The use of nuclear magnetic resonance in study of structurally dynamic membrane proteins

Phillips, Margaret

2016

Phillips, M. (2016). The use of nuclear magnetic resonance in study of structurally dynamic membrane proteins. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/68476>

<https://doi.org/10.32657/10356/68476>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**The use of Nuclear Magnetic Resonance in
study of structurally dynamic membrane
proteins**

MARGARET PHILLIPS
SCHOOL OF BIOLOGICAL SCIENCES
2016

**The use of Nuclear Magnetic Resonance in
study of structurally dynamic membrane
proteins**

MARGARET PHILLIPS

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2016

Acknowledgments

I would like to express my deep gratitude to my supervisor, Associate Professor Konstantin Pervushin for giving me an opportunity to work in his group. I greatly appreciate all the support he has given to me throughout my Ph.D days. His guidance and suggestions have encouraged me to learn more about NMR and helped me improve my scientific aptitude. I would also like to sincerely thank him for patiently discussing my projects and providing valuable insights.

I would also like to thank Asst. Prof. Said Eshaghi, our collaborator on the STIM1 project for giving me an opportunity to work with his group in Biopolis, Singapore. I have learned most of my protein expression and purification skills while working with his group. My heartfelt gratitude to Newsha, Aline, Jonanthan, Nurhuda, Zhang Ai, Vicky and Xia Yu for providing an excellent lab environment for learning and discussions. My appreciation to Prof Pär Nordlund, SBS, NTU for giving me an opportunity to work on the exciting PEN-2 project and for assisting me in keeping the project on schedule. I would like to thank Prof. Jaume Torres, SBS, NTU, for giving me an opportunity to work on Aquaporin Z, without this project my thesis would be incomplete. With this I would also like to extend my appreciation to Dr Toshio Yamazaki from Riken, Japan for helping with solid-state NMR experiments. His patience and guidance has helped me improve my knowledge in solid-state NMR technique.

My heartfelt appreciation to all my past and current colleagues: Rubing, Dr. Leo, Dr LiYan, Dr BaiYang and Dr Russell, Dr Alistair, Vidhya, Yossa, Dr Joshi, Bhargy and Bhuvaneshwari providing an encouraging environment and helping with my daily struggles. Special thanks to Mukesh, Shubhadra and Areetha for helping me on several occasions and for giving some excellent suggestions.

My deepest gratitude and love to all my friends in Singapore who made my stay here a wonderful experience. Archana, Divya, Hari, Laxmi, Manish, Mrinal, Nandhini, Naren, Reelina, Vinay and Siddharth, without you all, my journey would have been less exciting. Thank you for standing by me through all times.

Finally I would like to thank my family for their never ending support and love. I am indebted to you for life. My husband, Dr Ravishankar deserves special mention for not only being by my side whenever I needed his support but also for taking care of me when I was struggling with my work. His love and care has helped me come this far and I owe him the successful completion of my thesis.

In the end I would like to thank School of Biological Sciences, NTU, for selecting me in the Ph.D. program and providing me with the scholarship.

Table of Contents

ACKNOWLEDGEMENTS.....	I
TABLE OF CONTENTS.....	III
LIST OF FIGURES.....	VIII
LIST OF TABLES.....	XI
ABBREVIATIONS.....	XII
ABSTRACT.....	XV
1 INTRODUCTION	1
1.1 Types of proteins.....	3
1.2 Importance of Membrane Protein structure determination	4
1.3 Techniques for protein structure determination	5
1.4 Producing membrane protein for NMR studies	6
1.4.1 Using E.coli system for membrane protein expression	6
1.4.2 Using Cell Free system for membrane protein expression	7
1.5 Choice of a suitable membrane mimetic system for NMR	8
1.5.1 Micelles	8
1.5.2 Bicelles	10
1.5.3 Recently developed membrane simulating systems for use in Solution-NMR	10
1.6 Challenges in membrane protein structure determination	13
1.7 Overcoming NMR challenges of membrane protein studies	16
1.7.1 The TROSY effect.....	16
1.7.2 Deuteration of large proteins	17
1.7.3 Selective methyl labelling.....	177

1.7.4	Managing conformational exchange in solution.....	18
1.8	Structure Calculation from solution NMR data	199
1.8.1	Secondary structure prediction from NMR data.....	20
1.8.2	Distance restraints from Nuclear Overhauser Effect (NOE)	21
1.8.3	Distance restraints from Paramagnetic Relaxation Enhancement (PRE) measurements	22
1.8.4	Distance restraints from Residual Dipolar Couplings	22
1.9	Alpha-helical Membrane protein structures solved with solution state NMR -.....	23
1.9.1	Mitochondrial uncoupling protein UCP2	23
1.9.2	Human Glycine Receptor α -1 transmembrane domain	25
1.10	Beta-barrel membrane protein structures solved with solution state NMR –	28
1.10.1	Voltage-dependent anion channel VDAC-1.....	28
1.10.2	Neisserial outer membrane protein Opa ₆₀	30
1.11	Membrane protein structures solved with solid state NMR.....	32
1.11.1	Sample preparation for solid-state NMR.....	33
1.11.1.1	Using microcrystalline sample: Yersinia enterocolitica adhesin A (YadA).	33
1.11.1.2	Using precipitated sample: α -crystallin domain of α B-crystallin oligomer	35
2	STRUCTURAL AND FUNCTIONAL ANALYSIS OF THE HUMAN STROMAL INTERACTION MOLECULE 1 (HSTIM1).....	38
2.1	Introduction.....	39
2.1.1	Ca ²⁺ Release-Activated Ca ²⁺ channel (CRAC).....	39
2.1.2	STIM1 is the endoplasmic reticulum sensor.....	39
2.1.3	Orai1 CRAC channel protein.....	43
2.1.4	CRAC channel activation	43
2.1.5	Significance of CRAC channel activity.....	44
2.1.6	Aim of the Project.....	45
2.2	Materials and Methods.....	46
2.2.1	Cloning	46
2.2.2	Chemicals	46
2.2.3	Instruments for protein extraction and purification	46

2.2.4	Cloning	47
2.2.5	Expression, extraction and purification of hSTIM1 from <i>E.coli</i> cells	47
2.2.6	Circular Dichroism Spectroscopy	48
2.2.7	Buffer optimization screen	49
2.2.8	Thermal Shift Assay	49
2.2.9	NMR Spectroscopy and sample preparation	50
2.3	Results	52
2.3.1	Purification of hSTIM1 fragments	52
2.3.2	Buffer Screening for hSTIM1 fragment containing the transmembrane domain. ...	56
2.3.3	Secondary structure Analysis of hSTIM1 fragments	57
2.3.4	Temperature dependent unfolding of hSTIM1	59
2.3.5	NMR Study of hSTIM1 fragments	60
2.3.5.1	Structural characterization of hSTIM1 fragments	60
2.3.5.2	Interaction between CRAC Activating Domain of hSTIM1 and hOrai1 C-terminus	64
2.3.5.3	Effect of Ca ²⁺ ions on hSTIM1	67
2.4	Discussion	68
2.4.1	hSTIM1 coiled coil domain 1 is susceptible to oligomerization in solution	69
2.4.2	NMR characterization of hSTIM1 CAD containing domain	70
2.4.3	NMR characterization of the N-terminus S58-Q360 fragment of hSTIM1	71
3	NMR STUDIES OF PRESENILIN ENHANCER- 2 (PEN-2) SUB-UNIT OF GAMMA SECRETASE COMPLEX	73
3.1	Introduction	74
3.1.1	Role of gamma secretase in the pathophysiology of Alzheimer's disease	74
3.1.2	Role of PEN-2 in gamma-secretase complex activity	76
3.1.3	PEN-2 structural studies	76
3.1.4	Aim of study	78
3.2	Materials and Methods	80
3.2.1	Cloning	80
3.2.2	Chemicals	80
3.2.3	Instruments	80
3.2.4	Cloning	81
3.2.4.1	Full length PEN-2 (human) and hPEN-2 fragments	81

3.2.4.2	Cell-free Expression of PEN-2 (human)	81
3.2.5	Expression, extraction and purification of hPEN-2 from <i>E.coli</i> cells.....	81
3.2.6	Circular Dichroism Spectroscopy	83
3.2.7	Nuclear Magnetic Resonance Spectroscopy and sample preparation.....	84
3.2.7.1	One dimensional (1D) NMR spectroscopy	84
3.2.7.2	Multidimensional NMR spectroscopy	85
3.3	Results.....	87
3.3.1	Cloning, Expression and Purification of hPEN-2.....	87
3.3.2	Cell-free expression.....	89
3.3.3	Secondary structure determination of wild type hPEN-2	89
3.3.4	Nuclear Magnetic Resonance Spectroscopy.....	90
3.3.4.1	Detergent screening using 2D Heteronuclear NMR Spectroscopy	90
3.3.4.2	NMR characterization of full length his-tagged hPEN-2 in solution.....	94
3.3.4.3	Secondary structure prediction from NMR data	98
3.3.4.4	Methyl-TROSY NMR experiments of full length hPEN-2	99
3.3.4.5	Amide hydrogen-deuterium exchange studies using NMR	101
3.3.4.6	PRE measurements of full length hPEN-2.....	103
3.3.4.7	Progressive unfolding of hPEN-2 in solution studied by NMR.....	104
3.3.5	Truncated hPEN-2 fragments for structural studies	108
3.3.5.1	Structural Analysis of hPEN-2 fragments.....	111
3.4	Discussion	114
3.4.1	NMR characterization of the hPEN-2 in solution.....	115
3.4.2	Topology of hPEN-2 in solution.....	117
3.4.3	NMR characterization of hPEN-2 fragments.....	117
4	SOLID-STATE NMR STUDY OF AQUAPORIN Z	119
4.1	Introduction.....	120
4.1.1	Physiological significance of Aquaporins	120
4.1.2	Structure of Aquaporins.....	124
4.1.3	Structure of <i>E.coli</i> Aquaporin Z (AqpZ)	127
4.1.4	The “Open” and “Closed” state of AqpZ water channel	131
4.1.5	Aim of the project.....	133
4.2	Materials and Methods.....	134
4.2.1	Cloning, Expression and Purification of Aquaporin Z from <i>E.coli</i>	134

4.2.2	Nuclear Magnetic Resonance Spectroscopy sample preparation	134
4.2.3	Solid-state MAS NMR spectroscopy	135
4.3	Results	137
4.3.1	Good Spectral Quality of solid-state NMR experiments	137
4.3.2	Solid-state NMR characterization of AqpZ reconstituted in <i>E.coli</i> lipids	138
4.3.3	Secondary structure analysis of AqpZ	145
4.3.4	Conformational Analysis of AqpZ by NMR data	146
4.3.5	Characterization of the AqpZ - drug interaction site by solid-state NMR spectroscopy	148
4.4	Discussion	151
4.4.1	Aquaporin Z reconstituted into liposomes exists in an “Open” state conformation	151
4.4.2	The drug binds to AqpZ in an ‘open’ state and inhibits channel activity	152
5	CONCLUSION	155
	Bibliography	160
	Appendix	174

List of Figures

CHAPTER 1:

Figure 1.1: A schematic representation of the methods used to solve the structure of human membrane proteins.....	6
Figure 1.2: Schematic representation of the general structure of the membrane mimetic systems used in solution state NMR.....	12
Figure 1.3: Solution structure of the mitochondrial uncoupling protein UCP2.....	25
Figure 1.4: Open channel solution NMR structure of the transmembrane domain of the human glycine receptor α -1 subunit).....	27
Figure 1.5: Solution NMR structure of VDAC-1 in LDAO detergent micelles.....	29
Figure 1.6: The structure of Neisserial Outer Membrane Protein Opa ₆₀ using NMR distance restraints and refinement by molecular modelling.....	31
Figure 1.7: MAS solid-state NMR structure of the transmembrane domain of the YadA.....	35
Figure 1.8: Solid-state NMR and SAXS structure of the α B-crystallin oligomers.....	37

CHAPTER 2:

Figure 2.1: Full length hSTIM1 showing the predicted domains with respective amino-acid length.....	40
Figure 2.2: Solution state NMR structure of the EF-hand SAM domain with bound calcium.....	41
Figure 2.3: Crystal structure of the mutant, STIM1-Orai1 Activating Region (human) comprising of the cytosolic STIM1 domains.....	42
Figure 2.4: Proposed model for CRAC channel activation.....	44
Figure 2.5: hSTIM1 fragments selected for NMR and other biophysical assays.....	52
Figure 2.6: Purification profile of hSTIM1 K246-K349.....	54
Figure 2.7: Purification profile of hSTIM1 fragment K246-S600 (¹⁵ N labelled).....	55
Figure 2.8: Purification profile of hSTIM1 fragment S58-Q360 (¹⁵ N, ¹³ C and ² H labelled)....	56
Figure 2.9: Effect of temperature on the secondary structure of hSTIM1 fragments K246-K349 and K246-Q360.....	58
Figure 2.10: hSTIM1 fragment S58-Q360 in 20 mM Tris, 150 mM NaCl, 0.1% FC12 and 0.5 mM TCEP at pH 7.0.....	58
Figure 2.11: Thermal stability of the hSTIM1 fragments as seen by the ThermoFluor Assay....	59
Figure 2.12: 2D ¹⁵ N- ¹ H HSQC spectrum of K246-K349 fragment acquired at 298K on a 600 MHz spectrometer.....	60
Figure 2.13: 2D ¹⁵ N- ¹ H HSQC spectrum of K246-Q360 fragment acquired at 298K on a 600 MHz spectrometer.....	61
Figure 2.14: 2D ¹⁵ N- ¹ H TROSY spectrum of hSTIM1 (K246-S600) recorded at 298K on a 600 MHz spectrometer.....	62
Figure 2.15: NMR studies of S58-Q360 fragment of hSTIM1.....	63
Figure 2.16: hSTIM1-ORAI 1 interaction followed by NMR.....	65
Figure 2.17: Overlay of ¹ H- ¹⁵ N, 2D HSQC spectra recorded on Leu-[¹⁵ N] labelled free hOrai1 C-terminus derived peptide (blue) and with peptide in the presence of hSTIM1 K246-S600 (red).....	66
Figure 2.18: The secondary structure of the S58-Q360 fragment is preserved even in the absence of calcium ions.....	67

CHAPTER 3:

Figure 3.1: Assembly of the active gamma-secretase complex.	75
Figure 3.2: Topology of hPEN-2 in the endoplasmic reticulum membrane.	77
Figure 3.3: Human gamma-secretase complex structure revealed by single particle Cryo- Electron microscopy.	78
Figure 3.4: Full length hPEN-2 containing N-terminus, 6XHistidine tag and a TEV cleavage site	87
Figure 3.5: Typical purification profile of full-length hPEN-2 isotopically labelled M9 minimal media for NMR studies.	88
Figure 3.6: Immunoblotting result for full length hPEN-2 expressed in cell-free system..	89
Figure 3.7: Effect of temperature on the secondary structure of hPEN-2 full length..	90
Figure 3.8: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, his-tagged, U- ^{15}N labelled hPEN-2 in 100 mM DPC.	91
Figure 3.9: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, U- ^{15}N labelled hPEN-2 in 100 mM SDS.	92
Figure 3.10: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, U- ^{15}N labelled hPEN-2 in 100 mM DHPC.	93
Figure 3.11: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, his-tagged, U- ^{15}N labelled hPEN-2 in 25 mM DMPC: 100 mM DHPC bicelles	94
Figure 3.12: 2D ^1H - ^{15}N Tr-HSQC spectrum of 0.3 mM full length hPEN-2, his-tagged, U- ^{15}N , ^{13}C and ^2H -labeled.	95
Figure 3.13: Strip-plot of 3D Tr-HNCA (blue) and 3D Tr-HNCOCA (red).	96
Figure 3.14: NMR characterization of full length hPEN-2.	97
Figure 3.15: Secondary Structure prediction using the $^{13}\text{C}_\alpha$ chemical shifts of the assigned residues.	98
Figure 3.16: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length hPEN-2 [U- ^2H , ^{15}N]; Ile δ 1-[$^{13}\text{CH}_3$]; Leu, Val-[$^{13}\text{CH}_3$, $^{12}\text{CD}_3$]	100
Figure 3.17: 2D ^{13}C - ^1H HSQC spectrum of full length hPEN-2 [U- ^2H , ^{15}N]; Ile δ 1-[$^{13}\text{CH}_3$]; Leu, Val-[$^{13}\text{CH}_3$, $^{12}\text{CD}_3$]	100
Figure 3.18: Amide hydrogen-deuterium exchange followed for full length, his-tagged hPEN- 2, u ^{15}N -labeled, in 20 mM sodium phosphate buffer in 100 mM DPC at 35°C using NMR ..	102
Figure 3.19: PRE measurements performed for full length hPEN-2 using MTSL as the paramagnetic probe	104
Figure 3.20: Progressive unfolding of full length, his-tagged hPEN-2 as shown by NMR and CD	107
Figure 3.21: Different fragments of full length hPEN-2 submitted to PPP	109
Figure 3.22: Purification profile of hPEN-2 fragments.	110
Figure 3.23: Second Transmembrane domain comprising fragment of hPEN-2 used for structural characterization	111
Figure 3.24: 2D ^{15}N - ^1H HSQC of his-tagged, U- ^{15}N labelled, hPEN-2 fragment R16-A61 ..	112

CHAPTER 4:

Figure 4.1: Comparison of E.coli Aquaporin Z (AqpZ) with human Aquaporin 1 (AQP1)... ..	123
Figure 4.2: Schematic representation of Aquaporin Z in the membrane:	128
Figure 4.3: Periplasmic view of the Aquaporin Z tetramer reproduced from the Crystal structure.	129

Figure 4.4: Closed state of AqpZ water channel.....	132
Figure 4.5: Side chain conformation of R189 showing opening and closing of AqpZ water channel	132
Figure 4.6: 1D, ^{13}C cross-polarization spectrum of ^{15}N , ^{13}C labelled AqpZ (in E.coli lipid)...	137
Figure 4.7: 2D ^{13}C - ^{13}C DARR spectrum of ^{15}N , ^{13}C labelled AqpZ (in E.coli lipid)	139
Figure 4.8: 2D ^{15}N - ^{13}C NCA spectrum showing correlation between ^{15}N and $^{13}\text{C}\alpha$ within the residues.	140
Figure 4.9: 2D ^{15}N - ^{13}CO , NCO spectrum showing correlation between ^{15}N of one residue with the ^{13}CO of the previous residue..	141
Figure 4.10: Strips from systems A152 and T153 used to demonstrate the efficacy of 3D NCACB.....	142
Figure 4.11: Sequential walk from residue G105 to F109 assigned with the help of 3D NCACX and 3D NCOCX.	143
Figure 4.12: Assignment of AqpZ using MAS NMR experiments.....	144
Figure 4.13: Chemical Shift Index analysis of $\text{C}\alpha$ resonances of AqpZ assigned residues.	146
Figure 4.14: Drug molecule 1-(4-methylphenyl)1H-pyrrole-2,5-dione (ZINC id 94134) used in the AqpZ-drug interaction studies.....	149
Figure 4.15: AqpZ-drug interaction studied by chemical shift perturbation of residues.	149
Figure 4.16: Chemical shift perturbations after addition of the drug.....	150
Figure 4.17: Hypothetic model of the AqpZ-drug interaction	154

List of Tables

CHAPTER 1:

Table 1.1: A comprehensive list of some important membrane protein structures solved by solution NMR using detergent.....9

Table 1.2: A summary of commonly used detergents in structural and functional assays of membrane proteins.....15

CHAPTER 3:

Table 3.1: Amino acid sequences of the two hPEN-2 fragments R16-A61 (containing first predicted transmembrane domain) and S51-S93 (containing second predicted transmembrane domain).. 113

CHAPTER 4:

Table 4.1: Aquaporin structures belonging to different class of organisms..... 126

Table 4.2: Comparison of the observed NMR, N chemical shift values of A117 and V193 with the “closed” state crystal structures of AqpZ.....147

Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
^1H	Protons
^2H	Deuterons
AD	Alzheimer's disease
Ala	Alanine
APH-1	Anterior Pharynx Defective
APP	Amyloid Precursor Protein
AQPZ	Aquaporin Z
AQP	Aquaporin
AQPG	Aquaglyceroporin
Arg	Arginine
ATP	Adenosine triphosphate
CAA	Cerebral amyloid angiopathy
CAD	CRAC activation domain
CCR5	C-C chemokine receptor type 5
CD	Circular dichroism
CHAPS	3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate
CHAPSO	3-[(3-Cholamidopropyl)-dimethylammonio]-2-hydroxy-1-propane sulfonate
CMC	Critical Micelle Concentration
CO	Carboxyl
CPMAS	Cross Polarization Magic Angle Spinning
CRAC	Ca^{2+} Release-Activated Ca^{2+} channel
Cryo-EM	Cryo electron microscopy
CSA	Chemical shift anisotropy
CTAB	Cetyltrimethylammonium bromide
DARR	Dipolar Assisted Rotational Resonance
DAGK	Diacylglycerol kinase
DDM	n-dodecyl-D-maltoside
DHPC	1,2-dihexanoyl- <i>sn</i> -glycero-3-phosphocholine
DM	n-decyl- β -maltoside
DMPC	1,2-dimyristoyl- <i>sn</i> -glycero-3-phosphocholine
DNA	Deoxyribonucleic acid
DOPC	1,2-dioleoyl- <i>sn</i> -glycero-3-phosphocholine
DOSY	Diffusion ordered NMR spectroscopy
DPC	n-dodecylphosphocholine
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
D_2O	Deuterium oxide
ECD	Extracellular domain
<i>E. coli</i>	<i>Escherichia coli</i>
ER	Endoplasmic reticulum
FC12	Fos-Choline-12
GDP	Guanosine diphosphate
GlpF	<i>E. coli</i> glycerol facilitator aquaporin channel

Gly	Glycine
GPCR	G-protein-coupled receptors
GS	Gamma secretase complex
HCN	Proton carbon nitrogen
His-tag	Histidine tag
HMQC	Heteronuclear Multiple-Quantum Correlation
hPEN-2	Human PEN2
HSQC	Heteronuclear Single Quantum Coherence
hSTIM1	Human STIM1
HTS	High throughput screening
ICD	Intracellular domain
Ile	Isoleucine
IMP	Integral membrane protein
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IUPAC	International union of pure and applied chemistry
IVL	Isoleucine, Valine, Leucine - methyl labelling
k Da	Kilo Dalton
KcsA	Potassium crystallographically-sited activation channel
LB	Luria Bertini
LDAO	N,N-dimethyldodecylamine N-oxide
Leu	Leucine
LIC	Ligation independent cloning
MAS	Magic angle spinning
MFR	Molecular fragment replacement
MHz	Mega hertz
MSP	Mammalian scaffold protein
MTSL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate
MW	Molecular weight
NCT	Nicestrin
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NPA	Asparagin-Proline-Alanine motif
OmpA	Outer membrane protein A
OG	n-octyl- β -D-glucoside
OmpX	Outer membrane protein X
Opa	Opacity associated
ORAI1	Calcium release-activated calcium channel protein 1
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
pLGIC	Pentameric, ligand-gated, anion-selective channel
PMP	Peripheral membrane protein
PPP	Protein production platform
PRE	Paramagnetic relaxation enhancement
pSRII	phototaxis receptor sensory rhodopsin II
PS 1/2	Presenilin 1 or 2
RNA	Ribonucleic acid
RDC	Residual dipolar coupling

SAD	Single wavelength anomalous dispersion
SAM	Sterile α -motif
SAXS	Small angle X-ray scattering
SCID	Severe Combined Immunodeficiency syndrome
SDS	Sodium dodecyl sulphate
SEC	Size Exclusion Chromatogram
Ser	Serine
SF	Selectivity Filter
SOAR	STIM1-Orai1 Activating Region
SOC	Store-operated Ca^{2+}
SOCE	Store Operated Calcium Entry
STIM1	Stromal interaction molecule 1
ssNMR	Solid state NMR
TAA	Autotransporter
TB	Terrific broth
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
Thr	Threonine
TM	Transmembrane
TMD	Transmembrane domain
TOCSY	Total Correlation Spectroscopy
TROSY	Transverse relaxation optimized spectroscopy
TTAB	Myristyltrimethylammonium bromide
UCP2	Uncoupling protein 2
UV	Ultraviolet
Val	Valine
VDAC	Voltage dependent anion channel
v-SNARE	v-Soluble NSF attachment protein receptor
WT	Wild type
YadA	<i>Yersinia</i> adhesin A
16 DSA	16-doxyl stearic acid

Abstract

Proteins are important macromolecules performing vital cellular functions possessing a wide range of intramolecular dynamic behaviour. Some functional proteins possess a much consolidated, crystal-like structure even in solution at ambient conditions (e.g. multimeric bacterial enzymes), while others might be completely lacking any defined three dimensional structure and notwithstanding that, capable of performing their biological functions such as intrinsically disordered proteins. In order to comprehend how proteins function in the cell and to elucidate their mechanism of action, we very often have to characterize the underlying structural and dynamic characteristics at atomic level at variable degree of resolution despite the presence of extensive intramolecular dynamics. Nuclear Magnetic Resonance (NMR) spectroscopy is a unique and powerful tool used to study both the structure and dynamics of proteins. An interesting aspect of NMR is that it in principle can be used for studying both crystal-like and intrinsically disordered proteins. However, for a large class of only partially structured or otherwise flexible and dynamic proteins, the methodological problems such as inhomogeneous resonance line broadening resulting in partial or complete loss of signal to background noise may appear insurmountable. These natively, partially structured proteins form a large portion of the eukaryotic proteome and play crucial roles in many important cellular processes. In this thesis, with the help of novel NMR methods and cost-effective sample preparation, we plan to explore the structure and dynamics of such proteins belonging to the essential class of adaptor coil-coiled proteins and membrane proteins which are either part of a dynamic complex or themselves form tetrameric channels. In particular our focus is on the proteins: human Stromal Interaction Molecule 1 (hSTIM1), human Presenilin Enhancer-2 (hPEN-2) and *E.coli* water channel Aquaporin Z (AqpZ).

