalanine's aromatic ring and methyl groups (and even α -protons) in leucine, threonine and alanine could be assigned, among which the leucine contact was especially strong. This was also manifested in chemical shift perturbation observed only for the leucine sidechain between solution (dilute phase) and the hydrogel state, and not for the other methyl groups. Although this could possibly be attributed to L3 and F4 being neighbors, no or only very weak contacts are observed for other neighboring sidechains, meaning the stronger interaction probably stems from leucine's hydrophobicity. Contacts between the methyl groups (e.g. L3H $\delta_{1,2}$ – T8/12H γ) as well as between methyl groups and methylene groups (e.g. T8/12H γ – Q9H γ) were also observed, a sign of hydrophobic interactions not involving the phenylalanine sidechain. All protons show contacts to water, indicating water accessibility throughout the sequence.

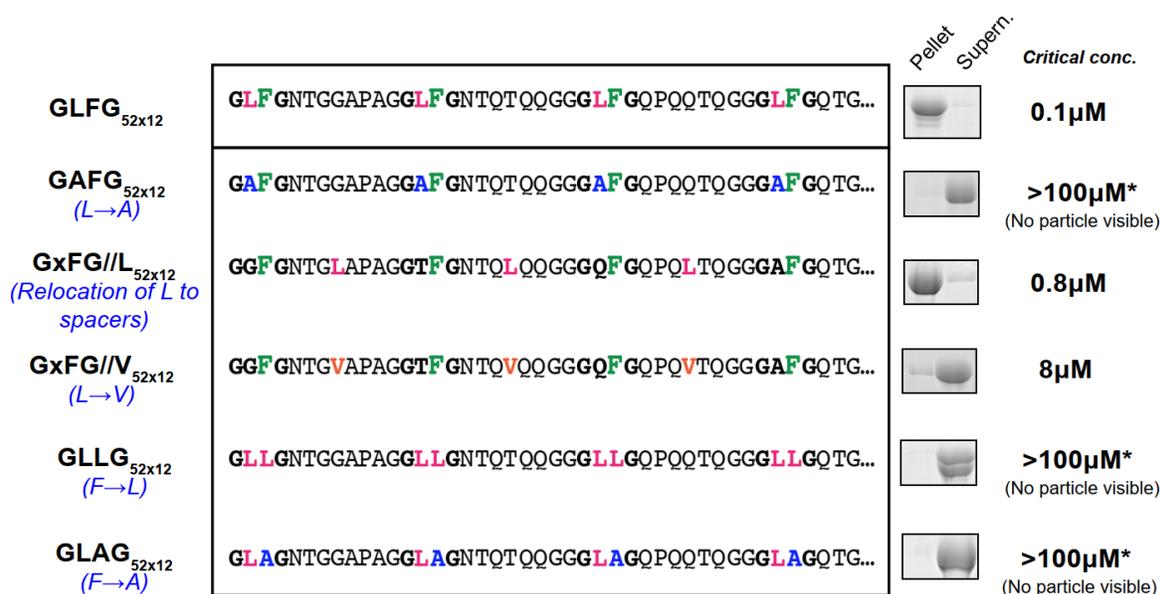


Figure 4-15: Influence of hydrophobic residues on phase separation of FG-domains was investigated on a series of GLFG_{52x12} mutants. The GLFG_{52x12} is an engineered FG domain variant which contains 52 GLFG motifs flanked by non-identical sequences (spacers) of equal length. The spacers do not contain Phe and Leu. Although not perfectly repetitive, the phase separation propensity of the GLFG_{52x12} was virtually identical to prf.GLFG_{52x12}. “//” indicates swapping of a residue with another one from the hydrophilic spacers. For clarity, the sequences up to the fourth FG motif are shown, while the C-terminal sequences (~620 residues) follow the same design strategy (see Appendix for complete sequences). The sequence of the N-terminal affinity tag was kept constant and is not shown. Phase separation of the variants were analyzed as described in Methods at room temperature, at [FG domain]=10 μ M, [NaCl]=150 mM. *The analyses for GAFG_{52x12}, GLLG_{52x12} and GLAG_{52x12} were further repeated at [FG domain] increased to 100 μ M, and no significant phase separation was observed. The absence of phase separation for a mutant is indicated with “No particle visible”.

The importance of non-aromatic hydrophobic residues in phase separation was tested through mutations of the leucine residue. The first mutant was prepared to investigate the significance of the leucine's presence, by mutating all leucine in the GLFG_{52x12} (sequences in Appendix) (in the GLFG motifs) to alanine (GAFG_{52x12}). The other mutant was designed to test for the significance of leucine's position within the sequence, so a residue in the middle of each spacer was swapped with the leucine in GLFG, resulting in GxFG motifs (**Figure 4-15**).

The centrifugation assay showed that removal of leucine resulted in an almost complete loss of phase separation propensity (GAFG_{52x12}), with almost all protein content appearing in the supernatant, while placing the leucine in the middle of the spacers barely influenced phase separation behavior (GxFG//L_{52x12}) (**Figure 4-15**). Remarkably, replacing leucine with only slightly less hydrophobic valine (one methylene group less) led to a 10-fold increase of critical concentration (GxFG//V_{52x12} mutant). Replacement of phenylalanine with alanine (GLAG_{52x12}) or even leucine (GLLG_{52x12}) led to a complete loss of phase separation propensity.

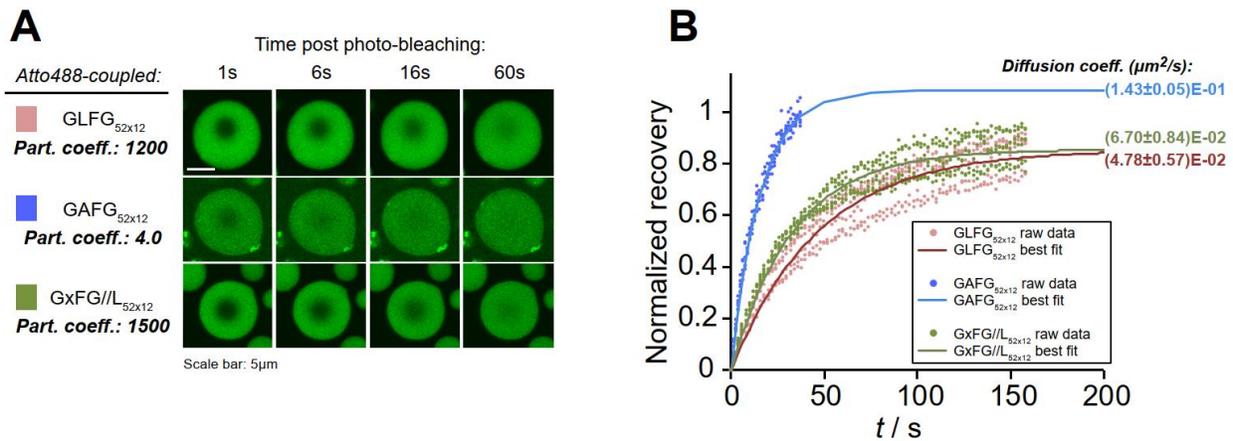


Figure 4-16: Influence of leucine on diffusion within the FG phase of prf.GLFG_{52x12} A) Unlabeled prf.GLFG_{52x12} was premixed with 2.5, 15 and 2.5% (by mole fraction) of Atto488-coupled GLFG_{52x12}, GAFG_{52x12} and GxFG//L_{52x12} respectively. Then the mixtures were allowed to phase-separate. Apparent partition coefficients of the Atto488-coupled species into the assembled FG particles were quantified by measuring the Atto488 signals IN:OUT of the FG particles. The FG particles were photo-bleached at 150 mM NaCl, the bleached area was circular, with a diameter of ~3 µm. Brightness of the image sets were adjusted individually. B) FRAP recovery curves of GLFG_{52x12}, GAFG_{52x12} and GxFG//L_{52x12} in prf.GFLG_{52x12} droplets. Signal recovery, scaled to 1 for 100% recovery of the signal before bleaching, was plotted against time post bleaching. Experiments were repeated with multiple FG particles (n=6). The translational diffusion coefficients are derived from fitting the recovery curves as described in Materials and Methods.

Partition coefficients of GLFG_{52x12} and its mutants were determined by confocal fluorescence microscopy, diffusion coefficients by fluorescence recovery after photobleaching (FRAP) (**Figure**

4-16). Solutions of fluorescently labeled GLFG_{52x12}, and both mutants were allowed to equilibrate with droplets of prf.GLFG_{52x12}, as described in the Methods section. While both the GLFG_{52x12} and its GxFG//L_{52x12} variant have similarly high partition coefficients (1200 and 1500, respectively), the partition coefficient of the GAFG_{52x12} mutant at 4 is ~300-fold lower. The recovery is much faster for the GAFG_{52x12} variant as for either of the other two, as can be seen from images recorded at various time points (**Figure 4-16A**), and the diffusion coefficients obtained from the FRAP recovery curves show the same behavior: the GAFG_{52x12} variant diffuses ~2-fold faster than the GLFG_{52x12} or the GxFG mutant, which again have very similar diffusion coefficients (**Figure 4-16B**).

It is clear that leucine is responsible for a relatively strong hydrophobic interaction, which significantly influences the diffusion of the protein chain inside the hydrogel. Since both sequences containing leucine show similar phase separation behavior and barrier-like properties in excluding inert species (**Figure 4-17**), and have very similar diffusion coefficients regardless of the leucine's position in the sequence, it is clear that it is the overall hydrophobicity of the sequence determining the properties of the selective phase, and not the distribution of the hydrophobic residues within the repeats.

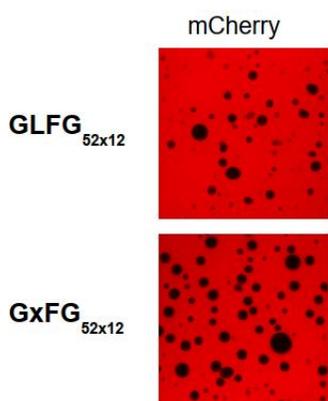


Figure 4-17: Dependence of hydrogel selectivity on the position of leucine within the GLFG_{52x12} sequence. Confocal laser microscopy images showing hydrogel particles prepared from the GLFG_{52x12} and the GxFG//L mutant, incubated with mCherry.

4.3.6 Effect of salt on phase separation and the mobility of the gel

To investigate the effect of salt concentration on phase separation, a centrifugation assay was performed at a series of salt concentrations, ranging from 75 mM to 600 mM of NaCl (**Figure 4-18**). Although all samples showed some degree of phase separation, at 75 mM NaCl, the majority of the protein content remained in the supernatant. With increasing salt concentrations, the protein in the pellet increased gradually, until at 600 mM NaCl, the phase separation was virtually complete.

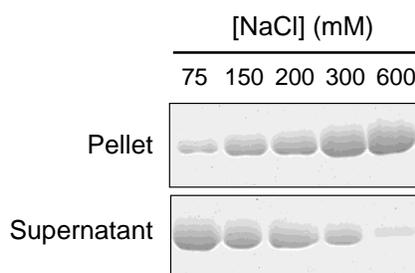


Figure 4-18: Influence of salt concentration on phase separation A centrifugation assay was performed with 10 μ M prf.GLFG_{52x12} at room temperature. The NaCl concentration of the buffer was varied between 75 mM and 600 mM.

We also observed the properties of the hydrogels under different salt conditions following phase separation. Changes in diffusion coefficient with salt concentration were followed with FRAP for two samples at 150 mM and 600 mM NaCl concentration (**Figure 4-19**).

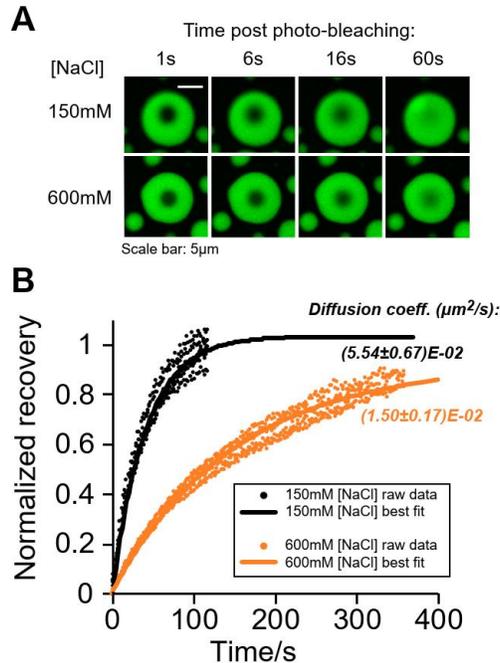


Figure 4-19: Influence of salt concentration on the self-diffusion of the prf.GLFG_{52x12} hydrogel. FG particles assembled from prf.GLFG_{52x12} containing 2.5% Atto488-coupled prf.GLFG_{52x12} molecules were photo-bleached in 150 mM or 600 mM NaCl and the fluorescence recovery after photo-bleaching (FRAP) was followed. A) Representative images showing the recovery of Atto488 signal (Bleached area: circular, with a diameter of $\sim 3 \mu\text{m}$). B) Signal recovery, scaled to 1 for 100% recovery of the signal before bleaching, was plotted against time post bleaching. Experiments were repeated with multiple FG particles ($n=6$). The translational diffusion coefficients are derived from fitting the recovery curves as described in Materials and Methods.

At higher concentration, the recovery of the bleached spot was much slower, and the diffusion coefficients were determined to be $(5.54 \pm 0.67) \cdot 10^{-2} \mu\text{m}^2/\text{s}$, and $(1.50 \pm 0.17) \cdot 10^{-2} \mu\text{m}^2/\text{s}$ for 150 mM and 600 mM NaCl, respectively.

We investigated influence of salt concentration on the dynamic properties of the hydrogel using [^1H - ^{15}N]-TRACT measurements. The overall pattern of the rotational correlation times when plotting τ_c against the residue number remains the same at each concentration; the maxima at L3-F4, and Q9-A11 are observable under all salt conditions, and an increase of salt concentration results in the increased τ_c of each residue. The relative increase in τ_c is similar throughout the sequence - not only the interacting hydrophobic residues, but the whole chain becomes more rigid at higher salt concentrations (**Figure 4-20**).

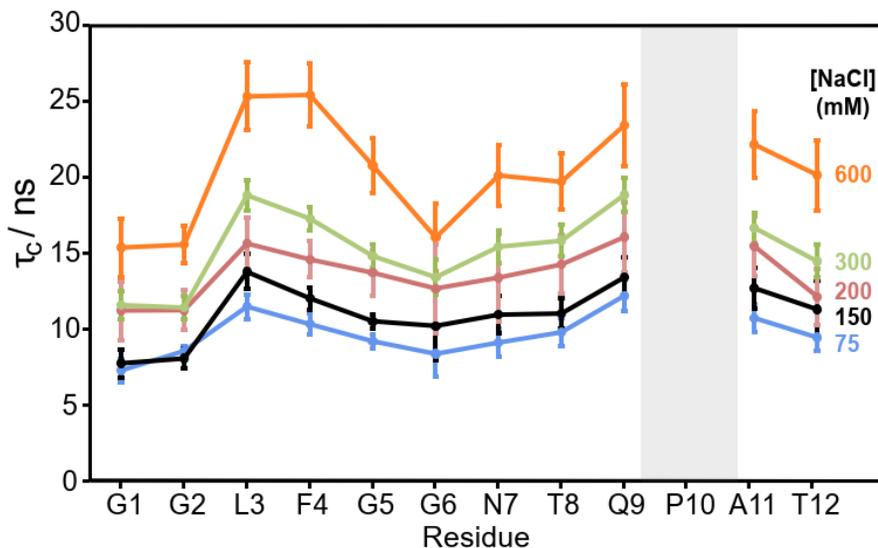


Figure 4-20: Influence of salt concentration on the residue-specific dynamics of the prf.GLFG_{52x12} Rotational correlation times (τ_c) under various NaCl concentrations were determined with the [¹H,¹⁵N] –TRACT experiment, as described in the Methods section. The temperature was set to 308.7 K for all measurements. Due to the absence of an amide proton, the rotational correlation time of proline could not be determined.

The increase in phase separation propensity and the slowing of the hydrogel’s dynamics at higher salt concentrations underlines the importance of hydrophobic interactions in phase separation of FG-domains.

4.3.7 Temperature-dependent phase separation of the prf.GLFG_{52x12}

A 15 μ M suspension of prf.GLFG_{52x12} droplets was prepared as previously described, and the effect of temperature on phase separation was investigated. Interestingly, at room temperature and above (37°C), the droplets formed a turbid colloid, however upon cooling to 4°C, a clear solution was obtained. The effect of cooling was found to be reversible, and upon heating the solution again became turbid (**Figure 4-21A**).

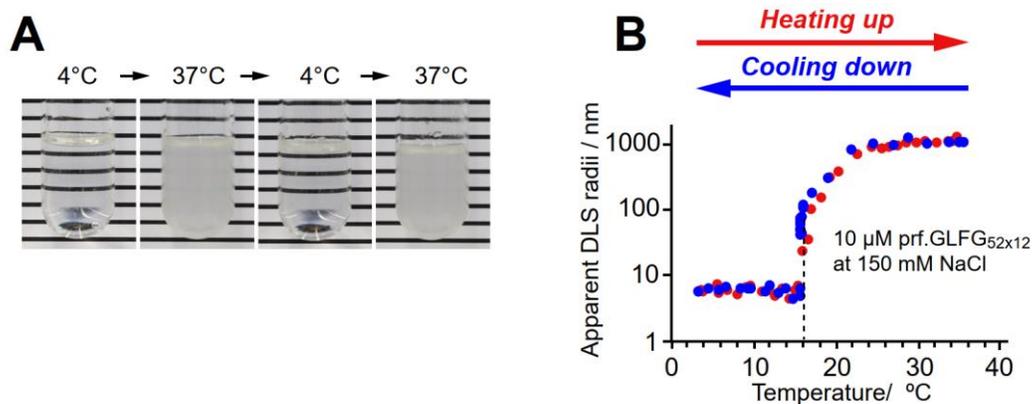


Figure 4-21: Temperature-dependent reversible phase separation of prf.GLFG_{52x12} A) A 15 μM ($[\text{NaCl}]=400 \text{ mM}$) dilution of prf.GLFG_{52x12} upon cooling to 4°C and warming to 37°C, repeatedly. B) A 10 μM dilution of prf.GLFG_{52x12} in buffer containing 150 mM NaCl was analyzed by Dynamic Light Scattering (DLS), with temperature increased from 3 to 37°C (red). A sharp increase in radii (from ~6 nm, sizes expected for monomers, to final radii ~1000 nm, sizes expected for the FG particles) indicates phase separation at the transition temperature ~16°C. The phase separation was reversible by cooling (blue).

The heating and cooling process was also followed with dynamic light scattering (DLS) (**Figure 4-21, Figure 4-21B**). A 10 μM suspension was gradually cooled from 37°C to 3°C, and the particle radii measured at each temperature. From 37°C to 20°C the particle radii remained constant, ~1 μm , where it suddenly dropped sharply, finally stabilizing at 16°C at ~6 nm. This corresponds roughly to the 7.9 nm hydrodynamic radius predicted for unfolded proteins of 58 kDa molecular weight (Hydrodynamic Radius Converter, 2020), which confirms that at low temperature the droplets do indeed become soluble. DLS also confirmed the reversibility of the process: upon heating the particle size increased once more to 1 μm .

The effect of temperature on phase separation could also be clearly seen by performing the centrifugation assay at various temperatures (**Figure 4-22**). While at 7°C most of the protein content can be found in the supernatant, corresponding to a critical concentration (the protein concentration remaining in the supernatant) of 80 μM , at 37°C almost all the protein content is pelleted upon centrifugation, and the critical concentration of the protein is 1.3 μM , showing that an increase in temperature drives phase separation.

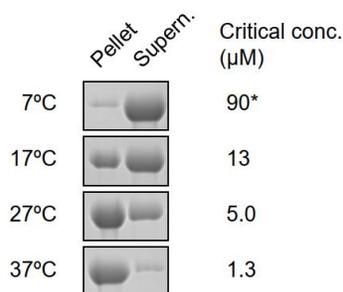


Figure 4-22: Centrifugation assay of the prf.GLFG_{52x12} at various temperatures Phase-separation of 20 μM prf.GLFG_{52x12} at 150 mM NaCl. Dilutions of prf.GLFG_{52x12} were prepared at room temperature and centrifuged at different temperatures. As described in the Methods section, samples of the obtained pellets (FG phase) and supernatants (soluble content) were loaded for SDS-PAGE, followed by Coomassie blue staining. Critical concentrations for individual conditions were defined by the concentrations of the supernatants, which were estimated with standard loading controls. *This set was repeated at 200 μM of prf.GLFG_{52x12} to determine the critical concentration (data not shown).

To better quantify the effect of temperature upon the phase separation of the prf.GLFG_{52x12}, a phase diagram was constructed from critical temperatures determined for protein concentrations between 2.5 μM to 500 μM (**Figure 4-23**). With increasing protein concentration, critical temperatures decrease exponentially, approaching ~273 K, or, inversely, increasing temperatures have exponentially decreasing critical concentrations in solution.

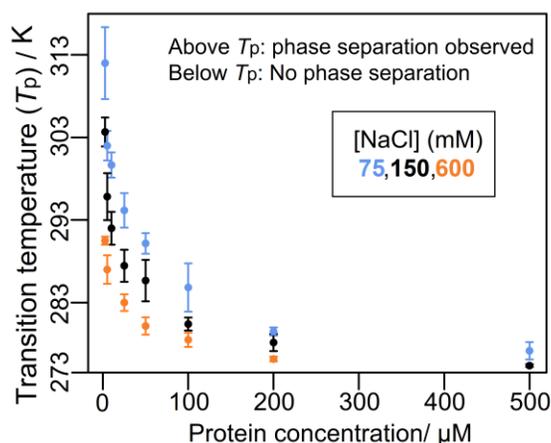


Figure 4-23: Phase diagram of the prf.GLFG_{52x12} DLS analysis (as shown in **Figure 4-21**) was repeated with different [prf.GLFG_{52x12}] and [NaCl] to determine transition temperatures, which were then plotted against [prf.GLFG_{52x12}].

A series of [¹H, ¹⁵N]-TRACT measurements were carried out to investigate the effect of temperature on backbone dynamics both in hydrogel and in solution (dilute phase). Although we had shown that the protein is disordered in both solution and in hydrogel form, a ~ 4-fold increase in backbone correlation times of the prf.GLFG_{52x12} was observed in the hydrogel state at 296.6 K (average τ_c

5 ns) when compared to the protein in solution (average τ_c 1.3 ns). To mostly eliminate the influence of intramolecular interactions, a [^{15}N]-prf.GLFG $_{7 \times 12}$ consisting of seven of the identical repeats found in the full-length protein was recombinantly expressed as a control molecule. The ^1H - ^{15}N correlation spectra of the fragment and the full-length construct showed excellent overlap (**Figure 4-25A**), eliminating the need for assignment of the 7x peptide. (The weaker peaks in the spectrum of the 7x peptide stem from the repeats at the edges of the peptide.) At the same temperature, the 7x peptide exhibits even faster nanosecond timescale motion than the full-length protein (the average τ_c being twofold lower, at 0.7 ns). This observation implies the presence of intramolecular interactions in the dilute phase, occurring due to the high local concentrations of the repeats in 52 repeat long protein. To check this, we measured the diffusion coefficients of prf.GLFG $_{7 \times 12}$, prf.GLFG $_{13 \times 12}$, prf.GLFG $_{18 \times 12}$, prf.GLFG $_{26 \times 12}$, prf.GLFG $_{37 \times 12}$, prf.GLFG $_{44 \times 12}$, and the full-length prf.GLFG $_{52 \times 12}$ peptides (**Figure 4-24**).

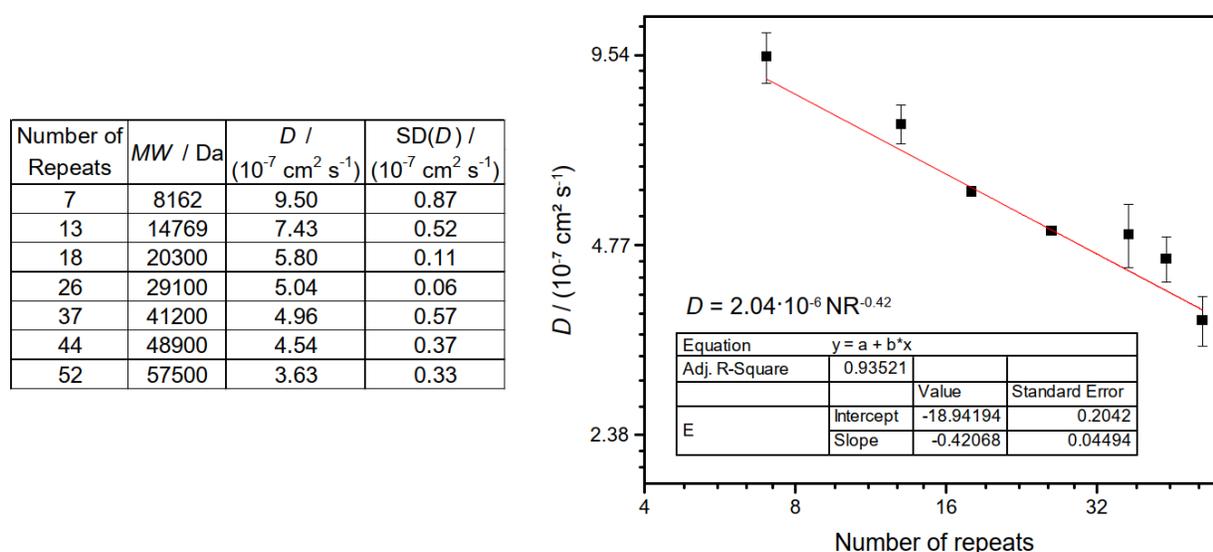


Figure 4-24: Diffusion coefficients of fragment peptides derived from prf.GLFG $_{52 \times 12}$. Since the number of perfect repeats in the peptide fragment (NR) is roughly proportional to its molecular weight (MW), and the NR is a more expressive quantity, the logarithm of the number of repeats was used for fitting. Diffusion coefficients and their standard deviations (SD) were obtained from the average of five independent dynamic light scattering (DLS) measurements at room temperature. The non-linear fitting was performed in the Origin software package.

Plotting $\log(D)$ against $\log(\text{NR})$, where D is the diffusion coefficient, and NR is the number of repeats (proportional to the molecular weight), we obtain linear correlation, where the slope is indicative of the protein's extended or globular conformation. According to empirical data by Dudas et al. (Dudas and Bodor, 2019), the slope is -0.381 for globular, and -0.507 for folded

proteins. The obtained value of -0.43 shows that while the protein is neither completely extended, nor completely globular, its conformation is somewhat closer to the latter, a direct sign of intramolecular interactions, even in solution.

Next, we determined rotational correlation times of each residue in the peptide for prf.GLFG_{7x12} in solution and the full-length prf.GLFG_{52x12} in both the dilute and the gel phase at two different temperatures (296.6 K and 308.7 K).

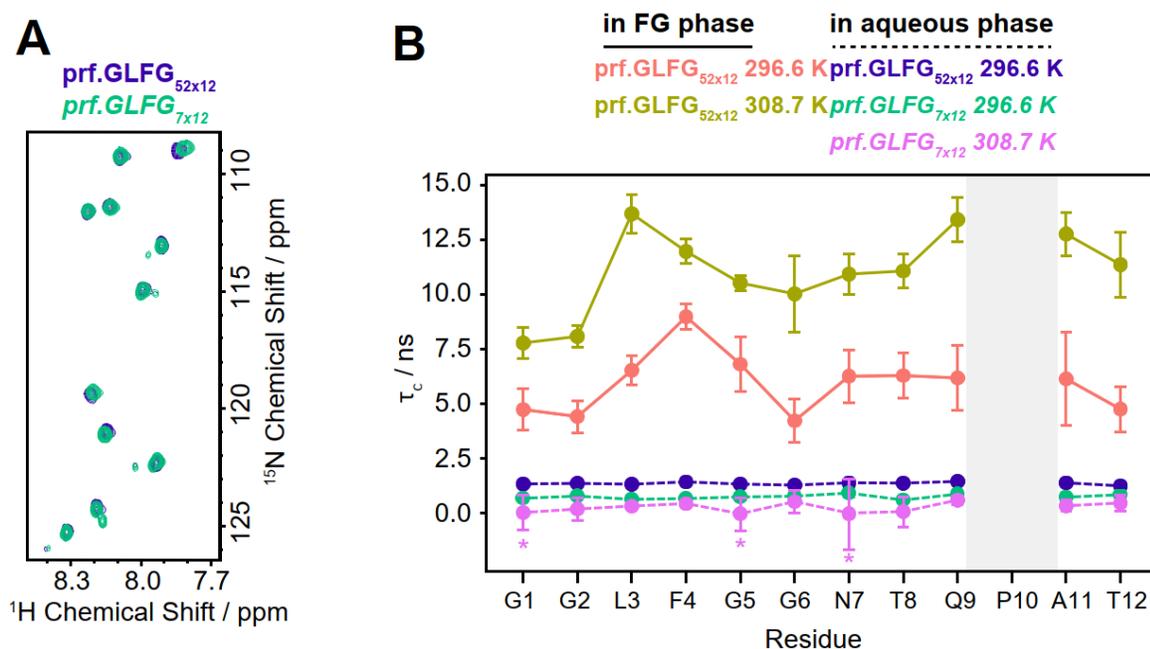


Figure 4-25: The effect of temperature on the dynamics of the prf.GLFG_{52x12}. A) ¹H-¹⁵N correlations spectrum of the prf.GLFG_{7x12} peptide (green) and the full-length (52x) prf.GLFG_{52x12} protein (blue) in solution at 296.6 K. B) Residue-specific correlation times of the full-length prf.GLFG_{52x12} hydrogel at 308.7 K (brown), and at 296.6 K (peach), both shown with solid lines. Dashed lines indicate measurements in solution: full-length prf.GLFG_{52x12} at 296.6 K (blue), prf.GLFG_{7x12} at 296.6 K (green), and prf.GLFG_{7x12} at 308.7 K (pink). The dynamics of the full-length construct could not be determined in solution at 308.7 K due to its extremely low critical concentration. Error bars correspond to σ . At 308.7 K, the exact correlation times of G1, G5, and N7 in the 7x peptide could not be determined, as the relaxation rates of α and β states of the amide ¹⁵N were similar within the range of errors. The τ_c values of these two residues are marked with an * and plotted as 0.

As expected, upon increasing the temperature to 308.7 K, the prf.GLFG_{7x12} peptide became even more mobile, showing a twofold decrease of backbone rotational correlation times to an average of ~0.3 ns, as determined by TRACT. (Correlation times of the full-length protein could not be determined at 308.7 K due to the extremely low, 1.3 μ M critical concentration at this temperature).

The effect of temperature on the hydrogel was also investigated (**Figure 4-26**). A suspension of hydrogel particles obtained after cleavage of SUMO-prf.GLFG_{52x12} were transferred into a 3 mm glass NMR tube, and allowed to settle with gravity. Upon heating to 60°C by immersion in a water bath, the resulting gel showed only slight shrinkage but no collapse. At 65°C, however the gel underwent immediate collapse, and this change was irreversible in the course of a few hours. However, after cooling on ice, then letting the gel warm up to room temperature, the volume around the shrunken gel became turbid, proving that the protein in the pellet shows phase separation even after heat treatment.

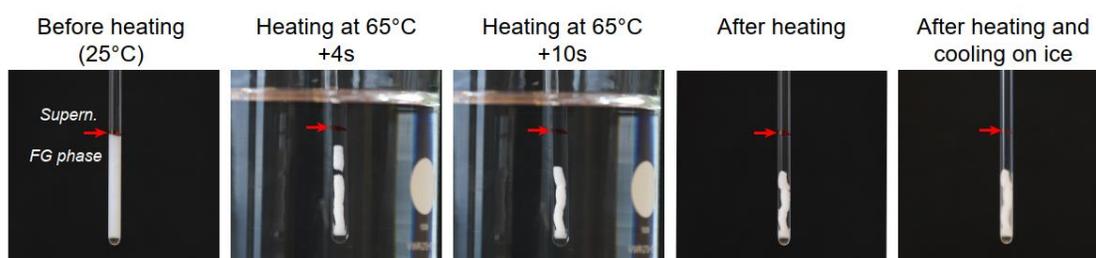


Figure 4-26: Gel collapse upon heat treatment 30 mg of prf.GLFG_{52x12} hydrogel particles were allowed to settle by gravity into a volume of ~150 μ l at 25°C and heated up to 65°C for 1 min and then cooled down on ice for 15 min. The red arrow marked the boundary between the condensed phase and solution before heating.