Human STIM1 is a type I transmembrane protein, present mainly in the endoplasmic reticulum (ER). It is involved in mediating calcium ion influx, when there is a depletion of calcium ions in the ER, by controlling the store operated Ca^{2+} ion channels, the SOCs. Using multiple, truncated fragments of full length protein, we studied the stability, dynamics and functional activity of hSTIM1 in solution. These fragments showed good alpha-helical secondary structure but lacked a well-defined

tertiary fold. Almost all fragments were disposed to significant conformational-exchange which manifested as a combination of sharp and broad signals in the NMR spectrum with a majority of signals broadened beyond detection. This prevented complete backbone assignment of hSTIM1. Nevertheless, the remaining visible resonances and the resonance from a peptide derived from ORAI1 (Calcium release-activated calcium channel protein 1) was employed to establish some structural characteristics of this very elusive complex. NMR was also used to establish the functional integrity of the calcium binding domains and the CRAC binding domains of hSTIM1.

Human PEN-2 is an essential subunit of the gamma secretase complex and is crucial to the maturation of the complex and hence its activity. The gamma secretase itself is involved in the proteolysis of transmembrane proteins like the Notch protein and the Amyloid Precursor Protein (APP) which has implications in Alzheimer's disease (AD). Currently, there is no high resolution structure of hPEN-2. The exact changes in the three dimensional structure of PEN-2 upon binding to its partner in the complex could serve as potential drug targets in the treatment of AD. Our efforts in this project were to characterize the structure of hPEN-2 in an isolated form in solution. Our results with PEN-2 in detergent micelles in solution indicate that it is a thermally stable membrane protein with intrinsically disordered domains. These domains are highly disposed to conformational changes on an intermediate timescale and possibly play an important role in protein-protein interactions. Our NMR experiments show that the backbone cross-peaks in the fingerprint TROSY spectra are broadened beyond detection for a majority of resonances coming from the transmembrane domains of the full length protein. With residue specific paramagnetic probe labelling along with Methyl-TROSY technology, we were able to define the fold and dynamics of hPEN-2 solubilized in detergents micelles despite the missing backbone resonances. Our results confirm the disordered nature of hPEN-2 in solution in the absence of its binding partner Presenilin 1.

The *E.coli* water channel Aquaporin Z has long been under study as a potential therapeutic drug target. In our study we have worked with the AqpZ sample reconstituted into liposomes for identifying the channel inhibition by small drug molecules using the solid-state NMR spectroscopy. With the help of backbone assignment of the full length protein in liposomes, we were able to identify the 'open' state of the channel. This observation was also confirmed by the drug-interaction

studies which displayed binding of the drug inside the channel cavity. The drug inhibits the channel activity by blocking the channel at the periplasmic entrance.

Chapter 1

1 Introduction

“In the drama of life on a molecular scale, proteins are where the action is.” - Lesk, A. M.

The classical view of the central dogma in molecular genetics explains that the genetic information coded in the DNA is transcribed into the RNA, which in turn is needed for the synthesis of proteins (Abraham, Cheng et al. 2015). Recent genomic studies have shed light on the many exceptions to this dogma (Jehle, Rajagopal et al. 2010). However, proteins are still widely accepted as the principal components of the molecular and enzymatic machinery which perform vital cellular functions.

Proteins are important components of all cellular systems and responsible for a wide variety of cellular activity. They are large macromolecules with a wide range of molecular weights, reaching up to tens of thousands of Daltons. All proteins are made up of polypeptide chains which consist of a sequence of 20 commonly occurring L-amino acids linked via peptide bonds. These peptide bonds are basically planar with the only freedom to rotate being around the bonds with the α carbon. These angles of rotations are referred to as phi (ϕ) for N-C α bond and psi (ψ) for the C α -C(O) bond. By finding out these two torsion angles of the backbone, the conformation of the whole polypeptide can be defined.

The amino acid sequence gives the “primary structure” of the protein. Certain regions within the polypeptide chain fold into stable “secondary structures,” primarily the α -helices or the β -sheets. Back in 1960, the first 3D structure of an “**Alpha-helix**” protein named Myoglobin was published by Kendrew *et. al.* (Kendrew, Dickerson et al. 1960). It is characterized by a right handed helical structure where there is a hydrogen bond between every fourth peptide bond, resulting in a complete turn every 3.6 residues with a translation per residue of 1.5 Å. There are also other types of helices present in a protein structure such as the 3_{10} helix and the π helix where hydrogen bonds are formed between the residue (i) – residue (i+3) and the residue (i) – residue (i+5) respectively.

Another kind of commonly found secondary structure in proteins is the “**Beta sheet**”. It is formed by several beta-strands held together by a network of hydrogen bonds. Marsh and Corey were the first to investigate the pleated sheet configuration of silk fibroin (Marsh, Corey et al. 1955). When the polypeptide chains run in the same

direction they form parallel sheets and when in opposite direction they form anti-parallel sheets.

When these secondary elements interact and fold into specific domains they give rise to the “**tertiary structure**” of a protein. Interactions between the amino acids in the polypeptide sequence bring about the stabilization of these tertiary structures. These interactions could be hydrophobic effects, formation of disulphide bridges or polar interactions which are visualized in the form of hydrogen bonds and ion-pair formations. A large protein composed of multiple subunits made up of two or more homo or hetero polypeptide chains arrange to form a “**quaternary structure**”.

1.1 Types of proteins

In order to complete its biological role, a protein must fold correctly into its required three-dimensional structure. In fact, a detailed understanding of the exact role of a protein in a biological system requires a complete knowledge of its three dimensional structure. Informally, the proteins can be divided into three classes, namely, globular proteins, membrane proteins and fibrous proteins.

The term ‘globular’ refers to “globe-like” and is commonly used for proteins possessing a sphere-like structure. They are different from fibrous proteins in the fact that they are more water soluble. Many enzymes, biological messengers and transporters are examples of globular proteins. Fibrous proteins, also known as Scleroproteins, are a class of structural or storage proteins. They are less water soluble due to a higher percentage of hydrophobic side chains and are thermally more stable than globular proteins (Miroshnikov, Marusich et al. 1998).

Membrane proteins are always associated with cellular membranes and are structurally and functionally very diverse. Some membrane proteins are attached to the surface of membranes whereas some are embedded in the membranes. Most of them have domains in either the extracellular or cytosolic part of a cell or organelle. The extracellular domains are mostly involved in binding cell-signalling ligands or other interaction proteins while the cytosolic domains may play various functions like anchoring or intracellular signalling. The transmembrane part of membrane proteins may be involved in channel formation allowing movement of ions across the cell.

Depending upon the topology of a membrane protein, it can be further classified into either an integral membrane protein (IMP) or peripheral membrane protein (PMP). Integral membrane proteins (e.g. Rhodopsin and Porins) are permanently attached to the cell membrane whereas the peripheral membrane proteins (e.g. cytochrome c) are temporarily attached to the membrane or to the integral protein.

A fraction of the proteins encoded by the genome lack a well-defined tertiary structure but still play a vital role in many cellular functions (Dyson and Wright 2005). These proteins belong to the Intrinsically/natively Disordered Protein class, IDPs. Since more than one third of the eukaryotic proteins contain the intrinsically disordered region, the IDPs are quite prevalent and form an important class of proteins (Li, Berthold et al. 2007). The lack of a proper three dimensional structure in IDPs can have several advantages namely, larger surface area for interactions with its partners and a high conformational flexibility which allows ‘interaction prone structural motifs’ within the IDPs to interact with several proteins in its environment with often very high specificity and low affinity (Marsh, Corey et al. 1955; Wishart, Sykes et al. 1992; Saier, Reddy et al. 2014). However, the altered availability of these IDPs have also been linked to several disorders and diseases such as cancer and neurodegenerative diseases (Davey 2004)

1.2 Importance of Membrane Protein structure determination

A genome wide analysis of the integral membrane protein sequences has revealed that up to 30% of all open reading frames (ORFs) are predicted to encode for membrane proteins (Wallin and von Heijne 1998). About 50% of all drugs are targeted towards membrane proteins, especially GPCRs and ion channels, owing to their critical role in several vital cellular functions (Terstappen and Reggiani 2001; Davey 2004). Some studies also indicate that about 1200 membrane proteins can still be classified as future drug targets making them the largest class of proteins with drug discovery potential (Almen, Nordstrom et al. 2009).

Though most of the drugs in the market today have been designed using computational methods and high-throughput screening (HTS), the drug interaction or its binding to the protein can only be visualized with the help of the three dimensional structure of the protein. In order to progress with better drug selectivity, stronger

binding and improved pharmacological properties, we need to understand the structure of the protein target against which the drugs are designed (Franzin, Gong et al. 2007; Nam, Jeon et al. 2009).

To date, the protein data bank (PDB) contains about 113816 biological macromolecular structures (as on 21.11.2015)(Takegoshi, Nakamura et al. 2001). About 25.6% of these proteins belong to the human proteome (as on 21.11.2015), and these structures were solved using different experimental methods as shown in figure 1.1. Only about 205 of these structures belong to the membrane protein class (as on 21.11.2015). This is less than 1% of the total human protein structures deposited in the PDB to date (Takegoshi, Nakamura et al. 2001).

1.3 Techniques for protein structure determination

To examine protein structure, techniques widely under use are X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and Electron Microscopy. Both X-ray and NMR techniques have been well established and allow for an atomic resolution of the protein structure. These techniques share certain similarities in that both determine time and spatial average of the molecules in a protein and offer a description of an ensemble average of the macromolecule. Nevertheless, these techniques are also dissimilar in many ways mainly with respect to their differences in sample preparation and the time scales accessible in each method. A unique advantage of NMR is that it helps to determine the atomic resolution structure of partially unfolded proteins that pose a problem for other techniques. The relaxation measurement experiments in NMR are useful in describing the internal dynamics of a protein molecule. Another important use of NMR is its ability to identify very weak protein-ligand interactions by following the chemical shift perturbations of the protein or ligand signals. A major advantage of the X-ray technique is its ability to study large biological macromolecules and complexes which cannot be studied by NMR. Both methods have their own advantages and disadvantages and can be used as complimentary techniques to obtain a high resolution three dimensional structure of a protein (Yee, Savchenko et al. 2005; Doerr 2006)

Cryo Electron Microscopy (Cryo-EM) is an emerging field in the experimental determination of protein structure. It is more commonly used for large proteins or

protein complexes. This technique offers only a low resolution structure as compared to X-ray crystallography or NMR. It involves reconstructing a model of the protein structure by observing many two-dimensional projections. Cryo-EM has its own shortcomings in that the relative orientation for each projection cannot be determined precisely and the images suffer from noise (Jonic and Venien-Bryan 2009). However with more research being done on developing algorithms to improve image resolution and accurate estimation of the viewing angles, there is a possibility that Cryo-EM may soon become an excellent tool complimentary to NMR and X-ray crystallography in studying protein structures.

Figure 1.1: A schematic representation of the methods used to solve the structure of proteins in the human proteome. Experimental methods used to determine the structures of different human membrane proteins with the numbers indicating the protein structures solved.

1.4 Producing membrane protein for NMR studies

1.4.1 Using E.coli system for membrane protein expression

One of the most frequently used protein expression system is the bacterium *Escherichia coli* (*E.coli*). Usually, the target membrane protein is incorporated into the cell membrane at the time of expression and later extracted by solubilizing the membrane in detergents (Duquesne and Sturgis 2010; Mancina and Love 2010). The successful expression and solubilisation of several small membrane proteins have been

demonstrated in various studies (Zhou, Cierpicki et al. 2008; Van Horn, Kim et al. 2009). However for larger membrane proteins with multiple transmembrane domains, the overexpression of the protein is not well tolerated by the *E.coli* cells (Wagner, Bader et al. 2006). To overcome this problem, new strains of *E.coli* have been developed which are more tolerant to the aforementioned stress (Miroux and Walker 1996). Another strategy to overexpress the membrane protein without stressing the bacterial cells is by inducing the target protein to move into inclusion bodies. These inclusion bodies are less toxic to the cells and are more resistant to proteolytic cleavage and hence allow for greater production of the membrane protein (Tate 2001). However extraction of membrane proteins from the inclusion bodies with the use of denaturing conditions, might pose a problem since refolding of the membrane protein to its native state may not be easy.

Another major disadvantage with the overexpression of membrane proteins in *E.coli* is the lack of post-translational modification mechanisms in bacteria which might affect the structure of eukaryotic membrane proteins. The absence of such modification may render the target protein functionally inactive and not suitable for interaction studies. However, with successful structure calculation of some large beta-barrel membrane proteins such as VDAC-1, OmpA, OmpG and OmpX which were overexpressed in inclusion bodies, the potential for *E.coli* as the expression system of choice still remains high.

1.4.2 Using Cell Free system for membrane protein expression

Another method to avoid the problems of refolding or degradation of membrane proteins inside the host cell is the use of cell-free systems. These expression systems consist of bacterial extracts (Knapp and Swartz 2004) or wheat-germ extracts (Vinarov, Lytle et al. 2004) that is added to the reaction mixture containing amino acids, RNAs, enzymes etc. With the availability of high purity isotopically labelled amino acids, it is now possible to even label membrane proteins with the cell free system. Also by adding detergents or lipids to the reaction mixture while expressing the membrane protein, it is possible to produce soluble proteins that are more suitable for solution NMR (Torizawa, Shimizu et al. 2004). Recent developments have allowed for the re-solubilisation of the precipitated membrane proteins expressed in

the absence of lipids or detergents. (Berrier, Park et al. 2004). Another advantage of the cell free system is the higher purity of the protein obtained. It also allows for specific amino-acid labelling which is not very convenient in *E.coli*, as some of the amino acids are selectively broken down in the bacterial metabolic pathway. A selective combinatorial labelling method in a cell-free system (Parker, Aulton-Jones et al. 2004) has enabled the NMR backbone assignment of several integral membrane proteins within a short span of time (Klammt, Maslennikov et al. 2012).

1.5 Choice of a suitable membrane mimetic system for NMR

In its native membrane environment, membrane proteins are surrounded by different kinds of lipids namely phospholipids, glycolipids and sometimes cholesterol (Dowhan 1997). The extraction and purification of membrane proteins from different expression systems involves their “reconstitution” into a membrane mimicking environment. These could be simple detergent micelles or bilayer vesicles such as liposomes. Most of these vesicles are too large in molecular weight and hence tumble too slowly in the NMR timescale causing signal broadening and other problems. Hence there is an urgent need to study other methods of membrane protein solubilisation for NMR investigations.

1.5.1 Micelles

The most commonly used method of membrane protein reconstitution for NMR structural characterization has been the use of detergents. Detergents form a nearly spherical envelope around the hydrophobic part of the membrane protein with the polar heads facing the solvent and the non-polar tails facing the hydrophobic core (figure 1.2A). A list of some important membrane proteins solubilised in detergent micelles which have been used for structure calculation in solution NMR have been given in table 1.1.

The use of detergents is sometimes detrimental for functional studies, especially when the solubilisation causes disruption of native contacts within the membrane protein (Otzen 2002). Another disadvantage is the need to acquire NMR

spectra at high temperatures. This could be a reason why even with the availability of several detergents, only a select few have been repeatedly used for NMR structural studies (table 1.1). Most detergents require high concentration of detergent molecules in solution to overcome the critical concentration barrier. This poses a problem since the excess detergent molecules interact with the hydrophilic residues and might interfere with the protein function (Stouffer, Acharya et al. 2008). Another significant problem encountered in detergent micelle reconstitution is the bending of the micellar surface to accommodate the hydrophobic length of the helix, known as the hydrophobic mismatch (Lau, Partridge et al. 2008).

Membrane Protein	Detergent	PDB id	Reference
Outer membrane protein A	DPC/DHP C	2KOL, IG90	(Renault, Saurel et al. 2009) (Arora, Abildgaard et al. 2001)
Human VDAC-1	LDAO	2K4T, 2JK4	(Hiller, Garces et al. 2008) (Bayrhuber, Meins et al. 2008)
Influenza M2 proton channel residues 18-60 with rimantadine inhibitor	DHPC	2RLF	(Schnell and Chou 2008)
KcsA with Charybdotoxin	DPC	2A9H	(Yu, Sun et al. 2005)
Diacylglycerol Kinase (DAGK)	DPC	2KDC	(Van Horn, Kim et al. 2009)
Mitochondrial uncouplingprotein 2	DPC, DMPC	2LCK	(Berardi, Shih et al. 2011)
v-SNARE Synaptobrevin 2	DPC	2KOG	(Ellena, Liang et al. 2009)
ZetaZeta transmembrane domain	DPC:SDS	2HAC	(Call, Schnell et al. 2006)
Outer membrane protein G	DPC	2JQY	(Liang and Tamm 2007)

Table 1.1: A comprehensive list of some important membrane protein structures solved by solution NMR using detergent.

1.5.2 Bicelles

Recently small bicelles have become increasingly popular for structural studies of membrane proteins in solution NMR (Poget and Girvin 2007). They consist of a mixture of short chain lipids 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) or detergents such as 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS), taken in almost five molar excess over long chain lipids such as 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) or 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) (Vold, Prosser et al. 1997). It has been known that the size and shape of bicelles mostly depends on the composition of the bicelles, defined as:

$$q = [\text{detergent}]_{\text{bicelles}} = [\text{detergent}]_{\text{total}} - \text{cmc}$$

Bicelles having $q \sim 0.25 - 0.5$ have a disk-shaped morphology (figure 1.2B) with the long-chain lipids forming the distinct bilayer and the short chain lipids forming the edges (Luchette, Vetman et al. 2001). Due to the larger size of the bicelles, the solution NMR spectra suffer from line broadening. However in some cases, only bicelles can ensure functional integrity of the membrane protein sample in solution (Poget, Cahill et al. 2007). An important aspect of working with bicelles is to ensure the stability of the bicelles itself. The average shelf life of proteins in bicelles is ~ 1 week. A major disadvantage of using bicelles is the presence of monomeric detergent molecules that are capable of binding to the exposed hydrophobic residues which might destabilize the protein.

1.5.3 Recently developed membrane simulating systems for use in Solution-NMR

To date less than 30 unique membrane protein structures have been determined using solution state NMR technique. This is mainly due to the lack of effective methods to mimic the membrane conditions and procedures to ensure the stability of the membrane protein. This clearly shows a need to improvise on the membrane mimicking conditions for the solubilisation and stability of the membrane proteins for structural studies. Recent developments have led to some new systems that have shown promising results. These include – Nanodiscs, Amphipols and reverse micelles.

Nanodiscs are assemblies of lipid molecules (130-160 in number) in combination with an engineered ‘mammalian scaffold protein’ (MSP) in water. This fusion takes the shape of a disc, with the lipids forming the bilayer and two molecules of MSP covering the hydrophobic part of the lipid tails, thus forming a belt like structure holding the bilayers together (figure 1.2D). Incorporation of an integral membrane protein into the nanodisc requires the protein to be first solubilised in a detergent and then incubated with the detergent solubilised lipids in the presence of the MSP. Addition of adsorbent beads to the incubating samples will cause the detergent to be adsorbed out of the system resulting in the formation of a lipid bilayer with the integral membrane protein and the MSP holding the bilayer together (Bayburt and Sligar 2010). Due to the large size of the nanodisc, it is more suitable for solid state NMR experiments than solution as its slow tumbling in solution leads to extensive line broadening of NMR signals. However, there have been studies involving the GPCR CC-chemokine receptor 5 (CCR5) in nanodisc using solution state NMR that has proved the utility of Nanodiscs in studying protein-ligand interactions (Yoshiura, Kofuku et al. 2010). Still, the utility of nanodisc is limited due to its large size and further developments are needed to improve its use in solution state NMR.

Amphipols are amphipathic polymers which have a hydrophilic backbone combined with hydrophobic groups. Overall, the amphipols resemble the micelles with the exception of having covalently linked polar head groups (Popot, Berry et al. 2003). This allows only one or two amphipols to envelop and stabilise the integral membrane protein (figure 1.2E). A good example is the ‘amphipol-assisted solubilization of GPCR’ (Dahmane, Damian et al. 2009). Another membrane protein studied via solution state NMR in amphipols was the outer membrane protein OmpA (Zoonens, Catoire et al. 2005). Initial results of the usefulness of amphipols for membrane protein structural studies have been promising and with improvement in stability and homogeneity of the resulting sample, this system could become a useful tool in membrane protein structural studies.

Reverse micelles were first studied by Wand et al. (Wand, Ehrhardt et al. 1998). It is prepared by solubilizing membrane proteins in the presence of a detergent and a co-surfactant such as CTAB (cetyltrimethylammonium bromide). This mixture is then dehydrated, followed by addition of a ‘low viscosity’ organic solvent, e.g. liquid pentane. This causes a ‘flip’ of the micelle configuration such that the hydrophobic domains of the membrane proteins are still in contact with the organic solvent while

the hydrophilic regions of the protein are in contact with the polar head groups of the detergents (figure 1.2C). The advantage of using reverse micelles is that it allows the large membrane protein to tumble at a rate much faster than in detergent micelles and helps improve the NMR spectral quality. A truncated KcsA potassium channel membrane protein was used with the reverse micelle system to be studied in solution NMR (Kielec, Valentine et al. 2009). The study revealed enhanced ^{15}N transverse relaxation times of the 'rigid' domains of the protein in reverse micelles, possibly due to faster tumbling as compared to when in detergents. Overall, the use of reverse micelle in the structural and functional studies of membrane proteins is quite promising and can serve as an advantage for solution state NMR.

Figure 1.2: Schematic representation of the general structure of the membrane mimetic systems used in solution state NMR. A) Micelles, B) Bicelles, C) Reverse micelles, D) Nanodiscs and E) Amphipols. Figure modified from (Qureshi and Goto 2012).

1.6 Challenges in membrane protein structure determination

Even with their evident importance, the paucity in the high resolution three dimensional structures of membrane proteins arise due to several bottlenecks such as their highly hydrophobic nature, high flexibility of some domains and their unstable nature in solution. A majority of membrane proteins produced via recombinant techniques also suffer from poor expression levels in different cellular and *in vitro* systems. The main obstacles faced when studying membrane proteins via solution state NMR are:

- **Large size of the protein-detergent complex.** In NMR, there is a direct correlation between the rotational correlation time (τ) and the line-width of the NMR signal (Wuthrich 2003). The large size of the membrane protein-detergent complex causes slow rotation of the complex in solution leading to severe line broadening problems in NMR spectra.
- **Internal protein dynamics.** Membrane proteins when extracted from their native membrane environment tend to have a fairly 'loose' tertiary structure. This is the leading cause of conformational exchange induced line broadening in NMR which often results in the complete disappearance of critical resonances.
- **Incomplete distance restraints.** The measurement of distance restraints obtained from Nuclear Overhauser effect (NOE) experiments, residual dipolar couplings and J-couplings become difficult in the presence of broadened or absent signals.

All this together with the narrow proton signal dispersion characteristic of the alpha-helical proteins and the high proton density in large protein-detergent complexes leads to further difficulties in structure calculation.

For structural studies, membrane proteins are widely produced using recombinant expression systems like bacteria - *Escherichia coli*, yeasts - *Pichia pastoris*, insect cells and sometimes mammalian cell lines (Junge, Schneider et al. 2008). For NMR studies, large quantities of membrane proteins are required and for this reason expression using *E.coli* is preferred as it is fairly less expensive, expression time is less, extraction from *E.coli* is easier and several constructs can be produced within a shorter time span.

When expressing eukaryotic membrane proteins, several expression systems are usually experimented with, before proceeding to large scale expressions. The

membrane proteins need a cell membrane to fold correctly where the lipid environment of the transmembrane domain plays an important role in its stability. Some membrane proteins are also post-translationally modified, requiring eukaryotic cell lines for their expression since the *E.coli* cells do not perform these modifications.

Another obstacle in the preparation of membrane protein samples for structural studies is the need to solubilize the proteins in native membrane like conditions to maintain their stability and functionality. To overcome this problem, most membrane proteins are isolated from the cells expressing them using detergents which help stabilize their hydrophobic domains. A standard rule of thumb is to screen several detergents and then choose the one which helps with the extraction of the maximum quantity of soluble, functional, homogenous and stable membrane protein. A detailed description of the detergents used in solubilization of membrane proteins have been mentioned in the Table 1.2.

Name of detergent	CMC (mM)	Chemical Structure	Hydrophilic group
<i>n</i> -Dodecyl β-D-maltoside	0.15		Non-ionic
Digitonin	<0.5		Non-ionic
N,N-dimethyldodecylamine N-oxide (LDAO)	1		Zwitterionic
3-[(3-Cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS)	6		Zwitterionic
3-[(3-Cholamidopropyl)-dimethylammonio]-2-hydroxy-1-propane sulfonate (CHAPSO)	8		Zwitterionic
Sodium dodecyl sulfate (SDS)	7-10	$\text{CH}_3(\text{CH}_2)_{10}\text{CH}_2\text{O}-\text{S}(=\text{O})_2-\text{ONa}$	Anionic
Myristyltrimethylammonium bromide (TTAB)	4-5	$\text{CH}_3(\text{CH}_2)_{12}\text{CH}_2-\text{N}^+(\text{CH}_3)_3 \text{ Br}^-$	Cationic
Triton X-100	0.2-0.9		Cationic

Table 1.2: A summary of commonly used detergents in structural and functional assays of membrane proteins.

1.7 Overcoming NMR challenges of membrane protein studies

1.7.1 The TROSY effect

Till two decades ago, solution NMR was unsuitable for large membrane and soluble proteins having molecular weight >25 kDa. With the introduction of the transverse relaxation optimized spectroscopy (TROSY) by K. Pervushin *et.al.* study of large macromolecules and membrane proteins with the help of solution-state NMR became possible (Pervushin, Riek et al. 1997). The main principle behind the experiment was exploiting the effects of the two main mechanisms of relaxation, viz. dipole–dipole coupling and chemical shift anisotropy (CSA). The TROSY-type pulse sequences are designed such that, only the slow relaxing component of the coherence pathway is preserved which gives rise to sharp resonances, while the other components are discarded. The gain in resolution and sensitivity in such type of experiments comes from the elimination of the fast relaxing coherences.

The introduction of the TROSY-type pulse scheme in the triple resonance NMR experiments, has allowed the backbone assignment of large soluble proteins (>40 kDa) and membrane protein-detergent complexes (Yang and Kay 1999). The experiments work well in the presence of deuteration of the sample. However, for most membrane proteins where the ‘loose’ tertiary structure of the protein allows for ‘intermediate exchange’ the signal attenuation cannot be compensated by even the TROSY-type experiments (Tian, Karra et al. 2005; Page, Moore et al. 2006). Recently Kay *et.al.* have shown the usefulness of TROSY experiments in studying specifically methyl labeled protein samples (Ollerenshaw, Tugarinov et al. 2003). The methyl groups serve as excellent tools for studying the dynamics and topology of large protein complexes (Ruschak and Kay 2010). Samples prepared in a uniformly deuterated background with only the methyl groups of Ile, Leu and Val isotopically labeled have been used for assisting in sequence assignment of large protein having a molecular weight > 60 kDa (Tugarinov and Kay 2003).

1.7.2 Deuteration of large proteins

Large proteins and protein-detergent complexes suffer from poor spectral resolution mostly due to high proton density which causes fast relaxation of proton spins and hence shorter (T_2) transverse relaxation time. Since there is a direct relation between the signal line width and relaxation time (width at half height $\sim 1/T_2$), a shorter T_2 leads to line broadening (Keeler 2005). To overcome the adverse relaxation effect of protons in the protein sample, it is now possible to replace all non-exchangeable protons (bound to carbon atoms) with deuterium (Goto and Kay 2000). Since deuterium (^2H) has a much lower gyromagnetic ratio ($\gamma \sim 6.7$ fold less than ^1H) it helps attenuate most relaxation pathways that cause broadening in protonated samples. Studies have shown a sensitivity gain in deuterated samples of soluble and membrane proteins (Gardner and Kay 1998). Most solution-state NMR studies of large membrane protein samples use the deuteration technique which greatly helps the assignment procedure. However, since the triple resonance experiments require the amide proton for generation and transfer of magnetization, the need to exchange the deuterated amide to proton amide must first be accomplished. Usually for soluble proteins, almost all deuterated amides successfully back exchange with protons from water in the purification buffer. However, in membrane proteins with a well-defined hydrophobic core which is well protected from the aqueous solvent, this back-exchange can be tricky. To overcome this problem deuterated glucose and amino acids can be used. An example is the tetrameric potassium channel protein KcsA which was studied in batches, where one batch had the protein growing in the deuterated media and purified in aqueous buffers, while the other batch was grown in media supplemented with deuterated amino acids (Chill, Louis et al. 2006). The resulting purified protein was exchanged into a deuterated buffer and only the protected sites gave rise to signal in the NMR spectrum. This method is useful in identifying domains with different protection from the solvent.

1.7.3 Selective methyl labelling

The use of perdeuterated samples helps simplify the NMR spectrum to a great extent but is not very useful when acquiring NOE based distance restraint. Therefore to

maintain the advantages of deuteration while maintaining protons at the protein core, several other labelling techniques have been developed.

Methyl groups which are present at the hydrophobic core of a protein, serve as excellent source of methyl-methyl NOE based distance restraints. Also, presence of three symmetric protons on the methyl group, gives rise to a single intense proton signal with a narrow line width. Thus in a slow tumbling protein background, selectively labelled fast rotating methyl groups act as excellent tools for investigating the dynamics and global fold of a large protein (Gardner, Rosen et al. 1997).

The most common method for methyl group labelling is the use of IVL method where methyl groups of only Ile, Leu and Val remain protonated while the rest of the protein is deuterated (Tugarinov and Kay 2004). It involves growing the protein in deuterated media supplemented with deuterated glucose and adding α -keto acid precursors which have protonated methyl groups of IVL. This method has been used to study various membrane proteins like – OmpX in DHPC micelles (Hilty, Fernandez et al. 2002) and KpOmpA (Renault, Saurel et al. 2009) where use of selective methyl labelling has witnessed a sharp increase in the NOE distance restraints(>400 restraints). However, the labelling technique proved less useful for alpha-helical integral membrane proteins where narrow proton signal dispersion associated with alpha-helix secondary structure hinders the assignment of NOE signals of the methyl groups. In some cases such as the pSRII GPCR (sensory rhodopsin II, G-protein coupled receptor) an increase in the NOE signals coming from inter-helix methyl groups proved helpful in the three dimensional structural assignment (Gautier, Mott et al. 2010). Thus the methyl labelling method is a complimentary tool to other labelling procedures and has proved promising in structural characterization of many membrane proteins.

1.7.4 Managing conformational exchange in solution

For membrane protein samples in solution that are prone to conformational exchange it is quite likely that two or more conformations are interconverting at different rates in the given sample conditions. Each of these conformers has its own resonance frequency and when the difference between these resonance frequencies is more than the rate at which the two conformers are exchanging, they appear as two

separate peaks in the NMR spectrum. Each peak corresponds to a particular conformational state of the protein in solution. When the conformational exchange is fast, only the average of two conformers is seen. Substantial line broadening is observed when the conformational exchange happens at a rate similar to the difference in the resonance frequencies of two states. This exchange is often termed as ‘intermediate exchange’ and it is one of the leading causes of line broadening of partially unfolded proteins or membrane proteins with dynamic domains.

Generally to circumvent the problem of conformational exchange, the sample conditions are altered such that majority of protein population exists in a single conformational state. For membrane proteins solubilised in detergents or lipids, several methods have been tried, including:

- Changing the temperature at which the NMR experiment is recorded. This affects the rotational time of the protein-detergent complex in solution. Higher temperatures will allow faster tumbling and result in narrower line widths of the signals as compared to lower temperatures.
- If the membrane protein is known to interact with a ligand, then recording NMR experiments in the presence of the interacting partner might help in stabilising the protein in one conformation and improve spectral resolution.
- Sometimes, proteins with specific mutations end up having either more stability or give a better spectrum as compared to the wild-type.
- Modifying the buffer conditions of the protein (such as pH and salt) along with the detergent screening might also help stabilize the protein.

1.8 Structure Calculation from solution NMR data

Solution-state NMR is generally used for the structure determination of membrane protein-detergent complexes. The usual strategy for the backbone assignment involves preparing uniformly ^{15}N and ^{13}C labelled protein samples in acquisition of Heteronuclear triple resonance NMR experiments. Nowadays, TROSY versions of the 3D experiments are being routinely used for structural characterization of large proteins. These include 3D TROSY-HNCA, 3D TROSY-HNCOCA, 3D TROSY-HNCACB and 3D TROSY-CBCACONH. When used together with ^{15}N - ^1H TROSY-HSQC, these experiments help to identify the backbone amide H^{N} , $^{\text{H}}\text{N}$, the C_{α}

and C_β of the 'i' spin-system and the C_α and C_β of the 'i-1' spin-system (Salzmann, Wider et al. 1999; Yang and Kay 1999). For side-chain assignments, the most commonly used 3D experiments are the ^1H - ^{13}C HSQC-NOESY (aliphatic and aromatic) and the ^1H - ^{13}C HCCH-TOCSY. NOE assignments are also needed for the modelling of the three dimensional structure of the protein. The ^{13}C resolved NOESY is required for identifying the methyl-methyl interactions and prove more helpful in characterizing the dynamic domains within the protein.

When working with large, dynamic protein-detergent complexes, uniform labelling of N and C atoms is not enough. Uniform deuteration of the sample is also essential to overcome the signal overlap problem. Many times, deuterated detergents are also required to identify the protein-detergent interaction sites. To complete the backbone assignment, the protein spectra should show good dispersion and optimum signal intensity. Once the backbone sequence is assigned, the restraints obtained from other NMR experiments are implemented into the software (e.g. CYANA) to calculate an ensemble of twenty lowest energy conformers which represent the three dimensional structure of the protein in solution.

1.8.1 Secondary structure prediction from NMR data

It is known that chemical shifts of nuclei act as sensitive probes for the chemical environment of the NMR sensitive nuclei. Fluctuations in the chemical shifts arise not only due to local changes within the protein structure but also due to interaction of the nuclei with its environment. The chemical shifts are closely related to the secondary structure of the protein in solution and have been used frequently to predict the backbone dihedral angles (ϕ and ψ) (Wishart, Sykes et al. 1992).

The chemical shift values obtained from the C_α and C_β resonances are used for the 'secondary structure' predictions (Wishart and Sykes 1994) by using the equation:

$$\Delta\delta = \delta_{\text{observed}} - \delta_{\text{random coil}} \text{ [where } \delta = \text{chemical shift (ppm)}]$$

Alpha-helical proteins give a positive secondary shift for C_α whereas beta-strand is characterized by a negative C_α secondary shift.

The dihedral angles can also be predicted from some commonly used software such as TALOS (Shen, Delaglio et al. 2009)(<http://spin.niddk.nih.gov/bax/software/TALOS/>) or PREDITOR (Berjanskii, Neal et al.

2006)(http://wishart.biology.ualberta.ca/shiftor/cgi-bin/predictor_current.py). Recently the use of CS-ROSETTA (Shen, Vernon et al. 2009) (http://spin.niddk.nih.gov/bax/software/CSROSETTA/#Preparation_of_input) has also proved very helpful in predicting good quality structures derived from the H^N , H^N , the C_α , H_α and C_β chemical shift values of the assigned sequence. This structure can then be further refined with the help of distance restraints from NOE data.

1.8.2 Distance restraints from Nuclear Overhauser Effect (NOE)

The standard method of protein structure determination using NMR involves the use of distance restraints obtained from the Nuclear Overhauser Experiments (NOE) (Anderson and Freeman 1962; Kaiser 1963). Since NOE originates from polarization transfer among spins through cross-relaxation, the intensity of the NOE signals depends upon the spatial proximity between the two neighbouring protons (Intensity $\propto 1/d^6$ where d = distance between the spins). Long range NOEs are crucial in identifying the global fold of the protein tertiary structure. For alpha-helical proteins, amide H^N - H^N correlations seen as NOE peaks are limited in number and not very useful in determining the overall fold of the protein. However, in beta-barrel proteins, the amide H^N - H^N correlations provide an intensive network of contacts and are very useful in identifying parallel and anti-parallel nature of the beta-strands.

For good quality NMR spectra of small proteins, distance restraints obtained from the NOESY type experiments are sufficient for getting high resolution NMR structure. However in large proteins or membrane proteins having >1 transmembrane helices, the NOE assignment is not easy due to signal overlap problem. Deuteration of sample results in loss of signal from non-exchangeable protons and hence not advisable. To improve the assignment strategies, methyl labelled IVL samples in a uniformly deuterated background are very useful for large protein molecules (Tugarinov, Kanelis et al. 2006).

1.8.3 Distance restraints from Paramagnetic Relaxation Enhancement (PRE) measurements

Paramagnetic Relaxation Enhancement (PRE) techniques have frequently been used for obtaining long range restraints in the absence of long range NOEs for protein structure determination (Battiste and Wagner 2000). They provide useful information both in the presence and absence of RDC (residual dipolar coupling) data. The technique involves the addition of paramagnetic probes; chemical agents with unpaired electrons to the protein sample and recording different 2D and 3D heteronuclear NMR experiments. The unpaired electron, having a large dipole, causes a distance dependent ($\text{Intensity} \propto 1/d^6$ where d = distance between the spins) relaxation of the NMR active nuclei in close proximity (12 - 23Å) to the probe and manifests as complete disappearance or line broadening of the resonances of those nuclei.

Several paramagnetic probes are available based on the nature and type of interaction under study. Most commonly used are reagents with a thiol-group (Girvin and Fillingame 1995; Gillespie and Shortle 1997; Battiste and Wagner 2000), reagents soluble in water (Hilty, Wider et al. 2004; Iwahara, Schwieters et al. 2004) or even specific to the lipids (Hilty, Wider et al. 2004). These paramagnetic probes offer good insight into not only the overall topology of the protein but also help in identifying the residues at the surface and residues buried within the transmembrane domains. A recent discovery of chelated Ni^{2+} probe allows for acquisition of multidimensional NMR experiments in the presence of the probe which reduces the recycle delay for proton magnetization recovery (Cai, Seu et al. 2006). This helps in reducing the NMR acquisition time without compromising on sensitivity.

1.8.4 Distance restraints from Residual Dipolar Couplings

In solution state NMR, the fast tumbling of the protein molecules in solution allows for longer T_2 relaxation resulting in sharp peaks and good spectral resolution. However, this tumbling causes averaging out of dipolar interactions and the chemical shift anisotropy. In order to preserve these orientation-dependent parameters which are needed to obtain maximum distance restraints for structure calculation, the protein is subjected to a weakly aligned medium (Tjandra and Bax 1997). The most commonly

used media for membrane protein alignment is the strained polyacrylamide gel (Tycko, Blanco et al. 2000). Other types of aligning media used frequently are charged polyacrylamide gels (Meier, Haussinger et al. 2002), the filamentous bacteriophage (Hansen, Hanson et al. 2000) and even bicelles (Bax and Tjandra 1997). To overcome problems in membrane protein sample preparation for RDC calculations, some developments have been made which allow the direct introduction of an SDS-solubilized sample in the gel polyacrylamide through electrophoresis (Jones and Opella 2004). Examples of membrane protein structures solved with the help of RDC restraints include the eight beta-strand OmpA (Cierpicki and Bushweller 2004) and the four transmembrane domain DsbB (Zhou, Cierpicki et al. 2008).

RDC experiments are measured by selecting a target spin-system which consists of covalently linked backbone atoms, usually the amide protons and nitrogen ($^1\text{H}^{\text{N}} - ^{15}\text{N}^{\text{H}}$), the aliphatic carbons ($^1\text{H}_{\alpha} - ^{13}\text{C}_{\alpha}$) and the peptide bond $^{15}\text{N} - ^{13}\text{CO}$ and $^{13}\text{CO} - ^{13}\text{C}_{\alpha}$ (Cierpicki, Liang et al. 2006; Ma, Tillman et al. 2008; Zhou, Cierpicki et al. 2008). These give an insight into the orientation of the corresponding peptide plane and are obtained by NMR experiments designed to measure the RDCs (Prestegard, Bougault et al. 2004). The RDCs alone cannot help determine the three dimensional structure of a protein but are used along with the NOE restraints to refine the structure.

1.9 Alpha-helical Membrane protein structures solved with solution state NMR -

1.9.1 Mitochondrial uncoupling protein UCP2

UCP2 is an integral membrane, anion-carrier channel protein found in the inner membrane of the mitochondria. It mediates the movement of small molecules across the mitochondrial membrane. Amidst other functions of the UCP2 such as glucagon secretion it was recently linked to reprogramming in cancer cells (Samudio, Fiegl et al. 2009).

Chou *et.al.* in 2011 studied and characterized the structure of the mitochondrial uncoupling protein UCP2 in DPC detergent, in solution, using NMR (Berardi, Shih et al. 2011). They reconstituted recombinantly expressed and purified UCP2 in DPC micelles and a small quantity of lipid, and used Guanosine diphosphate

(GDP) to separate the native protein. The initial 2D NMR experiments gave decent spectral quality as shown in figure 1.3A. However, the distance restraints obtained from NOE data between the amide-proton and the side-chain protons was not sufficient for unambiguous assignment for local structure determination. Residual Dipolar couplings obtained from the NMR experiments was then utilised for structure calculation of UCP2. With limited RDCs available per residue, it was difficult to define the local structure; hence the group used the molecular fragment replacement (MFR) method which involves exhaustive search of the Protein Data Bank for fragments which have structures similar to the protein under study based on the experimental data. They used a database of >100,000 seven residue fragments obtained from structures of soluble and membrane proteins. Using the UCP2 sequence and the RDC data they were able to short list fragments that best fit the corresponding protein fragment. The 'gaps' in the regions were filled by using fragments from the overlapping backbone dihedral angles with the two termini regions which provided the best fit. This method helped in the successful assignment of at least 15 fragments of the UCP2 backbone. PRE measurements were also useful in obtaining further distance restraints. They used nitroxide spin label (in MTSL) covalently attached to the cysteine residues and gave > 400 PRE distance restraints (figure 1.3C). The final structure of UCP2 (figure 1.3B) was calculated from the 15 MFR fragments and combined distance restraints from RDCs and PRE measurements.

Figure 1.3: Solution structure of the mitochondrial uncoupling protein UCP2 (Berardi, Shih et al. 2011). A) 2D ^{15}N - ^1H Tr- HSQC spectrum of ^{15}N -, ^{13}C - and ^2H -labelled UCP2 reconstituted in DPC micelles (with 5mM GDP), recorded at pH 6.5, 33°C. B) Side view of the ribbon representation of UCP2. C) Topology and sequence of UCP2 highlighting the paramagnetic probe positions in green. The red lines represent the long range PRE between the amide and the paramagnetic probe. Figures modified from (Berardi, Shih et al. 2011).

1.9.2 Human Glycine Receptor α -1 transmembrane domain

Glycine receptors are the major inhibitory receptors found in the cells of the human spinal cord and the brain stem. They are involved in facilitating fast synaptic transmissions in the peripheral and central nervous system. It forms a pentameric, ligand-gated, anion-selective channel (pLGICs) from five monomers. Each monomer

consists of a transmembrane domain (TMD), with an extracellular domain (ECD) and an intracellular domain (ICD). Some studies suggest that ICD is involved in trafficking and modulation by secondary messengers and does not contribute much to the ion channel activity (Jansen, Bali et al. 2008). The ECD binds agonistically to the channel and allows the flow of selected ions across the membrane potential (Betz and Laube 2006).

Mowery *et.al.* in 2013, described the open channel structure of the full length TMD of the human glycine receptor $\alpha 1$ subunit in lysophospholipid - LPPG in solution via NMR. Electron microscopy and functional studies showed that the channel is pentameric and conducts chloride ions. CYANA calculations indicated the four helix bundle structure of the monomeric hGlyR- $\alpha 1$ TMD (figure 1.4A). Distance restraints obtained from NOE connectivity and PRE measurements helped identify the packing of the transmembrane helices (figure 1.4B). Some unique features of the channel structure were identified which were previously unreported. First, the TM2 helix which lines the pore has a shorter stretch of amino acids (P250-S267) as compared to the other pLIGCs. The orientation of the TM2 helix against the membrane was obtained from the residual dipolar coupling studies. The second feature was the ‘kink’ in the TM3 helix at A288. Third, the conserved residue L261 was replaced by T262 in facing the open lumen. The last unique feature was the ‘cone-shaped’ pore of the hGlyR- $\alpha 1$ TMD which has a diameter of 6.2 Å at its narrowest point.

Figure 1.4: Open channel solution NMR structure of the transmembrane domain of the human glycine receptor α -1 subunit (Mowrey, Cui et al. 2013). A) Side view of the monomeric transmembrane domain of the hGlyR- α 1 TMD. B) Top view of the pentameric hGlyR- α 1 TMD obtained from CYANA using distance restraints from solution NMR data and EM data. Figures modified from (Mowrey, Cui et al. 2013).

1.10 Beta-barrel membrane protein structures solved with solution state NMR –

1.10.1 Voltage-dependent anion channel VDAC-1

VDAC is known to facilitate the trafficking of small metabolites across the mitochondrial outer membrane (Hodge and Colombini 1997). It is highly conserved across eukaryotes with 30% sequence similarity between yeast and human. The anion channel exists in three isoforms – VDAC-1, VDAC-2 and VDAC-3. Recent studies have reported the involvement of the channel in mitochondrial apoptosis (Brustovetsky, Dubinsky et al. 2003).

The three dimensional solution structure of human VDAC in detergent micelles was reported by Hiller *et.al.* in 2008 (Hiller, Garces et al. 2008). Selectively labelled sample was used for recording TROSY-type triple resonance experiments in high-field strength allowing for 80% sequence assignment (figure 1.5A). The structure revealed the presence of a short alpha-helix at the N-terminus followed by 19 beta-strands. Specifically isotope-labelled methyl groups of Ile, Leu and Val helped in describing the overall tertiary structure. Fully deuterated samples in deuterated LDAO detergents were used for recording 3D and 4D NOESY experiments (more than 600 NOE contacts) which helped confirm the beta-barrel topology of VDAC in LDAO detergent micelles and the orientation of the beta-strands (figure 1.5B). The beta-strands are tilted at an angle of about 45° to the main axis of the barrel. The tertiary structure of the barrel was consistent with the diffusion function of its wide pore and no contacts were observed between residues across the pore. The use of 16-doxyl stearic acid (16 DSA) paramagnetic probe helped identify the complete coverage of the VDAC-1 barrel periphery within the micelles.

Figure 1.5: Solution NMR structure of VDAC-1 in LDAO detergent micelles (Hiller, Garces et al. 2008). A) 2D ^{15}N - ^1H TROSY spectrum of $\text{U-}^2\text{H}$, ^{15}N , VDAC-1 in LDAO micelles. B) Ribbon representation of the side view of the 3D conformer of VDAC-1 in LDAO detergent. Figures modified from (Hiller, Garces et al. 2008).