4.4 Discussion

4.4.1 Native conformation of FG-domains

The investigation of FG-domains in hydrogel state at atomic-resolution has long been intractable due to their degenerate, highly repetitive sequence. The solid-state NMR study of the hydrogel formed by the FG repeat domain of the yeast nucleoporin Nsp1p showed the existence of both hydrophobic contacts and fibril-like structures in the hydrogel, however a complete assignment of the protein was not feasible due to low resolution, and the dynamics of the hydrogel were also not characterized in detail. Shorter segments of FG-domains had been previously studied by solution NMR, such as the PxFG-rich domain of Nup153 (Milles et al., 2015), or the in-cell study on fragments of Nsp1 (the N-terminal FG-N segment, and the central, charged, FSFG-K segment) (Hough et al., 2015; Pakravan et al., 2020). While these studies gave valuable insights into the binding of NTRs to FG-domains (Milles et al., 2015), and the effects of crowding agents and competitors in a cellular milieu (Hough et al., 2015), the protein concentrations used were more than an order of magnitude too low to reproduce the FG repeat concentration of up to 260 mM *in vivo* (Aramburu and Lemke, 2017).

Designer FG-domains, where the composition of the sequence mirrors that of the wild-type proteins (Fragasso et al., 2020) provide an excellent opportunity for constructing sequences amenable to atomic resolution characterization by NMR. Through simplification of the sequence of *Tetrahymena thermophila* MacNup98A (TtMacNup98A), Ng and Görlich obtained a perfectly repetitive sequence that could entirely recapitulate the permselectivity properties of the full-length protein and of nuclear pores (Ng and Görlich, 2021) (unpublished). We found that this construct reproduced the spectral properties and the mobility of the original sequence (**Figure 4-7**), and was also tractable by NMR spectroscopy, making it an ideal system for residue-specific investigations.

Although the hydrogel of Nsp1 also exhibited fibril-like structures, and these were hypothesized to be relevant for selective transport (Ader et al., 2010), various biophysical methods (gel filtration, CD and FTIR spectroscopy) had proven that FG repeat regions in nucleoporins are natively disordered (Denning et al., 2003). Remarkably, the composition of Nup98 FG-domains, showing a low fraction of charged residues along with high hydrophobicity, is distinct from that of both

globular and intrinsically disordered proteins (Schmidt and Görlich, 2015). This, however was found not to result in a globular fold, rather it was suspected that the protein exhibits cohesive interactions between the FG motifs. Indeed, we found a high degree of mobility in the protein backbone (discussed below), and that secondary chemical shifts in the prf.GLFG_{52x12} confirm disorder in the protein, evidence that the presence of fibrils is not necessary to achieve selectivity of the hydrogel phase (cf Ader et al. (Ader et al., 2010)). This is in accordance with the fact that even in the absence of NQ-rich spacers prone to forming fibrils, hydrogel particles from *Xenopus tropicalis* Nup98 (XtNup98) still showed permselectivity (Labokha et al., 2013). $\Delta\delta C\alpha$ - $\Delta\delta C\beta$ values were, however, not evenly distributed around zero, all being slightly negative, though not reaching the values expected from a fully formed β -sheet. This could confirm that the protein favors extended conformations, as observed in MD simulations for peptides derived from the non-aggregating C-terminal sequence of yeast Nsp1p (Dolker et al., 2010).

4.4.2 High degree of mobility in the hydrogel phase

We found liquid-like properties of the TtMacNup98A and the prf.GFLG_{52x12} hydrogel particles by i) observing narrow linewidths in ¹H- detected NMR spectra on the fully deuterated protein even at low (20 kHz) magic angle spinning frequencies, ii) measuring low CP efficiencies in the hydrogel phase, and finding an optimal Hartmann-Hahn condition independent of spinning frequency iii) observing self-diffusion in fluorescence recovery after photobleaching (FRAP) experiments. Furthermore, Ng et al. had previously also observed fusion of the prf.GLFG_{52x12} particles (Ng and Görlich, 2021) (unpublished). The measured self-diffusion coefficient of $\sim 5 \cdot 10^{-2} \mu\text{m}^2\text{s}^{-1}$ for prf.GLFG_{52x12}, although more than two orders of magnitude slower than expected for proteins of similar size in water (Dudas and Bodor, 2019), is comparable to diffusion rates seen in crowded, fluid, cellular environments, such as membranes (Milo and Phillips, 2015). Diffusion coefficients on the same order were measured in other liquid condensed phases: e.g. for the survival motor neuron protein (SMN) in phase-separated droplets of FUS (Murakami et al., 2015), phase-separated Ddx4 (Nott et al., 2015), and similarly rapid FRAP recovery was also seen in droplets of tau (Wegmann et al., 2018).

Rapid, nanosecond timescale motion was detected through measuring ^{15}N transverse cross-correlated relaxation rates using the TRACT experiment both in solution and in hydrogel state. These rapid motions are characteristic of highly dynamic molecules, such as IDPs (Khan et al., 2015; Abyzov et al., 2016; Wong et al., 2020), and were also observed for FG-domains in buffer (Milles et al., 2015; Wall and Hough, 2018). High mobility of the protein backbone was maintained for several disordered proteins upon phase separation, e.g. in droplets of Ddx4 (Brady et al., 2017), tau (Majumdar et al., 2019), or SLP65 (Wong et al., 2020).

The rotational correlation times (τ_c) we observed for the perfectly repetitive sequence were slightly dependent on the length of the construct – the prf.GLFG_{52x12} protein showing roughly twice slower motion than the prf.GLFG_{7x12} peptide, most probably a sign of transient, intramolecular interactions. Upon phase separation, backbone motions slowed down significantly, about ~2.5-fold for residues least affected. This could partially be due to viscosity effects, since – though more prominently observed for folded proteins (Luchinat and Banci, 2016) – in-cell studies of IDPs have also demonstrated faster transverse relaxation rates and broader lines compared to *in vitro* conditions (e.g. α -synuclein (Theillet et al., 2016)). This effect was even observed for FG-domains, where rotational correlation times in a cellular environment increased 1.5-fold on average for residues less affected in the FSFG-K segment of Nsp1 at 298 K (Wall and Hough, 2018).

4.4.3 Hydrophobic interactions in phase separation

Upon phase separation, the relaxation profile across the sequence was altered: rotational correlation times exhibited distinct maxima around the FG motifs, for both leucine and phenylalanine. This increase in the correlation times of certain residues usually stems from nascent secondary structures (Salvi et al., 2017) or the presence of active sites participating in binding, (Nyqvist and Dogan, 2019) as was observed in cell for the FSFG-K fragment (Wall and Hough, 2018). In the hydrogel, since there are no binding partners present, the change in the τ_c profiles clearly signals inter- and intramolecular association among particular residues. Transient interactions were previously observed in NOESY spectra between the aromatic ring in phenylalanine and aliphatic sidechains (mostly methyl groups in alanine, threonine, and leucine) (Ader et al., 2010), however we not only observed these $\text{CH}\cdots\pi$ interactions between the aliphatic donors and the aromatic π -acceptor

(Brandl et al., 2001) (**Figure 4-13**), but also interactions between purely hydrophobic, non-aromatic groups (**Figure 4-14**).

The importance of hydrophobic contacts had been previously demonstrated by showing that i) addition of hexane-diols disrupts the selective barrier of NPCs (Ribbeck and Görlich, 2002) and interferes with formation of FG particles (Schmidt and Görlich, 2015). Hexane-diols are widely used tools used for probing liquid-liquid phase separation both *in vivo* and *in vitro*, as they disrupt (amongh others) weak hydrophobic interactions (Posey et al., 2018). ii) both F→A and L→A mutations in bead-immobilized yeast FG-domains (Nup100 and Nup116) eliminate binding of wild-type proteins (CFP-FG-domain fusions) (Patel et al., 2007) iii) phase separation behavior can be eliminated by replacing all hydrophobic residues in yeast Nsp1 (including phenylalanine) with polar serine (Ader et al., 2010). So far, the direct influence of non-aromatic, hydrophobic residues upon phase separation had not been previously observed. Here we showed that replacing leucine with less hydrophobic alanine eliminates hydrogel formation. Relocating leucine within the sequence, though it has been speculated that the exact motifs in which phenylalanines appear have an influence on the binding of NTRs (Denning and Rexach, 2007), however does not have a significant effect on phase separation or permeselectivity (critical concentrations being sub-micromolar in either case), proving that overall hydrophobicity determines phase separation behavior of a sequence (**Figure 4-15, Table 4-2**). Indeed, although GLFG motifs appear to be evolutionarily conserved (Denning and Rexach, 2007), leucines (and other hydrophobic residues) are found dispersed throughout the sequences of FG-domains, e.g. in TtMacNup98A only two-thirds of leucines have phenylalanine as a neighboring residue.

The overall hydrophobicity of a sequence shows good correlation with its phase separation behavior, with a tendency for more hydrophobic sequences to have lower critical concentrations (**Table 4-2**). This is true regardless of the hydrophobicity scale being used to characterize the sequence's hydrophobicity (Appendix 5.4.4). Apparently, phase separation behavior of the sequence is relatively sensitive even to small changes in hydrophobicity: replacement of all leucine residues with valine (containing only one methylene group less in their sidechains) led to a ten-fold increase in critical concentration. An interesting exception to this rule seem to be aromatic residues: while the GLLG_{52x12} sequence has a higher hydrophobicity than the GxFG//V_{52x12}

sequence according to its hydrophobic index (regardless of the hydrophobicity scale, see Appendix 5.4.4), it does not show phase separation.

Table 4-2: Hydrophobicity and critical concentrations of designer 12-mer sequences shown in **Figure 4-15**. Percentages of hydrophobic residues in each sequence (F-phenylalanine, FILMV – phenylalanine, isoleucine, leucine, methionine, and valine) were calculated based on the sequences shown in the Appendix 5.4.1. Hydrophobic indices were calculated using the parameters published by Fauchere and Pliska (Fauchere and Pliska, 1983). A greater hydrophobic index corresponds to a more hydrophobic sequence. Hydrophobic indices based on (Wimley and White, 1996; Zhao and London, 2006; Moon and Fleming, 2011) show a very similar tendency for the hydrophobicity of the various sequences, and are shown in the Appendix, section 5.4.4.

	F	FILMV	Hydrophobic index	$c_{crit} / \mu\text{M}$
GLFG _{52x12}	8.10%	17.40%	0.31	0.1
GAFG _{52x12}	8.10%	9.60%	0.20	>100
GxFG//L _{52x12}	8.10%	17.40%	0.31	0.8
GxFG//V _{52x12}	8.10%	17.40%	0.27	8
GLLG _{52x12}	0.30%	17.40%	0.30	>100
GLAG _{52x12}	0.30%	9.60%	0.20	>100

Although a rather crude model system, around room temperature, the Gibbs energy for the dehydration of aliphatic molecules (ethane, propane, butane) was found to be negative. For aromatic molecules (benzene, toluene), however, dehydration is accompanied by an increase in Gibbs energy due to the large enthalpic contribution of hydrogen bonds formed between water and the aromatic group (Makhatadze and Privalov, 1994), meaning that replacement of an aromatic group with an aliphatic one should favor phase separation. The opposite behavior we observe indicates that the enthalpic contribution of π - π interactions between aromatic rings also might play a role in the cohesive behavior of a sequence. The direct observation of these (possibly transient) contacts is an important point for future investigations.

Hydrophobic interactions in the hydrogel, however, are most probably only transient, as indicated by the nanosecond-timescale mobility of the hydrogel (**Figure 4-11**), and by the random coil-like chemical shifts of leucine methyl groups even in the hydrogel state (**Figure 4-12**). This is in accordance with the MD simulations, where the majority of phenylalanines were found to be of a dynamic nature (Dolker et al., 2010).

While we have shown that hydrophobic (and most likely π - π) interactions in a fully disordered sequence are sufficient to obtain NPC-like permselectivity, fibrilization has been observed in

several FG-NUPs. Neither the purpose of the β -sheet structures or the driving forces for their formation are clear. Even homologs of the same nucleoporin might exhibit differing structural preferences in different species. For example, Labokha et al. had previously found that XtNup98 gave strong CP signals, characteristic of a structure rigid on the ms timescale, and several resonances clearly indicated the presence of a β -strand conformation (Labokha et al., 2013). On the other hand, we had observed the sequence of TtMacNup98A to be completely disordered, and giving only very weak CP signal. Previously the N and Q content had been suggested to determine whether the sequence forms fibrils (Michelitsch and Weissman, 2000; Ader et al., 2010), however Schmidt et al. found that this alone is not a reliable predictor for the formation of fibrils, and suggested that glycine content could also determine (counteract) the formation of β -sheets (Schmidt and Görlich, 2015). Indeed, XtNup98 has a lower ratio of N and Q in its sequence than TtMacNup98A (or the model sequence prf.GLFG_{52x12}). Upon inspection of the sequence composition, however, it immediately meets the eye that XtNup98 has a much lower ratio of proline and glycine than the TtMacNup98A (**Table 4-3**). Since both proline and glycine are residues known to break regular secondary structures (Levitt, 1978), this could explain the tendency of XtNup98 to fibrilize. Indeed, upon the deletion of prolines from the perfectly repetitive sequence (Pro-free-prf.GLFG_{52x12}), we also observed the formation of fibrils, supported by efficient CP, a spinning frequency-dependent Hartmann-Hahn condition, and ThT staining of the aggregates (**Table 4-3, Table 4-4**) (Ng and Görlich, 2021) (unpublished). Surprisingly, deletion of prolines from the TtMacNup98A sequence does not lead to fibril formation (Ng and Görlich, 2021) (unpublished). The reason for this might lie in the sequence's still very high glycine content (**Table 4-3**).

This evidence suggests that the glycine and proline content in a nucleoporin's sequence play a crucial role in determining the sequence's structural preference, and is in line with previous observations (Schmidt and Görlich, 2015). Of course, the repetitive nature of a sequence might also promote fibrilization, as indicated by comparison of the proline-free GLFG_{52x12}, and proline-free TtMacNup98A, which are similar in all calculated parameters, however only the perfectly repetitive sequence shows ThT staining.

Table 4-3: Composition of wild-type FG-domain sequences with a focus on fibrilization-inducing asparagine and glutamine, fibril-breaking proline and glycine, and hydrophobic residues. Percentages of residues in each sequence (P-proline, G-glycine, N-asparagine, Q-glutamine, F-phenylalanine, FILMV-phenylalanine, isoleucine, leucine, methionine, and valine) were calculated based on the sequences shown in Appendix 5.4.1. The hydrophobic indices are calculated with parameters published by Fauchere and Pliska (Fauchere and Pliska, 1983). A higher hydrophobic index indicates greater hydrophobicity. The presence or absence of Thioflavin T (ThT) staining for the sequences is also indicated. Sequences are listed in the Appendix. Whether the XtNup98 hydrogel shows ThT staining is currently unknown. Hydrophobic indices based on (Wimley and White, 1996; Zhao and London, 2006; Moon and Fleming, 2011) are shown in the Appendix, section 5.4.4. The FG-domains from *S. cerevisiae* Nup116 and *C. elegans* Nup98 are also included in the analysis.

	TtMacNup98A	Pro-free TtMacNup98A	XtNup98	ScNup116	CeNup98
P	4.60%	0.00%	5.40%	4.10%	4.30%
G	31.90%	33.60%	21.30%	19.30%	18.50%
P+G	36.50%	33.60%	26.70%	23.40%	22.80%
N	9.70%	11.00%	8.20%	13.60%	10.50%
Q	8.00%	8.90%	3.70%	13.00%	8.10%
N+Q	17.70%	19.90%	11.90%	26.60%	18.60%
F	7.00%	7.40%	11.70%	8.30%	7.90%
FILMV	16.70%	17.70%	18.30%	17.00%	17.00%
Hydrophobic index	0.30	0.27	0.37	0.22	0.26
ThT	-	-	?	+	+

Table 4-4: Composition of designer FG-domain sequences with a focus on fibrilization-inducing asparagine and glutamine, fibril-breaking proline and glycine, and hydrophobic residues. Percentages of residues in each sequence (P-proline, G-glycine, N-asparagine, Q-glutamine, F-phenylalanine, FILMV-phenylalanine, isoleucine, leucine, methionine, and valine) were calculated based on the sequences shown in Appendix 5.4.1. The hydrophobic indices are calculated with parameters published by Fauchere and Pliska (Fauchere and Pliska, 1983). A higher hydrophobic index indicates greater hydrophobicity. The presence or absence of Thioflavin T (ThT) staining for the sequences is also indicated. Sequences are listed in the Appendix. Hydrophobic indices based on (Wimley and White, 1996; Zhao and London, 2006; Moon and Fleming, 2011) are shown in the Appendix 5.4.4.

	prf.GLFG _{52x12}	Pro-free- prf.GLFG _{52x12}
P	8.30%	0.20%
G	33.30%	31.70%
P+G	41.60%	31.90%
N	8.33%	8.20%
Q	8.33%	8.20%
N+Q	16.66%	16.40%
F	8.30%	8.10%
FILMV	16.60%	17.40%
Hydrophobic index	0.35	0.29
ThT	-	+

4.4.4 Entropy-driven phase separation of FG-domains

LCST behavior has been observed mostly in polymers and proteins rich in hydrophobic residues, including elastin-like peptides (ELPs) (Li et al., 2014; Zhao et al., 2016) and the yeast heat shock granule-associated poly-A binding protein (Pab1) (Riback et al., 2017). We showed for the first time that FG-nucleoporins exhibit LCST behavior, showing phase separation at higher temperatures. Considering the hydrophobic nature of FG-domains, and the importance of hydrophobic interactions in their hydrogel state (Ribbeck and Görlich, 2002; Patel et al., 2007; Ader et al., 2010; Schmidt and Görlich, 2015) (**Figure 4-14**, **Figure 4-15**), the lower critical solution temperature behavior was not unexpected.

Upon coacervation, the hydrophobic sidechains in the dense phase form interactions, resulting in the release of water molecules from their hydration shells (Martin and Mittag, 2018). While the exact nature of this hydration shell is unclear, and various models have been proposed to describe it (Frank and Evans, 1945; Pratt, 2002; Godec and Merzel, 2012; Grdadolnik et al., 2017), the role of entropy in the hydrophobic effect is certain. Entropically driven solvent release is the basis of LCST-behavior (Martin and Mittag, 2018): upon the clustering of hydrophobic residues (Phe, Leu), the ordered water shell around the hydrophobic groups is released, resulting in a higher translational and rotational freedom for the water molecules. The deficiency in hydrophilic and charged residues in certain FG-domains and also in ELPs (Widder et al., 2017) most probably also promotes this effect, as this way the enthalpic term stemming from favorable hydrogen bonding interactions with water is lower, and can more easily be counteracted by the favorable entropy of solvent release. This interpretation is also consistent with earlier observations that engineered GFPs enriched in Arg partition into FG phase much stronger than GFPs enriched in Lys, as the guanidinium group of Arg is more hydrophobic than the ammonium group of Lys, leading to a lower dehydration (entropic) penalty upon interactions with the FG phase (Frey et al., 2018).

The temperature-dependence of phase separation has in some cases been suggested to play a regulatory role, for example in the formation of stress granules in response to environmental stress involving the Pab1 protein also showing LCST behavior (Lau et al., 2020). Although observing LCST behavior is an important indicator of the dominance of hydrophobic interactions in the hydrogel, considering the very high (2-3 orders of magnitude) oversaturation of FG-domains in the NPCs (Schmidt and Görlich, 2015), and assuming a similar temperature-dependence of the phase

separation for Nup98-like wild-type FG-domains as in the model sequence (**Figure 4-23**), the LCST behavior of these FG-domains is unlikely to play such a role in the physiologically relevant temperature range.

Notably, however, while an increase in temperature led to increased mobility in solution state for the GLFG_{7x12} peptide, in the hydrogel state of the prf.GFLG_{52x12} it resulted in a more rigid protein chain. The rigidified protein backbone directly shows the strengthening of the entropy-driven hydrophobic interactions at higher temperatures. It is an interesting question whether the temperature-sensitive dynamic behavior of the Nup98-like FG-domains influence transport rates, and perhaps plays a regulatory role in animals that experience a large temperature range (e.g. amphibians), however due to technical difficulties this has not yet been investigated.

We also observed higher rotational correlation times upon an increase in salt concentration. Addition of salt is known to disrupt hydration layers around hydrophobic residues, drawing water into the bulk solution. Hydrophobic groups thereby become exposed, and tend to cluster together, a process macroscopically manifesting itself as the precipitation of proteins with exposed hydrophobic groups, known as salting-out (Doran, 2013). Considering that phase separation of FG-domains is driven by hydrophobic interactions, it thus fits coherently into the picture that on the nanoscale a higher concentration of NaCl in the buffer led to increased rotational correlation times in the protein backbone (**Figure 4-20**) and slower self-diffusion of the hydrogel (**Figure 4-19**), while on the macroscopic scale critical concentrations decreased with increasing salt concentrations (**Figure 4-23**).

The change of critical concentration with respect to salt concentration is usually characteristic of the UCST or LCST behavior of the system. UCST systems, where phase separation is dominated by electrostatic interactions, such as Ddx4 (Brady et al., 2017) or LAF-1 (Elbaum-Garfinkle et al., 2015), tend to show an increase of critical concentrations (measured at the same temperature) at higher salt. The opposite is true for LCST systems, such as ELPs (Cho et al., 2008). Interestingly, while the Tau-RNA complex also exhibits LCST behavior, addition of salt increases the critical concentration. Contrary, however, to our system, Tau-RNA phase separation is an example of complex coacervation, where upon phase separation assemblies of charged polyelectrolytes are formed, held together by electrostatic interactions (Zhang et al., 2017). Thus, while the complex

itself is hydrophobic (causing LCST behavior), the higher salt concentration disrupts the association of the components.

4.4.5 FG-domains form a selective, mobile, hydrophobic phase

The detailed investigations performed on TtMacNup98A and the prf.GLFG_{52x12} as a model sequence for FG-domains contribute to a coherent picture for the basis of cohesiveness and permselectivity. In the selective phase model (Ribbeck and Görlich, 2001), hydrophobic clusters between the FG motifs form a mesh, which is the basis of selectivity, and NTRs melt their way through this mesh upon transport, transiently opening the Phe-Phe knots.

We, however found both TtMacNup98A and the prf.GLFG_{52x12} to be highly dynamic (**Figure 4-7**), and for the latter, we determined rotational correlation times to be in the nanosecond range for the protein backbone (**Figure 4-11**), which is inconsistent with the picture of a permanently cross-linked mesh. However, based on increased rotational correlation times for hydrophobic residues (**Figure 4-11**), and NOESY spectra (**Figure 4-14**), we could identify hydrophobic contacts between Phe and other hydrophobic residues (previously also observed by Ader et al (Ader et al., 2010)), as well as between non-aromatic hydrophobic sidechains. The identification of Phe-Phe contacts was not possible, due to their overlapping signals, but their presence is highly likely. We also showed that the deletion of leucine abolishes selective behavior, but the position of leucine within the sequence is of no importance in determining cohesiveness or permselectivity (**Figure 4-15**).

This evidence indicates a liquid-like selective phase, where the sum of hydrophobicity determines the FG-domain's cohesive and selective behavior, and “knots” do not exist on a relevant timescale. Instead interactions are in the form of transient hydrophobic contacts, constantly breaking and forming again. This is also supported by the self-diffusion observed by FRAP in the FG-phase of certain cohesive FG-domains (Celetti et al., 2020) (**Figure 4-16**). The high inherent mobility of the peptide chains in the selective phase could be an advantage when it comes to ensuring rapid transport of large cargos through the NPC and of resealing the barrier after transport (Frey and Görlich, 2009). From a macroscopic point of view, transport through the selective phase formed by FG-domains would be similar to diffusion of molecules from a polar phase (e.g. water) into an

apolar phase (e.g. oil), where it is the hydrophobic character of a molecule that determines whether it can diffuse in or not. This is completely in line with the fact that NTRs are the most hydrophobic soluble proteins in eukaryotic cells, and that surface hydrophobicity is a major sorting criterion for entry into the NPCs (Ribbeck and Görlich, 2002; Frey et al., 2018).

4.5 Conclusions and perspectives

We have extensively investigated the macro- and nanoscale properties of hydrogels formed by TtMacNup98A and designer sequences that recapitulate properties of the selective phase (Ng and Görlich, 2021) (unpublished) but are amenable to study by structural methods. We found both TtMacNup98A and the prf.GLFG_{52x12} cohesive FG-domains to remain highly mobile on the nanosecond timescale even upon phase separation, and showed self-diffusion of hydrogel particles, however a dramatic decrease of dynamics was found upon removal of prolines from the sequence, demonstrating proline's essential role in maintaining disorder and mobility in FG-domains. We identified hydrophobic interactions between aromatic and aliphatic hydrophobic groups, as was previously also observed, as well as between non-aromatic hydrophobic sidechains. We showed that non-aromatic hydrophobic amino acids are crucial for cohesive behavior, however that the motif in which these residues occur does not alter the behavior of the selective phase. We found that FG-domains exhibit a lower critical solution temperature (LCST behavior), and that phase separation in these systems is driven by entropy. By mapping the behavior of the hydrogel under various temperature and salt conditions, and determining a phase diagram under different salt conditions, the behavior of the FG-hydrogel was found consistent with the dominance of hydrophobic interactions in the selective phase.

This evidence allowed us to hypothesize a model of a selective phase, where it is the overall hydrophobicity of the hydrogel that determines its cohesiveness and permselectivity, and interactions not only with NTRs, but also within the hydrogel are transient, breaking and forming constantly.

In terms of future perspectives, the measurement of Phe-Phe contacts would complete the picture of transient structures in the hydrogel. This is theoretically possible using a mixed-isotopic labeling approach in an F1 ¹³C-filtered, F2 ¹³C-edited NOESY experiment (Brady et al., 2017), however has so far been technically difficult due to the absence of pulsed field gradients in solid-state NMR probes. Furthermore, if the protofibrils indicated by secondary chemical shifts do exist in the hydrogel, the weak hydrogen bonds connecting the protein chains might be visible in an INEPT-based HNhhNH experiment (similar in concept to the one described in Chapter 1) correlating amide groups in close proximity. So far, we have used backbone dynamics measurements to determine

interaction sites in the hydrogel, but characterization of sidechain dynamics would be a more direct probe of the influence of hydrophobic interactions. An interesting, though technically somewhat difficult way to probe the influence of aromaticity vs. hydrophobicity on cohesiveness and permselectivity of the hydrogel would be to produce a prf.GLFG_{52x12} containing cyclohexyl-alanine instead of phenylalanine. This way the overall size of hydrophobic groups would be constant, but the contribution of aromaticity would be eliminated.

Finally, though interaction of NTRs with FG-domains in solution has been previously investigated in detail, addition of NTRs to the selective hydrogel phase and elucidating their binding and the changes in dynamics for both binding partners by NMR spectroscopy would be necessary to complete the picture describing rapid transport through the NPCs.

4.6 References

- Abyzov, A., Salvi, N., Schneider, R., Maurin, D., Ruigrok, R.W.H., Jensen, M.R., and Blackledge, M. (2016). Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation. *J. Am. Chem. Soc.* *138*, 6240-6251.
- Ader, C., Frey, S., Maas, W., Schmidt, H.B., Görlich, D., and Baldus, M. (2010). Amyloid-like interactions within nucleoporin FG hydrogels. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 6281-6285.
- Akey, C.W., and Radermacher, M. (1993). Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. *J. Cell Biol.* *122*, 1-19.
- Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., and Zweckstetter, M. (2017). Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.* *8*, 275.
- Ando, D., and Gopinathan, A. (2017). Cooperative Interactions between Different Classes of Disordered Proteins Play a Functional Role in the Nuclear Pore Complex of Baker's Yeast. *PLOS ONE* *12*.
- Ando, D., Zandi, R., Kim, Y.W., Colvin, M., Rexach, M., and Gopinathan, A. (2014). Nuclear pore complex protein sequences determine overall copolymer brush structure and function. *Biophys. J.* *106*, 1997-2007.
- Andrade, M.A., and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. *Nat. Genet.* *11*, 115-116.
- Aramburu, I.V., and Lemke, E.A. (2017). Floppy but not sloppy: Interaction mechanism of FG-nucleoporins and nuclear transport receptors. *Semin. Cell Dev. Biol.* *68*, 34-41.
- Balu, R., Dutta, N.K., Choudhury, N.R., Elvin, C.M., Lyons, R.E., Knott, R., and Hill, A.J. (2014). An16-resilin: an advanced multi-stimuli-responsive resilin-mimetic protein polymer. *Acta Biomater.* *10*, 4768-4777.
- Bennet-Clark, H.C., and Lucey, E.C. (1967). The jump of the flea: a study of the energetics and a model of the mechanism. *J. Exp. Biol.* *47*, 59-67.
- Bennett, A.E., Rienstra, C.M., Auger, M., Lakshmi, K.V., and Griffin, R.G. (1995). Heteronuclear Decoupling in Rotating Solids. *J. Chem. Phys.* *103*, 6951-6958.
- Bischoff, F.R., Krebber, H., Smirnova, E., Dong, W., and Ponstingl, H. (1995). Co-activation of RanGTPase and inhibition of GTP dissociation by Ran-GTP binding protein RanBP1. *EMBO J.* *14*, 705-715.
- Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van den Bosch, L., *et al.* (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* *28*, 420-435.

- Bonner, W.M. (1975). Protein migration into nuclei. I. Frog oocyte nuclei in vivo accumulate microinjected histones, allow entry to small proteins, and exclude large proteins. *J. Cell Biol.* *64*, 421-430.
- Brady, J.P., Farber, P.J., Sekhar, A., Lin, Y.H., Huang, R., Bah, A., Nott, T.J., Chan, H.S., Baldwin, A.J., Forman-Kay, J.D., *et al.* (2017). Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U. S. A.* *114*, E8194-E8203.
- Brandl, M., Weiss, M.S., Jabs, A., Suhnel, J., and Hilgenfeld, R. (2001). C-H... π -interactions in proteins. *J. Mol. Biol.* *307*, 357-377.
- Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* *324*, 1729-1732.
- Bui, K.H., von Appen, A., DiGuilio, A.L., Ori, A., Sparks, L., Mackmull, M.T., Bock, T., Hagen, W., Andres-Pons, A., Glavy, J.S., *et al.* (2013). Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* *155*, 1233-1243.
- Cautain, B., Hill, R., Pedro, N., and Link, W. (2015). Components and regulation of nuclear transport processes. *FEBS Journal* *282*, 445-462.
- Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M., Skelton, N.J. (2007). *Protein NMR Spectroscopy*, Second Edition edn (Burlington, MA: Elsevier Academic press).
- Celetti, G., Paci, G., Caria, J., VanDelinder, V., Bachand, G., and Lemke, E.A. (2020). The liquid state of FG-nucleoporins mimics permeability barrier properties of nuclear pore complexes. *J. Cell Biol.* *219*.
- Chen, X.D., Wu, X.D., Wu, H.W., and Zhang, M.J. (2020). Phase separation at the synapse. *Nat Neurosci* *23*, 301-310.
- Cho, Y., Zhang, Y., Christensen, T., Sagle, L.B., Chilkoti, A., and Cremer, P.S. (2008). Effects of Hofmeister anions on the phase transition temperature of elastin-like polypeptides. *J. Phys. Chem. B* *112*, 13765-13771.
- Chug, H., Trakhanov, S., Hulsmann, B.B., Pleiner, T., and Görlich, D. (2015). Crystal structure of the metazoan Nup62*Nup58*Nup54 nucleoporin complex. *Science* *350*, 106-110.
- Cronshaw, J.A., Krutchinsky, A.N., Zhang, W.Z., Chait, B.T., and Matunis, M.J. (2002). Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* *158*, 915-927.
- Cushman, I., Palzkill, T., and Moore, M.S. (2006). Using peptide arrays to define nuclear carrier binding sites on nucleoporins. *Methods* *39*, 329-341.
- Davis, L.I. (1995). The nuclear pore complex. *Annu Rev Biochem* *64*, 865-896.