1.10.2 Neisserial outer membrane protein Opa₆₀

Gonorrhoea and meningococcal meningitis are diseases caused by gram negative bacteria - *Neisseria gonorrhoeae* (NG) and *Neisseria meningitides*. The ‘opacity associated’ (Opa) integral membrane protein consisting of eight beta-strands and four loops is found on the outer membrane of these bacteria. Opa is responsible for receptor-mediated phagocytosis of these bacteria by the host cells. The sequence diversity in *Neisseria* is probably the reason it eludes host immune response. Thus a better understanding of the interaction between the Opa-receptor and the host is of utmost importance in understanding the *Neisseria* pathogenesis.

Fox *et.al.* in 2014, described the structure of the Opa₆₀ in solution using a hybrid method involving distance restraints obtained from NMR experiments of Opa in detergent micelle (figure 1.6A) along with the molecular dynamic simulation using a lipid bilayer (Fox, Larsson et al. 2014). Due to long unstructured loops in the beta-barrel, the backbone assignment proved to be difficult and different strategies were applied to achieve maximum assignment. The beta-barrel shape was determined from the amide proton and aromatic-ring proton NOEs. TALOS was used to determine the backbone dihedral angles. Absence of resonances from loops 1, 2 and 4 affected the assignment. A synthetic peptide derived from the loop region was used for assignment. Structure refinement was done using molecular dynamics. The study revealed the dynamic nature of the loop region and offers a good example for the use of NMR in studying not only the structure but also the dynamics of membrane proteins (figure 1.6B).

1.11 Membrane protein structures solved with solid state NMR

Solid-state NMR offers a unique method to study the membrane protein structure and dynamics in membrane environment. In the past, several important membrane proteins have been studied using the Magic-Angle Spinning (MAS) solid-state NMR technique. Some of these membrane proteins include the potassium channel KcsA-Kv1.3 (Lange, Giller et al. 2006; Korukottu, Schneider et al. 2008), the structure and dynamics of the GPCR-rhodopsin (Struts, Salgado et al. 2011), the influenza M2 proton channel (Cady, Schmidt-Rohr et al. 2010) and the fibril formation of the amyloid beta (Ahmed, Davis et al. 2010).

The resolution and sensitivity of the solid-state NMR, unlike the solution-state NMR is influenced by several NMR parameters such as the dipolar couplings, the chemical shift anisotropy (CSA) and quadrupolar couplings etc. Since these cannot be averaged out by the restricted molecular motions in the solid-state samples the spectra will suffer from severe line broadening. To circumvent these problems, the MAS (magic-angle spinning) technique is used.

The method involves spinning the sample (inside the rotor) at a high frequency at an angle of 54.7° to the static magnetic field. This causes an averaging out of the CSA and dipolar couplings. As the frequency of the spinning sample is increased, the resolution and intensity of the solid-state spectra increases.

Most solid-state NMR experiments make use of the heteronuclear decoupling where the proton coupling to the lower gyromagnetic ratio nuclei (such as ^{13}C and ^{15}N) is removed by using a strong radio frequency pulse on proton spins. This irradiation causes the protons to rotate at high rate causing averaging out of its interaction with other spins (Mehring 1983).

To improve the NMR sensitivity most MAS experiments include a Cross-Polarization scheme in the pulse programme. This cross-polarization method involves the magnetization transfer from the greatly abundant, high gyromagnetic nuclei – ^1H to the less abundant, low gyromagnetic nuclei ^{13}C and ^{15}N (Pines, Gibby et al. 1973; M, A et al. 1974).

With rapid development in the field of solid-state NMR, new experiments involving up to four-dimensional correlations can now be achieved (McDermott 2004). These ^{13}C - ^{13}C and ^{13}C - ^{15}N correlation, magic-angle spinning NMR experiments have

been useful in assignment of several uniformly ^{13}C - ^{15}N labelled membrane proteins. Combined use of cross-polarization and dipolar-recoupling pulse sequences allow for the correlation between residues in a protein backbone. NCACX, allow identification of C_α , C_β and C' resonances of a spin-system. The resonances are then correlated to the C_α , C_β and C' resonances of the previous residue (obtained from NCOCX) through the CANCO experiment (Hong 1999; Rienstra, Hohwy et al. 2000). The chemical shift values obtained from the assigned ^{13}C and ^{15}N resonances can be used to predict the backbone dihedral angles. The long-range distance restraints between non-sequential amino-acids are then obtained from ^{13}C and ^1H spin-diffusion experiments (Lange, Luca et al. 2002), which help in identifying the protein fold.

Apart from structure determination, solid-state NMR experiments have also been used to study protein-ligand interactions with the help of homonuclear and heteronuclear experiments like the REDOR (Lipari, Szabo et al. 1982).

1.11.1 Sample preparation for solid-state NMR

1.11.1.1 Using microcrystalline sample: *Yersinia enterocolitica adhesin A* (YadA).

A more established way of generating a well resolved solid-state NMR spectrum involves the preparation of two dimensional or three dimensional crystals. These crystals can be generated by either the hanging drop vapour diffusion method or by using nanocrystalline material which is unsuitable for single crystal diffraction studies. One example of a protein studied in its crystal state is the YadA membrane protein (crystallized from detergent, as seen in figure 1.7A) (Wollmann, Zeth et al. 2006). Due to high resolution of NMR spectra, the crystals offer a good role in methodology development and can provide useful information about the structure and dynamics that might be difficult to obtain from X-ray crystallography.

The *Yersinia* adhesin A (YadA) is a trimeric, transmembrane adhesin, functioning as an autotransporter (TAA) (Linke, Riess et al. 2006). YadA is commonly found in the *Yersinia* species and acts as the virulence factor for *Y. enterocolitica* (El Tahir and Skurnik 2001). All TAAs have trimeric domains with the crucial element being the ‘translocator’ domain which fastens the protein to the outer

membrane of the bacteria (adhesin) and transports the other domains to the extracellular region (autotransporter).

In 2012, Shahid *et.al.* reported the high resolution structure of the transmembrane domain of the *Yersinia enterocolitica* YadA (figure 1.7C) obtained from detergent purified, microcrystalline sample (Shahid, Bardiaux et al. 2012). The entire structure was solved using only uniformly ^{13}C , ^{15}N labelled sample (figure 1.7B). The structure reveals the presence of a trimeric, alpha-helical, coiled-coil, N-terminus passing through the trimeric beta-barrel. Each monomer consists of an alpha-helical N-terminus which ends in four stranded beta-barrel structure. The entire protomer is tilted from the main axis. The hydrophobic surface of the beta-barrel faces the membrane and the barrel itself is formed by hydrogen bonds between the beta-strands 1 and 4. The inner packing of the barrel with smaller amino acids allows for the coiled-coil N terminus to be placed inside the space. The presence of the ASSA motif as seen from the NMR data is suggested to be involved in the autotransport of the trimeric adhesin (TAA). This study is an excellent example of utilising small crystals which do not give good diffractions in X-ray for solid-state NMR studies.

Figure 1.7: MAS solid-state NMR structure of the transmembrane domain of the Yada (Shahid, Bardiaux et al. 2012). A) The Electron micrograph image of the YadaA microcrystals. The crystals appear to be needle shaped. B) Overlaid 2D ^{13}C - ^{13}C CHHC spectrum (black) recorded with 200 µs mixing time and the 2D ^{13}C - ^{13}C DARR spectrum (red) recorded at 25 ms mixing time. All the labeled resonances show long range distance restraints which helped identify the overall structure. Figures modified from (Shahid, Bardiaux et al. 2012).

1.11.1.2 Using precipitated sample: α-crystallin domain of αB-crystallin oligomer

Some proteins like the α-spectrin SH3 domain (Pauli, Baldus et al. 2001) and the αB-crystallin (Jehle, Rajagopal et al. 2010) have been precipitated from their buffers. This method of sample preparation has proven useful in oligomer forming proteins.

The structure of the α B-crystallin oligomer solved using solid-state NMR data has been discussed in the following section.

The α B-crystallin is a subunit of the α -crystallin protein which plays an important role in lens transparency in the eyes. It belongs to the class of small heat shock proteins which perform major roles in protein unfolding, refolding thereby preventing the accumulation of misfolded proteins in the cell which could lead to cell death. Defects in α B-crystallin protein function, have been implicated in many disorders including neurodegenerative diseases, cardiomyopathy, eye defects etc. Understanding the structure of α B-crystallin has been difficult due to the polydisperse character of the protein and its existence as a complex with other proteins. Early efforts to study the structure of this protein utilized truncated constructs containing only the α -crystallin domain of human α B-crystallin. Models were also elucidated from the structures of the crystallin domain of rat HSP20 and other monodisperse heat shock proteins. These studies revealed that the α -crystallin domain exists as a dimer consisting of β 2- β 9 beta-strands.

In 2010, Jehle *et al.* described the structure of the full-length human α B-crystallin using solid-state NMR with SAXS (small angle X-ray scattering) data (Jehle, Rajagopal *et al.* 2010). Previous have discussed the structure of the monomeric α -crystallin domains but fail to explain the heterogeneous native oligomer (Kim, Kim *et al.* 1998; van Montfort, Basha *et al.* 2001). However, with the combined use of MAS NMR and SAXS experiments new insights were gained on the oligomeric α -crystallin domains. The solid-state NMR structure reveals the ‘curved’ nature of the α B oligomers at pH 7.5 and suggests the existence of a ‘pH switch’ regulating the quaternary structure (figure 1.8A). The distance restraints and chemical shifts of the assigned residues helped gain insight into atomic-level resolution of the α B oligomers. Identification of the site of oligomerization (S59 and L70) within the domain was possible due to heterogeneous NMR signals observed from this segment. Finally the SAXS data from two different pH conditions (7 and 6.5) revealed the expansion of the oligomeric structure by release of the C-terminal IXI motif from the substrate binding cleft (figure 1.8B).

Thus use of solid-state NMR along with the complimentary SAXS data was crucial in characterizing a highly dynamic, oligomeric and functionally active state of this chaperone.

Figure 1.8: Solid-state NMR and SAXS structure of the α B-crystallin oligomers (Jehle, Rajagopal et al. 2010). A) Cartoon representation of the α B-crystallin dimer solved by solid-state NMR. B) A model of the α B-crystallin dimers, linked via the C-terminus generated from the SAXS data. Figures modified from (Jehle, Rajagopal et al. 2010).

Solid-state NMR is fast emerging as an excellent tool for the study of large proteins, especially membrane proteins. It serves as a wonderful complimentary tool for X-ray diffraction studies. Since X-ray requires well ordered, high-quality crystals, the lesser quality microcrystals cannot be used for diffraction studies but serve well as solid-state NMR samples. MAS NMR experiments allow the study of dynamics within the membrane domains of the transmembrane proteins which are usually missing from solution-state NMR structures and sometimes even X-ray data.

Chapter 2

2 Structural and Functional Analysis of the Human STromal Interaction Molecule 1 (hSTIM1)

Collaboration with Assoc. Prof. Said Eshaghi (NTU).

Work published in: How J, Zhang A, Phillips M, Reynaud A, Lu SY, Pan LX, et al. (2013). Comprehensive Analysis and Identification of the Human STIM1 Domains for Structural and Functional Studies. PLoS ONE. (How, Zhang et al. 2013).

2.1 Introduction

Calcium ions are required by almost all types of cells for a wide range of functions such as exocytosis, cell differentiation etc. (Feske 2007; Scharenberg, Humphries et al. 2007). One of its major functions is the activation and proper functioning of the immune cells. The levels of Ca^{2+} ions within the T-lymphocytes and the mast cells is maintained by a specialised type of Store-operated Ca^{2+} (SOC) channel known as the Ca^{2+} Release-Activated Ca^{2+} channel (CRAC) (Zweifach and Lewis 1993) (Hoth and Penner 1992). When the Ca^{2+} ion concentrations within the cell drop, the SOC is activated. In the past years, two main components of the SOC induction were discovered, namely Stromal Interaction Molecule 1 (STIM1), which mostly acts as the Ca^{2+} ion sensor in the endoplasmic reticulum (Liou, Kim et al. 2005) and the CRACM1 (Ca^{2+} release-activated Ca^{2+} modulator 1), also known as ORAI1 located in the plasma membrane (Feske, Gwack et al. 2006).

2.1.1 Ca^{2+} Release-Activated Ca^{2+} channel (CRAC)

The first model of CRAC channel proposed that the ER luminal Ca^{2+} is constantly monitored by sensors present in the ER membrane itself. These sensors operate the Ca^{2+} channels present in the plasma membrane. A fall in the luminal Ca^{2+} concentration triggers the activation of the Ca^{2+} channel in the plasma membrane and allows the influx of Ca^{2+} into the cell (Putney 1986). Later studies have revealed that there are two major components of CRAC channel - STIM1 and ORAI1 (Roos, DiGregorio et al. 2005) (Feske, Gwack et al. 2006). STIM1 is responsible for sensing Ca^{2+} concentration in the ER and in response to low levels of Ca^{2+} activates ORAI1, a pore forming channel in the plasma membrane.

2.1.2 STIM1 is the endoplasmic reticulum sensor

STIM1 (Stromal Interaction Molecule 1) is the activator of the CRAC channel. In humans, it is a 77 kDa single pass transmembrane protein. It is made up of several functional domains, such as a pair of EF-hand motif, a sterile α -motif (SAM), protein-

protein interacting coiled-coil domain1 (CRAC activation domain-CAD), a serine/proline rich segment and a poly lysine tail (Lewis 2007). The two EF-hand motifs along with the SAM domain reside in the ER lumen whereas the remaining domains lie in the cytosolic region (Liou, Kim et al. 2005). The single pass transmembrane region acts as a linker between the two termini. Figure 2.1 shows the domains within STIM1 offering a better understanding of the position of the domains.

Figure 2.1: Full length hSTIM1 showing the predicted domains with respective amino-acid length. The diagram gives an indication of the boundaries of each domain.

In 2008, the first high resolution NMR structure of the EF-hand SAM domain was published (Stathopoulos, Zheng et al. 2008). This structure revealed the existence of an extra, “hidden” EF-hand in the N-terminus of STIM1. This hidden EF-hand is not involved in calcium binding but is important for the stabilization of the second EF-hand SAM domain by intramolecular linking. This also explains why STIM1, having two EF-hand motifs, has low calcium affinity ($\sim 250 \mu\text{M}$). The predominantly hydrophobic interactions present between the SAM and EF-hand domains are also responsible for the stability of the STIM1 N-terminus. Recombinant human STIM1 (S58-G201), consisting only of the luminal part of STIM1, was used for NMR structural characterization studies and reveals a compact globular and predominantly alpha-helical nature of the fragment (figure 2.2).

Figure 2.2: Solution state NMR structure of the EF-hand SAM domain with bound calcium published by (Stathopulos, Zheng et al. 2008). The Ca^{2+} ion is shown as a pink sphere.

More recently another group published a high resolution crystal structure of the human SOAR (STIM1-Orai1 Activating Region) (Yang, Jin et al. 2012). Since the wild type protein was extremely difficult to purify, they expressed and purified several mutants of the wild type hSOAR (amino acids 345-442, which corresponds to CC2 and CC3 with extension). One mutant termed hSOAR-MAT (L374M, V419A and C437T) gave high quality crystals and was used for structural studies. Co-localization studies done on this mutant showed that the mutations did not affect the hSOAR function of relocating to the plasma membrane where ORAI1 channel is present. The overall structure of the hSOAR domain is shaped like the letter “R” and consists of four helices. The two termini contain the two long helices while the middle region is composed of two short helices. The crystal structure of the hSOAR-MAT (amino acids 345-442) studied with single wavelength anomalous dispersion (SAD) is shown in figure 2.3.

Figure 2.3: Crystal structure of the mutant, STIM1-ORAI1 Activating Region (human) comprising of the cytosolic STIM1 domains (hSOAR-MAT: L374M, V419A and C437T) published by (Yang, Jin et al. 2012). A) Monomeric hSOAR-MAT displaying the four helices. B) hSOAR-MAT exists mostly as a dimer in its inactive state.

Analytical ultracentrifuge showed that the wild type hSOAR exists mostly as a dimer and partially as a tetramer in solution. These results were contrary to the earlier hypothesis that the SOAR domain exists as tetramer in solution. The dimers is in the shape of the letter “V” and is formed mostly due to multiple hydrophobic interactions between residues present in the N and C termini of the two monomers.

2.1.3 ORAI1 CRAC channel protein

ORAI1 (Calcium release-activated calcium channel protein 1) is an ion channel that plays an important role in the influx of Ca^{2+} ions into the cell. It has four transmembrane domains and forms a tetramer with both N and C termini facing the cytosol (Prakriya, Feske et al. 2006). Monomeric ORAI1 has a molecular weight of 33 kDa. Recently published data shows that a poly acidic region in the putative coiled-coil domain of ORAI1 binds to the positively charged stretch of amino acids (384-386 residues) on hSTIM1 and acts as the gating site (Calloway, Holowka et al. 2010).

2.1.4 CRAC channel activation

In its resting state (in the presence of Ca^{2+} ions), the N terminus of hSTIM1 exists as a monomer while the C terminus occurs as a dimer, with the hSOAR domain blocked by the inhibitory helix (Yang, Jin et al. 2012). A drop in the Ca^{2+} level (from ~ 500 μM to ~ 200 μM , at 20°C) within the ER lumen causes the dissociation of the bound Ca^{2+} from the EF-hand of STIM1 and the oligomerization of the N terminus EF-hand SAM domain. This triggers a major conformational change in the C terminus of STIM1, causing the CC1 domains to stretch out, releasing the SOAR domain which leads the CAD domain to be exposed (Yang, Jin et al. 2012). After STIM1 oligomerization, translocation towards the plasma membrane (where ORAI1 is located) takes place. The co-localization of STIM1 and ORAI1 is termed as “puncta” and this is further stabilized by interaction of a polybasic stretch of amino acids on the cytosolic region of STIM1 with the negatively charged poly acidic region in the C terminus of ORAI1 (Calloway, Holowka et al. 2010). The CRAC channel activation takes place several seconds after the puncta formation (Wu, Buchanan et al. 2006). It has been shown that in resting cells, the autoinhibitory domain in the CC1 of hSTIM1 binds to CAD/SOAR domain to keep the CRAC channel inactive when the intracellular calcium levels are high (Korzeniowski, Manjarres et al. 2011).

Figure 2.4: Proposed model for CRAC channel activation, modified from (Yang, Jin et al. 2012). In the resting state hSTIM1 is a dimer. It oligomerizes upon Ca²⁺ depletion and binds to ORAI1 in the plasma membrane.

2.1.5 Significance of CRAC channel activity

CRAC channels have been the most widely studied and well characterized of the Store Operated Calcium Channels (SOCs). They are known for their involvement in a number of cellular functions such as cell proliferation, gene expression, contraction of muscles and most importantly in the immunological response of T-cells and mast cells (Feske 2007; Scharenberg, Humphries et al. 2007; Di Capite, Bates et al. 2011). The impairment of Ca²⁺ ion regulation leads to several autoimmune and inflammatory diseases. Abnormalities in the development of T and B cells lead to a rare immunodeficiency disease known as Severe Combined Immunodeficiency syndrome (SCID) making the patient highly vulnerable to infectious diseases. Amidst several genetic mutations that are known to cause SCID, a mutation discovered in the CRAC channel has been linked to T lymphocyte malfunction (Carroll, McNaul et al. 2006). Recent findings have also shown that the animal models in which the genes coding for the CRAC channel were deleted had severely compromised functioning of mast cells and allergic responses. This offers great potential for treating severe allergic

responses such as asthma. Most anti-allergy drugs that inhibit the CRAC channel are proving highly effective in the animal models (Di Capite, Bates et al. 2011).

2.1.6 Aim of the Project

The Store Operated Calcium Entry (SOCE) is crucial to the calcium signalling pathway and is found in several human cell types. Several previous studies have tried to understand the role of STIM1 and ORAI1 in the activation of the Calcium Release Activated Channel (CRAC). STIM1 which is present in the endoplasmic reticulum membrane senses the depletion of the Ca^{2+} ions in the ER lumen through its EF-SAM motif and oligomerizes and binds to the ORAI1 channel present in the plasma membrane by interacting through its C-terminal. It was demonstrated that certain residues in the coiled coil domains 1, 2 and 3 in the C terminus cytoplasmic domain of STIM1 are involved in oligomerization and subsequent activation of the ORAI1 channel and were therefore termed - STIM1-ORAI1 Activating Region (SOAR) or CAD (CRAC activating domain). The EF-SAM calcium binding motif structure determined by solution state NMR (Stathopoulos, Zheng et al. 2008) was the only high resolution structure of hSTIM1 at the time of our study. In fact, most of the hSTIM1 and ORAI1 functions were gathered from experiments *in vivo*. A deterrent to structural studies of hSTIM1 can be attributed to the inherent nature of STIM1 which makes it structurally flexible and prone to conformational exchange to allow its oligomerization and translocation to the plasma membrane. In addition to that, the full length hSTIM1 is a huge protein made up of 685 amino acids consisting of a single pass transmembrane domain with a long cytosolic chain (coiled coils 1, 2, 3 and extension).

Since very little information on the large scale production of hSTIM1 is available, we decided to explore the overexpression of hSTIM1 using *Escherichia coli* expression system to produce stable, pure and high quantities of the protein which will be further used for biochemical and biophysical characterization. For this purpose we designed, expressed and purified several fragments of hSTIM1 with varying amino acid composition, in *E.coli*. We then used solution state NMR for the structural characterization and validation of the integrity of the functional domains of hSTIM1. We have also used biophysical assays such as Circular Dichroism spectroscopy and Thermal stability analysis to corroborate our NMR findings.

2.2 Materials and Methods

2.2.1 Cloning

Vector - pNIC28-Bsa4

Escherichia coli expression strains - BL21 (DE3) Rosetta T1R, BL 21(DE3) Rosetta.

Miniprep Plasmid Kit - Qiagen (Hilden, Germany)

2.2.2 Chemicals

Luria Bertini (LB) Broth – Formedium, United Kingdom

Terrific Broth (TB) – Formedium, United Kingdom

Antibiotics - Kanamycin (Sigma, U.S.A), Chloramphenicol (Sigma, U.S.A)

IPTG - Thermoscientific, USA

Labeled salts - Cambridge Isotope Limited (CIL, U.S.A)

Buffers and reagents - Sigma, U.S.A, Calbiochem (Darmstadt, Germany), Fluka (Sigma, Buchs Germany).

Detergents - Cambridge Isotope Limited (CIL, U.S.A), Avanti Polar Lipids (USA), Sigma Aldrich (Singapore).

TCEP - Thermoscientific, USA

2.2.3 Instruments for protein extraction and purification

PANDA - Gea Niro Soavi, Italy

Ni²⁺-NTA – QIAGEN

Gravity flow columns - Bio-Rad, USA

HiLoad 16/60 Superdex 200 - GE Healthcare, USA

Akta Explorer - GE Healthcare, USA

Akta micro - GE Healthcare, USA

Syringe and needles - BD Biosciences

Vivaspin 6 and Vivaspin 20 - Sartorius Stedium Biotech, Germany

2.2.4 Cloning

A library comprising different truncated fragments of full length hSTIM1 (human) was submitted to Protein Platform Production group (NTU, SBS). In brief, the hSTIM1 fragments were cloned into the expression vector pNIC-28-*Bsa*I by ligation independent cloning (LIC). We also introduced DNA sequence for an N-terminal hexa-histidine tag paired together with a TEV protease cleavage site (MHHHHHHSSGVDLGTENLYFQSM). The cDNA were transformed into BL21 (DE3) *Escherichia coli* cells for protein expression.

For our studies we focused on four main fragments – K246-K349 (coiled-coil domain 1 – CC1), K246-Q360 (CC1 with extension), K246-S600 (CC 1, 2 and extension) and S58-Q360 containing the EF-hand, SAM domain, the transmembrane region and the coiled coil 1 domain (EF-SAM, TM, and CC1).

2.2.5 Expression, extraction and purification of hSTIM1 from *E.coli* cells

Small scale expression to test the optimum growth in induction conditions was done by the PPP. We used the optimized conditions to grow large scale cultures of labelled and unlabelled samples for our structural and functional studies respectively.

Unlabelled hSTIM1 protein expression

The overnight cultures were diluted 1:500 in 1 L of TB (Formedium, United Kingdom). The cells were then grown further at 37°C by shaking at 220 rpm until the OD₆₀₀ reached 0.6–0.7. Protein expression was induced by the addition of 0.2 mM IPTG and grown overnight at 18°C. The cell cultures were harvested by centrifugation at 10,000×g for 10 min, and the pellets were suspended in lysis buffer supplemented with a cocktail of protease inhibitors (Roche, Switzerland), 1 mg/mL lysozyme (Sigma-Aldrich, USA) and 10 units/mL Benzonase (Merck, USA). The cells were lysed by three cycles of freezing at –80°C and thawing at 25°C. Unbroken cells and cell debris were removed by centrifugation at 30,000×g for 20 minutes. The

supernatant was loaded onto 1 mL of Ni-NTA agarose beads (Invitrogen, USA), washed with 10 CV of wash buffer and then eluted with elution buffer. Purified protein fragments were then injected onto either a Superdex 75 or Superdex 200 16/60 size exclusion column (GE Healthcare). Prior to IEX, the samples were desalted using a PD-10 desalting column (GE Healthcare) into an initial buffer (20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM TCEP). The desalted samples were subsequently injected into a pre-equilibrated 1 mL HiTrap Q HP column (GE Healthcare, USA). A linear NaCl gradient from 50 mM to 1 M NaCl in an identical buffer was used for the elution of the target protein. All IEX and SEC were performed using an AKTAexplorer FPLC chromatography system controlled by Unicorn software version 5.20 (GE Healthcare), and the fractions were analysed by SDS-PAGE using NuPAGE 4%–12% Bis-Tris gels (Invitrogen) and stained with SimplyBlue Coomassie stain (Invitrogen).

Isotope labelled hSTIM1 for NMR studies

Fragments of hSTIM1 containing the domains coiled coil 1, 2 and 3 with an extension (K246-S600) were used for this study. ^{13}C , ^{15}N -isotope labelled hSTIM1 protein was obtained by growing *E. coli* cells in M9 minimal media containing ^{13}C isotope labelled glucose and ^{15}N labelled NH_4Cl (Cambridge Isotope Laboratories). Induction was performed using 0.5 mM IPTG for approximately 18 hours at 18°C and purified using identical chromatographic procedures as described previously. The ORAI1 C-terminus derived peptide (China Peptides) contained an identical sequence used previously except for the presence of six ^{15}N labelled Leu. For ^2H labelling, same reagents as above were used with the replacement of water by D_2O filtered through 0.2 μM filter.

2.2.6 Circular Dichroism Spectroscopy

This technique was used to study the secondary structure of the hSTIM1 fragments. Far-UV (190-250 nm) was used to acquire steady-state CD spectra (Chirascan spectropolarimeter, Applied Photophysics) of the samples. Simultaneous purging of system with nitrogen gas and acquisition of spectra was done. For each run a 60 μL quartz cell (from Hellma) having a path length of 0.1 mm was used. An

average of three spectra for each temperature scan was taken. Temperature was gradually stepped up from 20°C - 50°C for the study of K246-K349 (CC1 domain) and K246-Q360 (CC1 with extension) fragments. Protein sample had about 4 mg/mL of K246-K349 and ~ 8 mg/ml of K246-Q360 in 20 mM HEPES, 250 mM NaCl at pH 7.5. For fragment S58-Q360, the temperature scan was done from 20, 25, 30 and 40°C for ~ 0.1 mg/mL protein in 20 mM Tris, 200 mM NaCl, 0.1% FC12 and 0.5 mM TCEP at pH 7.5. For baseline correction, a spectrum of buffer was acquired and subtracted from all the respective protein spectra. Microsoft Excel 2007 was used to plot the CD curve for all spectra.

2.2.7 Buffer optimization screen

Optimization for buffer and detergent was performed using the Analytical Size Exclusion Chromatography, Superdex 5/150 (AKTA micro, GE Healthcare, USA). The fragment S58-Q360, which contains the transmembrane domain, was screened for several buffer conditions. These include:

1. Salt concentration – 100 mM, 150 mM, 200 mM and 250 mM, NaCl
2. pH - 6, 7 and 7.5
3. buffer salts - 20mM Tris, 20mM Sodium Phosphate, 50mM Arginine-50mM Glutamic acid and 20mM HEPES.

2.2.8 Thermal Shift Assay

The thermal shift assay offers rapid determination of thermal stability of the different fragments of hSTIM1 in various buffer conditions by monitoring its unfolding at increasing temperatures. The melting point of the protein was calculated based on the increase in the absorbance intensity of the fluorophore (SYPRO orange dye). It was performed using a 96-well plate in a regular qPCR machine. Wells in the first column (A1-H1) contained K246-Q360 (CC1 with extension) protein ~0.4 mg/ml in 20 mM HEPES, 200 mM NaCl, pH 7.5, while wells H1-H9 in another plate contained ~ 0.5 mg/ml of K246-S600 (CC1, 2 and extension) 20 mM HEPES, 200 mM NaCl, pH 7.5. Around 1 µL of SYPRO orange dye (100X diluted in protein buffer), was added to each well. Wells containing only buffer were used for baseline

correction. As the temperature is ramped from 25-95°C, the absorption curve for the fluorophore intensity is plotted against (y-axis) the temperature (in °C on the x-axis). The curve for each run was plotted using Microsoft Excel 07.

2.2.9 NMR Spectroscopy and sample preparation

All experiments were performed on a Bruker AVANCE II 600 MHz and 700 MHz NMR Spectrometer equipped with four RF channels and a 5 mm Z-gradient TCI Cryo-Probe. All NMR experiments were done at 298K (unless specified) and sweep widths for ^1H and ^{15}N were 9804 and 2412 Hz, respectively.

Two-dimensional ^{15}N - ^1H HSQC experiments for fragments K246-K349 (CC1 domain) and K246-Q360 (CC1 with extension), with protein concentration ~ 0.3 mM, ^{15}N labelled, in different buffer conditions and at different temperatures were tried to obtain best spectral resolution.

For K246-S600 (CC1, 2 and extension), a ^{15}N and ^{13}C labelled, 0.3 mM sample was prepared in 20 mM Tris, 250 mM NaCl, and 0.5 mM TCEP at pH 7.5 and used for 2D ^1H - ^{15}N TROSY experiment. Interaction of this domain of hSTIM1 with its interacting partner hORAI1 C terminus derived peptide was tested using 2D NMR experiments. For these studies, the ^{15}N labelled hSTIM1 K246-S600 was prepared in 20 mM Tris-HCl and 300 mM NaCl adjusted to pH 8.0. The peptide was synthesized by Peptide 2.0 (USA) and delivered in a lyophilized state. The sequence of the hORAI1 C-terminal peptide is as follows, the six Leucine residues were ^{15}N labelled for NMR studies: CRSLVSHKTDRQFQELNELAEFARLQDQLDHRGDHPLTPGSHYA. The peptide was prepared as a 50 mM stock solution in an identical buffer as that for the hSTIM1 protein. The addition of the hORAI1 C-terminus-derived peptide to the peptide-free hSTIM1 sample was performed stepwise to achieve a final concentration of hORAI1 peptide: hSTIM1 protein with a ratio of 1:1. The hSTIM1 sample without any hORAI1 peptide was used as the reference for observing any chemical shifts. A 2D ^1H - ^{15}N HSQC was also used to map the interaction site of the protein on the peptide by using unlabelled hSTIM1, K256-S600 and Leucine ^{15}N labelled hORAI1 C-terminus derived peptide.

The fragment S58-Q360, having a concentration of 180 μM (^{15}N and ^2H labelled) in 20 mM Tris, 200 mM NaCl, 0.1% FC12 and 0.5 mM TCEP at pH 7.5 was used to record 2D ^1H - ^{15}N TROSY experiment on the 700 MHz spectrometer.

DOSY experiments were performed at 25°C on a 600 MHz spectrometer for the S58-Q360 (EF-SAM, TM, and CC1) hSTIM1 fragment in the presence and absence of calcium ions. The pulsed magnetic field gradient strength was calibrated using the ^1H signal in a 99.9% D_2O sample with the self-diffusion coefficient of HDO $[(1.902 \pm 0.002) \times 10^{-9} \text{ m}^2/\text{s}]$. A series of 1D spectra were measured with a diffusion delay of 200 ms and increasing gradient strength in 16 steps from 2 to 95% of G_{max} . Diffusion coefficient D_s is calculated by fitting the curve of signal intensity versus variable gradient strength and was processed using Topspin 2.1, and the chemical shifts of the ^1H resonances were referenced directly to the DSS signal.

The sample concentration $\sim 0.2 \text{ mM}$ for the unlabelled hSTIM1, in D_2O containing 20 mM Tris, 250 mM NaCl, 0.1% FC12 and 0.5 mM TCEP at pH 7.5. Diffusion coefficient for calcium loaded sample was acquired by addition of 5 mM CaCl_2 whereas for calcium free samples, 5 mM EDTA was added to the sample to ensure removal of all free Ca^{2+} ions.

2.3 Results

2.3.1 Purification of hSTIM1 fragments

The hSTIM1 library was provided by the protein production platform. It consisted of more than 200 fragments of full length hSTIM1. For NMR characterization purpose, we decided to work with three important domains (four fragments) as shown below.

Figure 2.5: hSTIM1 fragments selected for NMR and other biophysical assays. The choice of the fragments was based on the biological significance of the domains and their good expression and better yield. The domains were predicted from SMART (Schultz, Milpetz et al. 1998).

K246-K349 (CC1) and K246-Q360 (CC1 with extension) contains the first coiled-coil domain with a few amino acid differences. At the start of our study, the domains were selected based on our hypothesis that the CC1 might be involved in interaction with its partner hORAI1. By structural characterisation of the CC1 in the presence of the interacting partner, we hoped to better understand the changes that take place in hSTIM1 upon binding to hORAI1. The fragment K246-S600 contains all three coiled-coil domains 1, 2, 3 which are also known as the Channel Activating Domain – CAD and have been shown to interact with the negatively charged poly acidic region in the C terminus of ORAI1 (Calloway, Holowka et al. 2010). Our main aim with this fragment was to understand the interaction between hSTIM1 and hORAI1. The fragment S58-Q360 contains the EF-hand, SAM domain along with the

trans-membrane domain and the coiled-coil 1 domain. The EF-SAM domain is already known to bind calcium ion (Stathopulos, Zheng et al. 2008). However, not much is known about the changes that take place within the transmembrane domain as well the CC1 domain upon binding or release of calcium ion. With this fragment we wanted to follow these changes.

All four fragments were cloned into pNIC28-Bsa4 vector with kanamycin resistance and transformed into BL21 (DE3) Rosetta cells. The glycerol stock obtained from PPP for the above fragments were used for large scale production of unlabelled as well as isotopically labelled (^{15}N and ^{13}C) samples for NMR. All proteins were purified in two steps. The first step involved passing the His-tagged protein through the Ni-NTA column and eluting with buffer containing high imidazole concentration. The eluate was then concentrated to about 1ml before injecting onto the respective-size exclusion column. Shown below are the Gel-filtration chromatogram profiles and SDS-PAGE analysis of all the four fragments. Proteins from each fraction under the main peak were loaded onto a 12% poly-acrylamide gel and checked for purity. Fractions showing highest purity were pooled together and concentrated for further studies. The hSTIM1 fragments K246-K349 and K246-Q360 gave similar Size Exclusion Chromatogram (SEC) profiles consisting of a single peak. The fractions collected from under the peak were run on the SDS-PAGE and the gel picture is shown on the right – figure 2.6 B. Based on the SEC column used (Superdex 75); the elution volume suggests the protein should be in its monomeric form ~ 15 kDa. The final concentration of pure protein was 0.3 mM as determined by the Nanodrop.

Figure 2.6: Purification profile of hSTIM1 K246-K349.

A) SEC profile of hSTIM1 K246-K349 grown in ^{15}N labelled M9 media.

B) SDS-PAGE of the fractions from the main peak. Lane 1 is the protein M.W. marker. Lanes 3, 5, 6 and 7 showing highest purity and more intense bands were pooled together, concentrated and used for further studies.

The hSTIM1 (K246-S600) containing the CAD domain was purified using 20 mM Tris buffer and concentrated to a final concentration of ~ 0.2 mM (^{15}N labelled). The SEC profile (refer to fig. 2.7A) showed more than one peak and fractions from all three peaks were run on an SDS-PAGE to identify and use the most pure fraction. The SDS-PAGE gel in figure 2.7B shows fractions collected from peak centred at ~ 70 ml (indicated by the arrow and is the peak of our target protein). Most of the fractions coming from within the same peak showed signs of degradation and poor purity. Only the purest fraction was concentrated and used for further studies. The significant intensity of the peak around 50 mL and the presence of higher molecular weight band on SDS-PAGE is an indication of possible oligomerization of the K246-S600 fragment of hSTIM1.

Figure 2.7: Purification profile of hSTIM1 fragment K246-S600 (^{15}N labelled).

A) SEC profile of hSTIM1 K246-S600 grown in ^{15}N labelled M9 media.

B) SDS-PAGE of the fractions (from the target protein) obtained from the main peak at ~ 70 ml (indicated by the arrow). Lane 1 is protein M.W marker. Fractions showing highest purity (lane 2 and 3) were concentrated and used for NMR studies

The ^{15}N , ^{13}C and ^2H labelled hSTIM1 S58-Q360, containing the EF-hand, SAM domain, the transmembrane domain and the coiled-coil 1 was purified in a final buffer containing 0.1% FC12 (Fos-Choline-12 or DPC) and were concentrated to ~ 0.1 mM for NMR studies. Figure 2.8 gives the purification analysis of the protein. Based on the SEC column used (Superdex 200); the elution volume suggests the protein should be in its monomeric form ~ 38 kDa.

Figure 2.8: Purification profile of hSTIM1 fragment S58-Q360 (^{15}N , ^{13}C and ^2H labelled).
A) SEC profile of hSTIM1 S58-Q360 grown in ^{15}N , ^{13}C and ^2H labelled M9 media.
B) SDS-PAGE of the fractions collected from the main peak. Lane 1 is the protein M.W marker. Fractions (lane 2-5) showing highest purity were pooled together, concentrated and used for NMR studies.

2.3.2 Buffer Screening for hSTIM1 fragment containing the transmembrane domain.

Buffer screening for hSTIM1 S58-Q360 was controlled by Size Exclusion Chromatography (Superdex 5/150 GL, GE healthcare, USA). A three step screening process was performed, where the first step involved choosing the optimum buffer (Appendix 1) while the second and third step involved screening different salt concentrations (Appendix 2) and finally the optimum pH (Appendix 3). The optimum conditions were needed to ensure the functional activity of the EF-SAM domain which is known to bind Ca^{2+} ions.

The best buffer conditions for this fragment was 20 mM Tris, 150 mM NaCl, pH 7.0 and 0.1% FC12 (Fos-Choline-12 or DPC) and was chosen based on the narrow linewidth of the absorption peak. We used this buffer condition to proceed with NMR characterization and functional assays.

2.3.3 Secondary structure Analysis of hSTIM1 fragments

To determine the secondary structure of the hSTIM1 fragments in different buffer conditions we used Circular Dichroism Spectroscopy. For K246-K349 and K246-Q360, a temperature scan from 20-50°C was explored and their unfolding curve obtained to get some insight into their thermal stability. The CD curves (as seen in figure 2.9), obtained by plotting the Molar Ellipticity of the protein with the wavelength, shows that both hSTIM1 fragments have high alpha-helical content and are stable at 25°C. Previous papers (Lau, Taneja et al. 1984) suggest that $\Theta_{(222/208)\text{nm}}$ ratio ≥ 1 for coiled coils and ≤ 0.86 for isolated helices. In our study we found $\Theta_{(222/208)\text{nm}}$ ratio for K246-K349 (figure 2.9A) at around room temperature (20°C) to be ~ 0.8 which is similar to isolated helices. With increase in temperature the $\Theta_{(222/208)\text{nm}}$ ratio decreases and the overall secondary structure tends to shift towards random coil values. For fragment K246-Q360 (figure 2.9B) the $\Theta_{(222/208)\text{nm}}$ ratio at room temperature (20°C) ~ 0.80 which is also similar to isolated alpha-helices, but as the temperature is increased a shift towards the coiled-coil state is observed as seen by the increase in the $\Theta_{(222/208)\text{nm}}$ ratio ~ 1.4 (30°C).

Here we conclude that the fragment K246-Q360, which is about 10 amino acids longer than K246-K349, is structurally more similar to coiled coil 1 in hSTIM1 in its native state at the physiologically relevant temperature of $\sim 30^\circ\text{C}$. Also the extra stretch of mostly hydrophobic amino acids (WLQLTHEVEVQ) in the fragment K246-Q360 could be providing hydrophobic interactions which are needed for formation of the coiled-coil super secondary structure. Without this stretch, the fragment K246-K349 is unable to form the coiled-coil and possibly functionally inactive.

Figure 2.10 gives the secondary structure analysis of the hSTIM1 fragment S58-Q360. Buffer conditions similar to the NMR studies were used to ensure that the protein has some secondary structure before using the sample for NMR structural characterization. The CD curve shows alpha-helix secondary structure at room temperature. Not much change in the secondary structure is observed upon the increase of temperature. Based on our CD data we conclude that the fragment S58-Q360 which contains the calcium ion binding motif (EF-SAM domain), in the presence of

detergent, maintains its alpha-helical secondary structure even at higher temperatures of up to 40°C.

Figure 2.9: Effect of temperature on the secondary structure of hSTIM1 fragments K246-K349 and K246-Q360. Both fragments show good alpha-helical content at room temperature. A) K246-K349 ~ 4 mg/ml in 20 mM HEPES, 250 mM NaCl at pH 7.5. B) K246-Q360 ~ 8 mg/ml in 20 mM HEPES, 250 mM NaCl at pH 7.5.

Figure 2.10: hSTIM1 fragment S58-Q360 in 20 mM Tris, 150 mM NaCl, 0.1% FC12 and 0.5 mM TCEP at pH 7.0. The protein shows alpha-helix secondary structure which is stable at 40°C.

2.3.4 Temperature dependent unfolding of hSTIM1

The assay is based upon the principle that a well folded globular protein has its hydrophobic core buried inside, which is exposed only at high temperatures when the protein begins to unfold. At this point the hydrophobic dye SYPRO orange binds to the hydrophobic core and causes an increase in fluorescence intensity (Phillips and de la Pena 2011).

The results for hSTIM1 fragment K246-Q360 and K246-S600 are shown below in Figure 2.11. For fragment K246-Q360 (figure 2.11A), the assay shows high fluorescence intensity even at room temperature. We propose that in the given buffer conditions the exposed hydrophobic region of the protein causes the dye to bind to this region causing high fluorescence intensity. When the temperature is further increased, the affinity of the dye for the exposed hydrophobic residues decreases, causing a gradual downward slope of the fluorescence intensity.

For fragment K246-S600 (figure 2.11B), the initial low fluorescence intensity followed by a “sigmoidal” curve, indicates gradual unfolding of the protein upon increase in temperature. This indicates that the fragment is stable and possibly well-folded in the given buffer conditions and begins to unfold only at ~ 45°C.

Figure 2.11: Thermal stability of the hSTIM1 fragments as seen by the Thermofluor Assay. A) K246-Q360 results display a possible lack of stable tertiary structure at even room temperature. (Protein ~ 0.41 mg/ml in 20 mM HEPES, 200 mM NaCl, pH 7.5). B) K246-S600 results display a well-defined tertiary structure at room temperature which is stable up to 40°C. (Protein ~ 0.5 mg/ml in 20 mM Tris, 300 mM NaCl, 0.5 mM TCEP at pH 8.0).

2.3.5 NMR Study of hSTIM1 fragments

2.3.5.1 Structural characterization of hSTIM1 fragments

Since the hSTIM1 fragments containing the CC1 domains were shown to be more stable and gave better yield and sample purity we continued with NMR studies of the fragments. First we recorded 2D ^{15}N - ^1H HSQC experiments for both hSTIM1 fragments, K246-K349 (figure 2.12) and K246-Q360 (figure 2.13). Both spectra suffered from poor spectral dispersion and resolution. Of the ~ 120 resonances expected, we observed only about 25%.

Figure 2.12: 2D ^{15}N - ^1H HSQC spectrum of K246-K349 fragment acquired at 298K on a 600 MHz spectrometer. Buffer condition used for the experiment contain Protons from aromatic side-chain of tryptophan appear around 10 ppm.

Figure 2.13: 2D ^{15}N - ^1H HSQC spectrum of K246-Q360 fragment acquired at 298K on a 600 MHz spectrometer. Buffer condition used for the experiment contain Protons from aromatic side-chain of tryptophan appear around 10 ppm. Side chain ^1H - ^{15}N resonances of Asn and Gln residues appear as doublets along 112 ppm ^{15}N dimension.

Next we analysed the K246-S600 fragment containing the CRAC Activation Domain (CAD) known to interact with hORAI1 channel. Figure 2.14 shows the 2D ^{15}N - ^1H TROSY spectrum of the fragment recorded in 20 mM Tris, 250 mM NaCl and 0.5 mM TCEP at pH 7.5 at 25°C (298K).

Figure 2.14: 2D ^{15}N - ^1H TROSY spectrum of hSTIM1 (K246-S600) recorded at 298K on a 600 MHz spectrometer. Peaks connected by lines are signals coming from protons in side chains of glutamine and asparagine residues.

For the 43 kDa K246-S600 fragment containing about 270 residues we expected to observe at least 70-80% of the total resonances; however our NMR spectrum shows only about 15% of the total resonances. Even these signals are a mix of sharp and broad resonances. The severe line broadening, overlapping of resonances and complete absence of most of the resonances along with the unstable nature of the protein at temperatures above 25°C was detrimental to the recording of triple resonance experiments for backbone assignment. However, with the limited information we obtained, we were able to tentatively assign the observed resonances to the C-terminal region of this hSTIM1 fragment.

The 2D NMR experiments for hSTIM1 (S58-Q360) containing the EF-SAM the transmembrane helix and the first coiled-coil domain, was recorded on 700 and 800 MHz spectrometer at 30°C. The 2D ^1H - ^{15}N HSQC spectrum (figure 2.15A) showed severe signal overlap and line broadening. Increase of temperature and change in pH conditions did not help improve the spectral quality. Another 2D ^1H - ^{15}N TROSY experiment at 30°C was also performed at 800 MHz spectrometer to rule out signal broadening due to slow tumbling of protein molecule in solution (figure 2.15B). For

this purpose we prepared a ^{15}N , deuterated ($\sim 80\%$) sample of the fragment containing about 5 mM CaCl_2 .

Figure 2.15: NMR studies of S58-Q360 fragment of hSTIM1. A) 2D ^{15}N - ^1H HSQC spectrum of hSTIM1-S58-Q360 recorded at 303K on 700 MHz spectrometer. B) 2D ^{15}N - ^1H TROSY spectrum of hSTIM1 S58-Q360 recorded on a 800 MHz spectrometer. Resonances highlighted in the boxes, belong to side-chain of Gln/Asn residues (112 ppm of ^{15}N) and the aromatic side chain of Trp (around 10 ppm ^1H). About 30% of the total expected resonances were observed.

A reduction in the proton density due to deuteration of the protein, helped resolve the spectral overlap problem to some extent and new resonances which were absent in the non-deuterated sample could be observed (figure 2.15 B). However, even with the use of TROSY type experiments most of the line broadening and absence of signals prevailed. A comparison of our 2D NMR spectra with the previously published 2D NMR spectrum of the recombinant human STIM1 EF-SAM (Ser58 to Gly201) fragment (Stathopoulos, Zheng et al. 2008) confirmed that most of the observed signals in our 2D spectrum is originating from the EF-hand domain. Absence of most resonances, especially from the C-terminus part of S58-Q360 fragment can be attributed to the ability of the CC1 domain to oligomerization resulting in a large complex which is not suitable for NMR studies.

2.3.5.2 Interaction between CRAC Activating Domain of hSTIM1 and hOrai1 C-terminus

To observe the interaction between hSTIM1 CAD domain (containing coiled-coil 1, 2 and 3 with extension) and hORAI1; a hORAI1 C-terminus derived, 40 residue peptide with specific Leu-[^{15}N] labelling was used. Sample containing 0.2 mM hSTIM1 (K246-S600) in 20 mM Tris, 300 mM NaCl and 0.1 mM TCEP at pH 7.5 was used to record a reference 2D ^{15}N - ^1H TROSY spectrum at 25°C. The spectrum is shown in red in figure 2.16. The peptide was added to the hSTIM1 sample in a 1:1 molar ratio. Final spectrum in blue is overlaid with the reference spectrum and shown in figure 2.16.