- DeGrasse, J.A., DuBois, K.N., Devos, D., Siegel, T.N., Sali, A., Field, M.C., Rout, M.P., and Chait, B.T. (2009). Evidence for a Shared Nuclear Pore Complex Architecture That Is Conserved from the Last Common Eukaryotic Ancestor. *Mol Cell Proteomics* 8, 2119-2130.
- Denning, D.P., Patel, S.S., Uversky, V., Fink, A.L., and Rexach, M. (2003). Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. U. S. A.* 100, 2450-2455.
- Denning, D.P., and Rexach, M.F. (2007). Rapid evolution exposes the boundaries of domain structure and function in natively unfolded FG nucleoporins. *Mol Cell Proteomics* 6, 272-282.
- Dolker, N., Zachariae, U., and Grubmuller, H. (2010). Hydrophilic linkers and polar contacts affect aggregation of FG repeat peptides. *Biophys. J.* 98, 2653-2661.
- Donoughe, S., Crall, J.D., Merz, R.A., and Combes, S.A. (2011). Resilin in dragonfly and damselfly wings and its implications for wing flexibility. *J. Morphol.* 272, 1409-1421.
- Doran, P.M. (2013). *Bioprocess Engineering Principles (Second Edition)*. In, P.M. Doran, ed. (San Diego, CA: Academic Press), pp. 445-595.
- Dudas, E.F., and Bodor, A. (2019). Quantitative, Diffusion NMR Based Analytical Tool To Distinguish Folded, Disordered, and Denatured Biomolecules. *Anal. Chem.* 91, 4929-4933.
- Dumetz, A.C., Chockla, A.M., Kaler, E.W., and Lenhoff, A.M. (2008). Protein phase behavior in aqueous solutions: crystallization, liquid-liquid phase separation, gels, and aggregates. *Biophys. J.* 94, 570-583.
- Elbaum-Garfinkle, S. (2019). Matter over mind: Liquid phase separation and neurodegeneration. *J. Biol. Chem.* 294, 7160-7168.
- Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7189-7194.
- Eykyn, T.R., Philp, D.J., and Kuchel, P.W. (2003). Selective cross-polarization in solution state nuclear magnetic resonance of scalar coupled spin 1/2 and quadrupolar nuclei. *J. Chem. Phys.* 118, 6997-7004.
- Fauchere, J.L., and Pliska, V. (1983). Hydrophobic Parameters-Pi of Amino-Acid Side-Chains from the Partitioning of N-Acetyl-Amino-Acid Amides. *Eur. J. Med. Chem.* 18, 369-375.
- Finlay, D.R., and Forbes, D.J. (1990). Reconstitution of Biochemically Altered Nuclear-Pores - Transport Can Be Eliminated and Restored. *Cell* 60, 17-29.
- Flather, D., and Semler, B.L. (2015). Picornaviruses and nuclear functions: targeting a cellular compartment distinct from the replication site of a positive-strand RNA virus. *Front. Microbiol.* 6, 594.
- Flory, P.J. (1942). Thermodynamics of high polymer solutions. *J. Chem. Phys.* 10, 51-61.

- Fragasso, A., de Vries, H., van der Sluis, E., Van der Giessen, E., Onck, P.R., and Dekker, C. (2020). A Designer FG-Nup that Reconstitutes the Selective Transport Barrier of the Nuclear Pore Complex. *Biophys. J.* *118*, 341a-342a.
- Frank, H.S., and Evans, M.W. (1945). Free Volume and Entropy in Condensed Systems .3. Entropy in Binary Liquid Mixtures - Partial Molal Entropy in Dilute Solutions - Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.* *13*, 507-532.
- Freitas, N., and Cunha, C. (2009). Mechanisms and signals for the nuclear import of proteins. *Curr. Genomics* *10*, 550-557.
- Frey, S., and Görlich, D. (2007). A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* *130*, 512-523.
- Frey, S., and Görlich, D. (2009). FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties. *EMBO J.* *28*, 2554-2567.
- Frey, S., and Görlich, D. (2014a). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J. Chromatogr. A* *1337*, 95-105.
- Frey, S., and Görlich, D. (2014b). Purification of protein complexes of defined subunit stoichiometry using a set of orthogonal, tag-cleaving proteases. *J. Chromatogr. A* *1337*, 106-115.
- Frey, S., Rees, R., Schunemann, J., Ng, S.C., Funfgeld, K., Huyton, T., and Görlich, D. (2018). Surface Properties Determining Passage Rates of Proteins through Nuclear Pores. *Cell* *174*, 202-217 e209.
- Frey, S., Richter, R.P., and Görlich, D. (2006). FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* *314*, 815-817.
- Frottin, F., Schueder, F., Tiwary, S., Gupta, R., Korner, R., Schlichthaerle, T., Cox, J., Jungmann, R., Hartl, F.U., and Hipp, M.S. (2019). The nucleolus functions as a phase-separated protein quality control compartment. *Science* *365*, 342-347.
- Galkin, O., Chen, K., Nagel, R.L., Hirsch, R.E., and Vekilov, P.G. (2002). Liquid-liquid separation in solutions of normal and sickle cell hemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 8479-8483.
- Gilbert, W. (1978). Why genes in pieces? *Nature* *271*, 501.
- Godec, A., and Merzel, F. (2012). Physical origin underlying the entropy loss upon hydrophobic hydration. *J. Am. Chem. Soc.* *134*, 17574-17581.
- Gontan, C., Güttler, T., Engelen, E., Demmers, J., Fornerod, M., Grosveld, F.G., Tibboel, D., Görlich, D., Poot, R.A., and Rottier, R.J. (2009). Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors. *J. Cell Biol.* *185*, 27-34.
- Gorb, S.N. (2004). The jumping mechanism of cicada *Cercopis vulnerata* (Auchenorrhyncha, Cercopidae): skeleton-muscle organisation, frictional surfaces, and inverse-kinematic model of leg movements. *Arthropod Struct. Dev.* *33*, 201-220.

Görlich, D., and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. *Annual Review of Cellular and Developmental Biology* *15*, 607-660.

Görlich, D., Pante, N., Kutay, U., Aebi, U., and Bischoff, F.R. (1996). Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J.* *15*, 5584-5594.

Grdadolnik, J., Merzel, F., and Avbelj, F. (2017). Origin of hydrophobicity and enhanced water hydrogen bond strength near purely hydrophobic solutes. *Proc. Natl. Acad. Sci. U. S. A.* *114*, 322-327.

Grossman, E., Medalia, O., and Zwerger, M. (2012). Functional architecture of the nuclear pore complex. *Annu. Rev. Biophys.* *41*, 557-584.

Haas, F., Gorb, S., and Blickhan, R. (2000). The function of resilin in beetle wings. *Proc. Biol. Sci.* *267*, 1375-1381.

Hahn, S., and Schlenstedt, G. (2011). Importin beta-type nuclear transport receptors have distinct binding affinities for Ran-GTP. *Biochem Biophys Res Co* *406*, 383-388.

Hinshaw, J.E., Carragher, B.O., and Milligan, R.A. (1992). Architecture and design of the nuclear pore complex. *Cell* *69*, 1133-1141.

Hoh, J.H. (1998). Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. *Proteins* *32*, 223-228.

Holehouse, A.S., and Pappu, R.V. (2015). Protein polymers: Encoding phase transitions. *Nat. Mater.* *14*, 1083-1084.

Hopper, A.K., Traglia, H.M., and Dunst, R.W. (1990). The Yeast Rna1 Gene-Product Necessary for Rna Processing Is Located in the Cytosol and Apparently Excluded from the Nucleus. *J. Cell Biol.* *111*, 309-321.

Hough, L.E., Dutta, K., Sparks, S., Temel, D.B., Kamal, A., Tetenbaum-Novatt, J., Rout, M.P., and Cowburn, D. (2015). The molecular mechanism of nuclear transport revealed by atomic-scale measurements. *eLife* *4*.

Huggins, M.L. (1942). Some properties of solutions of long-chain compounds. *J Phys Chem-US* *46*, 151-158.

Hulsmann, B.B., Labokha, A.A., and Görlich, D. (2012). The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell* *150*, 738-751.

Hydrodynamic Radius Converter (2020). Hydrodynamic Radius Converter (Fluid Analytics).

Jacob, J., Duclouhier, H., and Cafiso, D.S. (1999). The role of proline and glycine in determining the backbone flexibility of a channel-forming peptide. *Biophys. J.* *76*, 1367-1376.

Kanaan, N.M., Hamel, C., Grabinski, T., and Combs, B. (2020). Liquid-liquid phase separation induces pathogenic tau conformations in vitro. *Nat. Commun.* *11*, 2809.

- Kapinos, L.E., Schoch, R.L., Wagner, R.S., Schleicher, K.D., and Lim, R.Y. (2014). Karyopherin-centric control of nuclear pores based on molecular occupancy and kinetic analysis of multivalent binding with FG nucleoporins. *Biophys. J.* *106*, 1751-1762.
- Kau, T.R., Way, J.C., and Silver, P.A. (2004). Nuclear transport and cancer: from mechanism to intervention. *Nat. Rev. Cancer* *4*, 106-117.
- Keminer, O., Siebrasse, J.P., Zerf, K., and Peters, R. (1999). Optical recording of signal-mediated protein transport through single nuclear pore complexes. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 11842-11847.
- Khan, S.N., Charlier, C., Augustyniak, R., Salvi, N., Dejean, V., Bodenhausen, G., Lequin, O., Pelupessy, P., and Ferrage, F. (2015). Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophys. J.* *109*, 988-999.
- Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W.Z., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., *et al.* (2018). Integrative structure and functional anatomy of a nuclear pore complex. *Nature* *555*, 475-482.
- Kim, Y.H., Han, M.E., and Oh, S.O. (2017). The molecular mechanism for nuclear transport and its application. *Anat. Cell. Biol.* *50*, 77-85.
- Kite, G.L. (1913). The relative permeability of the surface and interior portions of the cytoplasm of animal and plant cells - (A preliminary paper). *Biol. Bull.* *25*, 1-7.
- Konishi, H.A., and Yoshimura, S.H. (2020). Interactions between non-structured domains of FG- and non-FG-nucleoporins coordinate the ordered assembly of the nuclear pore complex in mitosis. *Faseb J.* *34*, 1532-1545.
- Krishnan, V.V., Lau, E.Y., Yamada, J., Denning, D.P., Patel, S.S., Colvin, M.E., and Rexach, M.F. (2008). Intramolecular cohesion of coils mediated by phenylalanine--glycine motifs in the natively unfolded domain of a nucleoporin. *PLOS Computational Biology* *4*, e1000145.
- Kutay, U., Bischoff, F.R., Kostka, S., Kraft, R., and Görlich, D. (1997). Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. *Cell* *90*, 1061-1071.
- Labokha, A.A., Gradmann, S., Frey, S., Hulsmann, B.B., Urlaub, H., Baldus, M., and Görlich, D. (2013). Systematic analysis of barrier-forming FG hydrogels from *Xenopus* nuclear pore complexes. *EMBO J.* *32*, 204-218.
- Lau, Y., Oamen, H.P., and Caudron, F. (2020). Protein Phase Separation during Stress Adaptation and Cellular Memory. *Cells* *9*.
- Lee, D., Hilty, C., Wider, G., and Wuthrich, K. (2006). Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.* *178*, 72-76.
- Levitt, M. (1978). Conformational preferences of amino acids in globular proteins. *Biochemistry* *17*, 4277-4285.

- Li, L., and Kiick, K.L. (2013). Resilin-based Materials for Biomedical Applications. *ACS Macro Lett.* *2*, 635-640.
- Li, N.K., Garcia Quiroz, F., Hall, C.K., Chilkoti, A., and Yingling, Y.G. (2014). Molecular description of the LCST behavior of an elastin-like polypeptide. *Biomacromolecules* *15*, 3522-3530.
- Li, P.L., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., *et al.* (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature* *483*, 336-340.
- Lim, R.Y.H., Fahrenkrog, B., Koser, J., Schwarz-Herion, K., Deng, J., and Aebi, U. (2007). Nanomechanical basis of selective gating by the nuclear pore complex. *Science* *318*, 640-643.
- Lin, D.H., and Hoelz, A. (2019). The Structure of the Nuclear Pore Complex (An Update). *Annu Rev Biochem* *88*, 725-783.
- Lin, D.H., Stuwe, T., Schilbach, S., Rundlet, E.J., Perriches, T., Mobbs, G., Fan, Y., Thierbach, K., Huber, F.M., Collins, L.N., *et al.* (2016). Architecture of the symmetric core of the nuclear pore. *Science* *352*.
- Lipowsky, G., Bischoff, F.R., Schwarzmaier, P., Kraft, R., Kostka, S., Hartmann, E., Kutay, U., and Görlich, D. (2000). Exportin 4: a mediator of a novel nuclear export pathway in higher eukaryotes. *EMBO J.* *19*, 4362-4371.
- Luchinat, E., and Banci, L. (2016). A Unique Tool for Cellular Structural Biology: In-cell NMR. *J. Biol. Chem.* *291*, 3776-3784.
- Macara, I.G. (2001). Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.* *65*, 570-594.
- MacArthur, M.W., and Thornton, J.M. (1991). Influence of proline residues on protein conformation. *J. Mol. Biol.* *218*, 397-412.
- Majumdar, A., Dogra, P., Maity, S., and Mukhopadhyay, S. (2019). Liquid-Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules. *J. Phys. Chem. Lett.* *10*, 3929-3936.
- Makde, R.D., England, J.R., Yennawar, H.P., and Tan, S. (2010). Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* *467*, 562-566.
- Makhatadze, G.I., and Privalov, P.L. (1994). Energetics of interactions of aromatic hydrocarbons with water. *Biophys. Chem.* *50*, 285-291.
- Marfori, M., Mynott, A., Ellis, J.J., Mehdi, A.M., Saunders, N.F., Curmi, P.M., Forwood, J.K., Boden, M., and Kobe, B. (2011). Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim. Biophys. Acta* *1813*, 1562-15677.

- Marsh, J.A., Singh, V.K., Jia, Z., and Forman-Kay, J.D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.* *15*, 2795-2804.
- Martin, E.W., and Mittag, T. (2018). Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* *57*, 2478-2487.
- Martinez, C.R., and Iverson, B.L. (2012). Rethinking the term "pi-stacking". *Chem. Sci.* *3*, 2191-2201.
- Michelitsch, M.D., and Weissman, J.S. (2000). A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 11910-11915.
- Milles, S., and Lemke, E.A. (2011). Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153. *Biophys. J.* *101*, 1710-1719.
- Milles, S., Mercadante, D., Aramburu, I.V., Jensen, M.R., Banterle, N., Koehler, C., Tyagi, S., Clarke, J., Shammas, S.L., Blackledge, M., *et al.* (2015). Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors. *Cell* *163*, 734-745.
- Milo, R., and Phillips, R. (2015). *Cell Biology by the Numbers* (New York: Garland Science).
- Mingot, J.M., Kostka, S., Kraft, R., Hartmann, E., and Görlich, D. (2001). Importin 13: a novel mediator of nuclear import and export. *EMBO J.* *20*, 3685-3694.
- Moon, C.P., and Fleming, K.G. (2011). Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 10174-10177.
- Moore, M.S. (1998). Ran and nuclear transport. *J. Biol. Chem.* *273*, 22857-22860.
- Moussavi-Baygi, R., and Mofrad, M.R. (2016). Rapid Brownian Motion Primes Ultrafast Reconstruction of Intrinsically Disordered Phe-Gly Repeats Inside the Nuclear Pore Complex. *Sci. Rep.* *6*, 29991.
- Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T., Michel, C.H., *et al.* (2015). ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* *88*, 678-690.
- Najbauer, E.E., Movellan, K.T., Schubeis, T., Schwarzer, T., Castiglione, K., Giller, K., Pintacuda, G., Becker, S., and Andreas, L.B. (2019). Probing Membrane Protein Insertion into Lipid Bilayers by Solid-State NMR. *ChemPhysChem* *20*, 302-310.
- Ng, S.C., and Görlich, D. (2021). Recapitulation of selective nuclear transport with a perfectly repeated 12mer GLFG peptide. Manuscript in preparation.

Ng, S.C., Güttler, T., and Görlich, D. (2021). Recapitulation of selective nuclear import and export with a perfectly repeated 12mer GLFG peptide. *Nat. Commun.* *12*, 4047.

NMR Thermometer (2020). NMR Thermometer.

Noble, W., Hanger, D.P., Miller, C.C., and Lovestone, S. (2013). The importance of tau phosphorylation for neurodegenerative diseases. *Front. Neurol.* *4*, 83.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., *et al.* (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* *57*, 936-947.

Nyqvist, I., and Dogan, J. (2019). Characterization of the dynamics and the conformational entropy in the binding between TAZ1 and CTAD-HIF-1 alpha. *Sci. Rep.* *9*.

Owen, I., and Shewmaker, F. (2019). The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins. *Int. J. Mol. Sci.* *20*.

Paine, P.L., Moore, L.C., and Horowitz, S.B. (1975). Nuclear envelope permeability. *Nature* *254*, 109-114.

Pakravan, D., Orlando, G., Bercier, V., and Van Den Bosch, L. (2020). Role and therapeutic potential of liquidliquid phase separation in amyotrophic lateral sclerosis. *J. Mol. Cell Biol.*

Pante, N., and Kann, M. (2002). Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. *Mol Biol. Cell.* *13*, 425-434.

Parrini, C., Taddei, N., Ramazzotti, M., Degl'Innocenti, D., Ramponi, G., Dobson, C.M., and Chiti, F. (2005). Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. *Structure* *13*, 1143-1151.

Patel, S.S., Belmont, B.J., Sante, J.M., and Rexach, M.F. (2007). Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* *129*, 83-96.

Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997). 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.

Peters, R. (2005). Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. *Traffic* *6*, 421-427.

Peters, R. (2009). Translocation through the nuclear pore: Kaps pave the way. *Bioessays* *31*, 466-477.

Platzer, G., Mayer, M., Beier, A., Bruschiweiler, S., Fuchs, J.E., Engelhardt, H., Geist, L., Bader, G., Schorghuber, J., Lichtenecker, R., *et al.* (2020). PI by NMR: Probing CH- π Interactions in Protein-Ligand Complexes by NMR Spectroscopy. *Angew. Chem. Int. Ed. Engl.* *59*, 14861-14868.

Posey, A.E., Holehouse, A.S., and Pappu, R.V. (2018). Phase Separation of Intrinsically Disordered Proteins. *Intrinsically Disordered Proteins* 611, 1-30.

Pratt, L.R. (2002). Molecular theory of hydrophobic effects: "She is too mean to have her name repeated.". *Annu. Rev. Phys. Chem.* 53, 409-436.

Quiroz, F.G., and Chilkoti, A. (2015). Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nat. Mater.* 14, 1164-1171.

Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., Mahato, J., Panigrahi, R., Navalkar, A., Mehra, S., *et al.* (2020). alpha-Synuclein aggregation nucleates through liquid-liquid phase separation. *Nat. Chem.* 12.

Reddin, A.L.C. (2008). *Solid State NMR Investigation of Protein-Based Biomaterials* (Vancouver: University of British Columbia).

Reichelt, R., Holzenburg, A., Buhle, E.L., Jarnik, M., Engel, A., and Aebi, U. (1990). Correlation between Structure and Mass-Distribution of the Nuclear-Pore Complex and of Distinct Pore Complex Components. *J. Cell Biol.* 110, 883-894.

Riback, J.A., Katanski, C.D., Kear-Scott, J.L., Pilipenko, E.V., Rojek, A.E., Sosnick, T.R., and Drummond, D.A. (2017). Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* 168, 1028-1040.

Ribbeck, K., and Görlich, D. (2001). Kinetic analysis of translocation through nuclear pore complexes. *EMBO J.* 20, 1320-1330.

Ribbeck, K., and Görlich, D. (2002). The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *EMBO J.* 21, 2664-2671.

Ribbeck, K., Lipowsky, G., Kent, H.M., Stewart, M., and Görlich, D. (1998). NTF2 mediates nuclear import of Ran. *EMBO J.* 17, 6587-6598.

Riddick, G., and Macara, I.G. (2007). The adapter importin-alpha provides flexible control of nuclear import at the expense of efficiency. *Mol. Syst. Biol.* 3, 118.

Rout, M.P., Aitchison, J.D., Magnasco, M.O., and Chait, B.T. (2003). Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol.* 13, 622-628.

Rout, M.P., Aitchison, J.D., Suprpto, A., Hjertaas, K., Zhao, Y.M., and Chait, B.T. (2000). The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J. Cell Biol.* 148, 635-651.

Rout, M.P., and Wentz, S.R. (1994). Pores for thought: nuclear pore complex proteins. *Trends Cell Biol.* 4, 357-365.

Salvi, N., Abyzov, A., and Blackledge, M. (2017). Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Progress in nuclear magnetic resonance spectroscopy* 102-103, 43-60.

Schmidt, H.B., and Görlich, D. (2015). Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *eLife* 4.

Schmidt, H.B., and Görlich, D. (2016). Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends Biochem. Sci.* 41, 46-61.

Schoch, R.L. (2017). Resolving conformational changes in FG nucleoporins due to multivalent Karyopherin binding (Basel: University of Basel).

Shaka, A.J., Keeler, J., and Freeman, R. (1983a). Evaluation of a New Broad-Band Decoupling Sequence - Waltz-16. *J. Magn. Reson.* 53, 313-340.

Shaka, A.J., Keeler, J., Frenkiel, T., and Freeman, R. (1983b). An Improved Sequence for Broad-Band Decoupling - Waltz-16. *J. Magn. Reson.* 52, 335-338.

Strawn, L.A., Shen, T., Shulga, N., Goldfarb, D.S., and Wentz, S.R. (2004). Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat. Cell Biol.* 6, 197-206.

SwissModel (2020). Principles of Protein Structure, Comparative Protein Modelling, and Visualisation. In *Principles of Protein Structure, Comparative Protein Modelling, and Visualisation*.

Taylor, N.O., Wei, M.T., Stone, H.A., and Brangwynne, C.P. (2019). Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophys. J.* 117, 1285-1300.

Terry, L.J., and Wentz, S.R. (2009). Flexible Gates: Dynamic Topologies and Functions for FG Nucleoporins in Nucleocytoplasmic Transport. *Eukaryot. Cell* 8, 1814-1827.

Tetenbaum-Novatt, J., Hough, L.E., Mironska, R., McKenney, A.S., and Rout, M.P. (2012). Nucleocytoplasmic transport: a role for nonspecific competition in karyopherin-nucleoporin interactions. *Mol Cell Proteomics* 11, 31-46.

Theillet, F.X., Binolfi, A., Bekei, B., Martorana, A., Rose, H.M., Stuijver, M., Verzini, S., Lorenz, D., van Rossum, M., Goldfarb, D., *et al.* (2016). Structural disorder of monomeric alpha-synuclein persists in mammalian cells. *Nature* 530, 45-50.

Tjin, M.S., Low, P.L., and Fong, E. (2014). Recombinant elastomeric protein biopolymers: progress and prospects. *Polym. J.* 46, 444-451.

Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M.S., Mading, S., Maziuk, D., Miller, Z., *et al.* (2008). Biological Magnetic Resonance Databank. *Nucleic Acids Res.* 36, D402-D408.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41, 415-427.

Wagner, R.S., Kapinos, L.E., Marshall, N.J., Stewart, M., and Lim, R.Y.H. (2015). Promiscuous binding of Karyopherin β 1 modulates FG nucleoporin barrier function and expedites NTF2 transport kinetics. *Biophys. J.* *108*, 918-927.

Wall, K.P., and Hough, L.E. (2018). In-Cell NMR within Budding Yeast Reveals Cytoplasmic Masking of Hydrophobic Residues of FG Repeats. *Biophys. J.* *115*, 1690-1695.

Watson, M.L. (1955). The Nuclear Envelope - Its Structure and Relation to Cytoplasmic Membranes. *J Biophys Biochem Cy* *1*, 257-270.

Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K.M., Bennett, R.E., Dujardin, S., Laskowski, P.R., MacKenzie, D., Kamath, T., Commins, C., *et al.* (2018). Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.* *37*.

Wei, M.T., Elbaum-Garfinkle, S., Holehouse, A.S., Chen, C.C., Feric, M., Arnold, C.B., Priestley, R.D., Pappu, R.V., and Brangwynne, C.P. (2017). Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.* *9*, 1118-1125.

Wente, S.R., and Rout, M.P. (2010). The nuclear pore complex and nuclear transport. *Cold Spring Harb. Perspect. Biol.* *2*, a000562.

Wheeler, R.J. (2020). Therapeutics-how to treat phase separation-associated diseases. *Emerg. Top. Life Sci.* *4*, 307-318.

Widder, K., MacEwan, S.R., Garanger, E., Nunez, V., Lecommandoux, S., Chilkoti, A., and Hinderberger, D. (2017). Characterisation of hydration and nanophase separation during the temperature response in hydrophobic/hydrophilic elastin-like polypeptide (ELP) diblock copolymers. *Soft Matter* *13*, 1816-1822.

Wimley, W.C., and White, S.H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* *3*, 842-848.

Wischnitzer, S. (1958). An electron microscope study of the nuclear envelope of amphibian oocytes. *J. Ultrastruct. Res.* *1*, 201-222.

Wolfram|Alpha (2018). Wolfram Alpha LLC.

Wong, L.E., Bhatt, A., Erdmann, P.S., Hou, Z., Maier, J., Pirkuliyeva, S., Engelke, M., Becker, S., Plitzko, J., Wienands, J., *et al.* (2020). Tripartite phase separation of two signal effectors with vesicles priming B cell responsiveness. *Nat. Commun.* *11*, 848.

www.cryst.bbk.ac.uk (1996). Principles of Protein Structure - Proline (1996 August 12: Birkbeck College).

Yamada, J., Phillips, J.L., Patel, S., Goldfien, G., Calestagne-Morelli, A., Huang, H., Reza, R., Acheson, J., Krishnan, V.V., Newsam, S., *et al.* (2010). A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol Cell Proteomics* *9*, 2205-2224.

Yang, W. (2011). 'Natively unfolded' nucleoporins in nucleocytoplasmic transport: clustered or evenly distributed? *Nucleus* *2*, 10-16.

Zhang, X., Lin, Y., Eschmann, N.A., Zhou, H., Rauch, J.N., Hernandez, I., Guzman, E., Kosik, K.S., and Han, S. (2017). RNA stores tau reversibly in complex coacervates. *PLOS Biol.* *15*, e2002183.

Zhao, B., Li, N.K., Yingling, Y.G., and Hall, C.K. (2016). LCST Behavior is Manifested in a Single Molecule: Elastin-Like polypeptide (VPGVG)_n. *Biomacromolecules* *17*, 111-118.

Zhao, G., and London, E. (2006). An amino acid "transmembrane tendency" scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. *Protein Sci.* *15*, 1987-2001.

Zhou, D.H., and Rienstra, C.M. (2008). High-performance solvent suppression for proton detected solid-state NMR. *J. Magn. Reson.* *192*, 167-172.

5 Appendix

5.1 Chapter 1: Investigation of VDAC structure, gating, and interactions in a lipid bilayer

5.1.1 Assignment of E73VC127AC232S-hVDAC1 in a DMPC lipid bilayer

Residue	Chemical shift (ppm)					
	N	CA	H	CB	CO	HA
P5	-	60.90	-	30.98	174.0	-
T6	105.3	59.96	7.232	66.43	175.8	3.526
Y7	121.7	61.77	7.638	38.78	178.4	2.905
A8	118.3	53.36	8.235	17.58	177.5	3.438
D9	114.5	53.47	7.047	39.82	177.2	3.894
L10	124.6	55.77	7.222	-	178.1	3.381
G11	113.0	44.44	8.664	-	175.2	2.604
K12	121.7	59.29	7.110	32.13	176.5	-
S13	114.8	62.67	8.064	59.25	176.2	3.341
A14	121.2	54.79	8.084	17.19	178.8	3.595
R15	116.2	59.67	7.018	29.16	180.2	-
D16	121.1	56.75	8.924	38.82	179.2	3.901
V17	119.9	66.23	7.330	30.20	178.7	3.066
F18	115.9	62.29	7.429	40.25	176.3	3.520
T19	107.6	63.35	7.836	70.72	176.2	4.127
K20	123.7	57.33	8.164	30.87	176.6	-
G21	107.2	44.20	8.025	-	171.6	2.971
Y22	113.0	55.52	6.476	39.39	175.1	3.854
G23	110.5	45.25	7.428	-	174.3	2.917
F24	119.0	55.90	6.911	45.60	177.0	3.594
G25	109.5	45.25	7.835	-	172.5	3.760
L26	117.4	52.60	7.165	-	176.2	5.108
I27	121.3	59.51	8.508	-	175.0	4.765
K28	125.8	54.63	8.556	36.25	173.1	-
L29	124.3	52.34	8.606	45.22	-	4.304
T42	-	60.58	-	71.31	172.7	-
S43	120.0	56.05	8.742	64.78	173.0	-
S44	120.2	56.40	8.536	65.37	174.6	4.161
G45	109.3	44.22	8.872	-	173.3	-
S46	116.1	56.21	8.794	65.58	171.8	-

Chemical shift (ppm)						
Residue	N	CA	H	CB	CO	HA
A47	123.7	49.40	9.005	21.38	172.0	-
V54	124.5	60.27	8.855	-	-	-
T55	119.1	60.38	8.589	-	-	-
G56	112.3	45.46	9.300	-	-	-
E59	123.6	54.17	8.706	33.01	-	-
T60	120.1	60.44	8.874	70.52	175.2	4.752
K61	126.2	53.89	8.731	36.06	-	-
Y62	123.5	57.22	9.097	38.21	173.2	4.496
R63	126.4	54.13	7.430	-	173.7	-
W64	127.2	53.87	9.021	-	177.0	4.767
T65	119.4	65.88	8.196	67.95	177.5	-
E66	122.1	58.69	9.344	28.15	175.9	3.494
Y67	112.9	55.67	6.477	39.08	173.6	4.112
G68	104.7	46.63	7.398	-	175.2	3.854
L69	119.6	53.53	7.119	45.69	176.4	5.144
T70	120.0	61.92	9.004	70.38	172.5	-
F71	126.4	55.37	9.104	42.18	173.7	5.097
T72	120.7	60.95	8.953	70.36	173.2	-
V73	125.9	60.40	8.844	33.27	173.3	4.557
K74	125.0	54.29	8.885	35.79	174.1	-
W75	-	52.73	-	31.12	172.5	-
N76	118.0	59.56	8.722	-	-	-
G82	-	-	-	-	172.2	-
T83	113.1	59.40	8.613	71.50	172.6	-
E84	124.2	54.23	8.690	33.05	175.0	-
I85	126.5	59.76	9.136	40.86	174.2	4.732
T86	122.7	60.39	8.797	71.65	172.8	-
V87	124.3	59.36	9.213	33.64	174.1	4.712
E88	126.5	54.83	8.902	32.80	175.5	4.780
D89	117.1	59.31	8.593	38.26	-	4.601
A92	115.8	50.68	7.479	21.10	176.3	-
R93	123.1	57.32	8.825	28.22	177.4	-
G94	115.0	44.82	9.422	-	173.8	4.126
L95	122.2	53.77	7.658	43.91	175.3	5.041
K96	131.0	55.09	9.350	35.49	173.9	-
L97	126.0	53.56	8.705	44.71	174.6	4.821
T98	119.0	61.09	9.012	71.84	172.7	-
F99	126.3	55.73	9.317	41.05	172.8	4.917
D100	128.7	52.03	8.492	43.91	174.3	4.869
S101	117.3	59.30	8.515	-	-	-
N111	122.5	53.12	8.357	-	173.8	5.112
A112	122.6	51.34	8.366	21.51	174.8	4.758
K113	119.2	54.47	8.847	36.12	175.2	-

Chemical shift (ppm)						
Residue	N	CA	H	CB	CO	HA
I114	120.1	59.05	8.618	40.36	174.7	4.711
K115	125.7	54.42	9.139	33.85	175.6	-
T116	115.2	59.38	8.636	71.23	174.3	-
G117	111.9	44.62	8.949	-	170.9	3.457
Y118	120.6	56.76	8.822	41.04	171.5	5.021
K119	124.0	52.08	6.784	34.14	171.7	-
R120	114.7	53.99	8.055	27.60	174.1	-
E121	118.6	58.87	8.341	28.10	178.3	3.045
H122	114.0	58.29	8.198	32.49	173.5	-
I123	112.9	58.67	7.604	42.40	173.3	5.333
N124	120.4	53.16	8.460	40.50	174.7	4.309
L125	127.1	53.10	9.208	45.11	176.4	4.076
G126	110.3	45.46	9.080	-	170.9	3.701
A127	120.1	51.29	8.199	20.65	175.1	-
D128	122.9	52.11	8.863	42.26	174.7	5.057
M129	123.6	54.22	9.390	34.90	172.9	4.559
D130	124.8	52.55	8.467	42.54	175.6	4.565
I133	-	59.68	-	46.47	180.2	-
S137	124.5	56.85	8.891	64.86	172.7	-
I138	122.1	59.35	8.949	40.58	173.5	-
R139	128.0	53.58	8.756	31.50	176.1	-
G140	113.8	43.63	8.702	-	170.8	2.583
A141	123.2	50.78	8.970	22.68	174.8	4.909
L142	120.7	53.55	8.802	45.65	173.9	4.793
V143	124.3	60.82	8.604	34.12	174.0	4.676
L144	125.2	51.91	9.070	43.52	175.8	4.719
G145	106.4	43.59	8.631	-	172.8	3.368
Y146	121.8	58.10	8.282	-	177.2	4.385
W149	120.6	57.28	8.476	29.70	177.3	-
L150	123.4	53.16	9.051	46.18	175.1	5.090
A151	121.9	50.26	8.895	20.98	175.1	5.056
G152	107.6	44.58	9.589	-	171.1	3.198
Y153	122.1	55.05	9.025	42.27	172.0	-
Q154	126.8	52.75	8.201	31.12	172.7	-
M155	121.8	52.70	8.720	35.12	-	4.734
N168	123.9	51.13	8.985	41.22	172.1	4.926
F169	117.3	56.10	8.429	43.50	173.5	5.074
A170	121.5	50.80	8.742	22.91	176.1	4.946
V171	117.5	60.20	8.051	34.62	175.2	4.798
G172	113.3	45.26	9.473	-	170.0	4.032