Figure 2.16: hSTIM1-ORAI 1 interaction followed by NMR. Superimposed 2D ^{15}N - ^1H TROSY spectra (amide region) of free hSTIM1, K246-S600 (red) and hSTIM1 in the presence of hORAI1 C-terminus derived peptide (blue), at 25°C.

Since the spectrum had poor signal dispersion and severe overlap problems to begin with, the addition of hORAI1 C-terminus derived peptide could not be monitored accurately. We thus proceeded with recording a 2D ^{15}N - ^1H HSQC of the free peptide and peptide in the presence of unlabelled hSTIM1 K246-S600 fragment. Due to specific Leu- ^{15}N labelling, we observed only 6 resonances as expected from the six leucine in the peptide amino acid sequence. Upon addition of the unlabelled hSTIM1 protein, changes in the position of resonances of hORAI1 derived peptide were observed. As seen in figure 2.17, four out of the six resonances experienced significant chemical shift perturbation of more than 1 ppm. The resonance labelled 1 (8.5 ppm ^1H , 128 ppm ^{15}N) disappeared in the presence of the STIM1 fragment. These significant changes in the Leucine resonances of the hORAI1 C-terminus derived peptide are an indication that the hORAI1 C terminus undergoes local structural changes upon interaction with the hSTIM1 CAD domain. We suspect, based on the large chemical shift perturbations (>1.0 ppm) along ^1H dimension that the presence of the hSTIM1 fragments causes a conformational change in the secondary structure of the C-terminus of the hORAI1 channel in solution.

Figure 2.17: Overlay of ^1H - ^{15}N , 2D HSQC spectra recorded on Leu- ^{15}N labeled free hOrai1 C-terminus derived peptide (blue) and with peptide in the presence of hSTIM1 K246-S600 (red) on 700 MHz spectrometer at 25°C in 20 mM HEPES, 300 mM NaCl and 0.1 mM TCEP at pH 7.5.

2.3.5.3 Effect of Ca^{2+} ions on hSTIM1

The EF hand-SAM motif is known to interact with the Ca^{2+} ion and senses the drop in calcium ion concentration within the ER lumen causing oligomerization of STIM1. Our hSTIM1 fragment, S58-Q360 was designed to contain the EF hand-SAM domain along with the transmembrane domain and coil 1 region. To monitor the structural transition of this hSTIM1 fragment from monomer to a higher oligomer, translational diffusion coefficients of S58-Q360 fragment in the presence (addition of 5 mM CaCl_2) and absence of calcium ions (addition of 5 mM EDTA) were calculated using DOSY NMR experiments.

Diffusion coefficient D_S (with Ca^{2+} ions) = $7.426\text{e-}11 \text{ m}^2/\text{s}$ (SD = $2.591\text{e-}02$);

Diffusion coefficient D_S (without Ca^{2+} ions) = $6.680\text{e-}11 \text{ m}^2/\text{s}$ (SD = $4.927\text{e-}02$).

We assume that the decrease in the diffusion coefficient of the fragment in the absence of Ca^{2+} ions might be caused by the oligomerization of the hSTIM1 fragment containing the calcium binding EF-motif. Since the changes in the hydrodynamic radius can also be caused by a change in the overall structure of the protein complex, we decided to analyse the calcium free (with 5 mM EDTA) and calcium loaded (with 5 mM CaCl_2) NMR samples by CD spectroscopy. CD spectroscopy experiments would give a good estimate of the changes within the protein secondary structure level. As seen from figure 2.18, the secondary structure of the fragment S58-Q360 is not altered in the absence or presence of calcium ions. Hence, we conclude that the hSTIM1 fragment-S58-Q360, containing the functionally active Ca^{2+} binding EF-SAM motif exists as a possible oligomer in the absence of Ca^{2+} ions.

Figure 2.18: The secondary structure of the S58-Q360 fragment is preserved even in the absence of calcium ions. Spectra in the presence and absence of calcium ions are similar.

2.4 Discussion

Full-length STIM1 (human) is a 77 kDa single pass transmembrane protein consisting of amino acid 685 residues. It belongs to an essential class of adaptor coiled-coil proteins known to possess high intrinsic dynamics. It is an important component of the SOCE and is responsible for activating the ORAI1 channel to allow the calcium entry into the cells. Its N terminus, consisting of two EF-hand motif and a SAM domain which act as Ca^{2+} ion sensors, is present within the ER lumen (Stathopulos, Zheng et al. 2008) while the C-terminus which resides in the cytoplasm consists of an extended coiled coil region with a serine/proline rich tail (Fahrner, Muik et al. 2009). The N and C terminus are joined together by a single pass transmembrane region (~ 20 amino acid stretch) which is located within the ER membrane.

At the time of our study there was limited knowledge on the mechanism of hSTIM1 activation, mostly due to absence of structural information on the cytoplasmic region (the CAD domain). NMR solution state structure of the N-terminus hSTIM1 was the only high resolution structure of the hSTIM1 at the time of our study (Stathopulos, Zheng et al. 2008). The structures of the other domains of hSTIM1 were not available till 2012, when Yang *et.al* described the first crystal structure of the hSOAR-MAT mutant domain of hSTIM1 (Yang, Jin et al. 2012). The mutations helped stabilize the otherwise extremely difficult to purify, wild-type hSTIM1 protein. Most of our understanding of the hSTIM1 functions is derived using *in-vivo* studies. The limited information on the structure of STIM1 and *in vitro* studies was attributed to the unique nature of STIM1: a large, transmembrane protein, containing an extended C-terminus with long coiled coil domain and an inherent nature to oligomerize in order to bind to its interacting partner. This nature of hSTIM1 poses a challenge in the overexpression of the protein, as was observed in our study with several constructs containing different domains of hSTIM1 overexpressed in *E.coli* system. Most of the fragments (constructs) showed poor stability, low yield and high tendency towards aggregation or even degradation (How, Zhang et al. 2013).

After an exhaustive high-throughput study involving over 200 fragments of hSTIM1, we were able to identify the crucial regions and boundaries in the full length hSTIM1 necessary for high quality and stability of the hSTIM1 sample (How, Zhang et al. 2013). To study and characterize the cytoplasmic domain of hSTIM1, which is

known to be containing the ORAI1 interacting domain (CAD/SOAR) and for which very little structural information was available, we designed multiple hSTIM1 fragments with various truncated hSTIM1 C-terminus domains to be used for both biophysical and NMR studies.

2.4.1 hSTIM1 coiled coil domain 1 is susceptible to oligomerization in solution

We used the hSTIM1, K246-K349 and K246-Q360 fragments containing the first coiled-coil domain of the hSTIM1 C-terminus, for NMR studies. Two dimensional heteronuclear NMR experiments like ^1H - ^{15}N HSQC were acquired at different temperatures and in different buffer conditions to obtain the best spectral resolution for structure determination. Low pH conditions were limited due to the low isoelectric point of the protein (~6.0) and could not be used for NMR. However the best NMR spectra that we were able to obtain also suffered from severe line broadening and signal overlap. The narrow amide proton dispersion and absence of a majority of peaks is mostly due to the high alpha helical content and possible aggregation or partial unfolding of the protein. Even in the absence of NMR resonances from the coiled-coil domains, the CD data clearly indicates the presence of alpha-helical and even coiled-coil secondary structure in the two hSTIM1 fragments (K246-K349 and K246-Q360) and that these secondary structures are stable over a range of temperature and buffer conditions in solution.

Together with fluorescence based thermal stability assay used to monitor the temperature-dependent unfolding of the two proteins, we hypothesize that, the first coiled-coil domain, mostly involving residues 246-360, by itself, has good alpha helical secondary structure but lacks a well-defined three-dimensional tertiary structure. We also hypothesize that on its own; the coiled coil 1 domain is not very stable and prone to oligomerization, as is its inherent nature. Our findings are supported by a recent study involving single wavelength anomalous dispersion (SAD) method and analytical ultracentrifuge technique, revealing an extended long helix of the wild type coiled coil 1 fragment forming an oligomer in solution (Cui, Yang et al. 2013).

2.4.2 NMR characterization of hSTIM1 CAD containing domain

According to studies done by Hoover et al, the hSTIM1 CAD domain interacts with the ORAI1 to activate the channel (Park, Hoover et al. 2009). To study the interaction between the CAD domain (K246-S600) as well as the hORAI1 C-terminus derived peptide, the resonances from the protein and the peptide were obtained from NMR studies. For this purpose we expressed and purified the hSTIM1 K246-S600 containing the coiled coil domains (CC1, IH) and the cytosolic tail. The hSTIM1 fragment had a molecular weight of ~ 43 kDa and contained 380 residues. Presence of extra peaks in the size exclusion chromatogram profile and additional bands on the polyacrylamide gel is an indication of degradation as well as aggregation of the fragment. The thermal stability assay gave a typical S-shaped curve characteristic of a well-folded protein melting at higher temperatures indicating the preservation of a tertiary fold, unlike K246-K349 and K246-Q360.

The 2D NMR spectrum of free hSTIM1 (K246-S600) showed significant line broadening and absence of a majority of resonances even in different buffer conditions. Based on our previous results with the CC1 containing hSTIM1 fragments we hypothesize that the residues present in the coiled-coil structures are involved in hydrophobic interactions and cause aggregation of the protein in solution. The resulting aggregate has a higher molecular weight tumbles slowly in solution and causes line broadening of the NMR spectra due to more effective relaxation (Wagner 1993). To overcome the problem of high molecular weight related line broadening, we performed 2D ^1H - ^{15}N TROSY NMR experiment with the K246-S600 hSTIM1 fragment. This spectrum showed a mixture of sharp and broad resonances with a majority of resonances still absent. We hypothesize that hydrophobic interaction between the residues (hydrophobic) in the coiled coil domains of the K246-S600 fragment cause the aggregation of the protein in solution. This results in lowering of the tumbling rate of the protein molecule which manifests as severe line broadening in the NMR spectra. The C-terminus of the fragment (devoid of the CC domains) is relatively unstructured and therefore has a higher tumbling rate than the bulkier CC domain and is observed as sharper resonances in the NMR spectra. Another possibility for the poor spectral resolution could be the presence of conformational changes within the coiled coil domains of STIM1 occurring on an intermediate NMR timescale.

The addition of known binding partner hORAI1 C-terminus derived peptide to the K246-S600 C-terminus of hSTIM1 was monitored by acquiring 2D ^1H - ^{15}N TROSY experiments. Even though no new resonances were observed in the presence of the peptide, we did observe some chemical shift perturbations in the overlapped central region of the spectrum. Since poor spectral quality of the K246-S600 fragment was a limitation to our NMR studies, we decided to monitor the hSTIM1-hORAI1 interaction by using ^{15}N Leu labelled hORAI1 C-terminus derived peptide for recording 2D HSQC NMR experiment. A comparison of the spectra of the free and bound peptide indicates that at least five out of the six labelled leucine are either directly involved in the STIM1-ORAI1 C-terminus interaction or are near the binding site. Based on the significant chemical shift perturbation observed in the ^{15}N resonances of the leucine residues of the bound peptide, we propose that hORAI1 C-terminus undergoes a significant conformational change upon binding with hSTIM1. Our initial NMR findings confirm the functionality of the hSTIM1 K246-S600 fragment in solution and are in agreement with the results published by Calloway et al. in 2010, which showed that a poly-acidic stretch on the Orai1 C-terminus binds to a poly lysine stretch (384-387) on hSTIM1 cytosolic tail (Calloway, Holowka et al. 2010).

2.4.3 NMR characterization of the N-terminus S58-Q360 fragment of hSTIM1

Based on the study performed by Huang et al. in 2009 that showed that the EF-hand of hSTIM1 is responsible for sensing the calcium ions in the ER lumen, we designed several fragments around the N-terminus and transmembrane domain of the hSTIM1. These were intended to help understand the conformational changes occurring within the CC1 of STIM1, in the presence and absence of Ca^{2+} ions and hence provide a better understanding of the oligomerization of STIM1 upon channel activation. One of these fragments, S58-Q360, containing the EF-SAM motif along with the transmembrane and first coiled-coil domain was used for NMR studies. The 2D HSQC spectrum of the fragment in solution gave a mixture of sharp and broad resonances with only about 15-20% of the total expected resonances observed at 30°C. The use of deuterated hSTIM1 fragment helped improve the line broadening problem

to some extent. The appearance of new resonances in the deuterated sample confirmed that the absence of some signals due to severe line broadening was caused due to high proton density in large protein molecules. Initial calcium binding studies done using CD and DOSY NMR experiments have confirmed the presence of a higher oligomeric state of the hSTIM1 fragment in the absence of Ca^{2+} ions. Thus to rule out the possibility of oligomerization in the NMR sample of hSTIM1 S58-Q360, due to the absence of calcium, we added 5 mM CaCl_2 to the NMR sample. The presence of calcium did not help improve the spectral resolution and the spectrum still showed absence of most resonances and signal overlap. The narrow dispersion along the H^{N} dimension in the NMR spectra is attributed to the presence of supercoiled alpha-helix domains (246-342). The presence of detergent in the sample, to stabilize the transmembrane domain also adds to the increase in molecular weight and causes further line broadening. Despite the absence of NMR resonances from the transmembrane and CC1 domains, the secondary structure determination using CD showed that the protein is mostly alpha-helical and stable at high temperatures of up to 40°C. Overall, results with the S58-Q360 hSTIM1 fragment suggest an interaction between the coiled coil domain (CC1) with one another in the given NMR conditions possibly resulting in a high molecular weight aggregate. The interaction of the transmembrane domain with the detergent (DPC) micelles also adds to the bulkiness of the protein. This causes the protein molecule to tumble at a slow rate resulting in severe broadening of the NMR signals. However, even with limited NMR characterization of the hSTIM1 fragment, we were able to establish the secondary structure of the protein and its ability to bind to calcium ions.

Chapter 3

3 NMR studies of Presenilin Enhancer- 2 (PEN-2) sub-unit of Gamma Secretase complex

Collaboration with Prof. Pär Lennart Nordlund

Manuscript in preparation

3.1 Introduction

Gamma-secretase is an intra-membrane aspartyl protease transmembrane protein consisting of four components: Presenilin 1 or 2 (PS 1/2), Anterior Pharynx Defective (APH-1), Nicastrin (NCT) and Presenilin Enhancer 2 (PEN-2) in a ratio of 1:1:1:1 (figure 3.1) (Dries and Yu 2008). Presenilin (1 or 2) is a polytopic membrane protein (~ 9 TMDs) which is cleaved into N and C terminal fragments of about 30 kDa and 20 kDa respectively, each containing a catalytic aspartate site. Presenilin is the catalytic site for the complex with Nicastrin possibly serving as a substrate binder and Aph-1 and PEN-2 serving important roles in the assembly and stabilization of the active complex.

3.1.1 Role of gamma secretase in the pathophysiology of Alzheimer's disease

Gamma secretase has been known to cleave a number of different substrates, one of which is the Amyloid Precursor Protein (APP). The APP is first cleaved by α or β -secretase which results in two fragments: the soluble APP which is in the extra cellular domain and the membrane containing fragment which can be either C88 or C99 depending upon whether it is cleaved by either α or β -secretase, respectively. Both C88 and C99 serve as a substrate for γ -secretase complex. When C99 is cleaved by the Presenilin subunit of the complex, the resulting fragments includes a 6 kDa known as the APP-Intracellular Domain(AICD) and the A β peptide (Kreft, Martone et al. 2009). Thus the main complex involved in the generation of the infamous A β peptide is the gamma-secretase complex. Figure 3.1 shows the active gamma-secretase complex with all four subunits assembled and in position for the catalytic activity of PS1.

Figure 3.1: Assembly of the active gamma-secretase complex as proposed by Dries *et. al.* The active complex consists of Presenilin (PS, green), Nicastrin (Nct, light blue), Aph-1 (dark blue), and Pen-2 (pink). On the left is shown the TM6 and TM7 of the Presenilin subunit cleaving the amyloid precursor protein resulting in the formation of the A β peptides. The numbers are indicative of the approximate molecular weight. The figure is modified from (Dries and Yu 2008).

Alzheimer's disease (AD) was first described in 1906, by a German psychiatrist, *Alois Alzheimer*. It is a common and destructive neurodegenerative disease affecting millions of humans across the globe. It is characterized by loss of memory (amnesia) and neuropsychiatric symptoms (Angelucci, Spalletta et al. 2010). Every year millions of dollars are invested in the Alzheimer's research to look for a cure. Current line of treatment for AD includes acetylcholinesterase inhibitors and the *N*-methyl-d-aspartate antagonist called memantine, which have been known to ameliorate the cognitive disorder (Greenblatt, Kryger et al. 1999).

Analysis of the human AD brains, have indicated the accumulation of extracellular plaques – Amyloid Beta and fibrillary tangles of aberrant filaments of τ protein (Goedert 2009). The A β is composed of A β peptide which itself is a 38-43 amino acid peptide originating from the amyloid precursor protein (APP) (Hardy and Selkoe 2002). Inside the plaques, the A β peptide is present as an insoluble, aggregated and oligomeric form (Kayed, Head et al. 2003). Brain injury is caused by the deterioration of neurons and gliosis surrounding the extracellular plaques (McGeer and McGeer 1995). In many cases of AD, cerebral amyloid angiopathy (CAA) is also observed which is caused by the aggregation of A β peptide in the blood vessel walls.

This may lead to several life threatening complications in AD patients (Smith and Greenberg 2009). Various studies have suggested the role of A β peptide and its role in initiating Alzheimer's disease pathogenesis and the role of τ protein in the progression of the disease. Under normal conditions, τ protein stabilizes the intracellular microtubules by binding to tubulin but in AD, τ undergoes hyper phosphorylation leading to self-aggregation (Kidd 1963).

3.1.2 Role of PEN-2 in gamma-secretase complex activity

Human Presenilin Enhancer 2 (hPEN-2) is a 101 amino-acid, integral trans-membrane protein and a subunit of the gamma secretase complex (GS) (Prokop, Haass et al. 2005). Knockdown studies on PEN-2 have shown that it plays an important role in stabilizing the γ -secretase complex and the endoproteolytic cleavage of Presenilin (Mao, Cui et al. 2012). The GS complex is degraded by the proteasome in the absence of PEN-2. The stability of the N and C fragments of the Presenilin subunit of the complex also depends on the length and sequence of the C terminus of PEN-2 (Prokop, Haass et al. 2005). The residues 90-94 in the PEN-2 C terminus are critical for binding to the Gamma Secretase complex (Hasegawa, Sanjo et al. 2004). The "NF" motif in the fourth transmembrane domain of the Presenilin subunit is crucial for binding with PEN-2 (Kim and Sisodia 2005).

3.1.3 PEN-2 structural studies

Currently there is no high resolution structure of human PEN-2 (hPEN-2). However, the topology of hPEN-2 within the endoplasmic reticulum was determined using a mutant, engineered with artificial glycosylation sites suggesting a hair-pin model for the transmembrane protein. In this model its two transmembrane helices span the ER membrane whereas the N and C termini protrude into the ER lumen (Crystal, Morais et al. 2003). On the contrary, studies by Bammens *et.al.* 2011, suggested that some residues in the hydrophobic domain 1 and loop region of hPEN-2 lie in a water containing cavity and thus gave a new model for the topology of hPEN-2

in the ER, where the transmembrane helix 1 is partly unstructured and transmembrane helix 2 is a regular alpha-helix (Bammens, Chavez-Gutierrez et al. 2011).

Figure 3.2: Topology of hPEN-2 in the endoplasmic reticulum membrane. A) Topology model of hPEN-2 as proposed by Crystal et al suggesting two parallel transmembrane helices with N and C termini in the lumen (Crystal, Morais et al. 2003). B) And C) Topology model of hPEN-2 as proposed by Bammens *et. al.* suggesting the existence of a hydrophilic cavity with residues from transmembrane helix 1 and loop region residing in this cavity (Bammens, Chavez-Gutierrez et al. 2011). Figure modified from (Bammens, Chavez-Gutierrez et al. 2011).

Recently, a 4.32Å resolution three dimensional structure of the entire human γ -secretase complex was determined using Single particle, Cryo-EM technique (Sun, Zhao et al. 2015). The structure was obtained after fusing a T4 lysozyme to the N-terminus of the PS1 subunit of the γ -secretase complex to unambiguously identify the transmembrane helix 1 of PS1 even in low resolution. Even with moderate resolution, the 20 transmembrane domains could be unambiguously assigned to the four subunits of the complex. In this model they propose that the membrane region contains centrally located PS1 with the N-terminus interacting with the PEN-2 subunit while the C-terminus fragment of PS1 interacts with APH-1. Most of the extracellular domain of the γ -secretase complex is formed by the extracellular domain of the subunit Nicastrin. Its only transmembrane helix interacts with APH1 and with PEN-2 in the extracellular domain. The entire transmembrane part of the complex resembles a horse-shoe with close proximity of the two ends of the horse-shoe. The TM6 and TM7 helices of PS1 which have been reported to contain the catalytic Aspartate residues, are present on the “convex side” of the horse-shoe transmembrane. An interesting discovery was made with regards to the topology of PEN-2. The structure reveals the presence of three transmembrane helices instead of two, as reported earlier. The first two transmembrane helices are half helices forming a hairpin and are a part of the previously termed TM1. The N-terminus of PEN-2 originates in the intracellular cavity and loops into the

second transmembrane domain. The loop itself lies on the intracellular side and extends into the third TM helix (TM3) which ends into the extracellular side. The structure also indicated that in its bound form, PEN-2, TM1 and TM3 interact closely with the TM4 of PS1 while the C-terminus of PEN-2 also interacts with the extracellular domain of Nicastrin. Thus the N-terminus and loop lie on the intracellular side while the C-terminus lies on the extracellular side. This topology is in contrast to the two previous models suggesting that both N and C terminus of PEN-2 lie in the extracellular cavity.

Figure 3.3: Human Gamma-Secretase complex structure, revealed by single particle Cryo-Electron microscopy (Sun, Zhao et al. 2015). Left hand side is the 4.32 Å resolution structure of the whole human γ -secretase complex with T4 lysozyme attached to the intracellular side. On the right hand side is the topology of PEN-2 within the complex, bound to Nicastrin on the extracellular side and interacting with the TM4 of PS1 within the membrane. Figures modified from (Sun, Zhao et al. 2015).

3.1.4 Aim of study

With the discovery of the role of the gamma-secretase complex in the synthesis of A β peptides, there has been a high interest to study the γ -complex as a potential Alzheimer's therapeutic target. Since inhibiting the production of A β peptides from the Amyloid Precursor Protein (APP) requires manipulation of gamma-secretase activity; a detailed insight into the overall structure of the active complex would be

indispensable. PEN-2 is an important subunit of the γ -secretase complex as it activates the Presenilin subunit which is the catalytic site for the complex.

At the time of our study, there was no high resolution structure of the PEN-2 subunit alone or together with the complex. Many important questions regarding assembly of the γ -secretase complex *in vivo* and the structural changes within the complex upon activation were still unanswered. Even though PEN-2 is crucial to the activation of the catalytic site of the complex, little is known about its three dimensional structure. Being a small protein with only 101 amino acids, it serves as an excellent drug target. Altering the structure of PEN-2 with drug inhibitors might be useful in inhibiting the activation of the complex and hence inability to cleave the amyloid precursor protein.

Our aim was to study the structure of PEN-2 in solution using Nuclear Magnetic Resonance Spectroscopy as a tool. NMR is a useful tool for studying drug targets like PEN-2 as it not only provides the atomic resolution three dimensional structure of the protein but also gives an insight into the dynamics of the protein, such as protein-ligand interactions at the atomic level. This would help gain insight into not only the structure of PEN-2 but also help us understand the dynamics of PEN2 upon interaction with a possible drug with relevance to AD. In the following chapters we discuss the methods and results obtained from our NMR studies in the process of identifying the structure and topology of the full length hPEN-2.

3.2 Materials and Methods

3.2.1 Cloning

Vector - pNIC28-Bsa4

Escherichia coli expression strains - BL21 (DE3) Rosetta T1R, BL21 (DE3) Rosetta.

Miniprep Plasmid Kit - Qiagen (Hilden, Germany)

3.2.2 Chemicals

Luria Bertini (LB) - Formedium, United Kingdom

Terrific Broth (TB) - Formedium, United Kingdom

Antibiotics - Kanamycin (Sigma, U.S.A), Chloramphenicol (Sigma, U.S.A)

IPTG - Thermoscientific, USA

Labeled salts - Cambridge Isotope Limited (CIL, U.S.A)

Buffers and reagents - Sigma, U.S.A, Calbiochem (Darmstadt, Germany), Fluka (Sigma, Buchs Germany).