Chemical shift (ppm)						
Residue	N	CA	H	CB	CO	HA
Y173	118.8	56.70	8.404	41.79	173.6	4.269
K174	127.8	55.31	8.032	35.14	173.8	-
T175	117.2	59.24	8.393	71.10	174.4	-
D176	117.9	57.07	8.477	39.85	177.2	3.968
F178	117.3	56.83	8.512	-	173.4	4.324
Q179	121.7	53.63	8.695	31.52	173.9	-
L180	126.5	53.01	9.208	44.76	174.4	-
H181	127.4	56.00	9.040	33.44	173.1	-
T182	117.6	59.61	7.463	71.55	171.5	4.179
N183	116.0	52.68	8.532	42.54	172.4	5.059
V184	118.0	59.52	8.704	32.68	175.2	4.643
N185	127.9	51.05	9.127	37.93	175.2	4.959
D186	126.5	54.83	8.962	-	175.9	4.373
G187	105.3	46.01	8.836	-	173.0	3.714
T188	108.5	62.78	7.513	71.32	172.7	-
E189	121.8	54.60	7.665	30.05	174.1	4.494
F190	125.0	56.12	8.699	39.97	175.3	4.329
G191	109.3	43.95	7.894	-	171.3	3.333
G192	105.4	45.27	8.247	-	171.1	3.429
S193	114.1	55.80	9.729	66.63	173.4	-
I194	117.4	59.29	8.557	41.24	173.5	-
Y195	127.0	55.11	8.589	41.95	173.3	5.210
Q196	123.9	54.15	8.155	31.73	173.5	-
K197	126.0	54.78	8.116	30.69	175.7	-
K200	-	58.81	-	31.33	177.1	-
K201	115.6	56.79	8.231	33.86	175.1	-
L202	121.7	54.05	7.470	44.10	174.0	-
E203	125.2	53.79	8.735	34.09	174.9	4.766
T204	110.7	58.90	8.619	71.95	172.4	-
A205	120.1	50.72	8.879	23.04	174.5	5.280
V206	118.0	59.42	8.612	35.20	174.2	4.974
N207	123.4	51.71	8.990	42.29	173.1	5.144
L208	120.7	53.39	8.971	46.36	176.6	5.105
A209	122.6	52.14	9.114	22.31	175.7	4.820
W210	117.2	57.80	8.523	30.49	173.0	4.664
T211	115.3	60.18	9.414	71.38	175.3	-
A212	127.1	53.11	9.237	18.31	178.5	-
G213	109.1	45.10	8.874	-	172.9	3.780
N214	118.5	51.63	7.632	40.35	175.2	4.660
S215	119.3	59.31	8.651	63.24	174.4	3.865
N216	118.3	53.23	8.185	38.95	176.7	4.341
T217	118.4	62.33	8.414	71.12	173.3	-
R218	126.5	54.52	9.355	31.08	175.2	-

Chemical shift (ppm)						
Residue	N	CA	H	CB	CO	HA
F219	116.0	55.75	7.703	42.76	172.8	5.568
G220	106.6	44.27	8.912	-	171.8	3.865
I221	120.1	60.40	8.949	41.21	172.9	4.736
A222	126.0	50.17	8.901	21.57	175.0	4.758
A223	118.0	50.02	8.970	22.90	176.6	4.844
K224	122.2	55.92	8.722	35.06	173.9	-
Y225	130.2	55.15	9.759	41.30	173.4	5.181
Q226	130.8	54.10	8.916	26.82	173.6	3.709
I227	130.0	64.61	7.778	38.66	175.8	2.970
D228	117.2	52.42	8.368	39.89	173.7	4.192
P229	121.8	65.65	-	30.91	176.1	-
D230	113.5	53.65	8.355	42.61	175.0	4.368
A231	123.7	51.58	7.604	21.88	175.2	-
S232	113.4	56.33	8.658	65.80	170.8	4.800
F233	120.6	55.44	8.953	43.54	173.0	5.395
S234	123.9	57.18	9.157	64.80	170.5	4.771
A235	121.9	50.50	7.903	22.92	175.3	5.138
K236	114.1	54.25	8.928	35.39	174.1	-
V237	113.4	58.31	8.966	34.60	174.6	5.168
N238	118.5	51.06	7.718	40.84	175.2	5.038
N239	114.3	52.77	8.148	-	175.4	3.722
S240	124.0	60.80	8.826	-	174.7	4.703
L242	-	-	-	-	176.5	-
I243	122.9	59.27	8.587	40.50	174.3	4.519
G244	113.6	43.84	9.276	-	171.2	2.599
L245	121.2	53.28	9.133	46.10	176.1	5.162
G246	110.5	44.46	9.064	-	170.5	2.865
Y247	124.4	55.69	8.961	42.68	174.1	5.151
T248	124.2	60.54	8.363	71.27	171.7	-
Q249	123.7	52.75	8.792	-	175.3	-
I255	-	-	-	-	174.4	-
K256	126.6	54.64	8.605	35.44	174.3	-
L257	127.6	52.74	8.676	45.48	174.4	4.924
T258	122.3	61.01	9.455	70.79	173.5	-
L259	128.5	52.62	9.267	44.76	174.5	4.896
S260	113.5	57.06	8.645	66.12	172.7	-
A261	119.5	51.14	9.365	22.34	173.2	5.069
L262	124.3	53.21	8.567	42.15	173.6	-
A270	-	-	-	-	175.0	-
L275	-	-	-	-	177.8	-
G276	110.1	45.08	8.903	-	171.1	3.519

Chemical shift (ppm)						
Residue	N	CA	H	CB	CO	HA
L277	120.6	53.99	8.129	45.75	174.9	4.986
G278	115.6	44.19	9.516	-	171.5	3.198
L279	125.0	52.86	8.605	45.00	174.5	4.899
E280	126.4	54.07	8.793	33.06	-	-

5.1.2 Intensity values for VDAC mutants and ligand bound samples

Intensity values in *italics* are uncertain due to signal overlap and were not used for evaluation.

	E73V/C127A/ /C232S	A2C/E73V/ C127A/C232S- MTSL	D2O exchange	G21V/G23V/ E73V/C127A/ /C232S	E73V/C127A/ C232S + Cholesterol	E73V/C127A/ C232S + G3139
T6	1728618	912648	98460	1545109	1606695	1463999
Y7	2056904	889265	502424	1522337	1853972	1788179
A8	<i>3031575</i>	<i>1362676</i>	<i>426668</i>	<i>2075694</i>	<i>3032107</i>	<i>2380931</i>
D9	1858389	675716	270611	1425905	1450710	1273483
L10	1005689	505736	531869	991609	920863	493012
G11	511170	85130	503210	510626	428383	41142
K12	1857165	802030	-211935	1288597	1740004	1322622
S13	1739415	733786	-12848	1417932	1477489	1225599
A14	1865723	1311011	1421552	1500620	1665015	1732962
R15	1714673	830774	-81155	1360179	1690004	1097632
D16	1806060	1262833	1296144	778446	1903261	1955397
V17	1748672	979639	-456740	273811	2014654	1093143
F18	1084028	591444	218621	-6762	980579	786233
T19	1185397	621070	405992	-276768	1173806	-118326
K20	988685	737856	661332	-183807	903931	187576
G21	1092294	517535	857600	59814	1219957	852415
Y22	<i>1381828</i>	<i>585982</i>	<i>78463</i>	<i>101037</i>	<i>1304766</i>	<i>971018</i>
G23	1070568	455690	377925	-129482	1163374	389402
F24	1008118	522094	1023729	393617	1291685	690503
G25	457329	639626	-410044	-77094	1181095	64514
L26	697732	483074	-458142	-32633	963769	462739
I27	459175	666430	1256843	298664	780915	477516
K28	1498081	1774442	1445614	327791	2047355	1150418
L29	<i>1178078</i>	<i>823289</i>	<i>-285127</i>	<i>152917</i>	<i>1410279</i>	<i>978167</i>
S43	507658	622484	32181	974401	845847	557620
S44	754693	990560	673623	1101872	875469	594699
G45	575214	646725	-292607	821852	586052	452569
S46	662448	674385	17534	713569	609871	406433
A47	360459	80620	357958	141430	643804	362471
V54	600520	707322	1328132	879188	1055481	738962
T55	458539	469617	199616	691474	669393	500657
G56	515642	872185	-73820	1258228	552540	477813
E59	<i>931066</i>	<i>1226456</i>	<i>688003</i>	<i>1243650</i>	<i>951009</i>	<i>540177</i>
T60	<i>1069653</i>	<i>633976</i>	<i>1723616</i>	<i>1078349</i>	<i>1049918</i>	<i>670207</i>
K61	<i>1240667</i>	<i>1393866</i>	<i>42486</i>	<i>2055185</i>	<i>1910723</i>	<i>1265072</i>
Y62	837835	978795	1831369	733529	1233302	989123

	E73V/C127A/ /C232S	A2C/E73V/ C127A/C232S- MTSL	D2O exchange	G21V/G23V/ E73V/C127A/ /C232S	E73V/C127A/ C232S + Cholesterol	E73V/C127A/ C232S + G3139
R63	751501	794205	66043	953275	819423	842836
W64	806384	991407	-118566	1242478	1102776	1167107
T65	898526	367710	-88244	877246	740101	728013
E66	1301467	872964	-64665	1315823	1150322	1129361
Y67	<i>1222961</i>	<i>1129920</i>	<i>51221</i>	<i>841635</i>	<i>1314239</i>	<i>844878</i>
G68	1205421	1018878	162690	1208633	1641797	1368408
L69	994583	735525	235556	993129	1254071	781229
T70	870753	1060710	36845	1219507	1268427	472798
F71	<i>818758</i>	<i>1252921</i>	<i>1268638</i>	<i>1158431</i>	<i>1424900</i>	<i>1424163</i>
T72	491789	917472	162029	956208	1940610	532979
V73	642635	460169	1834174	804639	792491	827827
K74	<i>1182653</i>	<i>1697555</i>	<i>1260512</i>	<i>1580940</i>	<i>1699029</i>	<i>929796</i>
T83	801119	894767	243260	981696	1175259	986327
E84	2123127	2179479	2278559	1973644	2620796	1496106
I85	891836	1030389	1384672	1256010	1424894	999467
T86	<i>1630464</i>	<i>1555751</i>	<i>1689595</i>	<i>1613905</i>	<i>1595279</i>	<i>1185009</i>
V87	1282820	1009465	1836027	1290743	1622143	1126576
E88	<i>2035220</i>	<i>1468188</i>	<i>1491999</i>	<i>1416635</i>	<i>1857509</i>	<i>1731071</i>
D89	<i>769011</i>	<i>677835</i>	<i>1416956</i>	<i>966734</i>	<i>1202098</i>	<i>797134</i>
A92	434559	406976	527139	664452	813477	654608
R93	1615206	1085300	721963	1231150	1686074	1654526
G94	1898831	887970	460096	1110003	1777913	1767769
L95	1351086	760795	47771	1335242	1955104	1426961
K96	<i>1275967</i>	<i>997333</i>	<i>717264</i>	<i>801758</i>	<i>1826613</i>	<i>1261166</i>
L97	<i>1570553</i>	<i>1346274</i>	<i>-220491</i>	<i>1497191</i>	<i>2426505</i>	<i>1595992</i>
T98	1997446	1546964	1953648	1931471	1818568	1806894
F99	1246826	1155499	2744775	1434258	1775505	1609861
D100	1192472	1183692	836616	1333433	1347596	954550
S101	887228	<i>553560</i>	<i>804144</i>	<i>1114212</i>	<i>1532869</i>	<i>848125</i>
N111	730596	822183	1148474	1146900	959477	628457
A112	446825	670650	193978	790563	952891	649770
K113	1544924	1668085	-181175	1720399	2130140	1215956
I114	1096141	1091841	2546151	1054110	1784162	994595
K115	1767799	1827815	1479686	1773567	1965046	1684380
T116	1046195	518473	262061	953935	1819060	728093
G117	1506769	590110	439465	1308547	2183311	1256588
Y118	1361620	618475	1511925	921309	1650212	1955397
K119	1743121	448349	741537	904183	1914393	1272304
R120	1985574	90301	2463863	885329	2062219	1619741
E121	2022818	-113941	292113	1090844	1703406	1707478
H122	984670	51111	-784604	-337041	692310	958299
I123	934066	-122075	2826762	226333	1048880	846120

	E73V/C127A/ /C232S	A2C/E73V/ C127A/C232S- MTSL	D2O exchange	G21V/G23V/ E73V/C127A/ /C232S	E73V/C127A/ C232S + Cholesterol	E73V/C127A/ C232S + G3139
N124	1467880	158207	2391819	926495	1730380	1645986
<i>L125</i>	<i>2519821</i>	<i>1199275</i>	<i>3657369</i>	<i>1968326</i>	<i>3062771</i>	<i>1946699</i>
G126	992023	501104	1420741	775033	1434387	1199463
A127	1135286	466491	670428	772573	1943799	632203
D128	648101	176988	636780	1150691	1076210	639564
<i>M129</i>	<i>581315</i>	<i>149011</i>	<i>267647</i>	<i>117556</i>	<i>1264706</i>	<i>480282</i>
<i>D130</i>	<i>1295281</i>	<i>1585016</i>	<i>20552</i>	<i>1315539</i>	<i>1589638</i>	<i>1063427</i>
S137	661535	580112	-78292	1342354	630897	587646
I138	629049	818509	46416	946321	1220331	712121
R139	1244495	1340393	1644626	1224060	1776426	893936
G140	386992	454108	-381496	686282	775565	330221
A141	1756317	1375145	683035	1991913	2470935	1822173
<i>L142</i>	<i>828667</i>	<i>512985</i>	<i>2645367</i>	<i>909438</i>	<i>1768940</i>	<i>1103791</i>
V143	1613445	743440	1999744	1066832	1472884	1974804
L144	1030065	1252758	2824391	983562	1315228	1408425
G145	1004295	-41856	2092336	857378	1297850	1125717
Y146	701125	359706	52799	655358	958354	778656
W149	502704	317365	447146	159096	711785	531286
<i>L150</i>	<i>1191733</i>	<i>667937</i>	<i>2899454</i>	<i>1897659</i>	<i>1448438</i>	<i>1256180</i>
A151	839753	631318	627896	885141	1737199	1094576
G152	782354	216033	621189	432531	1371182	900189
Y153	527109	528311	1234942	883787	874360	595477
Q154	819889	567067	1487218	600935	1104520	836265
<i>M155</i>	<i>540109</i>	<i>697429</i>	<i>548922</i>	<i>502530</i>	<i>1063144</i>	<i>678829</i>
N168	775933	670629	507670	801679	1391641	731862
F169	861826	602446	296182	963612	1423868	834831
A170	1783986	812182	2032557	1614602	1849339	1652996
V171	1070975	180231	3184056	727506	1790149	1405099
G172	1444904	589309	1871797	727506	1725994	1564048
Y173	1351035	820281	2392731	1243425	1828166	1193116
K174	814799	-70331	145260	213682	959723	933511
T175	895458	186273	241746	133319	1003391	859876
D176	1572142	970234	-645	1861471	1959519	1259991
F178	1529438	1545236	-153304	1410318	2339235	944993
Q179	1414504	1297519	1745471	1874750	1486167	558765
<i>L180</i>	<i>909511</i>	<i>1294254</i>	<i>1718897</i>	<i>872935</i>	<i>1705580</i>	<i>1198114</i>
H181	724595	631237	1104893	802130	1330171	940144
T182	685071	752509	-49758	924161	1279744	408911
N183	1044334	291336	808010	706795	1694082	774985
<i>V184</i>	<i>2057453</i>	<i>1080661</i>	<i>1599174</i>	<i>1685841</i>	<i>2830726</i>	<i>1099023</i>

	E73V/C127A/ /C232S	A2C/E73V/ C127A/C232S- MTSL	D2O exchange	G21V/G23V/ E73V/C127A/ /C232S	E73V/C127A/ C232S + Cholesterol	E73V/C127A/ C232S + G3139
N185	1252151	775119	-321607	801629	1323645	1131686
<i>D186</i>	<i>1889846</i>	<i>1468188</i>	<i>1471599</i>	<i>1618222</i>	<i>2353742</i>	<i>1782166</i>
G187	919634	496513	308106	771710	1122150	741912
T188	691540	545211	240916	603244	628035	388899
E189	1155722	736421	177126	684424	1285210	973677
<i>F190</i>	<i>1108611</i>	<i>517404</i>	<i>1166208</i>	<i>773821</i>	<i>1373267</i>	<i>1161756</i>
G191	468680	362033	275051	599641	1480519	733683
G192	1228549	441457	559366	563940	1788122	670908
S193	1393801	750164	1713199	1093258	1345673	1226892
<i>I194</i>	<i>920289</i>	<i>746009</i>	<i>356584</i>	<i>1091369</i>	<i>1669843</i>	<i>776991</i>
Y195	1373039	1047083	2018778	1212652	1667584	1444426
Q196	887616	430473	1588758	777593	649678	1186371
K197	793485	123614	-204561	261431	1097900	1167425
K201	853696	672974	551106	802637	882729	703776
L202	935466	871817	190643	1063166	1159818	906587
E203	1710086	2232746	1862814	2046428	3018832	2651072
T204	1275814	69209	1679426	654969	1467058	1357915
A205	1542546	556274	1571334	879456	2236253	1588206
<i>V206</i>	<i>1758701</i>	<i>1071627</i>	<i>2113348</i>	<i>1563232</i>	<i>2242829</i>	<i>1383158</i>
N207	1439097	625946	2011577	653219	1701419	792005
<i>L208</i>	<i>1326412</i>	<i>774533</i>	<i>2492344</i>	<i>137404</i>	<i>1986710</i>	<i>1186401</i>
A209	2219810	1350343	2004323	1541429	2481609	1758505
W210	1725940	439911	2111385	948677	1903345	1350753
T211	1604484	390248	203017	278042	1537206	922364
<i>A212</i>	<i>2519821</i>	<i>1294254</i>	<i>3657369</i>	<i>1957230</i>	<i>3062771</i>	<i>1188608</i>
G213	616855	349873	494420	278042	961071	427180
N214	1372878	477155	344099	827966	1549637	769263
S215	755085	616253	238458	1371592	498372	811246
<i>N216</i>	<i>2327694</i>	<i>1362676</i>	<i>298298</i>	<i>1983330</i>	<i>3032107</i>	<i>2418691</i>
T217	1082145	78338	477876	854339	1386675	976113
R218	2018590	698678	-335960	1165524	2185283	1361423
F219	732042	44469	-99426	238477	1164190	635899
G220	1572436	519052	1997039	1035040	1620105	1071178
<i>I221</i>	<i>1312347</i>	<i>560371</i>	<i>2354057</i>	<i>1078349</i>	<i>1940610</i>	<i>1055644</i>
A222	1710246	1029818	1531491	1103519	1981266	1870097
A223	1766568	1393953	2913958	1184497	2400123	1999666
K224	1875252	1630293	1907697	1472857	2059346	2047763
Y225	1077081	599319	2563015	523465	1604546	1272803
Q226	2206599	1396449	-269795	1015386	2154073	2130389
<i>I227</i>	<i>1046787</i>	<i>702906</i>	<i>1606389</i>	<i>945855</i>	<i>1541811</i>	<i>1289252</i>
D228	856436	700073	-95865	901831	1330515	955748
D230	1070979	1038154	-638067	1158462	1269587	905026

	E73V/C127A/ /C232S	A2C/E73V/ C127A/C232S- MTSL	D2O exchange	G21V/G23V/ E73V/C127A/ /C232S	E73V/C127A/ C232S + Cholesterol	E73V/C127A/ C232S + G3139
A231	1034226	822436	843241	890997	1334183	1212238
S232	1668791	1393978	-36068	1513249	1627933	1281649
F233	1183044	1080117	2009524	1163209	1907112	1611434
<i>S234</i>	<i>1506371</i>	<i>978795</i>	<i>2619554</i>	<i>900855</i>	<i>1800095</i>	<i>1330907</i>
A235	1496972	958109	1640828	1149407	2294796	1718894
K236	1749004	994644	1949540	882414	2230285	1782453
V237	962155	410630	1768472	613761	1269264	790661
N238	1085672	137984	-171875	177426	1311536	188511
N239	688162	90770	306994	-102899	739370	217995
S240	676215	934049	255358	1275647	1123777	392408
I243	762228	284644	1252798	1142145	1294171	184102
G244	1410203	811740	1875725	1007151	1579730	1183353
L245	1302841	588066	2391270	731903	1796180	1403012
G246	2062425	1767150	2553725	687846	2158280	1867663
Y247	1322882	773998	2507757	771245	1722437	1639231
T248	1500109	1065302	1811024	225000	1786820	1632024
<i>Q249</i>	<i>556100</i>	<i>620880</i>	<i>2213151</i>	<i>758298</i>	<i>1013555</i>	<i>698559</i>
<i>K256</i>	<i>1004297</i>	<i>1083227</i>	<i>54458</i>	<i>1048440</i>	<i>2047355</i>	<i>1118792</i>
L257	730951	758688	2277489	278680	1126809	899470
T258	917640	733460	376149	809998	1016223	1055449
L259	613420	593698	1843353	32974	924077	466856
<i>S260</i>	<i>919129</i>	<i>771064</i>	<i>1547251</i>	<i>618115</i>	<i>1247474</i>	<i>1120453</i>
A261	1278854	893791	1637704	121012	1756500	1307266
<i>L262</i>	<i>789775</i>	<i>971553</i>	<i>2733799</i>	<i>1247931</i>	<i>1058934</i>	<i>731032</i>
G276	1001208	738394	1493035	351355	1577916	777609
L277	924817	671246	189229	-2254	1373251	1008724
G278	700821	485778	657987	65252	767882	829857
L279	986047	1059631	2467253	434799	1944130	903697
<i>E280</i>	<i>877100</i>	<i>1166276</i>	<i>608813</i>	<i>1675126</i>	<i>1916244</i>	<i>1321057</i>

5.2 Chapter 2: Probing insertion of membrane proteins into lipid bilayers

5.2.1 NMR pulse sequences

5.2.1.1 HhNH sequence

```
;written by E. Najbauer, K. Tekwani Movellan, L. Andreas
;Avance II+ version
;p1 : 13C 90 pulse for 180 in t1
;p111 : 13C power for 180
;p3 : proton 90 at power level p112
;p15 : contact time at p11 (f1) and p12 (f2)
;p11 : X power level during contact
;sp0 : proton power level during contact
;p12 : =120dB, not used
;p17 : 15N hard pulse power level
;p112 : decoupling power level (if not p113)
;p113 : water suppression power level
;p120 : 15N CP power
;d1 : recycle delay
;cnst21 : on resonance, usually = 0
;pcpd2 : pulse length in decoupling sequence
;pcpd4 : pulse length for X dec during acq
;cpdprg2 : cw, tppm (at p112), or lgs, cwlg. cwlg. (LG-decoupling)
;spnam0 : use e.g. ramp.100 for variable amplitude CP
;here p113 is used instead of p112)
;zgoptns : -Dfslg, -Dlacq, or blank
;d5 : NOE mixing time
;p30 : total water suppression time

;$COMMENT=basic cp experiment, arbitrary contact and decoupling schemes
;$CLASS=Solids
;$DIM=1D
;$TYPE=cross polarisation
;$SUBTYPE=simple 1D
;$OWNER=Bruker
prosol relations=<solids_cp>

#include <Avancesolids.incl>

"acqt0=1u*cnst11"
"in0=inf1"
"in30=inf1"
"in10=inf2"
"d10=0"
"d0=0"
aqseq 312

define delay mix
"mix=d5"
```

```

1 ze
  mix
2 d1 do:f2 do:f3
#include <p15_prot.incl>
#include <aq_prot.incl>
  (p1 p11 ph9):f1

  1u cpds2:f2 cpds3:f3
  d10
  (p1 p11 ph3):f1 (1u do):f2 (1u do):f3
5 d5
  (p1 p11 ph0):f1
  (p15 pl20 ph2):f2 (p15:sp0 ph10):f1
  1u cpds1:f1
if "p3*2 > d0" goto RAWEVOL
  (center (d0) (p3*2 ph0 pl11):f3)
if "p3*2 <= d0" goto DECOFF

RAWEVOL, 1u
  d0
DECOFF, 1u do:f1

  (p7 pl7 ph5):f2

  (p30*0.25 pl13 ph0):f1
  (p30*0.25
pl13 ph1):f1
  (p30*0.25
pl13 ph0):f1
  (p30*0.25
pl13 ph1):f1

  (p7 pl7 ph6):f2
  (p17 pl20 ph7):f2 (p17:sp10 ph11):f1

  1u cpds2:f2 cpds3:f3
  go=2 ph31
  1m do:f2 do:f3
  10m mc #0 to 2 ;F1PH(ip2,id0)
  F1PH(calph(ph2, +90), caldel(d0, +in0)) ;15N
  F2PH(calph(ph9, +90), caldel(d10, +in10)) ;1H

HaltAcqu, 1m ;jump address for protection files
exit ;quit

ph0=0
ph1 = 1
ph10 = 1
ph2 = 1
ph3 = 1
ph5 = 0
ph6 = 0 2
ph12 = 1
ph7 = 1
ph8 = 2
ph9 = 3 3 1 1
ph11 = 1
ph31 = 2 0 0 2

```

5.2.1.2 HhnCANH sequence

```
;HhnCaNH
;written by E. Najbauer, K. Tekwani Movellan, L. Andreas
;Avance III version
;parameters:
;p1 : 1H 90 pulse duration
;p3 : 13C 90 pulse duration
;p7 : 15N 90 pulse duration
;p30 : water suppression time (30-200 ms)
;cnst21 : CO offset in ppm (173.7)
;cnst22 : CA offset in ppm (53.7)
;cnst23 : CO/CO offset in ppm (113.7)
;d1 : recycle delay
;d5 : NOE mixing time
;d0 : 15N initial evolution time
;d10 : 1/2 of 13CO initial evolution time
;in0 : increment for 15N evolution
;in10 : 1/2 increment for 13CA evolution
;cpdprg1 : tppm (at pl13) or waltz (at pl13)
;cpdprg4 : cwY (at pl12)
;cpdprg5 : cwX (at pl12)
;pcpd1 : pulse length in decoupling sequence (2xtau_r for tppm, 25us for
10kHz waltz)
;pl1 : power level of 1H hard pulse
;pl12 : power level of decoupling (10-15 kHz)
;pl13 : power level for water suppression
;spnam1 : 1H shape for 1H->13CA CP (ramp 10-20%)
;spoal1 : N/A
;spoff1 : [ON/RES]
;sp1 : 1H power level during 1H->13CA CP
;spnam10 : 1H shape for 15N->1H(N) CP (ramp 10-20%)
;spoal10 : N/A
;spoff10 : [ON/RES]
;sp10 : 1H power level during 15N->1H CP
;cpdprg2 : 15N decoupling pattern during acq (waltz-16)
;p17 : contact time 15N->1H(N) CP (300-700 us)
;pcpd2 : pulse length in 15N decoupling sequence (25 us)
;p17 : power level for 15N hard pulse
;p116 : power level for 15N decoupling (corr. to 10 kHz)
;p12 : power level for 15N hard pulse
;p120 : 15N power level for 15N->1H CP
;spnam2 : 15N shape for 13CA->15N CP (tan-c100-w10pct)
;sp2 : 15N power level for 13CA->15N CP
;spoal2 : N/A
;spoff2 : [ON/RES]
;p15 : contact time 1H->15N CP (short, 300 - 700us)
;p14 : contact time for 15N->13CA CP (10ms)
;p16 : contact time 13CA->15N CP (10 ms)
;p18 : Q3 CO pulse duration
;p19 : Q3 CA pulse duration
;pcpd3 : pulse length in 13C decoupling sequence (25 us)
;cpdprg3 : 13CO/CA decoupling pattern during 15N evol (waltz-16)
;p13 : power level of 13C hard pulse [REFERENCE]
;p117 : power level for 13CO/CA decoupling (10 kHz)
;spnam9 : 13C shape for 13CA->15N CP (rectangle)
;sp9 : 13C power for 13CA->15N CP
```

```

;spoal9 : N/A
;spoff9 : [ON/RES CO]
;spnam18 : 13CO selective pulse shape (Q3)
;spoal18 : N/A
;spoff18 : [ON/RES CO]
;spnam19 : 13CA selective pulse shape (Q3)
;spoal19 : N/A
;spoff19 : [ON/RES CA]
;spnam29 : 13C shape for 1H->13CO CP (rectangle)
;sp29 : 13C power level for 1H->13CO CP
;spoal29 : N/A
;spoff29 : [ON/RES CO]
;zgoptns : -Dfslg, -Dlacq, or blank

;$COMMENT=Inverse Cp with INEPT CBCA mixing
;$CLASS=Solids
;$DIM=3D
;$TYPE=H detect
;$SUBTYPE=Heteronuclear
;$OWNER=CRMN

#include <Avancesolids.incl>

; Start evolutions from exactly 0
"d0=0.0"
"d10=0.0"
"d20=0.0"

; 1H settings
"spoal1=0.5" ; default value (irrelevant)
"spoff1=0.0" ; on-resonance
"spoal10=0.5" ; default value (irrelevant)
"spoff10=0.0" ; on-resonance

; 15N settings
"pcpd2=25" ;does not work!
"plw2=plw7"
"plw16=plw2*(pow(p7/25,2))" ; 15N waltz 10kHz decoupling power level
"spoal2=0.5" ; default value (irrelevant)
"spoff2=0.0" ; on-resonance

; 13C settings
"plw17=plw3*(pow(p3/25,2))" ; 13C waltz 10kHz decoupling power level

"cnst22 = (sfo3-bf3)*1000000/bf3" ; CA frequency offset (ppm)
"cnst21 = cnst22+(173.7-53.7)" ; CO frequency offset (ppm)
"cnst23 = cnst22+(113.7-53.7)" ; the offset half-way CO and CA (ppm)

"p18=3.412/(95.0*bf3/1000000)" ; 95 ppm bandwidth (safe)
"spw18=plw3*pow((0.5/(p18*0.1515))/(0.25/p3),2)" ; Q3 power level
"spoal18=0.5" ; default value (irrelevant)
"spoff18=bf3*((cnst21-cnst22)/1000000)"

"p19=3.412/(105.0*bf3/1000000)" ; 95 ppm bandwidth (safe)
"spw19=plw3*pow((0.5/(p19*0.1515))/(0.25/p3),2)" ; Q3 power level
"spoal19=0.5" ; default value (irrelevant)
"spoff19=0.0" ; CA frequency

```

```

"spoal9=0.5" ; default value (irrelevant)
"spoff9=0.0" ; on-resonance

"spoal8=0.5" ; default value (irrelevant)
"spoff8=0.0" ; on-resonance CA

"in0=inf1"
"in10=inf2/2"
"in20=inf3"

1 ze

2 d1 do:f2
#include <p15_prot.incl>
#include <aq_prot.incl>

    1u fq=cnst23(bf ppm):f3

    (p1 p11 ph4):f1

    1u cpds2:f2 cpds3:f3
    d20
    (p1 p11 ph8):f1 (1u do):f2 (1u do):f3

5 d5

    1u fq=cnst22(bf ppm):f3
    (p1 p11 ph0):f1

;HN cp
    (p15:sp0 ph1):f1 (p15 pl20 ph1):f2

;NCa
    (p14:sp8 ph15):f3 (p14:sp2 ph1):f2

    1u cpds1:f1
    (p18:sp18 ph18):f3 ;CO selective Pi
    1u
    (p19:sp19 ph19):f3 ;CA selective Pi
    1u
    d10
    (center (p7*2 ph0 pl7):f2 (p18:sp18 ph18):f3) ;CO selective Pi
    d10
    1u do:f1

    (p16:sp9 ph10):f3 (p16:sp2 ph2):f2

    1u fq=cnst23(bf ppm):f3
    1u cpds1:f1 cpds3:f3
    d0
    1u do:f1 do:f3
    (p7 p17 ph5):f2
;water suppression
    (p30*0.25 pl13 ph0):f1
    (p30*0.25 pl13 ph1):f1
    (p30*0.25 pl13 ph0):f1
    (p30*0.25 pl13 ph1):f1
;end water suppression

```

```
(p7 p17 ph6):f2
(p17 p120 ph7):f2 (p17:sp10 ph11):f1
```

```
1u cpds2:f2
go=2 ph31
1m do:f2
```

```
10m mc #0 to 2
F1PH(calph(ph2, +90), caldel(d0, +in0)) ;15N
F2PH(calph(ph15, -90), caldel(d10, +in10)) ;13C
F3PH(calph(ph4, +90), caldel(d20, +in20)) ;1H
```

```
HaltAcqu, 1m ;jump address for protection files
exit ;quit
```

```
ph0 = 0
ph1 = 1
ph3 = 0
ph4 = 2 2 0 0 ;proton flip down (FIRST PULSE), edge artifacts in 1H
;dimension without this phase cycle
ph2 = 1 1 3 3 3 3 1 1 ; 13C->15N CP (15N pulse)
ph5 = 0
ph6 = 0
ph7 = 1
ph8 = 2
ph10 = 0 2 ; 13C->15N CP (13C pulse)
ph11 = 1
ph15 = 0
ph18 = 0
ph19 = 0 ; 13CA(all aliph) selective pulse
;ph20 = 3 3 1 1 1 1 3 3
ph31 = 1 3 1 3 3 1 3 1 ; receiver phase for joint phase cycling
```

5.2.1.3 HhNCAtedorNH sequence

```
;HhNCAtedorNH
;Avance III version
;parameters:
;p1 : 1H 90 pulse duration
;p3 : 13C 90 pulse duration
;p7 : 15N 90 pulse duration
;p30 : water suppression time (30-200 ms)
;cnst21 : CO offset in ppm (173.7)
;cnst22 : CA offset in ppm (53.7)
;cnst23 : CO/CA offset in ppm (113.7)
;d1 : recycle delay
;d5 : NOE mixing time
;d0 : 15N initial evolution time
;d10 : 1/2 of 13CO initial evolution time
;in0 : increment for 15N evolution
;in10 : 1/2 increment for 13CA evolution
;cpdprg1 : tppm (at pl13) or waltz (at pl13)
;cpdprg4 : cwY (at pl12)
;cpdprg5 : cwX (at pl12)
;pcpd1 : pulse length in decoupling sequence (2xtau_r for tppm, 25us for
10kHz waltz)
;pl1 : power level of 1H hard pulse
;pl12 : power level of decoupling (10-15 kHz)
;pl13 : power level for water suppression
;spnam1 : 1H shape for 1H->13CA CP (ramp 10-20%)
;spoal1 : N/A
;spoff1 : [ON/RES]
;sp1 : 1H power level during 1H->13CA CP
;spnam10 : 1H shape for 15N->1H(N) CP (ramp 10-20%)
;spoal10 : N/A
;spoff10 : [ON/RES]
;sp10 : 1H power level during 15N->1H CP
;cpdprg2 : 15N decoupling pattern during acq (waltz-16)
;p17 : contact time 15N->1H(N) CP (300-700 us)
;pcpd2 : pulse length in 15N decoupling sequence (25 us)
;pl7 : power level for 15N hard pulse
;pl16 : power level for 15N decoupling (corr. to 10 kHz)
;pl2 : power level for 15N hard pulse
;pl20 : 15N power level for 15N->1H CP
;spnam2 : 15N shape for 13CA->15N CP (tan-c100-w10pct)
;sp2 : 15N power level for 13CA->15N CP
;spoal2 : N/A
;spoff2 : [ON/RES]
;p15 : contact time 1H->15N CP (short, 300 - 700us)
;p14 : contact time for 15N->13CA CP (10ms)
;p16 : contact time 13CA->15N CP (10 ms)
;p18 : Q3 CO pulse duration
;p19 : Q3 CA pulse duration
;pcpd3 : pulse length in 13C decoupling sequence (25 us)
;cpdprg3 : 13CO/CA decoupling pattern during 15N evol (waltz-16)
;pl3 : power level of 13C hard pulse [REFERENCE]
;pl17 : power level for 13CO/CA decoupling (10 kHz)
;spnam9 : 13C shape for 13CA->15N CP (rectangle)
;sp9 : 13C power for 13CA->15N CP
;spoal9 : N/A
```

```

;spoff9 : [ON/RES CO]
;spnam18 : 13CO selective pulse shape (Q3)
;spoal18 : N/A
;spoff18 : [ON/RES CO]
;spnam19 : 13CA selective pulse shape (Q3)
;spoal19 : N/A
;spoff19 : [ON/RES CA]
;spnam29 : 13C shape for 1H->13CO CP (rectangle)
;sp29 : 13C power level for 1H->13CO CP
;spoal29 : N/A
;spoff29 : [ON/RES CO]
;zgoptns : -Dfslg, -Dlacq, or blank

;l1 : to set mixing time
;l8 : z-filter time (number or rotor periods)
;l9 : fraction of rotation rate for t1 increment, about 20 for 55 kHz
;d23 : indirect dim incrementation

;$COMMENT=Inverse Cp with INEPT CBCA mixing
;$CLASS=Solids
;$DIM=3D
;$TYPE=H detect
;$SUBTYPE=Heteronuclear
;$OWNER=CRMN
#include <Avancesolids.incl>

; Start evolutions from exactly 0
"d0=0.0"
"d10=0.0"
"d20=0.0"

; 1H settings
"spoal1=0.5" ; default value (irrelevant)
"spoff1=0.0" ; on-resonance
"spoal10=0.5" ; default value (irrelevant)
"spoff10=0.0" ; on-resonance

; 15N settings
"pcpd2=25" ;does not work!
"plw2=plw7"
"plw16=plw2*(pow(p7/25,2))" ; 15N waltz 10kHz decoupling power level
"spoal2=0.5" ; default value (irrelevant)
"spoff2=0.0" ; on-resonance

; 13C settings
"plw17=plw3*(pow(p3/25,2))" ; 13C waltz 10kHz decoupling power level

"cnst22 = (sfo3-bf3)*1000000/bf3" ; CA frequency offset (ppm)
"cnst21 = cnst22+(173.7-53.7)" ; CO frequency offset (ppm)
"cnst23 = cnst22+(113.7-53.7)" ; the offset half-way CO and CA (ppm)

"p18=3.412/(95.0*bf3/1000000)" ; 95 ppm bandwidth (safe)
"spw18=plw3*pow((0.5/(p18*0.1515))/(0.25/p3),2)" ; Q3 power level
"spoal18=0.5" ; default value (irrelevant)
"spoff18=bf3*((cnst21-cnst22)/1000000)"

"p19=3.412/(105.0*bf3/1000000)" ; 95 ppm bandwidth (safe)
"spw19=plw3*pow((0.5/(p19*0.1515))/(0.25/p3),2)" ; Q3 power level

```

```

"spoal19=0.5" ; default value (irrelevant)
"spoff19=0.0" ; CA frequency

"spoal9=0.5" ; default value (irrelevant)
"spoff9=0.0" ; on-resonance

"spoal29=0.5" ; default value (irrelevant)
"spoff29=0.0" ; on-resonance CA

"in0=inf1"
"in20=inf3"

"d25=0.25s/cnst31-1u"
"d26=0.25s/cnst31-(p3)" ; one-quarter rotor cycle ...
"d27=0.25s/cnst31-(p7)"
"d28=0.75s/cnst31-p7"

"d23=1s*19/cnst31"

"l21=p18/(2*(1s/cnst31))+0.25s/cnst31+1"
"d29=l21*1.0s/cnst31-p18/2-0.25/cnst31"
;d29=0.75s/cnst31-(p3/2)" =p18/2+0.25rotor rounded up to even rotor period -
p18/2 -0.25rotor
"d31=1s/cnst31"
define pulse tauz
"tauz=18*1s/cnst31-p3"

define delay t1incr
"l22=((p18+p19+p18)/(1.0s/cnst31)+1)/2"
"d24=l22*2*1.0s/cnst31-p18-p19-p18"
"d2=1s/cnst31-p3-p7" ;d2 if not using selective pulses for J dec in 13C
evolution

"t1incr=(1s*19)/(cnst31)"
"in10=t1incr/2"
"inf2=in10*2"

define delay mix
"mix=4*2*11/cnst31"
t1incr
d23
mix
1 ze

2 d1 do:f2
#include <p15_prot.incl>
#include <aq_prot.incl>

1u fq=cnst23(bf ppm):f3

(p1 p11 ph4):f1

1u cpds2:f2 cpds3:f3
d20
(p1 p11 ph0):f1 (1u do):f2 (1u do):f3

3 d5
1u fq=cnst22(bf ppm):f3

```

```

(p1 p11 ph0):f1

;HN cp
  (p15:sp0 ph1):f1 (p15 pl20 ph0):f2

;NCa
  1u cpds1:f1
  d25 p17:f2 p13:f3 ; tppm decoupling at p112 if synchronized with TPPM use
;special decoupling sequence
5 d26 ; several F3 pulses ...
  (p3*2 ph8^):f3 ;at intervals of d26 one-half rotor cycle
  d26
  (p3*2 ph8^):f3 ; ... at intervals of ...
  d26 ; F1 pulse
  d26 ; several F3 pulses ...
  (p3*2 ph8^):f3 ; ... at intervals of ...
  d26 ; ... at intervals of ...
  d26 ; ... one-half rotor cycle
  (p3*2 ph8^):f3 ; ... at intervals of ...
  d26 ; F1 pulse
lo to 5 times l1
  d29
  (center (p7*2 ph0):f2 (p18:sp18 ph18):f3) ;CO selective Pi to NOT REDOR
    ; ... at intervals of ...
  d29 ; Hahn echo refocussing pulse on F1
6 d26 p13:f3 ; several F3 pulses ...
  (p3*2 ph9^):f3 ; ... at intervals of ...
  d26 ; ... at intervals of ...
  d26 ; ... one-half rotor cycle
  (p3*2 ph9^):f3 ; ... at intervals of ...
  d26 ;
  d26 ; several F3 pulses ...
  (p3*2 ph9^):f3 ; ... at intervals of ...
  d26 ; ... at intervals of ...
  d26 ; ... one-half rotor cycle
  (p3*2 ph9^):f3 ; ... at intervals of ...
  d26 ; F1 pulse
lo to 6 times l1
  d27
  (p7 ph0):f2 (1u do):f1 ; end of integer rotor period
  (tauz pl14 ph0):f1 ;(1u fq=cnst22(bf ppm)):f3 ;Z-filter integer rotor
;period
  (p3 p13 ph0):f3 (1u cpds1):f1 ; begin rotor period at beginning of pulse

  (p18:sp18 ph18):f3 ;CO selective Pi
  d24*.5
  (p19:sp19 ph19):f3 ;CA selective Pi
  d24*.5
  d10
  ;(p7*2 ph0 pl7):f2
  (center (p7*2 ph0 pl7):f2 (p18:sp18 ph18):f3) ;CO selective Pi
  d10

  (p3 p13 ph3):f3 (1u do):f1
  (d2 cpds1):f1 ;(1u fq=cnst23(bf ppm)):f3 ;for rotor synchronization
  (p7 pl7 ph0):f2 ; rotor period done at end of p1!
  1u

```

```

    d25 p17:f2 p13:f3 ; tppm decoupling at p112
7 d26 ; several F3 pulses ...
  (p3*2 ph10^):f3 ; ... at intervals of ...
  d26 ; ... one-half rotor cycle
  d26
  (p3*2 ph10^):f3 ; ... at intervals of ...
  d26
  d26 ; several F3 pulses ...
  (p3*2 ph10^):f3 ; ... at intervals of ...
  d26 ; ... one-half rotor cycle
  d26
  (p3*2 ph10^):f3 ; ... at intervals of ...
  d26
lo to 7 times l1
; d28
; (p7*2 ph0):f2
; d28
d29
  (center (p7*2 ph0):f2 (p18:sp18 ph18):f3) ; Hahn echo refocusing pulse
d29
8 d26 p13:f3 ;
several F3 pulses ...
  (p3*2 ph11^):f3 ; ... at intervals of ...
  d26 ; ... one-half rotor cycle
  d26
  (p3*2 ph11^):f3 ; ... at intervals of ...
  d26
  d26 ; several F3 pulses ...
  (p3*2 ph11^):f3 ; ... at intervals of ...
  d26 ; ... one-half rotor cycle
  d26
  (p3*2 ph11^):f3 ; ... at intervals of ...
  d26
lo to 8 times l1
d27 fq=cnst23(bf ppm):f3 ; Hahn echo occurs about now
  (p7 ph1):f2 (lu do):f1
  (tauz p114 ph0):f1
  (p7 ph2):f2 (lu cpds1):f1 (lu cpds3):f3

; (p16:sp9 ph10):f3 (p16:sp2 ph2):f2

d0
lu do:f1 do:f3
(p7 p17 ph5):f2
;water suppression
  (p30*0.25 p113 ph0):f1
  (p30*0.25 p113 ph1):f1
  (p30*0.25 p113 ph0):f1
  (p30*0.25 p113 ph1):f1
;end water suppression
  (p7 p17 ph6):f2
  (p17 p120 ph7):f2 (p17:sp10 ph13):f1

lu cpds2:f2
go=2 ph31
lm do:f2

10m mc #0 to 2

```

```
F1PH(calph(ph2, +90), caldel(d0, +in0)) ;15N
F2PH(calph(ph3, -90), caldel(d10, +in10)) ;13C
F3PH(calph(ph4, +90), caldel(d20, +in20)) ;1H
```

```
HaltAcqu, 1m ;jump address for protection files
exit ;quit
```

```
ph0 = 0
ph1 = 1
ph3 = 0 2 ;up after 13C evolution
ph4 = 2 2 0 0 ; 0 2 2 0 2 0 0 2 ;proton flip down (FIRST
PULSE) edge artifacts in 1H dimension without this phase cycle
ph2 = 0 0 0 0 2 2 2 2 ; down before 15N evolution
ph5 = 0
ph6 = 0
ph7 = 1
```

```
ph8= 0 1 0 1 1 0 1 0 2 3 2 3 3 2 3 2 ;xy16
ph9= 0 1 0 1 1 0 1 0 2 3 2 3 3 2 3 2
ph10= 0 1 0 1 1 0 1 0 2 3 2 3 3 2 3 2
ph11= 0 1 0 1 1 0 1 0 2 3 2 3 3 2 3 2
```

```
ph13 = 1
ph18 = 0
ph19 = 0 ; 13CA (all aliph) selective pulse
ph31 = 0 2 2 0 2 0 0 2 ; receiver phase for joint phase cycling
```

5.2.2 Supplementary figures

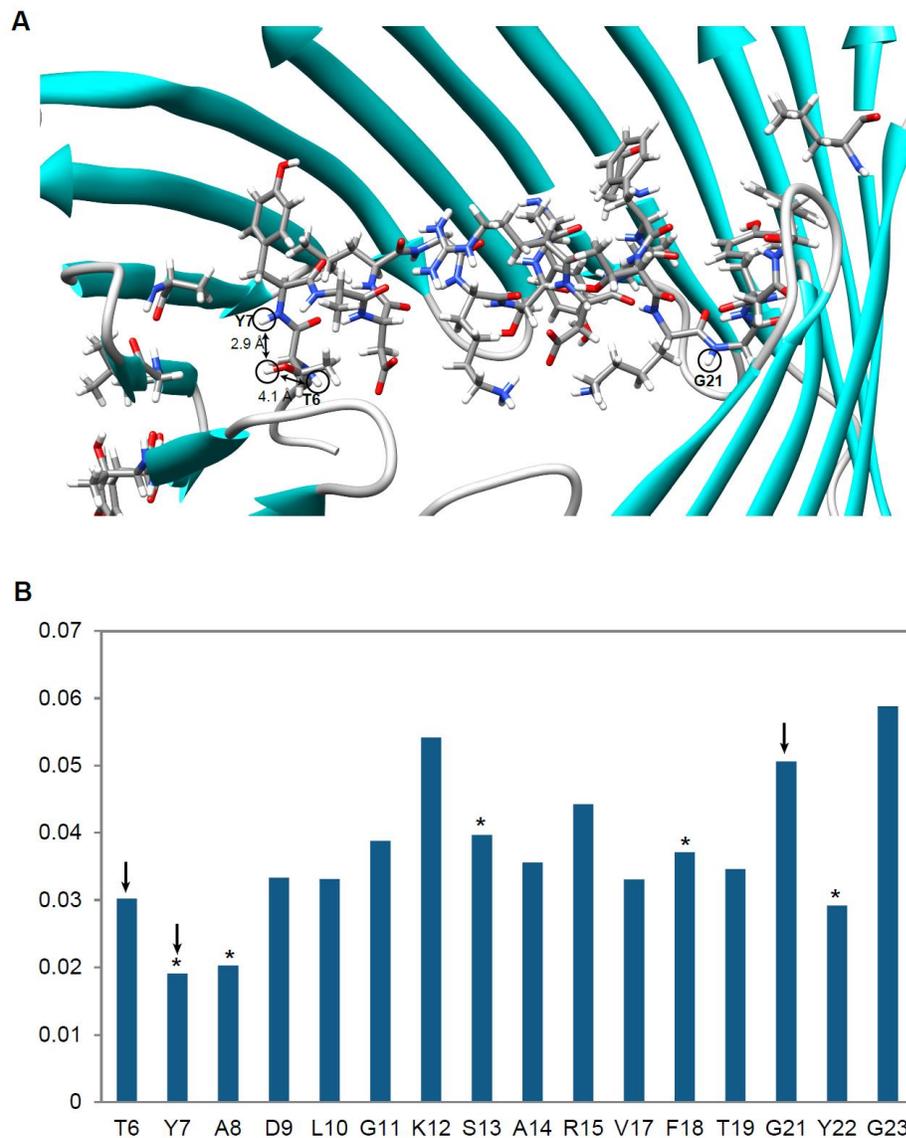


Figure S1: Proximity of exchanging residues has little or no effect on the observed buildup of the cross peak between a mobile species and an amide proton. A) displays the water-exposed N-terminal α -helix of VDAC (pdb: 3emn). B) shows the intensity of water-amide cross peaks for the α -helix at 50 ms mixing time, scaled with the amide peaks at 0 ms mixing time to remove differences in HN CP efficiency. 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.

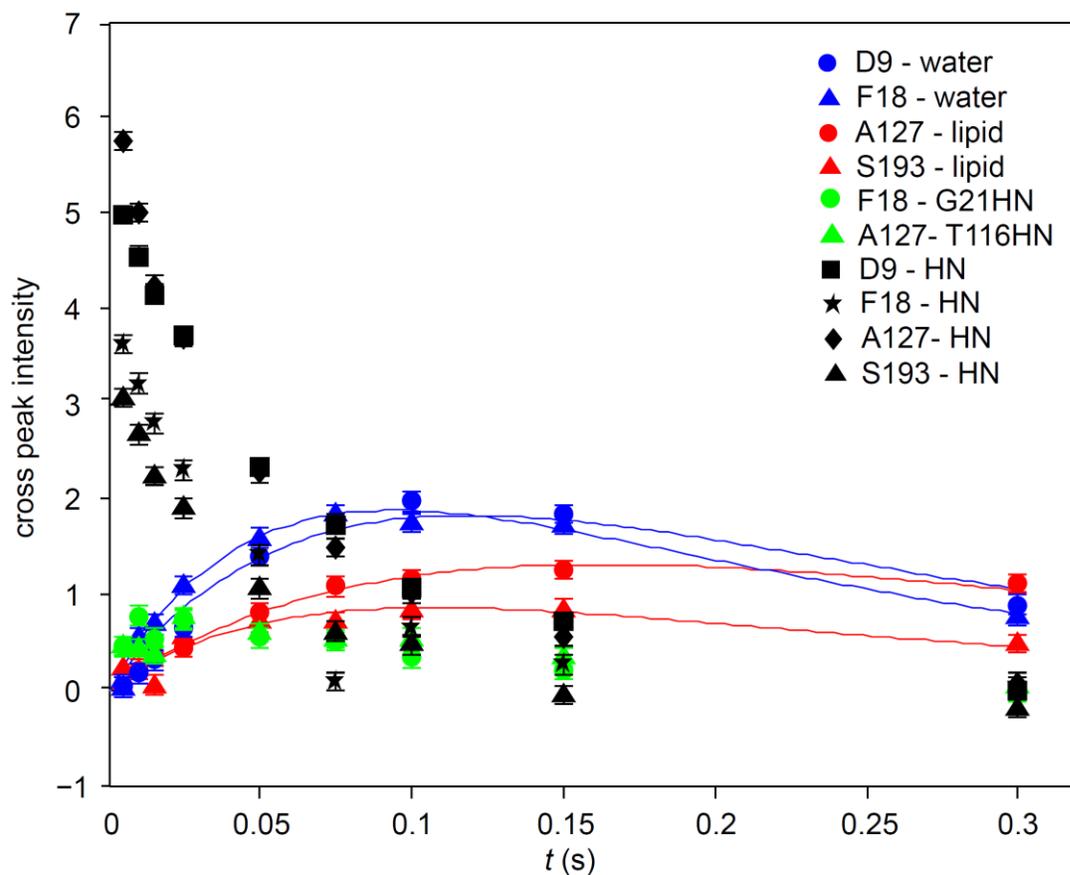


Figure S2: Decay of HN peaks (signal from amides, where magnetization did not get transferred during z-mixing) and buildup of cross peaks between amides and mobile species, as well as other residues in function of z-mixing times. The starting intensity of the HN peaks is significantly higher than that of the cross peaks building up through magnetization transfer to other residues, justifying the choice to neglect the effect of spin diffusion in the fitting.

5.3 Chapter 3: Compensating the effect of linear field drift

5.3.1 Script for compensating linear field drift

```
/*
*****
*/
/*      lindriftcomp              30.01.2019      */
/*
*****
*/
/*      Short Description :
/*      linear drift compenstaion using C binary file handling
/*      fopen, fread for file handling;  for 2D, 3D, 4D
/*      spectra using planes, points or NUS acquisition
/*      (NUS available only for 3D and 4D spectra)
/*
*****
*/
/*      Description/Usage :
/*      The AU program applies a linear phase ramp in both
/*      direct and indirect dimensions to correct for any
/*      occuring drift, assuming the drift is linear.
/*      Save the program (depending on your Topspin version)
/*      to Bruker/Topspin_*version*_exp/stan/nmr/au/src/user
/*      The script can be started by typing the name of the
/*      script at the command line. Tested in Topspin 3.5pl6
/*      The start shift value is applied to the first FID,
/*      and the end shift is applied to the last acquired FID.
/*      The shift for intermediate FIDs is determined by linear
/*      interpolation.
/*      Start and end shift values are each derived from 1D
/*      spectra measured at the start and end of the nD
/*      experiment.(E.g. this could be the distance of the
/*      water peak of the current spectrum from that of the
/*      reference.) Enter values in units of Hz for the
/*      detected nucleus.
/*      A positive shift will shift the spectrum to the left.
/*      The AU program does not overwrite the current dataset,
/*      but generates a new dataset, whose number should be
/*      entered by the user.
/*
*****
*/
/*      Author(s) :
/*      Name          : Eszter E. Najbauer
/*      Organisation  : Max Planck Institute for
/*                    Biophysical Chemistry
/*      Email        : esna@nmr.mpibpc.mpg.de
/*
/*      Name          : Loren B. Andreas
/*      Organisation  : Max Planck Institute for
/*                    Biophysical Chemistry
/*      Email        : land@nmr.mpibpc.mpg.de
/*      We thank Wolfgang Bermel for his technical assistance.
/*      Citation: Najbauer and Andreas: JMR 2019
/*
*****
*/
```

```
/*
*****PROGRAM OUTLINE:
      READ IN PARAMETERS
*/
```

```

SHIFT DATA
  PLANES
    2D (including Echo-Antiecho module)
    3D
      312 aqseq
      321 aqseq
    4D 4321 aqseq
  NUS
    3D
      312 aqseq
      321 aqseq
  POINTS
    3D
      321 aqseq
      312 aqseq
    4D 4321 aqseq  *****/

```

```

char infile[PATH_MAX], outfile[PATH_MAX];
char path[PATH_MAX + 64], modetext[80], autext[256];
double g_delay, ymin, ymax, bf1, bf2, bf3, bf_dir;
double dur = 0;
double dur2= 0;
double dur3= 0;
double dur4= 0;
double in_f1 =0;
double in_f2 =0;
double in_f3 =0;
float dw, start_c, end_c;
int *row1;
int *rowo;
int td, td1, td2, td3, tds, td1s, td2s, td3s, tot_points;
int i, j, k, l;
int oexpno, nexpno, byteorder, parmode, dspfv, dty, aqmod;
int fntype = 0;
int fnmode1s, fnmode2s, fnmode3s;
int aqseq = -1;
int arglim = 3;
int nus_points;
int cur_pointf1, cur_pointf2, cur_pointf3;
char datapath1[1024]; /*for nuslist*/
char* hashbuf;
const char* modepnt;
FILE *fpin, *fpout;
FILE *fp;

static const int decim_tab[] = {2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64, 96,
128,
          192, 256, 384, 512, 768, 1024, 1536, 2048};

static const int delay_tab_10[] = {179, 201, 533, 709, 1097, 1449, 2225, 2929,
4481,
          5889, 8993, 11809, 18017, 23649, 36065, 47329, 72161,
          94689, 144353, 189409, 288737};

static const int delay_tab_12[] = {184, 219, 384, 602, 852, 1668, 2292, 3368,
4616, 6768,
          9264, 13568, 18560, 27392, 36992, 55040, 73856, 110336,
          147584, 220928, 295040};

```

```

/***** get dataset and parameters *****/

FETCHPARS("PARMODE", &parmode)

if (parmode != 1 && parmode != 2 && parmode !=3)
    STOPMSG("Program is only suitable for 2D or 3D or 4D data")

FETCHPARS("BF1", &bf_dir)
FETCHPARS("DTYPA", &dtypa)

if (dtypa)
    STOPMSG("Program is only suitable for DTYPA 0 data")

FETCHPARS("AQ_mod", &aqmod)

if (aqmod < 0 || aqmod > 3)
    STOPMSG("acquisition status parameter AQ_mod out of range")

FETCHPARS("BYTORDA", &byteorder)
FETCHPARS("TD", &tds)

td = ((tds + 255) / 256) * 256;

if (td <= 0 || td > 64 * 1024 * 1024)
    STOPMSG("TD { F2 } direct dim")

if (parmode == 1) { /*get parameters for 2D data*/
    FETCHPAR1("TD", &td1)
    FETCHPAR1S("TD", &td1s)

    if (td1s <= 0 || td1s > 64 * 1024 * 1024)
        STOPMSG("TD{F1} out of range")

    td2 = td2s = 1;

    FETCHPARNS(1, "IN_F", &in_f1)
    in_f1 = in_f1 * (1.0f / 1000000.0f);
    FETCHPARNS(1, "BF1", &bf1)

    FETCHPARNS(1, "FnMODE", &fnmodels)
} else if (parmode == 2) { /*get parameters for 3D data*/
    FETCHPARN(1,"TD", &td1)
    FETCHPARNS(1,"TD", &td1s)

    FETCHPARN(2,"TD", &td2)
    FETCHPARNS(2,"TD", &td2s)

    if (td1s <= 0 || td1s > 64 * 1024 * 1024)
        STOPMSG("TD (td1s) out of range")

    if (td2s <= 0 || td2s > 64 * 1024 * 1024)
        STOPMSG("TD (td2s) out of range")

    FETCHPARNS(1, "IN_F", &in_f1)
    in_f1 = in_f1 * (1.0f / 1000000.0f);
    FETCHPARNS(1, "BF1", &bf1)

```

```

    FETCHPARNS(2, "IN_F", &in_f2)
    in_f2 = in_f2 * (1.0f / 1000000.0f);
    FETCHPARNS(2, "BF1", &bf2)

    FETCHPARS("AQSEQ", &aqseq) /*aqseq of 0 is 321, aqseq of 1 is 312, only
required for 3D data, 4D aqseq is always 4321*/

    if (aqseq != 0 && aqseq != 1)
        STOPMSG("AQSEQ out of range")

    FETCHPARNS(1, "FnMODE", &fnmodels)
    FETCHPARNS(2, "FnMODE", &fnmode2s)

    if (fnmodels == 6 || fnmode2s == 6)
        STOPMSG("Program does not handle 3D Echo-Antiecho data. For Echo-
Antiecho template, see 2D section of script.")
} else if (parmode == 3) { /*get parameters for 4D data*/
    FETCHPARN(1,"TD", &td1)
    FETCHPARNS(1,"TD", &td1s)

    FETCHPARN(2,"TD", &td2)
    FETCHPARNS(2,"TD", &td2s)

    FETCHPARN(3,"TD", &td3)
    FETCHPARNS(3,"TD", &td3s)

    if (td1s <= 0 || td1s > 64 * 1024 * 1024)
        STOPMSG("TD (td1s) out of range")

    if (td2s <= 0 || td2s > 64 * 1024 * 1024)
        STOPMSG("TD (td2s) out of range")

    if (td3s <= 0 || td3s > 64 * 1024 * 1024)
        STOPMSG("TD{F1} out of range")

    FETCHPARNS(1, "IN_F", &in_f1)
    in_f1 = in_f1 * (1.0f / 1000000.0f);
    FETCHPARNS(1, "BF1", &bf1)

    FETCHPARNS(2, "IN_F", &in_f2)
    in_f2 = in_f2 * (1.0f / 1000000.0f);
    FETCHPARNS(2, "BF1", &bf2)

    FETCHPARNS(3, "IN_F", &in_f3)
    in_f3 = in_f3 * (1.0f / 1000000.0f);
    FETCHPARNS(3, "BF1", &bf3)

    FETCHPARNS(1, "FnMODE", &fnmodels)
    FETCHPARNS(2, "FnMODE", &fnmode2s)
    FETCHPARNS(3, "FnMODE", &fnmode3s)
    if (fnmodels == 6 || fnmode2s == 6 || fnmode3s == 6)
        STOPMSG("Program does not handle 4D Echo-Antiecho data. For Echo-
Antiecho template, see 2D section of script.")
}

FETCHPARS("FnTYPE", &fnctype) /*planes: 0, points: 1, NUS: 2 */

```

```

if (fntype < 0 || fntype > 2)
    STOPMSG("FnTYPE out of range")

FETCHPARS("NusPOINTS", &nus_points) /*number of points for NUS sampled
data*/

FETCHPARS("DW", &dw)
dw = dw * (2.0f / 1000000.0f);

FETCHPARS("DSPFVS", &dspfvs) /*firmware version*/

if (dspfvs == 10 || dspfvs == 12) {
    int d_index;
    double decim = 0;
    const int* delay_tab = (dspfvs == 10) ? delay_tab_10 : delay_tab_12;

    FETCHPARS("DECIM", &decim)

    for (d_index = sizeof(decim_tab) / sizeof(decim_tab[0]) - 1; d_index >= 0
; d_index--)
        if (decim_tab[d_index] == decim)
            break;

    if (d_index < 0)
        STOPMSG("DECIM out of range")

    g_delay = (double)delay_tab[d_index] / decim;
} else if (dspfvs >= 20 && dspfvs <= 23) {
    double grpdlly = 0;

    FETCHPARS("GRPDLY", &grpdlly)
    g_delay = grpdlly * 2.0;
} else
    STOPMSG("DSPFVS out of range")

/***** get input *****/

GETFLOAT("Enter the start shift in Hz (positive values shift left):",
start_c);
GETFLOAT("Enter the end shift in Hz:", end_c);
GETINT("Please enter the new experiment number:", nexpno);

/**** read in NUS file****/
if ( fntype==2 ) {
    (void)sprintf(datapath1,ACQUPATH("nuslist"));
    fp= fopen(datapath1,"r");

    if (fp == 0) { /*file cannot be read*/
        (void)sprintf(path, "%s:\n%s", "cannot open file for reading",
infile);
        STOPMSG(path)
    }
}

/***** check files *****/

WRAPARAM(nexpno)