Detergents - Cambridge Isotope Limited (CIL, U.S.A), Avanti Polar Lipids (USA), Sigma Aldrich (Singapore).

TCEP - Thermoscientific, USA

3.2.3 Instruments

PANDA - Gea Niro Soavi, Italy

Ni²⁺-NTA – QIAGEN

Gravity flow columns - Bio-Rad, USA

HiLoad 16/60 Superdex 200 - GE Healthcare, USA

Akta Explorer - GE Healthcare, USA

Akta micro - GE Healthcare, USA.

3.2.4 Cloning

3.2.4.1 Full length PEN-2 (human) and hPEN-2 fragments

Full length (human) PEN-2 plasmid was provided by Prof. Paer Nordlund's group from Karolinska Institute, Sweden. The plasmid was cloned into pNIC28-Bsa4 vector with N-terminus 6X His-tag and carrying tetracycline resistance gene and was transformed into BL21 (DE3) *Escherichia coli* cells.

All hPEN-2 fragments comprising different domains were submitted to the Protein Platform Production group (NTU, SBS) for initial expression screening. The constructs were cloned into pNIC-Bsa4 vector, which contains a kanamycin resistant gene, and transformed into BL21 (DE3) Rosetta T1R cells. Each construct had a 6X His-tag at the N-terminus. SDS-PAGE and western blot analysis was done to identify the best expressing constructs. These constructs were then screened for optimal expression conditions. For structural studies only two constructs, S51-S93 and R16-A61 were chosen based on the good expression and purification yield.

3.2.4.2 Cell-free Expression of PEN-2 (human)

Human PEN-2 full length gene was cloned into pEXP5-NT/TOPO vector (Invitrogen, USA) for cell free synthesis. Cloning was performed using manufacturer's protocol. DNA samples were sent for sequencing for verification. Successfully cloned plasmids were stored at -80°C and used in the cell free production of PEN-2.

3.2.5 Expression, extraction and purification of hPEN-2 from *E.coli* cells

Small scale expression to test the optimum growth conditions were done by the PPP. We used the optimized conditions to grow large scale cultures of labelled and unlabelled samples for our structural and functional studies respectively. The procedure was consistent for all fragments of hPEN-2 and the full length.

Unlabelled hPEN-2 protein expression

The overnight cultures were diluted 1:500 in 1 L of TB (Formedium, United Kingdom). The cells were then grown further at 37°C by shaking at 220 rpm until the OD₆₀₀ reached 0.6 –0.7. Protein expression was induced by the addition of 0.2 mM IPTG and grown overnight at 18°C. The cell cultures were harvested by centrifugation at 10,000×g for 10 min, and the pellets were suspended in lysis buffer supplemented with a cocktail of protease inhibitors (Roche, Switzerland), 1 mg/mL lysozyme (Sigma-Aldrich, USA) and 10 units/mL Benzonase (Merck, USA). The cells were lysed by three cycles of French press [PANDA (Gea Niro Soavi, Italy)] at 4°C with ~ 700-800 bar pressure. The lysate was centrifuged at 4,500 g for 10 minutes at 4°C. The membranes were collected at 4°C by using ultra-centrifuge at 40,000 rpm for 60 minutes. The membrane pellets were solubilized in 1% FC12 (DPC) detergent in a ratio of 1:20 for membrane pellet: solubilization buffer (20 mM Tris, 300 mM NaCl, 0.5 mM TCEP, 20 mM imidazole, and pH 8.0). After 15 minutes the solubilized membrane was re-centrifuged at 40,000 rpm for 60 minutes at 4°C to remove any insolubilized matter. The supernatant was loaded onto 1 mL of Ni-NTA agarose beads (Invitrogen, USA), washed with 10 CV of wash buffer and then eluted with elution buffer. Purified protein constructs were then injected onto either a Superdex 75 or Superdex 200 16/60 size exclusion column (GE Healthcare). hPEN-2 SEC was performed in the buffer condition of 100 mM NaCl, 20 mM sodium phosphate, pH 5.5 to 6.5, 100 mM detergent (DPC/SDS) and 0.5 mM TCEP, using an AKTAexplorer FPLC chromatography system controlled by Unicorn software version 5.20 (GE Healthcare), and the fractions were analysed by SDS-PAGE using NuPAGE 4%–12% Bis-Tris gels (Invitrogen) and stained with SimplyBlue Coomassie stain (Invitrogen).

Isotope labelled hPEN-2 for NMR studies

All constructs of hPEN-2 were eventually ¹³C, ¹⁵N-isotope labelled for NMR studies. To obtain labelled samples, the hPEN-2 plasmid containing BL21 (DE3) Rosetta cells were grown in M9 minimal media supplemented with ¹³C isotope labelled glucose and ¹⁵N labelled ammonium chloride (Cambridge Isotope

Laboratories). Induction was performed using 0.5 mM IPTG for approximately 18 hours at 18°C and purified using identical chromatographic procedures as described previously. For ^2H labelling, same reagents as above were used with the replacement of water by D_2O filtered through a 0.2 μm filter paper.

Isotope labelling of selective methyl groups for NMR studies

The sample for Methyl-TROSY HSQC set of 2D ^1H - ^{15}N HSQC and HMQC experiments was prepared by following the labelling strategy published by Tugarinov *et.al.* (Tugarinov, Kanelis et al. 2006). Briefly, for the methyl labelling scheme ^{12}C ([U- ^{15}N , ^2H]; Ile δ 1- [$^{13}\text{CH}_3$]; Leu, Val-[$^{13}\text{CH}_3$, $^{12}\text{CD}_3$]) the starting compound was 2-keto-3- d_2 -4- ^{13}C , d_1 -butyrate dissolved in D_2O . The protocol for protein overexpression is the same as described previously, only difference is the use of uniformly deuterated ^{12}C -glucose and the addition of the above α -ketoisovalerate at an O.D. $\sim$ 0.25. The culture is allowed to grow till an O.D. $\sim$ 0.5-0.6 is reached before inducing with 0.3 mM IPTG. The remaining protocol for induction and purification is same as described previously. Final protein sample is exchanged into a buffer containing deuterated 100 mM DPC in 20 mM sodium phosphate, 150 mM NaCl, pH 6.0 and 0.5 mM TCEP.

Cell-free expression

The cell free expression kit was purchased from Life technologies, Invitrogen. The cell free expression of full length hPEN-2 was done according to the protocol mentioned in the ExpresswayTM Mini Cell-Free Expression System manual. Western blot assay was performed on the precipitated protein to check for expression and purity.

3.2.6 Circular Dichroism Spectroscopy

Far-UV (190-250 nm) was used to acquire steady-state CD spectra (Chirascan spectropolarimeter, Applied Photophysics) of the full length hPEN-2 for secondary structure determination. Simultaneous purging of system with nitrogen gas and

acquisition of spectra was done. For each run a 60 μ L quartz cell (from Hellma) having a path length of 0.1 mm was used, with data acquired at 25°C and a step resolution of 1 nm. For Thermo-stability experiment, the temperature of the sample holder controlled by an external water-bath was gradually increased from 25-45°C and spectra recorded at intervals of 5°C. An average of three spectra was taken for recording the ellipticity values. Two sample conditions with different detergents were chosen for the assay. One condition had ~1.0 mg/ml in 100 mM DPC (FC12) in 20 mM Sodium Phosphate, 100 mM NaCl, and 0.5 mM TCEP at pH 6.0. while the other condition had ~1.0 mg/ml in 100 mM SDS in 20 mM Sodium Phosphate, 100 mM NaCl, and 0.5 mM TCEP at pH 6.0. For baseline correction, a spectrum of protein buffer was acquired and subtracted from the protein spectra. Microsoft Excel 2007 was used to plot the CD curve for all spectra.

3.2.7 Nuclear Magnetic Resonance Spectroscopy and sample preparation

Nuclear Magnetic Resonance (NMR) spectroscopy is a tool to measure the three dimensional structure of proteins and nucleic acids at atomic resolution. It can also be used to study protein dynamics as well as the functional activity of proteins in solution.

3.2.7.1 One dimensional (1D) NMR spectroscopy

All experiments were performed on a Bruker AVANCE II 600 MHz and 700 MHz NMR Spectrometer equipped with four RF channels and a 5 mm Z-gradient TCI Cryo-Probe. All experiments were conducted with ~ 500 μ L of isotopically labelled samples containing 10% D₂O (v/v) and 0.1 mM DSS for internal ¹H signal referencing. The chemical shifts of ¹³C and ¹⁵N nuclei were calculated indirectly from the ¹H chemical shifts (Cavanagh, Fairbrother et al. 1995).

1D, ¹H NMR spectra contain valuable information regarding the stability of protein in the given buffer condition. Dispersion of the ¹H signals in the amide (6-8 ppm) and the methyl (negative ppm) regions are an indication of how well folded the protein is under given buffer and temperature conditions. The hPEN-2 purified protein

in varying buffer conditions and at different temperatures were screened using ^1H 1D NMR experiments (with using pre-saturation water suppression). Sample preparation for NMR involved addition of 10% D_2O (v/v) and 0.5 mM DSS to $\sim 500\ \mu\text{L}$ of unlabelled 0.2 mM protein in a 5 mm NMR tube.

3.2.7.2 Multidimensional NMR spectroscopy

Two dimensional ^{15}N Heteronuclear Single Quantum Coherence (HSQC) is one of the first and most essential experiments to be performed on any labelled protein sample. It gives the backbone amide H-N correlation and thus serves as the “fingerprint” of the protein. Dispersion of peaks in the spectrum give an indication of protein fold in given buffer conditions and helps in deciding if further isotopic labelling and hence triple resonance experiments will be required.

Sequence-specific assignment of backbone was achieved using standard triple resonance experiments (HNCA, HN(CO)CA, CBCACONH, HNCACB, HNCO) with 0.3 mM ^{15}N , ^{13}C , ^2H ($\sim 70\text{-}80\%$ deuteration) labelled full length, his-tagged hPEN-2 in 100 mM NaCl, 20 mM sodium phosphate, pH 5.5, 150 mM SDS and 0.5 mM TCEP at 45°C - 50°C .

For Methyl-TROSY HSQC set of experiments, $[\text{U-}^2\text{H}, ^{15}\text{N}]$; Ile $\delta 1$ - $^{13}\text{CH}_3$; Leu, Val- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$ hPEN-2 full length protein in 100 mM deuterated DPC was used for recording 2D heteronuclear experiments at 30°C on the 600 MHz spectrometer.

Hydrogen-deuterium experiments were performed to examine the integrity of the structural domains of the full length, his-tagged hPEN-2 sample in the given buffer conditions. Uniformly ^{15}N labelled, full length hPEN-2 in 100 mM NaCl, 20 mM sodium phosphate and 100 mM DPC was aliquot into three, each containing 500 μL of sample and lyophilized. The first sample was suspended in aqueous buffer and was used as a reference. A 2D ^{15}N - ^1H Tr-HSQC experiment was acquired at 35°C for the reference sample. The second sample was suspended in $\sim 60\%$ D_2O . A 2D ^{15}N - ^1H Tr-HSQC spectrum was acquired at 35°C after 10 minutes of incubation on ice. The third sample was suspended in 100% D_2O and a 2D Tr-HSQC recorded at 35°C . The intensity of the amide proton resonances of the deuterium exchanged spectrum was

compared to the reference spectrum to identify the fast exchanging and slow exchanging domains within the protein.

PRE measurements were performed with sample containing U-¹⁵N labelled, 0.2 mM full length hPEN-2 in 20 mM sodium phosphate buffer with 100 mM NaCl, 0.5 mM TCEP and 100 mM DPC. MTSL ((*S*-(2, 2, 5, 5-tetramethyl-2, 5-dihydro-1H-pyrrol-3-yl) methyl methanesulfonothioate) was used as the paramagnetic label which covalently binds to free cysteine in the protein. The reference spectrum was recorded at 40°C in the absence of any MTSL. After incubation of sample with MTSL (protein: MTSL concentration was 20:1 molar) at 37°C for one hour, the protein was then desalted and buffer exchanged into the NMR buffer to remove any excess spin label. A 2D Tr-HSQC of the spectrum was then recorded for the protein + paramagnetic probe condition.

For the denaturing of hPEN-2, we increased the concentration of urea in the reference sample in a stepwise manner from 2 M urea to 4 M urea. For each step, a 2D ¹⁵N-¹H Tr-HSQC spectrum was acquired, starting with the reference sample (no urea) – 0.2 mM U-¹⁵N labelled, full length hPEN-2 in 20 mM sodium phosphate buffer with 100 mM DPC at 40°C.

Fragments of hPEN-2 with U-¹⁵N labelling were also used for recording 2D heteronuclear NMR experiments at 30°C.

All NMR data was processed using TopSpin 2.1 and spectra were analysed using CARA software.

3.3 Results

3.3.1 Cloning, Expression and Purification of hPEN-2

For structural characterization of full length, wild type PEN-2 (human), the plasmid was provided by our collaborator Prof. Paer Nordlund's group (Karolinska Institutet, Sweden). Expression and purification of full length hPEN-2 was first screened in small scale. Based on the expression conditions, the purification yield and the stability of the purified protein, large scale expression and purification of protein in the tested optimum conditions was done to obtain samples for further studies.

MHHHHHHGSTSLYKKAGSENL**YFQGMNLERVSNEEKLNLCRK****YYLGGFAFLPFLWLVNI**
FWFFRE**AFLVPAYTEQSQIK****GYVWRS****AVGFLFWVIVLT****SWIT****IFQIYRPRWGALGDYLSFT**
IPLGTP

Figure 3.4: Full length hPEN-2 containing N-terminus, 6XHistidine tag and a TEV cleavage site. On top is the amino acid sequence of the full length wild-type hPEN-2 with a 6Xhistidine, N-terminus tag and TEV cleavage site. Highlighted residues are predicted transmembrane domains predicted by TMHMM (Schumann, Riepl et al. 2007). Molecular weight of the full protein with tag is 14.9 kDa.

BL21 (DE3) Rosetta cells transformed with the PEN-2 (human) full length plasmid was used for the large-scale production of the membrane protein. The final step of purification of the isotopically labelled full length hPEN-2 involved Size Exclusion Chromatography. The U-¹⁵N, U-¹³C and U-²H labelled protein collected from the fractions under the main peak was loaded onto a 12% SDS polyacrylamide gel to check for sample purity. Fractions showing highest purity were pooled together and concentrated. Around 7ul of purified and concentrated sample was used to perform Western Blot assay. Figure 3.5 shows a typical Size Exclusion chromatogram profile of full length PEN-2 and the inset shows the Western blot result with an intense band around 14 kDa, confirming the presence of N-terminus His-tagged full length hPEN-2.

The protein was concentrated and the concentration was measured using the Nanodrop. Final protein used for NMR structural characterization was labelled with ~ 0.3 mM ^{15}N , ^{13}C , ^2H (~70-80% deuteration) in 100 mM NaCl, 20 mM sodium phosphate, pH 5.5, 150 mM SDS and 0.5 mM TCEP.

Figure 3.5: A) Typical purification profile of full-length hPEN-2 isotopically labelled M9 minimal media for NMR studies. B) Western Blot analysis of the purified protein from the main peak is seen as an intense band around 14 kDa. C) Fractions from the peak ~ 80 mL were used to run the SDS-PAGE.

The full length hPEN-2 sample with $[\text{U-}^2\text{H}, ^{15}\text{N}]$; Ile δ 1- $^{13}\text{CH}_3$; Leu, Val- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$ labelling for the Methyl TROSY type NMR experiments were also prepared. These samples were used to study the global fold of the protein in the detergent complex where the labelled methyl groups would serve as indicators of the protein dynamics. For this study, we purified the protein using Size Exclusion Chromatography and ran the fractions on SDS-PAGE for purity analysis. The final hPEN-2 sample was buffer exchanged into 20 mM sodium phosphate, 100 mM NaCl; 100 mM deuterated DPC and 0.5 mM TCEP at pH 5.5.

3.3.2 Cell-free expression

The cell free expression of full length hPEN-2 was analysed using western blot assay. Anti-His antibody was used to detect the 6Xhistidine N-terminus tag on the protein. Another protein with a mol. wt. of ~ 19 kDa was also used as a control. As seen in the figure 3.6, greater intensity of the protein is observed for the fraction containing the precipitated hPEN-2 without the detergent (DDM), as compared to the protein in the micelles. We hypothesize that the presence of detergent in the reaction system is not conducive to the protein expression and hence the lesser yield of protein in the presence of the detergent.

Figure 3.6: Immunoblotting result for full length hPEN-2 expressed in cell-free system. Lane 1 is the protein ladder while lane 4 has the 19 kDa control protein and lanes 7 and 8 have hPEN-2 in detergent (DDM) and without detergent (~ 14 kDa) respectively.

3.3.3 Secondary structure determination of wild type hPEN-2

To examine the secondary structure of full length hPEN-2 in solution, we decided to use Circular Dichroism spectroscopy. A temperature scan from 25-45°C was performed with the protein in 100 mM DPC. As can be seen from the figure 3.7A, hPEN-2 in DPC detergent micelles is able to maintain an alpha-helix secondary structure even at temperatures of up to 45°C. As seen from figure 3.7B, hPEN-2 in 100 mM SDS detergent also shows alpha-helical content and is stable at higher temperatures of 45°C and 55°C.

The protein-detergent complex is a large molecular weight species which tumbles slowly in solution and causes pervasive line broadening in the NMR spectroscopic studies. To overcome this problem of slow tumbling protein molecule we acquired most of the NMR spectra at high temperatures without causing protein aggregation or unfolding. Our hPEN-2 study in both DPC and SDS detergent micelles show that the full-length protein has a well-defined alpha-helix secondary structure and is stable at high temperatures. These conditions will be essential for recording further NMR experiments for structural characterization studies of hPEN-2.

Figure 3.7: Effect of temperature on the secondary structure of hPEN-2 full length. A) Full length hPEN-2 in 100 mM DPC (FC12) showing alpha-helical secondary structure from temperature ranging between 25°C to 45°C. B) Full length hPEN-2 in 100 mM SDS showing alpha-helical secondary structure even at high temperature of 55°C.

3.3.4 Nuclear Magnetic Resonance Spectroscopy

3.3.4.1 Detergent screening using 2D Heteronuclear NMR Spectroscopy

Acquiring 2D ¹H-¹⁵N HSQC spectrum is the first step towards structural characterization of a protein. Each amide group in the amino acid (except proline), gives rise to a signal in the two dimensional spectrum resulting in a “fingerprint” of the protein. The resonances coming from protons in side chains of glutamine and asparagine residues appear as doublets around ~ 112 ppm (^{N^H}) and ~ 7 ppm (^{H^N}) whereas tryptophan side-chain protons appear at ~ 130 ppm (^{H^N}) and ~ 10 ppm (^{H^N}).

To mimic the native membrane topology of the full length hPEN-2, we decided to test different detergents such as SDS, DHPC, and even DMPC: DHPC bicelles. Our main aim with the screening was to find conditions which will maintain the solubility and stability of the protein at the same time allowing resonance assignment using NMR experiments.

Figure 3.8: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, his-tagged, U- ^{15}N labelled hPEN-2 in 100 mM DPC in buffer containing 100 mM NaCl and 20 mM sodium phosphate at pH 5.5 recorded at 40°C.

Figure 3.8 shows 2D ^1H - ^{15}N TROSY HSQC spectrum of full length (U- ^{15}N labelled) his-tagged hPEN-2 in 20 mM sodium phosphate buffer with 100 mM NaCl, 100 mM DPC and 0.5 mM TCEP at pH 5.5, recorded at 40°C. We observed about 28% of the total expected resonances (121 resonances coming from backbone amide of residues except proline). The spectrum suffered from narrow dispersion along the proton dimension, signal overlap and absence of several resonances.

Figure 3.9, shows 2D ^{15}N - ^1H TROSY HSQC spectrum of full length, his-tagged U- ^{15}N labelled hPEN-2 in 100 mM SDS, 20 mM sodium phosphate, 100 mM

NaCl and 0.5 mM TCEP at pH 5.5 recorded at 40°C. The spectrum showed improvement with the appearance of more new peaks. However, dispersion along the proton dimension was still less and the spectrum still suffered from signal overlap. We were able to detect about 40% of the total expected resonances.

Figure 3.9: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, U- ^{15}N labelled hPEN-2 in 100 mM SDS in buffer containing 20 mM sodium phosphate, 100 mM NaCl and 0.5 mM TCEP at pH 5.5 recorded at 40°C.

Figure 3.10 shows 2D ^{15}N - ^1H TROSY HSQC of the full length his-tagged hPEN-2, buffer exchanged into 100 mM DHPC. Although the spectrum shows better dispersion in the amide region, the total number of peaks visible was similar to the spectrum acquired with the protein in 100mM SDS. Peaks appearing in the central region still exhibited line broadening and overlapping with majority of peaks being undetectable.

Figure 3.10: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, U- ^{15}N labelled hPEN-2 in 100 mM DHPC in buffer containing 20 mM sodium phosphate, 100 mM NaCl and 0.5 mM TCEP at pH 5.5 recorded at 40°C.

Bicelles which are formed by a mixture of short and long chain lipids, have been known to provide a better membrane mimetic environment than detergent micelle for *in vitro* studies (Poget, Cahill et al. 2007). We recorded a 2D ^{15}N - ^1H TROSY HSQC experiment for the full length, his-tagged hPEN-2 ^{15}N labelled protein in buffer containing a mix of 25 mM DMPC, 100 mM DHPC bicelles in 20 mM sodium phosphate, 100 mM NaCl and 0.5 mM TCEP at pH 5.5 at 40°C (figure 3.11). We were now able to observe all the resonances in the glycine region (110 ppm ^{15}N , 8 ppm ^1H) coming from the ten glycine amino acids in the sequence. Out of the total 6 tryptophan in the amino acid sequence of the full length hPEN-2, we observed side chain resonances of at least five tryptophan residues (10 ppm ^1H). The signal overlap problem was also greatly reduced; however, the absence of a majority of signals could not be overcome. In fact the number of resonances detected in the central amide region

was less than what was observed when the protein was in detergents. We hypothesize that even though the bicelles (DMPC: DHPC) provide a more native membrane like environment to the protein; they increase the size of the protein-lipid complex causing line broadening of signals beyond detection. Due to this problem, we decided against using the bicelles for structural studies.

Figure 3.11: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length, his-tagged, U- ^{15}N labelled hPEN-2 in 25 mM DMPC: 100 mM DHPC bicelles. Spectrum was recorded at 40°C.

3.3.4.2 NMR characterization of full length his-tagged hPEN-2 in solution

The high proton density in the protein-detergent complex is one of the major causes of severe signal overlap. Large size of the complex also restricts the tumbling rate in solution and further causes line broadening due to relaxation of the proton

signals. To overcome these problems we decided to produce full length his-tagged hPEN-2 in uniformly ^{15}N , ^{13}C and deuterium media. The resulting protein - 0.3 mM, $\text{U-}^{15}\text{N}$, ^{13}C , ^2H (~70-80% deuteration) labelled in 100 mM NaCl, 20 mM sodium phosphate, pH 5.5, 150 mM SDS and 0.5 mM TCEP was used for recording the triple resonance experiments. The perdeuteration of the protein sample has been known to reduce the proton density and hence reduce the pervasive line broadening and signal overlap problem. We acquired 2D ^1H - ^{15}N TROSY HSQC spectrum of the sample at both 45°C and 50°C. Deuterated hPEN-2 in 150 mM SDS at 50°C produced the best resolved spectra (figure 3.12).

Figure 3.12: 2D ^1H - ^{15}N Tr-HSQC spectrum of 0.3 mM full length hPEN-2, his-tagged, $\text{U-}^{15}\text{N}$, ^{13}C and ^2H -labeled in 100 mM NaCl, 20 mM sodium phosphate, pH 5.5, 150 mM SDS and 0.5 mM TCEP at 50°C on a 700 MHz spectrometer. Residues identified have been labelled on the spectrum. The central amide region has been zoomed in to show the assignment. We achieved about 54% of the total assignment.