```

ERRORABORT

```
(void)strcpy(infile, ACQUPATH("ser"));
fpin = fopen(infile, "rb");
if (fpin == 0) {
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile);
    STOPMSG(path)
}

oexpno = expno;
expno = nexpno;
(void)strcpy(outfile, ACQUPATH("ser"));
expno = oexpno;
fpout = fopen(outfile, "wb");
if (fpout == 0) {
    (void)sprintf(path, "%s:\n%s", "cannot open file for writing", outfile);
    STOPMSG(path)
}

/*****SHIFT CORRECTION*****/

if (fntype == 0) { /****PLANES ACQUISITION MODE****/
    if (parmode == 1) { /****2D data****/
        tot_points = tdl1 / 2;

        row1 = (int*)malloc(td * tdl1 * sizeof(int));
        row0 = (int*)malloc(td * tdl1 * sizeof(int));
        fread(row1, sizeof(int), td * tdl1, fpin);

        for (i = 0; i < tdl1 / 2; i++) { /*indirect dim loop*/
            for (j = 0; j < td; j += 2) { /*direct dim loop*/
                int rr, ri, ir, ii;
                double arg1, c1, s1, arg2, c2, s2, crr, cri, cir, cii;

                if (fnmodels < 6) { /* not Echo-Antiecho*/
                    rr = row1[2*i*td+j];
                    ri = row1[2*i*td+j+1];
                    ir = row1[(2*i+1)*td+j];
                    ii = row1[(2*i+1)*td+j+1];
                }

                else if (fnmodels == 6) { /*Echo-Antiecho*/
                    ri = (row1[2*i*td+j] + row1[(2*i+1)*td+j] ) / 2; /* rr
and ri swapped and ri is inverted relative to other components to apply 90 deg
phase shift*/
                    rr = -(row1[2*i*td+j+1] + row1[(2*i+1)*td+j+1] ) / 2;
                    ir = -(row1[2*i*td+j] - row1[(2*i+1)*td+j] ) / 2;
                    ii = -(row1[2*i*td+j+1] - row1[(2*i+1)*td+j+1] ) / 2;
                }

                dur = - g_delay * dw / 2 + j * dw / 2;

                arg1 = (start_c + i*(end_c-start_c)/tot_points) * dur * 2 *
M_PI;

                c1 = cos(arg1);/*direct*/
                s1 = sin(arg1);/*direct*/

                crr = c1*rr - s1*ri;
```

```

        cri = s1*rr + c1*ri;
        cir = c1*ir - s1*ii;
        cii = s1*ir +c1*ii; /*corrected direct dim*/

        dur2 = in_f1*i;
        arg2 = (start_c + i*(end_c-
start_c)/tot_points)*dur2*2*M_PI*bf1/bf_dir;
        if (fnmodels == 1) { /*QF acquisition mode (pseudo 2D data)*/
            arg2 == 0;
        }
        c2 = cos(arg2); /*indirect*/
        s2 = sin(arg2); /*indirect*/

        rowo[2*i*td+j] = c2*crr - s2*cir;
        rowo[2*i*td+j+1] = c2*cri - s2*cii;
        rowo[(2*i+1)*td+j] = s2*crr + c2*cir;
        rowo[(2*i+1)*td+j+1] = s2*cri + c2*cii; /*corrected indirect
as well*/
    }
}
fwrite(rowo, sizeof(int), td * tdls, fpout);
if (fnmodels == 6) { /* status parameter changed to
States processing for Echo-Antiecho data to make write out uniform*/
    DATASET(name, nexpno, 1, disk, user)
    STOREPARNS(1,"FnMODE", 4)
}
/*end of 2D correction*/

} else if (parmode == 2 ) { /***3D data***/
    if (aqseq == 1){ /***3D case with aqseq of 312***/
        tot_points = tdls / 2 * td2s / 2;
        row1 = (int*)malloc(td * tdls * td2s * sizeof(int));
        rowo = (int*)malloc(td * tdls * td2s * sizeof(int));
        fread(row1, sizeof(int), td * tdls * td2s, fpin);
        for (j = 0; j < td2s / 2; j++) { /*indirect dim loop1, slow*/
            for (k = 0; k < tdls / 2; k++) { /*indirect dim loop1, fast*/
                for (l = 0; l < td; l += 2) { /*direct dim loop*/
                    int rrr, rri, rir, rii, irr, iri, iir, iii;
                    double arg1, arg2, arg3, c1, s1, c2, s2, c3, s3,
crrr,crri,crir,crii,cirr, ciri, ciir, ciii;
                    double ccrrr,ccrri,ccrir,ccrii,ccirr, cciri, cciir,
cciii;

                    rrr = row1[2*j*tdls*td + 2*k*td + l];
                    rri = row1[2*j*tdls*td + 2*k*td + l+1];
                    rir = row1[2*j*tdls*td + (2*k+1)*td + l];
                    rii = row1[2*j*tdls*td + (2*k+1)*td + l+1];
                    irr = row1[(2*j+1)*tdls*td + 2*k*td + l];
                    iri = row1[(2*j+1)*tdls*td + 2*k*td + l+1];
                    iir = row1[(2*j+1)*tdls*td + (2*k+1)*td + l];
                    iii = row1[(2*j+1)*tdls*td + (2*k+1)*td + l+1];

                    dur = - g_delay * dw / 2 + l * dw / 2;

                    arg1 = (start_c + (tdls / 2 * j + k)*(end_c-
start_c)/tot_points) * dur * 2 * M_PI;
                    c1 = cos(arg1);/*direct part*/
                    s1 = sin(arg1);/*direct part*/

```

```

crrr = c1*rrr - s1*rri;
crrr = s1*rrr + c1*rri;
crir = c1*rir - s1*rii;
crii = s1*rir + c1*rii;

cirr = c1*irr - s1*iri;
ciri = s1*irr + c1*iri;
ciir = c1*iir - s1*iii;
ciiii = s1*iir + c1*iii; /*corrected direct dim*/

dur2 = in_f1*k; /*fast indirect dim*/
arg2 = (start_c + (td1s / 2 * j + k )*(end_c-
start_c)/tot_points)*dur2*2*M_PI*bf1/bf_dir;
if (fnmodels == 1) { /*QF acquisition mode (pseudo 3D
data)*/
    arg2 = 0;
}
c2 = cos(arg2);
s2 = sin(arg2);

ccrrr = c2*crrr - s2*crir;
ccrri = c2*crrr - s2*crii;
ccrir = s2*crrr + c2*crir;
ccrii = s2*crrr + c2*crii;

ccirr = c2*cirr - s2*ciir;
cciri = c2*ciri - s2*ciii;
cciiir = s2*cirr + c2*ciir;
cciiii = s2*ciri + c2*ciii; /*corrected indirect dim
1*/

dur3 = in_f2*j; /*slow indirect dim. elapsed time
is the same: (td2s / 2 * j + k)*/
arg3 = (start_c + (td1s / 2 * j + k )*(end_c-
start_c)/tot_points)*dur3*2*M_PI*bf2/bf_dir;
if (fnmode2s == 1) { /*QF acquisition mode (pseudo 3D
data)*/
    arg3 = 0;
}
c3 = cos(arg3);
s3 = sin(arg3);

rowo[2*j*td1s*td + 2*k*td + 1] = c3*ccrrr - s3*ccirr
;
rowo[2*j*td1s*td + 2*k*td + 1+1]= c3*ccrri - s3*cciri;
rowo[2*j*td1s*td + (2*k+1)*td + 1]= c3*ccrir -
s3*cciiir;
rowo[2*j*td1s*td + (2*k+1)*td + 1+1]= c3*ccrii -
s3*cciiii;

rowo[(2*j+1)*td1s*td + 2*k*td + 1]= s3*ccrrr
+c3*ccirr;
rowo[(2*j+1)*td1s*td + 2*k*td + 1+1]=s3*ccrri
+c3*cciri;
rowo[(2*j+1)*td1s*td + (2*k+1)*td + 1]=s3*ccrir
+c3*cciiir;

```

```

rowo[(2*j+1)*td1s*td + (2*k+1)*td + l+1]=s3*ccrii
+c3*cciii;
        /*corrected indirect dim2*/
    }
}
fwrite(rowo, sizeof(int), td * td1s * td2s, fpout);
} else if (aqseq == 0 ) { /***3D data, aqsec of 321**/
    tot_points = td1s / 2 * td2s / 2;

    row1 = (int*)malloc(td * td1s * td2s * sizeof(int));
    rowo = (int*)malloc(td * td1s * td2s * sizeof(int));
    fread(row1, sizeof(int), td * td1s * td2s, fpin);

    for (j = 0; j < td1s / 2; j++) { /*indirect dim loop1,slow*/
        for (k = 0; k < td2s / 2; k++) { /*indirect dim loop1,fast*/
            for (l = 0; l < td; l += 2) { /*direct dim loop*/
                int rrr, rri, rir, rii, irr, iri, iir, iii;
                double arg1, arg2, arg3, c1, s1, c2, s2, c3, s3,
crrr,crri,crir,crii,cirr, ciri, ciir, ciii;
                double ccrrr,ccrri,ccrir,ccrii,ccirr, cciri, cciir,
cciii;

                rrr = row1[2*j*td2s*td + 2*k*td + l];
                rri = row1[2*j*td2s*td + 2*k*td + l+1];
                rir = row1[2*j*td2s*td + (2*k+1)*td + l];
                rii = row1[2*j*td2s*td + (2*k+1)*td + l+1];
                irr = row1[(2*j+1)*td2s*td + 2*k*td + l];
                iri = row1[(2*j+1)*td2s*td + 2*k*td + l+1];
                iir = row1[(2*j+1)*td2s*td + (2*k+1)*td + l];
                iii = row1[(2*j+1)*td2s*td + (2*k+1)*td + l+1];

                dur = - g_delay * dw / 2 + l * dw / 2;

                arg1 = (start_c + (td2s / 2 * j + k)*(end_c-
start_c)/tot_points) * dur * 2 * M_PI;
                c1 = cos(arg1); /*direct part*/
                s1 = sin(arg1); /*direct part*/

                crrr = c1*rrr - s1*rri;
                crri = s1*rrr + c1*rri;
                crir = c1*rir - s1*rii;
                crii = s1*rir + c1*rii;

                cirr = c1*irr - s1*iri;
                ciri = s1*irr + c1*iri;
                ciir = c1*iir - s1*iii;
                ciii = s1*iir + c1*iii; /*corrected direct dim*/

                dur2 = in_f2*k; /*fast indirect dim*/
                arg2 = (start_c + (td2s / 2 * j + k)*(end_c-
start_c)/tot_points)*dur2*2*M_PI*bf2/bf_dir;
                if (fnmode2s == 1) { /*QF acquisition mode (pseudo 3D
data)*/

                    arg2 = 0;
                }
                c2 = cos(arg2);
                s2 = sin(arg2);

```

```

        crrrr = c2*crrr - s2*crir;
        crrii = c2*crri - s2*cirii;
        ccrir = s2*crrr + c2*crir;
        cciii = s2*crri + c2*cirii;

        ccirr = c2*cirr - s2*ciir;
        cciri = c2*ciri - s2*ciiii;
        cciir = s2*cirr + c2*ciir;
        cciiii = s2*ciri + c2*ciiii; /*corrected indirect dim
1*/

        dur3 = in_f1*j; /*slow indirect dim. elapsed time is
the same: (td2s / 2 * j + k)*/
        arg3 = (start_c + (td2s / 2 * j + k)*(end_c-
start_c)/tot_points)*dur3*2*M_PI*bf1/bf_dir;
        if (fnmodels == 1) { /*QF acquisition mode (pseudo 3D
data)*/

            arg3 =0;
        }
        c3 = cos(arg3);
        s3 = sin(arg3);

        rowo[2*j*td2s*td + 2*k*td + 1] = c3*crrrr - s3*ccirr ;
        rowo[2*j*td2s*td + 2*k*td + 1+1]= c3*crrii - s3*cciri;
        rowo[2*j*td2s*td + (2*k+1)*td + 1]= c3*ccrir -
s3*cciir;
        rowo[2*j*td2s*td + (2*k+1)*td + 1+1]= c3*ccrii -
s3*cciii;

        rowo[(2*j+1)*td2s*td + 2*k*td + 1]= s3*crrrr
+c3*ccirr;
        rowo[(2*j+1)*td2s*td + 2*k*td + 1+1]=s3*crrii
+c3*cciri;
        rowo[(2*j+1)*td2s*td + (2*k+1)*td + 1]=s3*ccrir
+c3*cciir;
        rowo[(2*j+1)*td2s*td + (2*k+1)*td + 1+1]=s3*ccrii
+c3*cciii;
    }
}
    }
    fwrite(rowo, sizeof(int), td * td1s * td2s, fpout);
}
/*end parmode == 2, 3D case*/

} else if (parmode == 3 ) { /***4D case with planes acquisition, aqseq of
4321 (only possible option)***/
    tot_points = td1s / 2 * td2s / 2 * td3s / 2;

    for (i = 0; i < td1s / 2; i++) { /*read in 2 x 3D sized blocks to
avoid running out of memory*/
        row1 = (int*)malloc(td * td3s * td2s * 2 * sizeof(int)); /*
replace td1 with td3*/
        if (row1 == 0)
            STOPMSG("cannot get enough memory")
        rowo = (int*)malloc(td * td3s * td2s * 2 * sizeof(int));
        if (rowo == 0)
            STOPMSG("cannot get enough memory")

```

```

fread(row1, sizeof(int), td * td3s * td2s * 2, fpin);

for (j = 0; j < td2s / 2; j++) { /*indirect dim loop1,slow*/
    for (k = 0; k < td3s / 2; k++) { /*indirect dim loop1,fast*/
        for (l = 0; l < td; l += 2) { /*direct dim loop*/
            int rrrr, rrrr, rrrr, rrrr, rrrr, rrrr, rrrr, rrrr,
            irrr, irri, irir, irii, iirr, iiri, iirr, iiii;
            double crrrr, crrri, crrir, crrii, crirr, criri,
            criir, criii, cirrr, cirri, cirir, cirii, ciirr, ciiri, ciiir, ciiii;
            double ccrrrr, ccrrri, ccrrir, ccrrii, ccrrir, ccrriri,
            ccrrriir, ccrrriii, ccrrrir, ccrrrii, ccrrrirr, ccrrriri,
            ccrrriir, ccrrriii, ccrrrirr, ccrrrii, ccrrrirr, ccrrrii, ccrrrirr,
            ccrrrii, ccrrriir, ccrrriii;
            double arg1, arg2, arg3, arg4, c1, s1, c2, s2, c3, s3,
            c4, s4;

            /*remove i, since now reading in blocks of data*/
            rrrr = row1[2*j*td3s*td + 2*k*td + 1];
            rrrr = row1[2*j*td3s*td + 2*k*td + 1+1];
            rrrr = row1[2*j*td3s*td + (2*k+1)*td + 1];
            rrrr = row1[2*j*td3s*td + (2*k+1)*td + 1+1];

            rrrr = row1[(2*j+1)*td3s*td + 2*k*td + 1];
            rrrr = row1[(2*j+1)*td3s*td + 2*k*td + 1+1];
            rrrr = row1[(2*j+1)*td3s*td + (2*k+1)*td + 1];
            rrrr = row1[(2*j+1)*td3s*td + (2*k+1)*td + 1+1];

            irrr = row1[td2s*td3s*td + 2*j*td3s*td + 2*k*td + 1];
            irri = row1[td2s*td3s*td + 2*j*td3s*td + 2*k*td +
            l+1];
            irir = row1[td2s*td3s*td + 2*j*td3s*td + (2*k+1)*td +
            l];
            irii = row1[td2s*td3s*td + 2*j*td3s*td + (2*k+1)*td +
            l+1];

            iirr = row1[td2s*td3s*td + (2*j+1)*td3s*td + 2*k*td +
            l];
            iiri = row1[td2s*td3s*td + (2*j+1)*td3s*td + 2*k*td +
            l+1];
            iirr = row1[td2s*td3s*td + (2*j+1)*td3s*td +
            (2*k+1)*td + 1];
            iiii = row1[td2s*td3s*td + (2*j+1)*td3s*td +
            (2*k+1)*td + 1+1];

            dur = - g_delay * dw / 2 + l * dw / 2;

            arg1 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
            * j + k )*(end_c-start_c)/tot_points) * dur * 2 * M_PI;
            c1 = cos(arg1);/*direct part*/
            s1 = sin(arg1);/*direct part*/

            crrrr = c1*rrrr - s1*rrri;
            crrri = s1*rrrr + c1*rrri;
            crrir = c1*rrir - s1*rrii;
            crrii = s1*rrir + c1*rrii;

```

```

        crirr = c1*rirrr - s1*rirri;
        criri = s1*rirrr + c1*rirri;
        criir = c1*riirr - s1*riiii;
        criii = s1*riirr + c1*riiii;

        cirrr = c1*irrrr - s1*irrii;
        cirri = s1*irrrr + c1*irrii;
        cirir = c1*irrir - s1*iriii;
        cirii = s1*irrir + c1*iriii;

        ciirr = c1*iiirr - s1*iiiri;
        ciiri = s1*iiirr + c1*iiiri;
        ciiir = c1*iiiiir - s1*iiiiii;
        ciiii = s1*iiiiir + c1*iiiiii; /*corrected direct dim*/

        dur2 = in_f3*k; /*fast indirect dim*/
        arg2 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
* j + k )*(end_c-start_c)/tot_points)*dur2*2*M_PI*bf3/bf_dir;
        if (fnmode3s == 1) { /*QF acquisition mode (pseudo 4D
data)*/

            arg2 = 0;
        }
        c2 = cos(arg2);
        s2 = sin(arg2);

        crrrrr = c2*crrrrr - s2*crrrir;
        crrrri = c2*crrrri - s2*crrrii;
        crrrir = s2*crrrrr + c2*crrrir;
        crrrii = s2*crrrri + c2*crrrii;

        crirrr = c2*crirrr - s2*criirr;
        crirri = c2*crirri - s2*criiii;
        criirr = s2*crirrr + c2*criirr;
        criiii = s2*crirri + c2*criiii;

        ccirrr = c2*cirrrr - s2*cirrir;
        ccirri = c2*cirrii - s2*cirrii;
        ccirir = s2*cirrrr + c2*cirrir;
        ccirii = s2*cirrii + c2*cirrii;

        cciirr = c2*ciirr - s2*ciirr;
        cciiri = c2*ciiri - s2*ciiii;
        cciirr = s2*ciirr + c2*ciirr;
        cciiii = s2*ciiri + c2*ciiii; /*corrected indirect
dim 1*/

        dur3 = in_f2*j; /*next slowest indirect dim.
elapsed time is the same: (td2s / 2 * td1s / 2 * i + td1s / 2 * j + k )*/
        arg3 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
* j + k )*(end_c-start_c)/tot_points)*dur3*2*M_PI*bf2/bf_dir;
        if (fnmode2s == 1) { /*QF acquisition mode (pseudo 4D
data)*/

            arg3 = 0;
        }
        c3 = cos(arg3);
        s3 = sin(arg3);

```

```

ccccrrr = c3*ccccrrr - s3*ccccrri;
ccccrri = c3*ccccrri - s3*ccccrir;
ccccrir = c3*ccccrir - s3*ccccrii;
ccccrrii = c3*ccccrrii - s3*ccccriii;

ccccrirr = s3*ccccrrr +c3*ccccrri;
ccccriri = s3*ccccrri +c3*ccccrir;
ccccriir = s3*ccccrir +c3*ccccrii;
ccccriii = s3*ccccrrii +c3*ccccriii;

ccccirrr = c3*ccccirrr - s3*cccciirr;
ccccirri = c3*ccccirri - s3*cccciiri;
ccccirir = c3*ccccirir - s3*cccciiri;
ccccirii = c3*ccccirii - s3*cccciiii;

cccciirr = s3*ccccirrr +c3*cccciirr;
cccciiri = s3*ccccirri +c3*cccciiri;
cccciirir = s3*ccccirir +c3*cccciiri;
cccciiii = s3*ccccirii +c3*cccciiii;

/*slowest dim*/
dur4 = in_f1*i; /* elapsed time is the same */
arg4 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
* j + k )*(end_c-start_c)/tot_points)*dur4*2*M_PI*bf1/bf_dir;
if (fnmodels == 1) { /*QF acquisition mode (pseudo 4D
data)*/
    arg4 = 0;
}
c4 = cos(arg4);
s4 = sin(arg4);

/*remove i, since reading blocks of data*/
rowo[ 2*j*td3s*td + 2*k*td + 1] = c4*ccccrrr -
s4*ccccirrr ; /*rrrr*/
rowo[2*j*td3s*td + 2*k*td + 1+1]= c4*ccccrri -
s4*ccccirri; /*rrri*/
rowo[2*j*td3s*td + (2*k+1)*td + 1]= c4*ccccrir -
s4*ccccirir; /*rrir*/
rowo[2*j*td3s*td + (2*k+1)*td + 1+1]= c4*ccccrrii -
s4*ccccirii; /*rrii*/

rowo[(2*j+1)*td3s*td + 2*k*td + 1]= c4*ccccrirr -
s4*cccciirr; /*rirr*/
rowo[(2*j+1)*td3s*td + 2*k*td + 1+1]=c4*ccccriri -
s4*cccciiri; /*riri*/
rowo[(2*j+1)*td3s*td + (2*k+1)*td + 1]=c4*ccccriir -
s4*cccciirir; /*riir*/
rowo[(2*j+1)*td3s*td + (2*k+1)*td + 1+1]=c4*ccccriii -
s4*cccciiii; /*riii*/

rowo[td2s*td3s*td + 2*j*td3s*td + 2*k*td + 1] =
s4*ccccrrrr + c4*ccccirrr; /*irrr*/
rowo[td2s*td3s*td + 2*j*td3s*td + 2*k*td + 1+1]=
s4*ccccrrri + c4*ccccirri; /*irri*/
rowo[td2s*td3s*td + 2*j*td3s*td + (2*k+1)*td + 1]=
s4*ccccrrir + c4*ccccirir; /*irir*/

```

```

        rowo[td2s*td3s*td + 2*j*td3s*td + (2*k+1)*td + l+1]=
s4*cccrrii + c4*cccirri; /*irri*/

        rowo[td2s*td3s*td + (2*j+1)*td3s*td + 2*k*td + l]=
s4*cccrirr + c4*ccciirr; /*iirr*/
        rowo[td2s*td3s*td + (2*j+1)*td3s*td + 2*k*td +
l+1]=s4*cccriri + c4*ccciiri; /*iiri*/
        rowo[td2s*td3s*td + (2*j+1)*td3s*td + (2*k+1)*td +
l]=s4*cccriir + c4*ccciir; /*iir*/
        rowo[td2s*td3s*td + (2*j+1)*td3s*td + (2*k+1)*td +
l+1]=s4*cccriii + c4*ccciiii; /*iiii*/
    }
    }
    }
    fwrite(rowo, sizeof(int), td * td3s * td2s * 2, fpout);
    free(rowl);
    free(rowo);
}
}
/*end for data recorded in planes*/
} else if ( fntype==2 ) { /****NUS DATA****/
    if (parmode == 2 ) { /****3D data****/
        if (aqseq == 1) { /***3D case with aqseq of 312, not used in Topspin
3.5, always recorded as aqseq = 321***/
            tot_points = nus_points * 4;

            rowl = (int*)malloc( td * 4* nus_points * sizeof(int));
            rowo = (int*)malloc( td * 4* nus_points * sizeof(int));
            fread(rowl, sizeof(int), td *4* nus_points, fpin);

            for (j = 0; j < nus_points * 4; j+=4) { /*indirect dim
loop1,slow*/
                fscanf(fp, "%d %d", &cur_pointf2, &cur_pointf1); /*read in NUS
file*/

                for (l = 0; l < td; l += 2) { /*direct dim loop*/
                    int rrr, rri, rir, rii, irr, iri, iir, iii;
                    double arg1, arg2, arg3, c1, s1, c2, s2, c3, s3,
crrr,crri,crir,crii,cirr,ciri,ciir,cciii;
                    double ccrrr,ccrri,ccrir,ccrii,ccirr,cciri,cciiir,cciii;
                    rrr = rowl[j*td + l];
                    rri = rowl[j*td + l+1];
                    rir = rowl[(j+1)*td + l];
                    rii = rowl[(j+1)*td + l+1];
                    irr = rowl[(j+2)*td + l];
                    iri = rowl[(j+2)*td + l+1];
                    iir = rowl[(j+3)*td + l];
                    iii = rowl[(j+3)*td + l+1];

                    dur = - g_delay * dw / 2 + l * dw / 2;

                    arg1 = (start_c + j * (end_c-start_c)/tot_points) * dur *
2 * M_PI;

                    c1 = cos(arg1);/*direct part*/
                    s1 = sin(arg1);/*direct part*/

                    crrr = c1*rrr - s1*rri;
                    crri = s1*rrr + c1*rri;

```

```

        crir = c1*rir - s1*rii;
        crii = s1*rir + c1*rii;

        cirr = c1*irr - s1*iri;
        ciri = s1*irr + c1*iri;
        ciir = c1*iir - s1*iii;
        ciii = s1*iir + c1*iii; /*corrected direct dim*/

        dur2 = in_f1*cur_pointf1; /*fast indirect dim*/
        arg2 = (start_c + j * (end_c-
start_c)/tot_points)*dur2*2*M_PI*bf1/bf_dir;
        c2 = cos(arg2);
        s2 = sin(arg2);

        crrr = c2*crrr - s2*crir;
        crri = c2*crri - s2*crii;
        crir = s2*crrr + c2*crir;
        crii = s2*crri + c2*crii;

        ccirr = c2*cirr - s2*ciir;
        cciri = c2*ciri - s2*ciii;
        cciir = s2*cirr + c2*ciir;
        cciii = s2*ciri + c2*ciii; /*corrected indirect dim 1*/

        dur3 = in_f2*cur_pointf2; /*slow indirect dim. elapsed
time is the same: (td2s / 2 * j + k)*/
        arg3 = (start_c + j * (end_c-
start_c)/tot_points)*dur3*2*M_PI*bf2/bf_dir;
        c3 = cos(arg3);
        s3 = sin(arg3);

        rowo[j*td + 1] = c3*crrr - s3*ccirr ;
        rowo[j*td + 1 + 1]= c3*crri - s3*cciri;
        rowo[(j+1)*td + 1]= c3*crir - s3*cciir;
        rowo[(j+1)*td + 1 + 1]= c3*crii - s3*cciii;

        rowo[(j+2)*td + 1]= s3*crrr +c3*ccirr;
        rowo[(j+2)*td + 1 + 1]= s3*crri +c3*cciri;
        rowo[(j+3)*td + 1]= s3*crir +c3*cciir;
        rowo[(j+3)*td + 1 + 1]= s3*crii +c3*cciii;
        /*corrected indirect as well*/
    }
}
fwrite(rowo, sizeof(int), td * nus_points * 4, fpout);
free(row1);
free(rowo);
} else if (aqseq == 0) { /**3D case with aqseq of 321**/

    tot_points = nus_points * 4;

    row1 = (int*)malloc( td * 4* nus_points * sizeof(int));
    rowo = (int*)malloc( td * 4* nus_points * sizeof(int));
    fread(row1, sizeof(int), td *4* nus_points, fpin);

    for (j = 0; j < nus_points * 4; j+=4) { /*indirect dim
loop1,slow*/

```

```

fscanf(fp, "%d %d", &cur_pointf2, &cur_pointf1); /*read in NUS
file*/

for (l = 0; l < td; l += 2) { /*direct dim loop*/
    int rrr, rri, rir, rii, irr, iri, iir, iii;
    double arg1, arg2, arg3, c1, s1, c2, s2, c3, s3,
crrr, crri, crir, crie, cirr, ciri, ciir, ciii;
    double ccrrr, ccrrr, ccrir, ccree, ccirr, cciri, cciir, cciii;
    rrr = row1[j*td + l];
    rri = row1[j*td + l+1];
    rir = row1[(j+1)*td + l];
    rii = row1[(j+1)*td + l+1];
    irr = row1[(j+2)*td + l];
    iri = row1[(j+2)*td + l+1];
    iir = row1[(j+3)*td + l];
    iii = row1[(j+3)*td + l+1];

    dur = - g_delay * dw / 2 + l * dw / 2;

    arg1 = (start_c + j * (end_c-start_c)/tot_points) * dur *
2 * M_PI;

    c1 = cos(arg1);/*direct part*/
    s1 = sin(arg1);/*direct part*/

    crrr = c1*rrr - s1*rri;
    crri = s1*rrr + c1*rri;
    crir = c1*rir - s1*rii;
    crie = s1*rir + c1*rii;

    cirr = c1*irr - s1*iri;
    ciri = s1*irr + c1*iri;
    ciir = c1*iir - s1*iii;
    ciii = s1*iir + c1*iii; /*corrected direct dim*/

    dur2 = in_f2*cur_pointf2; /*fast indirect dim*/
    arg2 = (start_c + j * (end_c-
start_c)/tot_points)*dur2*2*M_PI*bf2/bf_dir;
    c2 = cos(arg2);
    s2 = sin(arg2);

    ccrrr = c2*crrr - s2*crir;
    ccrrr = c2*crri - s2*crie;
    ccrir = s2*crrr + c2*crir;
    ccree = s2*crri + c2*crie;

    ccirr = c2*cirr - s2*ciir;
    cciri = c2*ciri - s2*ciii;
    cciir = s2*cirr + c2*ciir;
    cciii = s2*ciri + c2*ciii; /*corrected indirect dim 1*/

    dur3 = in_f1*cur_pointf1; /*slow indirect dim. elapsed
time is the same: (td2s / 2 * j + k)*/
    arg3 = (start_c + j * (end_c-
start_c)/tot_points)*dur3*2*M_PI*bf1/bf_dir;
    c3 = cos(arg3);
    s3 = sin(arg3);

```

```

        rowo[j*td + 1] = c3*ccrrr - s3*ccirr ;
        rowo[j*td + 1 + 1]= c3*ccrri - s3*cciri;
        rowo[(j+1)*td + 1]= c3*ccrir - s3*cciiir;
        rowo[(j+1)*td + 1 + 1]= c3*ccrii - s3*cciii;

        rowo[(j+2)*td + 1]= s3*ccrrr +c3*ccirr;
        rowo[(j+2)*td + 1 + 1]= s3*ccrri +c3*cciri;
        rowo[(j+3)*td + 1]= s3*ccrir +c3*cciiir;
        rowo[(j+3)*td + 1 + 1]= s3*ccrii +c3*cciii;
        /*corrected indirect as well*/
    }
}
fwrite(rowo, sizeof(int), td * nus_points * 4, fpout);
free(row1);
free(rowo);
}
} else if (parmode == 3 ) {  /***4D NUS case, aqseq of 4321***/

    tot_points = nus_points * 8;

    row1 = (int*)malloc(td * nus_points * 8 * sizeof(int));
    if (row1 == 0)
        STOPMSG("cannot get enough memory")
    rowo = (int*)malloc(td * nus_points * 8 * sizeof(int));
    if (rowo == 0)
        STOPMSG("cannot get enough memory")
    fread(row1, sizeof(int), td * nus_points * 8, fpin);

    for (j = 0; j < nus_points * 8; j+=8) {
        fscanf(fp, "%d %d %d", &cur_pointf3, &cur_pointf2, &cur_pointf1);
/*read in NUS file*/
        for (l = 0; l < td; l += 2) { /*direct dim loop*/
            int rrrr, rrrr, rrrr, rrrr, rrrr, rrrr, rrrr, rrrr,
irri, irir, irii, iirr, iiri, iirr, iiii;
            double crrrr, crrri, crrir, crrii, crirr, criri, criir, crii,
cirrr, cirri, cirir, cirii, ciirr, ciiri, ciiir, ciiii;
            double ccrrrr, ccrrri, ccrrir, ccrrii, ccrrrr, ccrrri, ccrrir,
ccriii, ccirrr, ccirri, ccirir, ccirii, cciiir, cciiii;
            double cccrrrr, cccrrri, cccrrir, cccrrii, cccrrrr, cccrrri,
cccriir, cccriii, cccirrr, cccirri, cccirir, cccirii, ccciiir, ccciiir,
ccciirr, ccciiii;
            double arg1, arg2, arg3, arg4, c1, s1, c2, s2, c3, s3, c4, s4;

/*remove i, since now reading in blocks of data*/
            rrrr = row1[j*td + 1];
            rrrr = row1[j*td + 1+1];
            rrrr = row1[(j+1)*td + 1];
            rrrr = row1[(j+1)*td + 1+1];

            rrrr = row1[(j+2)*td + 1];
            rrrr = row1[(j+2)*td + 1+1];
            rrrr = row1[(j+3)*td + 1];
            rrrr = row1[(j+3)*td + 1+1];

            rrrr = row1[(j+4)*td + 1];
            rrrr = row1[(j+4)*td + 1+1];
            rrrr = row1[(j+5)*td + 1];

```

```

irii = row1[(j+5)*td + l+1];

iirr = row1[(j+6)*td + l];
iiri = row1[(j+6)*td + l+1];
iiir = row1[(j+7)*td + l];
iiii = row1[(j+7)*td + l+1];

dur = - g_delay * dw / 2 + l * dw / 2;

arg1 = (start_c + j*(end_c-start_c)/tot_points) * dur * 2 *
M_PI;
c1 = cos(arg1); /*direct*/
s1 = sin(arg1); /*direct*/

crrrr = c1*rrrr - s1*rrri;
crrri = s1*rrrr + c1*rrri;
crrir = c1*rrir - s1*rrii;
crrii = s1*rrir + c1*rrii;

crirr = c1*rirr - s1*riri;
criri = s1*rirr + c1*riri;
criir = c1*riir - s1*riii;
criii = s1*riir + c1*riii;

cirrr = c1*irrr - s1*irri;
cirri = s1*irrr + c1*irri;
cirir = c1*irir - s1*irii;
cirii = s1*irir + c1*irii;

ciirr = c1*iiir - s1*iiiri;
ciiiri = s1*iiir + c1*iiiri;
ciiir = c1*iiii - s1*iiii;
ciiii = s1*iiii + c1*iiii; /*corrected direct dimension*/

dur2 = in_f3*cur_pointf3;
arg2 = (start_c + j*(end_c-
start_c)/tot_points)*dur2*2*M_PI*bf3/bf_dir;
c2 = cos(arg2);
s2 = sin(arg2);

ccrrrr = c2*crrrr - s2*crrir;
ccrrri = c2*crrri - s2*crrii;
ccrrir = s2*crrrr + c2*crrir;
ccrrii = s2*crrri + c2*crrii;

ccrirr = c2*crirr - s2*criir;
ccriri = c2*criri - s2*criii;
ccriir = s2*crirr + c2*criir;
ccriii = s2*criri + c2*criii;

ccirrr = c2*cirrr - s2*cirir;
ccirri = c2*cirri - s2*cirii;
ccirir = s2*cirrr + c2*cirir;
ccirii = s2*cirri + c2*cirii;

cciiir = c2*ciirr - s2*ciiir;
cciiiri = c2*ciiri - s2*ciiii;

```

```

cciiir = s2*ciirr + c2*ciirr;
cciiii = s2*ciiri + c2*ciiii; /*corrected indirect dim3*/

dur3 = in_f2*cur_pointf2;
arg3 = (start_c + j*(end_c-
start_c)/tot_points)*dur3*2*M_PI*bf2/bf_dir;
c3 = cos(arg3);
s3 = sin(arg3);

ccrrrr = c3*ccrrrr - s3*ccrirr;
ccrrri = c3*ccrrri - s3*ccriri;
ccrrir = c3*ccrrir - s3*ccriir;
ccrrii = c3*ccrrii - s3*ccriii;

ccrirr = s3*ccrrrr +c3*ccrirr;
ccriri = s3*ccrrri +c3*ccriri;
ccriir = s3*ccrrir +c3*ccriir;
ccriii = s3*ccrrii +c3*ccriii;

cccirrr = c3*ccirrr - s3*cciiir;
cccirri = c3*ccirri - s3*cciiiri;
cccirir = c3*ccirir - s3*cciiir;
cccirii = c3*ccirii - s3*cciiii;

ccciirr = s3*ccirrr +c3*cciiir;
ccciiri = s3*ccirri +c3*cciiiri;
ccciir = s3*ccirir +c3*cciiir;
cciiii = s3*ccirii +c3*cciiii; /*corrected indirect dim2*/

/*slowest dim*/
dur4 = in_f1*cur_pointf1; /*elapsed time is the same*/
arg4 = (start_c + j*(end_c-
start_c)/tot_points)*dur4*2*M_PI*bf1/bf_dir;
c4 = cos(arg4);
s4 = sin(arg4);

rowo[j*td + 1] = c4*ccrrrr - s4*cccirrr ; /*rrrr*/
rowo[j*td + 1+1]= c4*ccrrri - s4*cccirri; /*rrri*/
rowo[(j+1)*td + 1]= c4*ccrrir - s4*cccirir; /*rrir*/
rowo[(j+1)*td + 1+1]= c4*ccrrii - s4*cccirii; /*rrii*/

rowo[(j+2)*td + 1]= c4*ccrirr - s4*ccciirr; /*rirr*/
rowo[(j+2)*td + 1+1]=c4*ccriri - s4*ccciiri; /*riri*/
rowo[(j+3)*td + 1]=c4*ccriir - s4*ccciir; /*riir*/
rowo[(j+3)*td + 1+1]=c4*ccriii - s4*cciiii; /*riii*/

rowo[(j+4)*td + 1] = s4*ccrrrr + c4*cccirrr; /*irrr*/
rowo[(j+4)*td + 1+1]= s4*ccrrri + c4*cccirri; /*irri*/
rowo[(j+5)*td + 1]= s4*ccrrir + c4*cccirir; /*irir*/
rowo[(j+5)*td + 1+1]= s4*ccrrii + c4*cccirii; /*irii*/

rowo[(j+6)*td + 1]= s4*ccrirr + c4*ccciirr; /*iirr*/
rowo[(j+6)*td + 1+1]= s4*ccriri + c4*ccciiri; /*iiri*/
rowo[(j+7)*td + 1]= s4*ccriir + c4*ccciir; /*iirir*/
rowo[(j+7)*td + 1+1]= s4*ccriii + c4*cciiii; /*iiii*/
/*corrected indirect dim1*/

```

```

    }
    }
    fwrite(rowo, sizeof(int), td * nus_points * 8, fpout);
    free(rowl);
    free(rowo);
}

} else if ( fntype==1 ) { /****DATA RECORDED IN POINTS****/
    if (parmode == 2 ) { /****3D data****/
        if (aqseq == 0) { /**aqseq of 321**/

            tot_points = td1s / 2 * td2s / 2 ; /*number of complex points*/

            int rrr, rri, rir, rii, irr, iri, iir, iii;
            double arg1, arg2, arg3, c1, s1, c2, s2, c3, s3,
crrr,crri,crir,crii,cirr, cirr, ciir, ciii;
            double ccrrr,ccrri,ccrir,ccrii,ccirr, cciri, cciir, cciii;

            rowl = (int*)malloc( td * 4 * sizeof(int));
            rowo = (int*)malloc( td * 4 * sizeof(int));

            for (j = 0; j < td1s / 2; j++) {
                for (k = 0; k < td2s / 2; k++) {
                    fread(rowl, sizeof(int), td * 4, fpin);
                    for (l = 0; l < td; l += 2) {

                        rrr = rowl[l];
                        rri = rowl[l+1];
                        rir = rowl[td + l];
                        rii = rowl[td + l+1];
                        irr = rowl[2*td + l];
                        iri = rowl[2*td + l+1];
                        iir = rowl[3*td + l];
                        iii = rowl[3*td + l+1];

                        dur = - g_delay * dw / 2 + l * dw / 2;

                        arg1 = (start_c + j * (end_c-start_c)/tot_points) *
dur * 2 * M_PI;

                        c1 = cos(arg1);/*direct*/
                        s1 = sin(arg1);/*direct*/

                        crrr = c1*rrr - s1*rri;
                        crri = s1*rrr + c1*rri;
                        crir = c1*rir - s1*rii;
                        crii = s1*rir + c1*rii;

                        cirr = c1*irr - s1*iri;
                        ciri = s1*irr + c1*iri;
                        ciir = c1*iir - s1*iii;
                        ciii = s1*iir + c1*iii; /*corrected direct dim*/

                        dur2 = in_f2*k; /*fast indirect dim*/
                        arg2 = (start_c + (td2s / 2 * j + k) * (end_c-
start_c)/tot_points)*dur2*2*M_PI*bf2/bf_dir;
                        if (fnmode2s == 1) { /*QF acquisition mode (pseudo 3D
data)*/

```

```

        arg2 = 0;
    }
    c2  = cos(arg2);
    s2  = sin(arg2);

    crrr = c2*crrr - s2*crir;
    crri = c2*crri - s2*cric;
    crir = s2*crrr + c2*crir;
    cric = s2*crri + c2*cric;

    ccir = c2*ccir - s2*ccri;
    ccri = c2*ccri - s2*ccir;
    ccir = s2*ccir + c2*ccri;
    ccir = s2*ccri + c2*ccir;

    ccir = c2*ccir - s2*ccri;
    ccir = c2*ccri - s2*ccir;
    ccir = s2*ccir + c2*ccri;
    ccir = s2*ccri + c2*ccir; /*corrected fast indirect
dim*/

        dur3 = in_f1*j; /*slow indirect dim. elapsed time
is the same: (td2s / 2 * j + k)*
        arg3 = (start_c + (td2s / 2 * j + k)*(end_c-
start_c)/tot_points)*dur3*2*M_PI*bf1/bf_dir;
        if (fnmodels == 1) { /*QF acquisition mode (pseudo 3D
data)*/

            arg3 = 0;
        }
        c3  = cos(arg3);
        s3  = sin(arg3);

        rowo[1] = c3*crrr - s3*ccir ;
        rowo[1+1]= c3*crri - s3*ccri;
        rowo[td + 1]= c3*crir - s3*ccir;
        rowo[td + 1+1]= c3*cric - s3*ccir;

        rowo[2*td + 1]= s3*crrr +c3*ccir;
        rowo[2*td + 1+1]= s3*crri +c3*ccri;
        rowo[3*td + 1]= s3*crir +c3*ccir;
        rowo[3*td + 1+1]= s3*cric +c3*ccir; /*corrected
slow indirect dim*/
    }
    fwrite(rowo, sizeof(int), td *4, fpout);
}
}
free(row1);
free(rowo);
}

if (aqseq == 1) { /**aqseq of 312**/

    tot_points = td1s / 2 * td2s / 2 ; /*number of complex points*/

    int rrr, rri, rir, rii, irr, iri, iir, iii;
    double arg1, arg2, arg3, c1, s1, c2, s2, c3, s3,
crrr,crri,crir,cric,ccir, ccir, ccir, ccir, ccir, ccir, ccir, ccir;
    double crrr,crri,crir,cric,ccir, ccir, ccir, ccir, ccir, ccir, ccir;

    row1 = (int*)malloc( td * 4 * sizeof(int));
    rowo = (int*)malloc( td * 4 * sizeof(int));

```

```

for (j = 0; j < td2s / 2; j++) {
  for (k = 0; k < td1s / 2; k++) {
    fread(row1, sizeof(int), td * 4, fpin);
    for (l = 0; l < td; l += 2) {

      rrr = row1[l];
      rri = row1[l+1];
      rir = row1[td + l];
      rii = row1[td + l+1];
      irr = row1[2*td + l];
      iri = row1[2*td + l+1];
      iir = row1[3*td + l];
      iii = row1[3*td + l+1];

      dur = - g_delay * dw / 2 + l * dw / 2;

      arg1 = (start_c + j * (end_c-start_c)/tot_points) *
dur * 2 * M_PI;

      c1 = cos(arg1);/*direct*/
      s1 = sin(arg1);/*direct*/

      crrr = c1*rrr - s1*rri;
      crri = s1*rrr + c1*rri;
      crir = c1*rir - s1*rii;
      crii = s1*rir + c1*rii;

      cirr = c1*irr - s1*iri;
      ciri = s1*irr + c1*iri;
      ciir = c1*iir - s1*iii;
      ciii = s1*iir + c1*iii; /*corrected direct dim*/

      dur2 = in_f1*k; /*fast indirect dim*/
      arg2 = (start_c + (td1s / 2 * j + k) * (end_c-
start_c)/tot_points)*dur2*2*M_PI*bf1/bf_dir;
      if (fnmodels == 1) { /*QF acquisition mode (pseudo 3D
data)*/
        arg2 = 0;
      }
      c2 = cos(arg2);
      s2 = sin(arg2);

      ccrrr = c2*crrr - s2*crir;
      ccrrr = c2*crrr - s2*crir;
      ccrrr = s2*crrr + c2*crir;
      ccrrr = s2*crrr + c2*crir;

      ccirr = c2*cirr - s2*ciir;
      ccirr = c2*cirr - s2*ciir;
      ccirr = s2*cirr + c2*ciir;
      ccirr = s2*cirr + c2*ciir;

      cciir = s2*ciri + c2*ciir;
      cciir = s2*ciri + c2*ciir;

      cciir = s2*ciri + c2*ciir;
      cciir = s2*ciri + c2*ciir; /*corrected indirect dim
1*/

      dur3 = in_f2*j; /*slow indirect dim. elapsed time is
the same: (td2s / 2 * j + k)*/
      arg3 = (start_c + (td1s / 2 * j + k) * (end_c-
start_c)/tot_points)*dur3*2*M_PI*bf2/bf_dir;

```



```

riir = row1[3*td + 1];
riii = row1[3*td + 1+1];

irrr = row1[4*td + 1];
irri = row1[4*td + 1+1];
irir = row1[5*td + 1];
irii = row1[5*td + 1+1];

iirr = row1[6*td + 1];
iiri = row1[6*td + 1+1];
iiir = row1[7*td + 1];
iiii = row1[7*td + 1+1];

dur = - g_delay * dw / 2 + 1 * dw / 2;

arg1 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
* j + k )*(end_c-start_c)/tot_points) * dur * 2 * M_PI;
c1 = cos(arg1);/*direct*/
s1 = sin(arg1);/*direct*/

crrrr = c1*rrrr - s1*rrri;
crrri = s1*rrrr + c1*rrri;
crrir = c1*rrir - s1*rrii;
crrii = s1*rrir + c1*rrii;

crirr = c1*rirr - s1*riri;
criri = s1*rirr + c1*riri;
criir = c1*riir - s1*riii;
criii = s1*riir + c1*riii;

cirrr = c1*irrr - s1*irri;
cirri = s1*irrr + c1*irri;
cirir = c1*irir - s1*irii;
cirii = s1*irir + c1*irii;

ciirr = c1*iirr - s1*iiiri;
ciiiri = s1*iirr + c1*iiiri;
ciiir = c1*iiir - s1*iiii;
ciiii = s1*iiir + c1*iiii; /*corrected direct dim*/

dur2 = in_f3*k; /*fast indirect dim*/
arg2 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
* j + k )*(end_c-start_c)/tot_points)*dur2*2*M_PI*bf3/bf_dir;
if (fnmode3s == 1) { /*QF acquisition mode (pseudo 4D
data)*/

    arg2 = 0;
}
c2 = cos(arg2);
s2 = sin(arg2);

ccrrrr = c2*crrrr - s2*crrir;
ccrrri = c2*crrri - s2*crrii;
ccrrir = s2*crrrr + c2*crrir;
ccrrii = s2*crrri + c2*crrii;

ccrirr = c2*crirr - s2*criir;
ccriri = c2*criri - s2*criii;

```

```

ccriir = s2*crirr + c2*criir;
ccriii = s2*criri + c2*criii;

ccirrr = c2*cirrr - s2*cirir;
ccirri = c2*cirri - s2*cirii;
ccirir = s2*cirrr + c2*cirir;
ccirii = s2*cirri + c2*cirii;

cciiir = c2*ciirr - s2*ciiii;
cciiiri = c2*ciiri - s2*ciiii;
cciiir = s2*ciirr + c2*ciiii;
cciiii = s2*ciiri + c2*ciiii; /*corrected fast
indirect dim*/

dur3 = in_f2*j; /*next slowest indirect dim.
elapsed time is the same: (td2s / 2 * td1s / 2 * i + td1s / 2 * j + k )*/
arg3 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
* j + k )*(end_c-start_c)/tot_points)*dur3*2*M_PI*bf2/bf_dir;
if (fnmode2s == 1) { /*QF acquisition mode (pseudo 4D
data)*/

arg3 = 0;
}
c3 = cos(arg3);
s3 = sin(arg3);

ccrrrr = c3*ccrrrr - s3*ccrirr;
ccrrri = c3*ccrrri - s3*ccriri;
ccrrir = c3*ccrrir - s3*ccriir;
ccrrii = c3*ccrrii - s3*ccriii;

ccrirr = s3*ccrrrr +c3*ccrirr;
ccriri = s3*ccrrri +c3*ccriri;
ccriir = s3*ccrrir +c3*ccriir;
ccriii = s3*ccrrii +c3*ccriii;

cccirrr = c3*cccirrr - s3*cciiir;
cccirri = c3*cccirri - s3*cciiiri;
cccirir = c3*cccirir - s3*cciiiiir;
cccirii = c3*cccirii - s3*cciiiiii;

ccciirr = s3*cccirrr +c3*ccciirr;
ccciiri = s3*cccirri +c3*ccciiri;
ccciir = s3*cccirir +c3*ccciir;
cciiii = s3*cccirii +c3*cciiii; /*corrected next
slowest indirect dim*/

/*slowest indirect dim*/
dur4 = in_f1*i; /* elapsed time is the same */
arg4 = (start_c + (td2s / 2 * td3s / 2 * i + td3s / 2
* j + k )*(end_c-start_c)/tot_points)*dur4*2*M_PI*bf1/bf_dir;
if (fnmode1s == 1) { /*QF acquisition mode (pseudo 4D
data)*/

arg4 = 0;
}
c4 = cos(arg4);
s4 = sin(arg4);

```

```

/*remove i, since reading blocks of data*/
rowo[1] = c4*cccrrrr - s4*cccirrr ; /*rrrr*/
rowo[1+1]= c4*cccrrri - s4*cccirri; /*rrri*/
rowo[td + 1]= c4*cccrrir - s4*cccirir; /*rrir*/
rowo[td + 1+1]= c4*cccrrii - s4*cccirii; /*rrii*/

rowo[2*td + 1]= c4*cccrrrr - s4*cccirrr; /*rirr*/
rowo[2*td + 1+1]= c4*cccrrri - s4*cccirri; /*riri*/
rowo[3*td + 1]= c4*cccrrir - s4*cccirir; /*riir*/
rowo[3*td + 1+1]= c4*cccrrii - s4*cccirii; /*riii*/

rowo[4*td + 1] = s4*cccrrrr + c4*cccirrr; /*irrr*/
rowo[4*td + 1+1]= s4*cccrrri + c4*cccirri; /*irri*/
rowo[5*td + 1]= s4*cccrrir + c4*cccirir; /*irir*/
rowo[5*td + 1+1]= s4*cccrrii + c4*cccirii; /*irii*/

rowo[6*td + 1]= s4*cccrrrr + c4*cccirrr; /*iirr*/
rowo[6*td + 1+1]=s4*cccrrri + c4*cccirri; /*iiri*/
rowo[7*td + 1]=s4*cccrrir + c4*cccirir; /*iirr*/
rowo[7*td + 1+1]=s4*cccrrii + c4*cccirii; /*iiii*/ /*
corrected slowest indirect dim*/
    }
    fwrite(rowo, sizeof(int), td * 8, fpout);
}
}
}
free(row1);
free(rowo);
}
}

fclose(fpin);
fclose(fpout);

if ( fntype==2 ) { /*close NUS file if NUS data*/
    fclose(fp);
}

Show_status("done");

REFRESHDATA
QUIT

```

5.3.2 Script for concatenation of 3D and 4D experiments

```
/*
*****
/*      concatenate_ser_nuslist                10.04.2019      */
/*
*****
/*      Short Description :                      */
/*      concatenates nuslists and ser files from two experiments*/
/*      available for 3D and 4D spectra          */
/*
*****
/*      Description/Usage :                     */
/*      The AU program concatenates the nuslist and ser file   */
/*      of a given measurement to the current dataset.        */
/*      If a 3D or 4D measurement has to be split up due to   */
/*      sampling limitation, it can be done by generating a    */
/*      nuslist for points sampling. After recording          */
/*      the blocks individually, drift correction can be      */
/*      applied and the nuslist files and ser files can be    */
/*      concatenated.                                         */
/*      The spectra can be processed by changing the status   */
/*      parameters of FnType to points, and the                */
/*      number of points to correspond to how the complete    */
/*      measurement was set up.                               */
/*      Save the program (depending on your Topspin version)  */
/*      to Bruker/Topspin_*version*_exp/stan/nmr/au/src/user  */
/*      The script can be started by typing the name of the   */
/*      script at the command line. Tested in Topspin 3.5pl6  */
/*      To concatenate multiple files, apply the AU program   */
/*      multiple times consecutively.                         */
/*      THE AU PROGRAM WILL OVERWRITE THE CURRENT DATASET!!!  */
/*      Note that the NUS option may ignore definition of aqseq*/
/*      in the pulse sequence, and record the default of 321.*/
/*      Verify by checking the status parameter "s aqseq".    */
/*      The NUS list must be set up accordingly.              */
/*      Example NUS list before splitting:                   */
/*      0 0; 1 0; 2 0;...; n 0; 0 1; 1 1; 2 1; ...; n m      */
/*      n = td(F2)/2 -1; m = td(F1)/2 - 1                    */
/*
*****
/*      Author(s) :                                         */
/*      Name          : Eszter E. Najbauer                   */
/*      Organisation  : Max Planck Institute for             */
/*                    Biophysical Chemistry                 */
/*      Email         : esna@nmr.mpibpc.mpg.de              */
/*
/*      Name          : Loren B. Andreas                    */
/*      Organisation  : Max Planck Institute for             */
/*                    Biophysical Chemistry                 */
/*      Email         : land@nmr.mpibpc.mpg.de              */
/*      We thank Wolfgang Bermel for his technical assistance.*/
/*      Citation: Najbauer and Andreas: JMR 2019            */
*****

```

```
char infile1[PATH_MAX], infile2[PATH_MAX];
char path[PATH_MAX + 64], modetext[80], autext[256];
int cur_expno = expno;
int byteorder, parmode, concatexpno;
int fntype = 0;
```

```

int fnmodels, fnmode2s, fnmode3s;
int nus_points;
int cur_pointf1, cur_pointf2, cur_pointf3;
int *row1;
int j;
char datapath1[1024]; /*for nuslist*/
char datapath2[1024];
const char* modepnt;
int td, tds;
int td2s, td3s;
FILE *fpnus1, *fpnus2;
FILE *fpser1, *fpser2;
char ch;

/***** get dataset and parameters *****/

FETCHPARS("PARMODE", &parmode)

if (parmode != 2 && parmode !=3)
    STOPMSG("Program is only suitable for 3D or 4D data")

FETCHPARS("TD", &tds)
td = ((tds + 255) / 256) * 256;

FETCHPARS("FnTYPE", &fnstype) /*planes: 0, points: 1, NUS: 2 */

if (fnstype != 2 )
    STOPMSG("This AU program only works on NUS data.")

/***** get input *****/

GETINT("Please enter the experiment number to concatenate to the current
one:", concatexpno);

/**** open files for first dataset ****/

(void)sprintf(datapath1, ACQUPATH("nuslist"));
fpnus1= fopen(datapath1,"a");

if (fpnus1 == 0) { /*file cannot be read*/
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", datapath1);
    STOPMSG(path)
}
(void)strcpy(datapath1, ACQUPATH("ser"));
fpser1 = fopen(datapath1, "ab");

if (fpser1 == 0) {
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", datapath1);
    STOPMSG(path)
}

/* open files for second dataset*/

DATASET(name, concatexpno, 1, disk, user);

(void)sprintf(datapath2, ACQUPATH("nuslist"));

```

```

fpnus2= fopen(datapath2,"r");

if (fpnus2 == 0) { /*file cannot be read*/
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", datapath2);
    STOPMSG(path)
}

(void)strcpy(datapath2, ACQUPATH("ser"));
fpser2 = fopen(datapath2, "rb");
if (fpser2 == 0) {
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", datapath2);
    STOPMSG(path)
}

if (parmode ==2) {
/*FETCHPARS("NusPOINTS", &nus_points) */ /*number of points for NUS sampled
data, only needed for second file, since first one is appended to*/
FETCHPARNS(2,"TD", &td2s) /*status parameter in td always correct, will
calculate nus_points from this!*/
}

if (parmode ==3) {
/*FETCHPARS("NusPOINTS", &nus_points) */ /*number of points for NUS sampled
data, only needed for second file, since first one is appended to*/
FETCHPARNS(3,"TD", &td3s) /*status parameter in td always correct, will
calculate nus_points from this!*/
}

/*concatenate files*/

DATASET(name, cur_expno, 1, disk, user);

if (parmode ==2) {
    nus_points = td2s / 4;
    for (j = 0; j < nus_points; j+=1) {
        fscanf(fpnus2, "%d %d", &cur_pointf2, &cur_pointf1); /*read in NUS
file*/
        fprintf(fpnus1, "%d %d\n", cur_pointf2, cur_pointf1 ); /*concatenate
to first NUS file*/
    }
    row1 = (int*)malloc( td * 4* nus_points * sizeof(int)); /*read in ser
file*/
    fread(row1, sizeof(int), td *4* nus_points, fpser2); /*concatenate ser*/
    fwrite(row1, sizeof(int),td * nus_points * 4, fpser1);
}

if (parmode == 3) {
    nus_points = td3s / 8;
    for (j = 0; j < nus_points; j+=1) {
        fscanf(fpnus2, "%d %d %d", &cur_pointf3, &cur_pointf2, &cur_pointf1);
/*read in NUS file*/
        fprintf(fpnus1, "%d %d %d\n", cur_pointf3, cur_pointf2, cur_pointf1 );
    }
    row1 = (int*)malloc( td * 8* nus_points * sizeof(int)); /*read in ser
file*/
    fread(row1, sizeof(int), td *8* nus_points, fpser2); /*concatenate ser*/
    fwrite(row1, sizeof(int),td * nus_points * 8, fpser1);
}

```

```
/*close NUS files and ser files*/  
fclose(fpser1);  
fclose(fpser2);  
fclose(fpnus1);  
fclose(fpnus2);  
  
free(row1);  
  
Show_status("done");  
  
REFRESHDATA  
QUIT
```

5.4 Chapter 4: Phase separation behavior of FG-domains

5.4.1 Amino acid sequences of engineered FG domains and reference wild-type MacNup98A domains:

Wild-type *Tetrahymena thermophile* macronuclear Nup98A FG-domain (MacNup98A)

The Gle-binding domain (GLEBS) is colored in red (44 residues). The sequence contains the His-tag.

```
MQHSHSHHGHHSHHGHHGHHGHHGHHGSMFGNTGGGGGLFGNTQTQQTGGGLFGQPQQTQFGQT
GATGGGLFGGATNTFGGGGGGLFGGNNNQQTNPAGGGIFGQGTTLGGAPAQTGGGLFGAPQNN
QGGGLFGGGTTTGGGMFGNQANTQTGGGLFGGPSQPTTQPPAFSLNNPTTGGGLFGQPANTMGG
NNGGLFGGQTNNSFGANNMGLGNNNRPQAGIFGGATTTAPTGNTGMFGGIGANNGGGLFGMNNNTNT
NPTGGFGATNPAGGGGLFGGGATTTGGGLFGGGNTQGGGLLGTANTTAGGLLGGGFNMNNNTGGI
LGQTNQFGLGSFGTNNNAAPFQPKASANGVLTKEKNLCYAINSGTDFCIFEALTQRKLVKAGQL
KPGAQQAGGMFGQPAQGGNGLFGGGGAATTPFGGAQNGNLFGGQNTQAQGGGLFGAPVNNAAATGA
GGGLFGAKPAATTTGGGLFGQMPAQTTGGFLGNTATQPAGGLFGGATTTQAPGGGGGGGLFGGNTTA
ATTGGGLFGGNTQTGGATGGLFGGQQPNQGLFLNTGNANNANTGGGLFGGATTPATGGGLFGGS
TNTQPGLATGGGLFGNNQGGASQPAQGGGLFGGAAPQQNSLFGGATAGGQTGLFGGATGATQQGG
GLFGQTASNPTQGGGLFGAANPGLGAAA
```

Perfectly repetitive GLFG_{52x12} (prGLFG_{52x12})

```
GGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGN
TQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGG
LFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQ
PATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLF
GGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPA
TGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGG
NTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATG
GLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNT
QPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGL
FGGNTQPATGGLFGGNTQPAT
```

prGLFG_{7x12} fragment (7x perfect repeats)

```
GGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGNTQPATGGLFGGN
TQPATGGLFGGNTQPAT
```


GxFG//L_{52x12}

The Gle-binding domain (GLEBS) is colored in red (44 residues).

GGFGNTGLAPAGGTFGNTQLQQGGGQFGQPQLTQGGGAFGQTGLTTGGGNFGGATLTAPGGGFGGG
GLNPTGGNFGGNNLQQTGTTFGQGTLQTGGGNFGAPQLNQQGGTFGGGTLLTGGGQFGANTLTGGG
GQFGGPSLPTTAGPFGSNNLTTGGGNFGQPALTNNGGNFGGQTLNQASGQFGANNLPPTNGKFGNNN
LPQTAGTFGGATLTGNTGNFGGANLTGGGGNFGNNTLNPTGGPFGATNLAGGGGTFGGGALTGGGGT
FGGGNLQTGGGTFGTANLTTAGGTFGGGNLQPQNGTFGNNNLPATGGNFGQTNLAAPQGNFGGTNLN
AASGAFGQKPLSANGVLT~~TKPNEKNLCY~~AINSGTDFCIFELALTQRKLVKAGQLKPGAQAGGQFGQPALNT
QGGAFGGGGLATTPGNFGGAQLNNTGGTFGGQNLQAGGGNFGAPNLAATGAFGAGNLTQGGAFGA
KPLATGGGQFGQPALTAAGGQFGNTALPAGGGTFGGATLTPGGGAFGGNTLATGGGQFGGNTLGATG
GPFGGQQLNNQGGAFGNTNLTGGGTFGGATLTTGGGGFGGSTLATGGGQFGGASLPAAGGPFGGAA
LQQNSGAFGGATLGGTGGQFGGATLQQGGGSFGQTALNPGGGAFGAANLTTQPGQFGGNNLAAT

GxFG//V_{52x12}

The Gle-binding domain (GLEBS) is colored in red.

GGFGNTGVAPAGGTFGNTQVQQGGGQFGQPQVTQGGGAFGQTGVTTGGGNFGGATVTAPGGGFGG
GGVNPTGGNFGGNNVQQTGGTFGQGTVQTGGGNFGAPQVNQQGGTFGGGT~~VTTGGGQFG~~ANTVTG
GGGQFGGPSVPTTAGPFGSNNVTTGGGNFGQPAVTNNGGNFGGQTVNQASGQFGANNVPPTNGKFG
NNNV~~PQTAGT~~FGGATVTGNTGNFGGANVTGGGGNFGNNTVNPTGGPFGATNVAGGGGTFGGGAVTG
GGGTFGGGNVQTGGGTFGTANVTTAGGTFGGGNVQPQNGTFGNNNV~~PATGGN~~FGQTNVAAPQGNFG
GTNVNAASGAFGQKPV~~SANGVLT~~TKPNEKNLCYAINSGTDFCIFELALTQRKLVKAGQLKPGAQAGGQFG
QPAVNTQGGAFGGGGVATTPGNFGGAQVNTTGGTFGGQNVQAGGGNFGAPNVAATGAFGAGNVNT
QGGAFGAKPVATGGGQFGQPAVTQAGGQFGNTAVPAGGGTFGGATVTPGGGAFGGNTVATGGGQFG
GNTV~~GATGGP~~FGGQQVNNQGGAFGNTNVNTGGGTFGGATVTTGGGGFGGSTVATGGGQFGGASVPA
AGGPFGGAAVQQNSGAFGGATVGGTGGQFGGATVQQGGGSFGQTA~~VNPGGGA~~FGAANVTTQPGQFG
GNNVAAT

GLLG_{52x12}

GLLGNTGGAPAGGLLGNTQTQQGGGLLGQPQQTQGGGLLGQTGATTGGGLLGATNTAPGGLLGGG
GNPTGGLLGGNNNQQTGGLLGGQTTQTGGGLLGA~~PQNNQGG~~GLLGGTTTTGGGLLGANTQTGGGGL
LGGPSQPTTAGLLGSNNPTTGGGLLGGPANTNNGGLLGGQTNQASGLLGGANNQPPTNGLLGN~~NNK~~PQ
TAGLLGGATTTGNTGLLGGANNTGGGGLLGNNTNPTGGLLGATNPAGGGGLLGGATTGGGGLLGG
GNTQTGGGLLGTANTTTAGGLLGGGNTQPQNGLLGNNTPATGGLLGGTNNAAPQGLLGGTN~~NNA~~ASG
LLGQKPASANGVLT~~TKPNEKNLCY~~AINSGTDFCIFELALTQRKLVKAGQLKPGAQAGGLLGGPAQNTQGG
LLGGGAATTPGLLGGAQNNTTGGLLGGQNTQAGGGLLGGAPNNAATGLLGGAGNANTQGGGLLGA
KPAAT
GGGLLGGQPAQTQAGGLLGN~~TAPAGG~~GLLGGATTTGGGLLGGNTAATGGGLLGGNTQGATGGLLGG
QQPNNQGLLGN~~TANTGG~~GLLGGATTTGGGLLGGSTGATGGGLLGGASQPAAGGLLGGAA~~PQ~~NS
GLLGGATAGQTGGLLGGATQQGGGLLGGQTASNPGGLLGGANATTQPGLLGGNNQAAT

GLAG_{52x12}

GLAGNTGGAPAGGLAGNTQTQQGGGLAGQPQQTQGGGLAGQTGATTGGGLAGGATNTAPGGLAGGG
GGNPTGGLAGGNNNQQTGGLAGQGTTQTGGGLAGAPQNNQGGGLAGGGTTTTGGGLAGANTQTGGG
GLAGGPSQPTTAGLAGSNNPTTGGGLAGQPANTNNGGLAGGQTNNQASGLAGANNQPPTNGLAGNNN
KPQTAGLAGGATTTGNTGLAGGANNTGGGGLAGNNTNNTGGLAGATNPAGGGGLAGGGATTGGGL
AGGNTQTGGGLAGTANTTTAGGLAGGGNTQPQNGLAGNNTNPATGGLAGQTNNAAPQGLAGGTNNN
AASGLAGQKPASANGVLTKPNEKNLCYAISNGTDFCIFEALTRKLVKAGQLKPGAQAGGLAGQPAQNT
QGGLAGGGGAATTPGLAGGAQNNTTGLAGGQNTQAGGGLAGAPNNAATGLAGAGNANTQGGLAG
AKPAATGGGLAGQPAQTQAGGLAGNTAQPAGGGLAGGATTTTGGGLAGGNTAATGGGLAGGNTQGAT
GGLAGGQQPNQGGLAGNTNANTGGGLAGGATTTTGGGLAGGSTGATGGGLAGGASQPAAGGLAGG
AAPQNSGLAGGATAGQTGGLAGGATQQQGGGLAGQTASNPGGGLAGAAANATTQPGLAGGNNQAAT

Proline-free Mac98A

The Gle-binding domain (GLEBS) is colored in red.

MFGNTGGGGLFGNTQTQQTGGGLFQAQQTQFGQTGATGGGLFGGATNTFGGGGGGLFGGNNNQ
QTNTAGGGIFGQTTGLGGAQAQTGGGLFAGQNNQGGGLFGGGTTTTGGMFGNQANTQTGGGL
FGGASQTTQTAFLNNTTTGGGLFQANANTMGGNNGGLFGGQTNSFGANNMLGNNNRGQAGI
FGGATTAATGNTGMFGGIGANNGGGGLFGMNNTNTNTTGGFGATNGTAGGGGLFGGGATTGGGL
FGGNTQGGGLLGTANTTAGLLGGGFNMNNTGILGQTNNQFGLGSFGTNNNAAGAFQNKASAN
GVLTKPNEKNLCYAISNGTDFCIFEALTRKLVKAGQLKPGAQQAAGMFGQMAQGGNGLFGGGGAATT
TTFGGAQNGNLFGGQNTQAQGGGLFGATVNNAATGAGGGLFGAKAATTTGGGLFGQMQAQTGGFLG
NTATQNAAGGLFGGATTTQAGGGGGGGGLFGGNTAATTGGGLFGGNTQTGGATGGLFGGQQTNNQ
GGLFLNTGNANNANTGGGLFGGATTTQATGGGLFGGSTNTQAGLATGGGLFGNNQASQGAQGGGLF
GGAANQQNSLFGGATAGGQTGGLFGGATGATQQQGGGLFGQTASNATQGGGLFGAANQGLGAAA

Wild-type *Saccharomyces cerevisiae* Nup116 FG domain

The Gle-binding domain (GLEBS) is colored in red.

MFGVSRGAFPSATTQPFSTGSTFGGQQQQQPVANTSALFSLSQNTTTQAPAFNFGNQTSNSPFG
MSGSTTANGTPFGQSQLTNNNASGSIFGGMGNNTALSAGSASVVPNSTAGTSIKPFTTFEEKDPTTGVIN
VFQSITCMPEYRNFSEELRFQDYQAGRKFGTSQNGTGTTFNNPQGTNTGFGIMGNNNSTTSATTGL
FGQKPATGMFGTGTGSGGGFGSGATNSTGLFGSSTNLGNSAFGANPATSGGLFGNTTNNPTNGTNN
TGLFGQQNSNTNGGLFGQQQNSFGANNVSNNGGAFGQVNRGAFFPQQQTQQGSGGIFGQSNANANGGA
FGQQQGTGALFGAKPASGGLFGQSAGSKAFGMNTNPTGTTGGLFGQTNQQQSGGGLFGQQQNSNAG
GLFGQNNQSQNSGLFGQQNSSNAFGQPQQQGLFGSKPAGGLFGQQQGASTFASGNAQNSIFGQ
NNQQQSTGGLFGQQNNQSQSPGGLFGQTNQNNQPFGQNLQPPQQNNSLFGAKPTGFGNTSLF
SNSTTNQSNIGSNNLQQQSGGLFQNKQPPASGGLFGSKPSNTVGGGLFGNNQVANQNNPASTSGGL
FGSKPATGSLFGGTNSTAPNASSGGIFGSNNASNTAATTNSTGLFGNKPVGAGASTSAGGLFGNNNNSS
LNNSNGSTGLFGSNNTSQSTNAGGLFQNTSTNTSGGGLFSQPSQSMAQSQUALQQQQQQQR

Caenorhabditis elegans Nup98

The Gle-binding domain (GLEBS) is colored in red.

MFGQNKSFGSSSFGGGSSGSGLFGQNNQNNQNKGLFGQPANNSGTTGLFGAAQNKPAGSIFGAASNT
SIFGSPQQPQNNQS SLFGGGQNNANRSIFGSTSSAAPASSSLFGNNANNTGTS SIFGSNNNAPSGGGL
FGASTVSGTTVKFEPPISSDTMMRNGTTQTISTKHMCI SAMS KYDGK SIEELRVEDYIANRKAPGTGTTST
GGGLFGASNTTNQAGSSGLFGSSNAQQKTSLFGGASTSSPFGGNTSTANTGS SLFGNNNANTS AASGS
LFGAKPAGSSSLFGSTATTGASTFGQTTGS SLFGNQQPQTNTGG SLFGNTQNQNQSG SLFGNTGTTGTG
FGQAQQQPQQSSGFSFGGAPAAATNAFGQPAAANTGG SLFGNTSTANTGS SLFGAKPATSTGF FGAT
QPTTTNAFGSTNTGGGLFGNNAAKPGGLFGNTTNTGTGGGLFGSQPQASSGGLFGSNTQATQPLNTGF
GNLAQPQIVMQQQ

Xenopus tropicalis Nup98

MFNKTFGSPFGTGNGAFGATSTFGQTTGFGTTPATAFGSAGFGTNTSTGGLFGNTQTKPGGLFGSTTFN
QPATSSSSSSGFGFGASTGTTNSLFGSTNTGSGLFATQSNAGFGQAKPTTFGNFGTSTSTGGLFGNTNTAN
PFGGTSASLFGASTFSAAPTGTVGAPTGTGLFGTSAATSSASTGIFGSTAANNSFSFAGNKTTFGTAGTG
AFGGNTGGLFGQPANQPAASLFNKPFGNATTTQSTGFSFGNTSTLGQPQTSTMGLFGANQPTQSGGLF
GTTTNTNATGAFGAGTSLFGQPNPAPFGTGSTLFGNKPAGFGTTTTSAPAFGTTTGGLFGNKPRTLGTN
TNTSNFGFGSNTAGTSLFGNKTATGTIGPSLGTGFGTALNPGQTSLFGSNQPKLTGTLGTGAFGNAGFN
STSAGLGFAGAPQ

5.4.2 Proteins and corresponding bacterial expression constructs used in this study

Protein name	Plasmid*	Encoding for	,Reference
TtMacNup98A FG (Mac98A)	pHBS418	His ₁₈ -TtMacNup98A ₁₋₆₆₆ -Cys	Schmidt & Görlich, 2015
prGLFG _{52x12}	pSNG064	His ₁₄ -ZZ-scSUMO-prGLFG _{52x12}	this study
prGLFG _{52x12} -Cys	pSNG102	His ₁₄ -ZZ-scSUMO-prGLFG _{52x12} -Cys	this study
Pro-free-prf.GLFG _{52x12}	pSNG037	His ₁₈ - Pro-free-prf.GLFG _{52x12} -Cys	Ng et al., 2021
prGLFG _{7x12} fragment	pSNG110	His ₁₄ -ZZ-scSUMO-prGLFG _{7x12} -Cys	this study
GLFG _{52x12}	pSNG036	His ₁₈ -GLFG _{52x12} -Cys	Ng et al., 2021
GLFG _{52x12}	pSNG114	His ₁₄ -ZZ-scSUMO-GLFG _{52x12} -Cys	this study
GAFG _{52x12}	pSNG072	His ₁₈ -GAFG _{52x12}	this study
GAFG _{52x12}	pSNG112	His ₁₄ -ZZ-scSUMO-GAFG _{52x12} -Cys	this study
GxFG//L _{52x12}	pSNG087	His ₁₈ -GxFG//L _{52x12}	this study
GxFG//L _{52x12}	pSNG113	His ₁₄ -ZZ-scSUMO-GxFG//L _{52x12} -Cys	this study
GxFG//V _{52x12}	pSNG125	His ₁₈ -GxFG//V _{52x12}	this study
GLLG _{52x12}	pSNG124	His ₁₄ -ZZ-scSUMO-GLLG _{52x12} -Cys	this study
GLAG _{52x12}	pSNG123	His ₁₄ -ZZ-scSUMO-GLAG _{52x12} -Cys	this study
rat NTF2	pDG2121	rNTF2	Frey et al., 2018
mCherry	pSF779	His ₁₄ -TEV-mCherry-Cys	Schmidt & Görlich, 2015
hslImportin β (hslmp β)	pDG2305	His ₁₄ -MBP-brSUMO-hslmp_beta	Frey et al., 2018

**Plasmid numbers are unique identifiers.*

5.4.3 Composition of M9 medium with isotope labels for bacterial culture

	Final concentration
Na ₂ HPO ₄	100 mM
KH ₂ PO ₄	20 mM
MgSO ₄	2 mM
CaCl ₂	0.2 mM
NTA·Na ₂	500 nM
FeCl ₃	100 nM
ZnCl ₂	10 nM
MnCl ₂	10 nM
H ₃ BO ₄	5 nM
NiSO ₄	0.5 nM
(NH ₄) ₆ Mo ₇ O ₂₄	0.5 nM
Na ₂ SeO ₃	0.5 nM
Na ₂ WO ₄	0.5 nM
CoSO ₄	1 nM
CuSO ₄	1 nM
Biotin	1 mg/L
Thiamin	1 mg/L
¹⁵ NH ₄ Cl (Sigma-Aldrich, Germany)	1.1 g/L
D-Glucose U- ¹³ C ₆ (Cambridge Isotope Laboratories, USA)	6.0 g/L

5.4.4 Hydrophobic indices of FG-domains

Hydrophobic indices were calculated based on (Fauchere and Pliska, 1983; Moon and Fleming, 2011; Zhao and London, 2006; Wimley and White, 1996). For the hydrophobicity scale by Moon and Flemming, lower values indicate greater hydrophobicity, for all other scales, higher values correspond to greater hydrophobicity. For all three groups of sequences (non-perfectly repetitive designer FG-domains; perfectly repetitive designer FG-domains, wild-type sequences), the rank of a sequence with respect to its hydrophobicity is shown in bold (1 corresponds to the most hydrophobic sequence within that group).

	Fauchere		Moon		Zhao		Wimley	
GLFG_{52x12}	0.31	1,2	1.08	1,2	-0.19	1,2	-0.03	1,2
GAFG_{52x12}	0.20	5,6	1.22	5	-0.30	5	-0.09	5
GxFG//L_{52x12}	0.31	1,2	1.08	1,2	-0.19	1,2	-0.03	1,2
GxFG//V_{52x12}	0.27	4	1.16	4	-0.21	4	-0.08	3,4
GLLG_{52x12}	0.30	3	1.11	3	-0.20	3	-0.08	3,4
GLAG_{52x12}	0.20	5,6	1.25	6	-0.31	6	-0.13	6
prGLFG_{52x12}	0.35	1	0.95	1	-0.18	2	-0.02	2
Pro-free-prf.GLFG_{52x12}	0.29	2	1.22	2	-0.13	1	-0.01	1
TtMacNup98A	0.30	2	1.20	2	-0.23	3	-0.07	3
Pro-free TtMacNup98A	0.27	3	1.24	4	-0.15	2	-0.03	2
XtNup98	0.36	1	0.97	1	-0.12	1	0.00	1
ScNup116	0.22	5	1.36	5	-0.42	5	-0.11	5
CeNup98	0.26	4	1.23	3	-0.33	4	-0.10	4

5.4.5 Pulse sequences used in MAS NMR experiments

5.4.5.1 Gradient free INEPT-based ^1H - ^{15}N correlation spectrum

```
;HSQC, 2D H-1/X correlation via double inept transfer
;phase sensitive
;with decoupling during acquisition
;Pell, Webber, Pintacuda, modified by E. Najbauer
;
;$CLASS=HighRes
;$DIM=2D

#include <Avance.incl>
#include <Delay.incl>

"p2=p1*2"
"p4=p3*2"
"d0=-1u"
"d11=30m"

"in0=inf1"
"in10=inf2/2"

"DELTA=d0*2+p2"

1 ze
  d11 ;p112:f2
2 d1 do:f2 do:f3

; -----1H to XY plane-----
3 (p1 p11 ph14):f1

; -----1H to 15N INEPT transfer-----
d4 ;p12:f2
(center (p1*2 p11 ph2):f1 (p7*2 p17 ph9):f2)
d4
(center (p1 p11 ph3):f1 (p7 p17 ph6):f2)
d14
(center (p1*2 p11 ph2):f1 (p7*2 p17 ph5):f2)
d14 ;p112:f2

; -----15N CS evolution-----

if "d0<0" goto 4
  0.5u cpds1:f1
  (center (d0) (p3*2 p111 ph0):f3)
  0.5u do:f1

; -----15N to z axis for water suppression-----
4 (p7 p17 ph10):f2

1u p113:f1
p30*0.25 ph0
p30*0.25 ph1
p30*0.25 ph0
p30*0.25 ph1
```

```

; -----15N to XY plane after water suppression-----
(p7 p17 ph11):f2

; -----15N to 1H INEPT transfer-----
d14
(center (p1*2 p11 ph2):f1 (p7*2 p17 ph5):f2 )
d14 ;p112:f2
(center (p1 p11 ph4):f1 (p7 p17 ph7):f2)
d4
(center (p1*2 p11 ph2):f1 (p7*2 p17 ph5):f2 )
d4 ;p112:f2

; -----detection-----
go=2 ph31 cpds2:f2 cpds3:f3
d1 do:f2 do:f3 mc #0 to 2
F1PH(calph(ph10, -90), caldel(d0, +in0))
exit

ph0=0
ph1=1
ph2=0
ph3=1
ph4=0
ph5=0
ph6=0 0 0 0 2 2 2 2
ph7=0 2
ph8=0
ph9=0
ph10=1
ph11=3
ph12=0
ph14=0 0 2 2
ph31=0 2 2 0 2 0 0 2

;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default)
;p112: f2 channel - power level for CPD/BB decoupling
;p113: water suppression
;p30 : total water suppression.
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel - 90 degree high power pulse
;p4 : f2 channel - 180 degree high power pulse
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I/O [30 msec]
;cnst2: = J(XH)
;inf1: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 4 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEQ
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

```

5.4.5.2 Gradient free ^1H - ^{15}N TROSY

```
;Gradient-free version by E. Najbauer 2017
;D. Lee, Nov. 2002
;K. Pervushin et al, PNAS, 94, 12366 (1997)

;p11 : power for 1H
;p12 : power for 13C
;p13 : power for 15N

;p1 : 90 degree hard pulse 1H
;p3 : 90 degree hard pulse 13C
;p5 : 90 degree hard pulse 15N

;d1 : relaxation delay
;d2 : INEPT delay (~2.7ms)

define delay INEPT1
define delay INEPT2
define delay INEPT3

#include <Avance.incl>

"p2=2*p1"
"p6=2*p5"

"in0=infl/2"

"d0=in0/2-(p5+p1)*2/3.14159265358"
"INEPT1=d2-(210u)-10u"
"INEPT2=d2-(210u)-10u"
"INEPT3=d2-(210u)-10u"

"l2 = 1"

1 10u ze
2 1m
   d1
   20u p11:f1
   20u p12:f2
   20u p13:f3
;-----first INEPT
   1m
5 10u
   (p1 ph20):f1
   10u
   INEPT1 p11:f1
   (center(p2 ph21):f1 (p6 ph20):f2)
   INEPT1
   5u
   5u p11:f1
   (p1 ph23):f1

;-----15N evolution
if "l2 %2 == 1" goto 31
   0.5u cpds3:f3
   (p5 ph2):f2
```

```

goto 32
31 0.5u cpds3:f3
    (p5 ph1):f2
32 d0
    d0

;-----second INEPT
    (p1 ph10):f1
    0.5u do:f3
    10u
    INEPT2 p11:f1
    (center(p2 ph20):f1 (p6 ph20):f2)
    INEPT2
    5u
    5u p11:f1
    (center(p1 ph20):f1 (p5 ph12):f2)
;-----WATERGATE
    INEPT3
    5u
    5u p11:f1
    (center(p2 ph20):f1 (p6 ph20):f2)
    10u
    INEPT3 ;LOCKH_OFF
    (p5 ph11):f2
;-----acquisition
    go=2 ph31
    1m mc #0 to 2 F1EA(ip10*2 & ip12*2 & ip17*2 & iu2,ip1*2 & ip2*2 & ip31*2 &
id0)
    10u do:f1
    10u do:f3
    10u do:f2
exit

ph1 =1 3 2 0
ph2 =1 3 0 2
ph31=1 3 2 0
ph10=3 3 3 3
ph11=0 0 0 0
ph12=3 3 3 3
ph17=1
ph20=0
ph21=1
ph22=2
ph23=3

```

5.4.5.3 Gradient-free HNCA sequence

```
;written by E. Najbuaer
;based on pp 7.31c, pp 620 in Cavanagh book: decoupled CT-HNCA
; phase cycle based on Bax paper, without water suppression
;avance-version (15/02/27)
;HNCA
;3D sequence with
;  inverse correlation for triple resonance using multiple
;    inept transfer steps
;
;    F1(H) -> F3(N) -> F2(Ca,t1) -> F3(N,t2) -> F1(H,t3)
;
;on/off resonance Ca and C=O pulses using shaped pulse
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho gradient selection (t2)
;using constant time in t2
;(use parameterset HNCAGP3D)
;
;S. Grzesiek & A. Bax, J. Magn. Reson. 96, 432 - 440 (1992)
;(J. Schleucher, M. Sattler & C. Griesinger, Angew. Chem. Int. Ed. 32,
; 1489-1491 (1993))
;(L.E. Kay, G.Y. Xu & T. Yamazaki, J. Magn. Reson. A109, 129-133 (1994))
;
;$CLASS=HighRes
;$DIM=3D
```

```
prosol relations=<triple>
```

```
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>
```

```
"p2=p1*2"
"p22=p21*2"
"p23=p30/12"
```

```
"d11=30m"
"d13=4u"
```

```
"d21=5387.9u"
"d26=2.7m"
```

```
"d0=0"
```

```
"d10=0"
"d30=d23"
```

```
"in0=inf1/2"
"in10=inf2/2"
"in30=inf2/2"
```

"DELTA2=d23-p19"

"DELTA3=d21"

"spoff2=0"

"spoff3=0"

"spoff5=bf2*(cnst21/1000000)-o2"

"spoff8=0"

aqseq 321

1 d11 ze

d11 p116:f3

2 d11 do:f3

3 d1 fq=0:f1 fq=0:f2 fq=0:f3

1u p11:f1

(p1 ph0):f1

d26 p13:f3

(center (p2 ph0):f1 (p22 ph2):f3)

d26

(center (p1 ph3):f1 (p21 ph4):f3)

(d10 p19:sp19 ph0 DELTA2):f2 (d21 p119 1u cpds1):f1

(center (p18:sp18 ph6):f2 (p22 ph5):f3)

d30

(p21 ph7):f3 ;nitrogen (f3) up

1u do:f1 p113:f1

(p23 ph0):f1

(p23 ph1):f1

(p23 ph0):f1

(p23 ph1):f1

1u p11:f1

(1u cpds1):f1

(p3 p12 ph8):f2 ;carbon (f2) down

(p19:sp19 ph0):f2

(p18:sp18 ph11):f2

d0

(center (p19:sp19 ph0):f2 (p22 ph0):f3)

d0

(p3 p12 ph0):f2 ;carbon back up

1u do:f1 p113:f1

(p23 ph0):f1

(p23 ph1):f1

(p23 ph0):f1

(p23 ph1):f1

1u p11:f1

(1u cpds1):f1

(p21 ph0):f3 ;nitrogen down

d23

(center (p18:sp18 ph0):f2 (p22 ph0):f3)

```

d23

; d10
; (p14:sp5 ph0):f2
; d29

(p21 ph1):f3 ;15N flip up
  lu do:f1 pl13:f1
  (p23 ph0):f1
  (p23 ph1):f1
  (p23 ph0):f1
  (p23 ph1):f1
lu pl1:f1

(p21 ph10):f3 ;15N flip down

d26
(center (p2 ph0) (p22 ph0):f3 )
d26
(center (p1 ph0):f1 (p21 ph9):f3 )
d26
(center (p2 ph0):f1 (p22 ph0):f3 )
d26 pl16:f3

go=2 ph31 cpds3:f3
d11 do:f3 mc #0 to 2
  F1PH(caliph(ph8, -90), caldel(d0, +in0))
  F2PH(caliph(ph4, +90), caldel(d10, +in10) & caldel(d30, -in30))
  ;F2EA(calgrad(EA) & caliph(ph6, +180), caldel(d10, +in10) & caldel(d29,
+in29) & caldel(d30, -in30))
exit

ph0=0
ph1=1
ph2=0 2
ph3=1 3
ph4=1
ph5=0 0 0 0 1 1 1 1; 2 2 2 2 3 3 3 3
ph6=0 0 0 0 0 0 0 0 ; 0 0 0 0 0 0 0 0
      2 2 2 2 2 2 2 2 ;2 2 2 2 2 2 2 2
ph7=1 1 1 1 1 1 1 1 ; 1 1 1 1 1 1 1 1
      3 3 3 3 3 3 3 3 ; 3 3 3 3 3 3 3 3
ph8=0 0 0 0 2 2 2 2
ph11=0 1 2 3
ph9=1
ph10=0

ph31= 1 0 0 3 1 0 0 3
      3 2 2 1 3 2 2 1 ; 1 3 3 1

;p10 : 0W
;p11 : f1 channel - power level for pulse (default)
;p13 : f3 channel - power level for pulse (default)
;p13: water suppression
;p16: f3 channel - power level for CPD/BB decoupling

```

```

;pl19: f1 channel - power level for CPD/BB decoupling
;sp1: f1 channel - shaped pulse 90 degree (H2O on resonance)
;sp2: f2 channel - shaped pulse 90 degree (Ca on resonance)
;sp3: f2 channel - shaped pulse 180 degree (Ca on resonance)
;sp5: f2 channel - shaped pulse 180 degree (C=O off resonance)
;sp8: f2 channel - shaped pulse 90 degree (Ca on resonance)
;
;      for time reversed pulse
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p13: f2 channel - 90 degree shaped pulse
;p19: f2 channel - 180 degree shaped pulse
;p21: f3 channel - 90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;p26: f1 channel - 90 degree pulse at pl19
;p30: total of water suppression
;d0 : incremented delay (F1 in 3D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d10: incremented delay (F2 in 3D) = d23/2-p14/2
;d11: delay for disk I/O [30 msec]
;d13: short delay [4 usec]
;d21: 1/(2J(NH)) [5.5 msec]
;d23: 1/(4J(NCa)) [12 msec]
;d26: 1/(4J'(NH)) [2.3 msec]
;d29: incremented delay (F2 in 3D) = d23/2-p14/2-p26-d21-4u
;d30: decremented delay (F2 in 3D) = d23/2-p14/2
;cnst21: CO chemical shift (offset, in ppm)
;cnst22: Calpha chemical shift (offset, in ppm)
;o2p: Calpha chemical shift (cnst22)
;inf1: 1/SW(Ca) = 2 * DW(Ca)
;inf2: 1/SW(N) = 2 * DW(N)
;in0: 1/(2 * SW(Ca)) = DW(Ca)
;nd0: 2
;in10: 1/(4 * SW(N)) = (1/2) DW(N)
;nd10: 4
;in29: = in10
;in30: = in10
;ns: 8 * n
;ds: >= 16
;td1: number of experiments in F1
;td2: number of experiments in F2 td2 max = 2 * d30 / in30
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: echo-antiecho in F2
;cpds1: decoupling according to sequence defined by cpdprg1
;cpd3: decoupling according to sequence defined by cpdprg3
;pcpd1: f1 channel - 90 degree pulse for decoupling sequence
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence

```

5.4.5.4 Gradient-free TRACT alpha sequence

```
;Edited by E. Najbauer 2017 for use on MAS probes (no gradients)
;based on L. Wong's sequence
```

```
;p11 : power for 1H
;p12 : power for 13C
;p13 : power for 15N
```

```
;p1 : 90 degree hard pulse 1H
;p3 : 90 degree hard pulse 13C
;p4 : 180 degree hard pulse 13C (225d for 5/600)
;p5 : 90 degree hard pulse 15N
```

```
;d1 : relaxation delay
;d2 : INEPT delay (~2.7ms)
;in0 : 1/(2 SW) (Hz)
```

```
define delay INEPT1
define delay INEPT2
define delay INEPT3
```

```
#include <Avance.incl>
```

```
"p2=2*p1"
"p4=2*p3"
"p6=2*p5"
```

```
"d0=3u"
"in0=inf2/2"
```

```
"d7=3u"
"in7=inf1/2"
```

```
"d31=2*(3u+p26+210u+p6+3u) "
```

```
"INEPT1=d2-(210u)-10u"
"INEPT2=d2-(210u)-10u"
"INEPT3=d2-(210u)-10u"
```

```
"12 = 1"
```

```
aqseq 312
```

```
1 10u ze
2 1m
d1
20u p11:f1
20u p12:f2
20u p13:f3
```

```
;-----first INEPT
```

```
5 10u
(p1 ph20):f1
10u p10:f1
10u
```

```

200u
  INEPT1 p11:f1
  (center(p2 ph21):f1 (p6 ph20):f3)
10u
200u
  INEPT1 p10:f1
  4u
  6u p11:f1
  (p1 ph23):f1
10u
200u
;-----15N evolution
if "12 %2 == 1" goto 31
  0.5u cpds2:f2
  (p5 ph2):f3
goto 32
31 0.5u cpds2:f2
  (p5 ph1):f3
32 d7
10u
200u
  (p5 ph8):f3
  3u
  (p6 ph9):f3
  3u
  (p5 ph8):f3
10u
200u
  d7
;no p2 pulse here!!!!!!
  d0
  d0
;-----second INEPT
  (p1 ph10):f1
  0.5u do:f2
  10u p10:f1
10u
200u
  INEPT2 p11:f1
  (center(p2 ph20):f1 (p6 ph20):f3)
10u
200u
  INEPT2 p10:f1
  4u
  6u p11:f1
  (center(p1 ph20):f1 (p5 ph12):f3)
;-----WATERGATE
  10u
  200u
  INEPT3 p10:f1
  4u
  6u p11:f1
  (center(p2 ph20):f1 (p6 ph20):f3)
  10u p10:f1
  6u
  200u p11:f1
  INEPT3
  (p5 ph11):f3

```

```

;-----acquisition
  go=2 ph31
  lm mc #0 to 2
    F1QF(id7)
    F2EA(rd7 & ip10*2 & ip12*2 & iu2, id0)
10u do:f1
10u do:f2
10u do:f3
10u LOCKH_OFF
exit
d31

ph1 =1 3 2 0
ph2 =1 3 0 2
ph8 =1 1 1 1 3 3 3 3
ph9 =0 0 0 0 2 2 2 2
ph10=3 3 3 3
ph11=0 0 0 0
ph12=3 3 3 3
ph15=1
ph16=0
ph17=3
ph18=2
ph19=3
ph20=0
ph21=1
ph22=2
ph23=3
ph24=0
ph26=2
ph31=1 3 0 2

```

5.4.5.5 Gradient-free TRACT beta sequence

```
;Edited by E. Najbauer 2017 for use on MAS probes (no gradients)
;based on L. Wong's sequence
;p11 : power for 1H
;p12 : power for 13C
;p13 : power for 15N

;p1 : 90 degree hard pulse 1H
;p3 : 90 degree hard pulse 13C
;p5 : 90 degree hard pulse 15N

;d1 : relaxation delay
;d2 : INEPT delay (~2.7ms)
;in0 : 1/(2 SW) (Hz)

define delay INEPT1
define delay INEPT2
define delay INEPT3

#include <Avance.incl>

"p2=2*p1"
"p4=2*p3"
"p6=2*p5"

"d0=3u"
"in0=inf2/2"

"d7=3u"
"in7=inf1/2"

"d31=2*(3u+p26+210u+p6+3u) "

"INEPT1=d2-(210u)-10u"
"INEPT2=d2-(210u)-10u"
"INEPT3=d2-(210u)-10u"

"12 = 1"

aqseq 312

1 10u ze
2 1m
   d1
   20u p11:f1
   20u p12:f2
   20u p13:f3
;-----first INEPT
5 10u
   (p1 ph20):f1
   10u :f1
   10u
   200u
   INEPT1 p11:f1
   (center(p2 ph21):f1 (p6 ph20):f3)
```

```

10u
200u
  INEPT1 p10:f1
  4u
  6u p11:f1
  (p1 ph23):f1
10u
200u
;-----15N evolution
if "12 %2 == 1" goto 31
  0.5u cpds2:f2
  (p5 ph2):f3
goto 32
31 0.5u cpds2:f2
  (p5 ph1):f3
32 d7
10u
200u
  (p5 ph8):f3
  3u
  (p6 ph9):f3
  3u
  (p5 ph8):f3
10u
200u
  d7
  (p2 ph20):f1
  d0
  d0
;-----second INEPT
  (p1 ph10):f1
  0.5u do:f2
  10u p10:f1
10u
200u
  INEPT2 p11:f1
  (center(p2 ph20):f1 (p6 ph20):f3)
10u
200u
  INEPT2 p10:f1
  4u
  6u p11:f1
  (center(p1 ph20):f1 (p5 ph12):f3)
;-----WATERGATE
  10u
  200u
  INEPT3 p10:f1
  4u
  6u p11:f1
  (center(p2 ph20):f1 (p6 ph20):f3)
  10u p10:f1
  6u
  200u p11:f1
  INEPT3 ;LOCKH_OFF
  (p5 ph11):f3
;-----acquisition
go=2 ph31
lm mc #0 to 2

```

```

        F1QF(id7)
        F2EA(rd7 & ip10*2 & ip12*2 & iu2, id0)
10u do:f1
10u do:f2
10u do:f3
10u LOCKH_OFF
exit
d31

ph1 =1 3 2 0
ph2 =1 3 0 2
ph8 =1 1 1 1 3 3 3 3
ph9 =0 0 0 0 2 2 2 2
ph10=3 3 3 3
ph11=0 0 0 0
ph12=3 3 3 3
ph15=1
ph16=0
ph17=3
ph18=2
ph19=3
ph20=0
ph21=1
ph22=2
ph23=3
ph24=0
ph26=2
ph31=1 3 0 2

```

5.4.5.6 Gradient-free INEPT-based HhCH sequence with NOE mixing

```
;HSQC, 2D H-1/X correlation via double inept transfer
;phase sensitive, decoupling during acquisition
;
;Pell, Webber, Pintacuda (le chef), modified by E. Najbauer
;
;$CLASS=HighRes
;$DIM=2D

#include <Avance.incl>
#include <Delay.incl>

"p2=p1*2"
"p4=p3*2"
"d11=30m"

"p22=d30/4"
"d0=1u"

"in0=inf1"
"in10=inf2"

"DELTA=d0*2+p2"
aqseq 312

1 ze
  d11 ;p112:f2
2 d1 do:f2 do:f3
3 (p1 p11 ph15):f1
  0.5u cpds3:f3 cpds2:f2
  d10 ;1H evolution
  (p1 p11 ph0):f1 (0.5u do):f3 (0.5u do):f2
  d20 ;NOE mixing

  (p1 p11 ph13):f1
  d4 ;p12:f2
  (center (p1*2 p11 ph2):f1 (p3*2 p111 ph9):f3)
  d4
  (center (p1 p11 ph3):f1 (p3 p111 ph6):f3)

  d14
  (center (p1*2 p11 ph2):f1 (p3*2 p111 ph5):f3 )
  d14 ;p112:f1

  1u cpds1:f1
  (center (d0) (p7*2 ph0 p17):f2)
  1u do:f1

  (p3 p111 ph10):f3

  1u p120:f1
  p22 ph0
  p22 ph1
  p22 ph0
  p22 ph1
```

```

(p3 pl11 ph11):f3
d14
(center (p1*2 pl1 ph2):f1 (p3*2 pl11 ph5):f3 )
d14 ;pl12:f2
(center (p1 pl1 ph4):f1 (p3 pl11 ph7):f3)
d4
(center (p1*2 pl1 ph14):f1 (p3*2 pl11 ph5):f3 )
d4 ;pl12:f1
go=2 ph31 cpds2:f2 cpds3:f3
d1 do:f2 do:f3 mc #0 to 2 F1PH(caliph(ph10, -90), caldel(d0, +in0))
  F2PH(caliph(ph15, +90), caldel(d10, +in10))

exit

ph0=0
ph1=1
ph13=0 2 0 2 2 0 2 0
ph14=0 1 0 1 1 0 1 0
ph15=0
ph2=0
ph3=1
ph4=0 0 2 2
ph5=0
ph6=0 2
ph7=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2
ph8=0
ph9=0
ph10=1
ph11=3
ph12=0
ph31=0 2 2 0 0 2 2 0 2 0 0 2 2 0 0 2

;p11 : f1 channel - power level for pulse (default)
;p12 : f2 channel - power level for pulse (default)
;p17 : f2 channel - power level for high power pulse
;p111: f3 channel - power level for high power pulse
;p112: proton decouple
;p120: water suppression power level
;p1 : f1 channel - 90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p3 : f2 channel - 90 degree high power pulse
;p4 : f2 channel - 180 degree high power pulse
;p7 : f2 channel - 90 degree high power pulse
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I/O [30 msec]
;cnst2: = J(XH)
;inf1: 1/SW(X) = 2 * DW(X)
;in0: 1/(2 * SW(X)) = DW(X)
;NS: 4 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEQ
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

```