However, the spectrum of his-tagged hPEN-2 in 150 mM SDS at 50°C still contained a mixture of sharp and broad resonances which could be an indication of the conformational exchange within the full length hPEN-2. With the help of 3D Heteronuclear experiments like – Tr-HNCA, Tr-HNCO, Tr-HNCACB and Tr-HNCOCA, we were able to assign 66 out of 121 expected resonances. Most of these assigned residues lie in the extra membrane loop region while a few are within the transmembrane domain. Figure 3.13 is a strip plot of one of the assigned fragment comprising W110-G114 using 3D Tr-HNCA (blue) for $^{13}\text{C}_\alpha$ and $^{13}\text{C}_{\alpha-1}$ resonances for the spin systems and 3D Tr-HNCOCA for confirming the $^{13}\text{C}_{\alpha-1}$.

Figure 3.13: Strip-plot of 3D Tr-HNCA (blue) and 3D Tr-HNCOCA (red) overlaid to show the connectivity between the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_{\alpha-1}$ of the fragment W110-G114 of his-tagged, full length U- ^{15}N , ^{13}C and ^2H labelled hPEN-2.

The two major bottlenecks preventing completion of the chemical shift assignments of the full length hPEN-2 were the chemical shift degeneracy, particularly of the ^{13}C spins and incomplete observation of all expected resonances. The assignment statistics are given in the Appendix 4.

Using I-TASSER (Yang, Yan et al. 2015) for the three dimensional structure predictions of the full length PEN-2; five final models were generated of which the model with the highest confidence score was selected. This model was used to map the assigned residues. A comparison with the published Cryo-EM structure of PEN-2 in the active gamma-secretase complex shows that there is possible interaction between TMD3 and TMD2 which could be the reason for the complete absence of resonances from these regions.

Figure 3.14: NMR characterization of full length hPEN-2. A) Assigned residues (red) mapped to the predicted structure of the full length hPEN-2 obtained from I-TASSER. Most of the assignments come from the N and C termini and part of the transmembrane domains. B) The Cryo-EM structure of PEN-2 obtained from (Sun, Zhao et al. 2015) showing the three transmembrane domains.

3.3.4.3 Secondary structure prediction from NMR data

Since the chemical shift of the $^{13}\text{C}_\alpha$ and ^1H resonances have a strong correlation to the secondary structure of the protein, it is possible to determine the “secondary chemical shift” by using the equation:

$$\Delta\delta = \delta_{\text{observed}} - \delta_{\text{random coil}}$$

Where we use the chemical shifts of the $^{13}\text{C}_\alpha$ resonances of all the 66 assigned resonances (Wishart and Sykes 1994). The $^{13}\text{C}_\alpha$ chemical shift values are positive in an alpha-helix environment whereas it is negative for beta-strands. Figure 3.15, below gives the secondary structure propensity of the assigned amino acids in his-tagged full length hPEN-2.

Figure 3.15: Secondary Structure prediction using the $^{13}\text{C}_\alpha$ chemical shifts of the assigned residues. Some residues from the N-terminus loop show propensity towards the alpha-helix, while assigned residues from the second transmembrane domain show high alpha-helix structure. Asterisk represents a break in the sequence. The sequence includes the 25 residues of the His tag and the TEV cleavage site.

As can be seen from the secondary chemical shift plot above, the residues in the second transmembrane domain possess a high alpha-helical content. The residues from the long N-terminus containing the histidine-tag and the TEV cleavage site also show some propensity towards alpha-helix. Most residues in the loop show tendency towards beta-strand or are unstructured. Overall, the predictions show good association with the predicted domains within the full-length hPEN-2.

3.3.4.4 Methyl-TROSY NMR experiments of full length hPEN-2

For large molecular weight protein-detergent complexes, line broadening caused by faster relaxing protons is detrimental to NMR studies. Perdeuteration is helpful as it reduces the proton density and purifying proteins in aqueous buffers ensures reintroduction of about 20% of the total protons at labile sites such as the backbone amide. This helps in improving the signal overlap and line broadening problems (Kerfah, Plevin et al. 2015). However we still need protons at the key sites due to their high NMR sensitivity. For this purpose we specifically label the methyl groups in a uniformly deuterated protein sample (Tugarinov, Kanelis et al. 2006). Methyl groups are selected owing to their small size which allows them to tumble fast and give rise to well-defined signals and their hydrophobic nature which is the reason why they are abundant in the hydrophobic domains of the protein. To help us understand the global fold of the full length hPEN-2 protein in its detergent environment, we decided to selectively methyl label the protein sample [U-²H, ¹⁵N]; Ile δ 1-[¹³CH₃]; Leu, Val-[¹³CH₃, ¹²CD₃] (Tugarinov, Kanelis et al. 2006). The first experiment we recorded was the 2D ¹⁵N-¹H TROSY HSQC in 100 mM deuterated DPC at 45°C (refer to figure 3.16). We observed a much improved spectral resolution as compared to previous spectrum with non-deuterated sample. A total of about 77 out of the expected 125 peaks (60%) could be observed.

To monitor the ¹³C-labeled methyl groups we acquired the 2D ¹³C-¹H HSQC of the [U-²H, ¹⁵N]; Ile δ 1-[¹³CH₃]; Leu, Val-[¹³CH₃, ¹²CD₃] his-tagged, full length hPEN-2. Since the focus was on the ¹³CH₃ groups of Isoleucine, Leucine and Valine, the decoupler offset was set at 35 ppm (¹³C) and the spectral width along the ¹³C dimension was set to 80ppm. Of the total 29 resonances expected (Ile-7, Leu-15 and Val-7) a majority were observed, although most of them suffered from signal overlap (figure 3.17). Theoretically there should be seven sharp resonances coming from the fast tumbling methyl side-chains of the seven isoleucine residues, however more than the expected resonances (figure 3.17, Ile δ 1 region) were observed. In the Leucine, Valine methyl side chain region (> 20 ppm, ¹⁵N) there was significant signal overlap and only a few well resolved signals were observed.

Figure 3.16: 2D ^{15}N - ^1H Tr-HSQC spectrum of full length hPEN-2 [$\text{U-}^2\text{H}$, ^{15}N]; Ile δ 1- $^{13}\text{CH}_3$; Leu, Val- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$] in 20 mM sodium phosphate, 100 mM NaCl and 100 mM deuterated DPC at pH 5.5 recorded at 45°C on 600 MHz spectrometer.

Figure 3.17: 2D ^{13}C - ^1H HSQC spectrum of full length hPEN-2 [$\text{U-}^2\text{H}$, ^{15}N]; Ile δ 1- $^{13}\text{CH}_3$; Leu, Val- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$] in 20 mM sodium phosphate buffer containing 100 mM NaCl and 100 mM DPC at pH 5.5 recorded at 30°C.

Overall, based on our NMR data we conclude that the perdeuteration is helpful in lowering signal overlap problem and can help resolve NMR spectra. With the use of ^{13}C -labeled methyl groups of IVL to record TROSY-type 2D HSQC experiments, we can hypothesize that full length hPEN-2 in solution is prone to conformational exchange on an intermediate timescale (>1 ms) the net effect of which is causing line broadening, presence of extra set of signals as well complete absence of some methyl resonances.

3.3.4.5 Amide hydrogen-deuterium exchange studies using NMR

Backbone amide hydrogen-deuterium exchange experiments have been widely used to identify the solvent protected backbone amide protons that maybe involved in hydrogen bonding in both soluble and membrane proteins (Henry and Sykes 1994). The experiments have also been long used for characterizing unfolded or disordered domains within well-folded or partly folded proteins (Baum, Dobson et al. 1989). We applied this method to elucidate dynamics of the full length hPEN-2 in detergent micelles.

Based on the backbone assignment we know that majority of the resonances observed in the 2D Tr-HSQC experiment belong to residues in the extra membrane loops. The reference spectrum (figure 3.18A) was acquired with the full length protein in aqueous buffer at 35 °C after lyophilization to ensure that absence of peaks is not due to experimental errors. After 10 minutes of incubation on ice, the lyophilized sample was dissolved in ~ 60% D_2O and another 2D Tr-HSQC was acquired at 35°C (figure 3.18B). About 20 resonances were still visible, while all others had been exchanged. Out of these visible resonances, some came from the side chain of Asparagine and Glutamine residues, while very few came from within the transmembrane domain (G46, G47, V95 and L96). The experiment was repeated with another sample lyophilized and dissolved in 100% D_2O (figure 3.18C). Within ten minutes, the protein showed complete exchange of all backbone - ^1H N with deuterium indicating very little protection of the residues observed in the 2D Tr-HSQC spectrum.

Figure 3.18: Amide hydrogen-deuterium exchange followed for full length, his-tagged hPEN-2, U- ^{15}N -labeled, in 20 mM sodium phosphate buffer in 100 mM DPC at 35°C using NMR. A) 2D Tr-HSQC of reference sample in aqueous buffer. Resonances still visible have been highlighted. B) 2D Tr-HSQC of sample suspended in 60% D_2O and incubated for 10 minutes. C) 2D Tr-HSQC of sample in 100% D_2O incubated for 10 minutes.

The hydrogen-exchange data agrees well with our backbone assignment. Most of the assigned residues belong to the loop regions which are less protected from the solvent, hence undergo fast exchange (<10 minutes) with deuterium resulting in signal

loss. The few residues in the transmembrane domains that have been assigned show little protection from the solvent and thus when deuterium strength is increased, these residues also exchange and a complete loss of signal is seen. Literature available on the hydrogen-deuterium exchange study of membrane proteins in micelles, studied using NMR, suggest that the amide protons of residues in the transmembrane regions exchange very slowly with the solvent (Deisenhofer, Epp et al. 1984; Terstappen and Reggiani 2001; Nam, Jeon et al. 2009; Murray and Blundell 2010). Our results are therefore unexpected since we expected most of the residues in the transmembrane region to have better protection against the solvent. However, when taken together with the complete absence of resonances from the transmembrane region, we can hypothesize that there is considerable dynamics in the transmembrane domains of the full length hPEN-2 in solution.

3.3.4.6 PRE measurements of full length hPEN-2

To get a better understanding of the topology of hPEN-2 in solution we decided to perform PRE measurements using the native Cysteine residue (C40) in his-tagged, full length hPEN-2. PRE involves the use of paramagnetic compounds often referred to as the “paramagnetic probe” which cause the enhancement of NMR relaxation rates of nuclei in its vicinity. PRE measurements utilize the conjugation of the paramagnetic probe to a solvent exposed site in the protein (Kosen 1989). In 2011, complete structural characterization of the Mitochondrial uncoupling protein 2 (UCP2) was achieved using PRE measurements (Berardi, Shih et al. 2011). The group used MTSL ((S-(2, 2, 5, 5-tetramethyl-2, 5-dihydro-1H-pyrrol-3-yl) methyl methanesulfonothioate) as the paramagnetic label which covalently binds to free cysteine in the sample. We decided to use the same probe for our PRE measurements. The 2D Tr-HSQC reference spectrum was recorded using U-¹⁵N labelled, 0.2 mM full length hPEN-2 in 20 mM sodium phosphate buffer with 100 mM NaCl, 0.5 mM TCEP and 100 mM DPC at 40°C. Another 2D Tr-HSQC experiment was then recorded in the presence of MTSL. Figure 3.19 gives the PRE measurement results.

Our PRE results indicate that the full length hPEN-2 has two highly dynamic, N and C-termini which are in close proximity with the C40 residue. We also hypothesize that since residues on the distal end of the cytoplasmic loop (A71, T73,

E74 and Q77) were also perturbed by the probe, the two putative TM helices are not parallel but are at an angle which allows the proximal (towards the loop) end of TM2 to be closer to the C40 residue. This can be visualised as in figure 3.19B. Only the key residues that were helpful in defining the Pen-2 topology have been highlighted in blue.

Figure 3.19: PRE measurements performed for full length hPEN-2 using MTSL as the paramagnetic probe. A) 2D Tr-HSQC of full length hPEN-2 in the presence (red) and absence (blue) of the paramagnetic probe. Highlighted assigned residues showed either a drop in signal intensity or complete disappearance of signal, indicating their close proximity to the C40-MTSL residue. B) Pen-2 structure as shown in the Cryo-EM gamma secretase complex (Sun, Zhao et al. 2015) has been used to map the key residues (highlighted in blue) that disappeared in the presence of MTSL probe (at residue C40 –Red). These residues help define the topology of Pen-2 in solution.

3.3.4.7 Progressive unfolding of hPEN-2 in solution studied by NMR

The addition of denaturing agents such as urea or guanidine hydrochloride is known to disrupt the tertiary structure and in large quantities even the secondary structure of soluble proteins. However in case of membrane proteins the denaturants tend to disrupt only the extra membrane regions whereas the transmembrane regions are not disturbed, this makes them more resistant to both chemical and thermal denaturation as compared to soluble proteins (Bowie 2004), (Haltia and Freire 1995).

The reason why transmembrane helices are generally very stable towards denaturation is due to the presence of detergents, which form micelles around the hydrophobic residues of the transmembrane domains and protect them from the denaturants in the solvent. These detergents therefore, under optimal concentrations, not only stabilize the secondary structure of the polypeptide backbone but also contribute towards thermostability and resistance towards denaturing agents. In NMR, most resonances coming from the transmembrane helices are broadened either due to large size of the protein-detergent complex or due to effective relaxation of protons in the high proton density environment and are not observed in the spectrum. One of the methods to validate this hypothesis is to add denaturants in the protein sample and record NMR experiments to monitor the progressive unfolding of the protein. We expect residues in the previously ordered domains (mostly transmembrane) to now begin to unfold and their resonances to appear in the NMR spectrum. These unfolded residues give rise to sharper resonances as there is greater mobility in the disordered regions.

For this purpose, we recorded the 2D Tr-HSQC of the full length, his-tagged hPEN-2, in 20 mM sodium phosphate, 100 mM NaCl, 0.5 mM TCEP and 100 mM DPC at 40°C (figure 3.20A). As seen previously only about 30% of the total expected resonances were observed. Next, we acquired 2D Tr-HSQC of the same sample after stepwise addition of urea from 2M – 4M. An examination of the 2D HSQC spectra of the samples with different concentrations of the denaturants revealed an interesting unfolding pattern. The full length his-tagged hPEN-2 contains 10 Glycine residues in the amino acid sequence. Of these, three lie in the N-terminus (G8, G17 and G25), two glycine lie in the first transmembrane domain (G46 and G47), two in the second transmembrane domain (G80 and G88) and the remaining three in the C-terminus (G111, G114 and G124). With our previous experiments we were able to observe only about six resonances (G8, G14, G25, G46, G111 and one unassigned resonance) in DPC detergent. With the addition of 4 M urea, new resonances were observed (figure 3.20B). We hypothesize that the addition of urea is causing the transmembrane helices to begin to unfold and allow the residues in that domain to rotate more freely than in the absence of denaturant. More resonances in the other part of the spectrum were also observed. CD was performed for the samples at each interval to monitor the secondary structure of the full length protein with the addition of denaturant. From the CD data (Figure 3.20C and D) there appears to be a decrease in intensity of the alpha-helical content of the protein with 4 M urea when compared to the sample lacking urea. A

shift towards random coil, characteristic of unfolding, is observed for hPEN-2 in 4 M urea (Figure 3.20D). Our NMR and CD results complement our backbone assignment where most of the assigned residues lay in the loop regions and that the resonances coming from residues within the transmembrane domains are broadened beyond detection. The residues especially within the second transmembrane domain are more prone to pervasive line broadening mostly because their interaction with the micelle decreases their mobility to time-scale (in milliseconds) and is not suitable for NMR (Opella, Nevzorov et al. 2002).

Figure 3.20: Progressive unfolding of full length, his-tagged hPEN-2 as shown by NMR and CD. A) 2D ^{15}N - ^1H Tr-HSQC of his-tagged full length U- ^{15}N labelled hPEN-2 in 100 mM DPC and sodium phosphate buffer without urea at 40°C. B) 2D ^{15}N - ^1H Tr-HSQC of protein + 4M Urea at 40°C. The comparison of the spectra indicates the appearance of new resonances in the presence of the denaturing agent, specially highlighted in the glycine region (105 ppm ^{15}N). C) CD spectrum of protein in the absence of denaturant at 40°C and D) CD spectrum of the protein with 4 M urea. A comparison of the two CD spectra indicates the unfolding of hPEN-2 in the presence of urea.

3.3.5 Truncated hPEN-2 fragments for structural studies

To overcome the line broadening and poor spectral resolution associated with the large size of protein-detergent complex, we decided to design smaller fragments of his-tagged full length hPEN-2. We designed several fragments containing different domains of the full length protein based on the transmembrane predictions by TMHMM software (Krogh, Larsson et al. 2001). A library of these hPEN-2 fragments was submitted to the Protein Platform Production group (NTU, SBS) for initial small scale screening. Out of a total of 14 constructs (fragments) submitted, only four (as seen in figure 3.21) showed good expression in small scale.

For structural studies only two out of the four fragments were selected based on their thermal stability and significance of the domain represented. These fragments were R16-A61 and S51-S93, comprising the first and the second transmembrane domains, respectively. Figure 3.22A and Figure 3.22B show the final step in the purification of the U-¹⁵N labelled hPEN-2, S51-S93 and R16-A61 fragments, respectively. Fractions collected from the main peak (around 90 ml) were checked for purity by SDS-PAGE. Fractions showing pure bands around 7 kDa for S51-S93 and ~ 8 kDa for R16-A61 were pooled together and concentrated for NMR experiments. Both fragments had low expression yields and the final concentration of U-¹⁵N labelled samples was ~ 100 μ M for S51-S93 and less than 100 μ M for R16-A61 fragment.

Figure 3.21: Different fragments of full length hPEN-2 submitted to PPP. A) All the designed fragments of hPEN-2 submitted to PPP for small scale screening. B) The four fragments showing good expression were chosen for further studies. TMD1 and TMD2 are Trans-Membrane Domains 1 and 2 respectively of hPEN-2. M.W. includes N-terminus histidine tag with TEV cleavage site.

Figure 3.22: Purification profile of hPEN-2 fragments. A) SEC profile of S51-S93 fragment with SDS-PAGE analysis of the protein fractions from the main peak. B) SEC profile of R16-A61 fragment with SDS-PAGE analysis of the protein fractions from the main peak.

3.3.5.1 Structural Analysis of hPEN-2 fragments

For hPEN-2, S51-S93 fragment in 100 mM DPC, 2D ^{15}N - ^1H HSQC was recorded at 40°C. The spectrum displayed about 45% of the total expected resonances. As seen from the figure 3.23A, the spectrum is well resolved and does not suffer from signal overlap problem. However, complete absence of a number of resonances could not be overcome by different buffer conditions.

Figure 3.23: Second Transmembrane domain comprising fragment of hPEN-2 used for structural characterization. A) 2D ^{15}N - ^1H HSQC of his-tagged, U- ^{15}N labelled S51-S93 fragment of hPEN-2 in 100 mM DPC recorded at 40°C. The spectrum looks well resolved with sharp resonances observed for about 30 out of the expected 67 resonances. B) CD spectrum of the same sample at 40°C showing a well-defined alpha-helix.

The protein was stable at high temperatures of even 40°C for several hours. With the help of Circular Dichroism spectroscopy, we were able to determine the secondary structure of the S51-S98 fragment in the NMR conditions. Figure 3.23B, shows the characteristic curve of alpha-helical protein with two minima at $\Theta_{222\text{nm}}$ and $\Theta_{209\text{nm}}$. Our results indicate that the second predicted transmembrane domain is mostly stable but due to a high content of hydrophobic residues (as seen in table 3.1) they are possibly involved in hydrophobic interactions causing the fragments to self-associate

to form some type of supercoiled structure as suggested by the $\Theta_{222\text{nm}} / \Theta_{209\text{nm}}$ value which is greater than 1. These supercoiled structures having high alpha-helical content and higher molecular weight along with the protein's inherent dynamic nature could be responsible for the complete absence of most residues. We presume that the resonances visible in the 2D spectrum are mostly coming from the extra membrane parts.

The 2D ^{15}N - ^1H HSQC of his-tagged, U- ^{15}N labelled, hPEN-2 fragment R16-A61 was also acquired at 30°C in 100 mM DPC. However, the sample was unstable at this temperature and showed signs of precipitation within less than an hour. The spectrum (figure 3.24) showed poor dispersion and signal overlap regions along with complete absence of a majority of expected signals.

Figure 3.24: 2D ^{15}N - ^1H HSQC of his-tagged, U- ^{15}N labelled, hPEN-2 fragment R16-A61 in 20 mM sodium phosphate, 100 mM NaCl and 100 mM DPC at pH 5.5 recorded at 30°C.

Our NMR data suggests that the predicted first transmembrane domain, by itself is highly unstable and prone to aggregation. This instability could be caused by hydrophobic interactions between the residues along the protein backbone. Another unique feature of this fragment is the presence of two oppositely charged residues R39 and E40 (highlighted in red in table 3.1) within the predicted transmembrane domain. Usually charged residues reside on the surface of proteins or in extra membrane loops of membrane protein and are involved in salt bridge formation. These oppositely

charged residues lying next to each other in the sequence could be adding to the protein instability.

R16-A61	<u>M</u> <u>H</u> <u>H</u> <u>H</u> <u>H</u> <u>H</u> <u>G</u> <u>S</u> <u>T</u> <u>S</u> <u>L</u> <u>Y</u> <u>K</u> <u>K</u> <u>A</u> <u>G</u> <u>S</u> <u>E</u> <u>N</u> <u>L</u> <u>Y</u> <u>F</u> <u>Q</u> <u>G</u> <u>R</u> <u>K</u> <u>Y</u> <u>Y</u> <u>L</u> <u>G</u> <u>G</u> <u>F</u> <u>A</u> <u>F</u> <u>L</u> <u>P</u> <u>F</u> <u>L</u> <u>W</u> <u>L</u> <u>V</u> <u>N</u> <u>I</u> <u>F</u> <u>W</u> <u>F</u> <u>F</u> <u>R</u> <u>E</u> <u>A</u> <u>F</u> <u>L</u> <u>V</u> <u>P</u> <u>A</u> <u>Y</u> <u>T</u> <u>E</u> <u>Q</u> <u>S</u> <u>Q</u> <u>I</u> <u>K</u> <u>G</u> <u>Y</u> <u>V</u> <u>W</u> <u>R</u> <u>S</u> <u>A</u>
S51-S93	<u>M</u> <u>H</u> <u>H</u> <u>H</u> <u>H</u> <u>H</u> <u>G</u> <u>S</u> <u>T</u> <u>S</u> <u>L</u> <u>Y</u> <u>K</u> <u>K</u> <u>A</u> <u>G</u> <u>S</u> <u>E</u> <u>N</u> <u>L</u> <u>Y</u> <u>F</u> <u>Q</u> <u>S</u> <u>Q</u> <u>I</u> <u>K</u> <u>G</u> <u>Y</u> <u>V</u> <u>W</u> <u>R</u> <u>S</u> <u>A</u> <u>V</u> <u>G</u> <u>F</u> <u>L</u> <u>F</u> <u>W</u> <u>V</u> <u>I</u> <u>V</u> <u>L</u> <u>T</u> <u>S</u> <u>W</u> <u>I</u> <u>T</u> <u>I</u> <u>F</u> <u>Q</u> <u>I</u> <u>Y</u> <u>R</u> <u>P</u> <u>R</u> <u>W</u> <u>G</u> <u>A</u> <u>L</u> <u>G</u> <u>D</u> <u>Y</u> <u>L</u> <u>S</u>

Table 3.1: Amino acid sequences of the two hPEN-2 fragments R16-A61 (containing first predicted transmembrane domain) and S51-S93 (containing second predicted transmembrane domain). Residues highlighted in blue are part of the N-terminus 6X histidine with TEV cleavage site tag. Hydrophobic amino acids have been underlined. Charged residues R39 (positively charged) and E40 (negatively charged) highlighted in red and the predicted transmembrane domains by TMHMM have been highlighted in yellow (as shown previously on page).

3.4 Discussion

Human PEN-2 is an integral membrane protein comprising 101 amino-acids with two putative transmembrane domains and is an important subunit of gamma-secretase complex. Most of our knowledge of the structure and function of hPEN-2 - PS1 interactions comes from *in vivo* studies.

At the time of our study, no high-resolution structure of any of the sub-units of the γ -secretase complex was known. The first low resolution (12 Å) structure of the gamma secretase complex determined by Cryo-EM was published in 2009 (Osenkowski, Li et al. 2009). Only in 2014, the first atomic-resolution (1.9Å) crystal structure of eukaryotic Nicastrin, the presumed substrate engaging subunit of the complex, was published (Xie, Yan et al. 2014). Due to the absence of high resolution structure of the complex, the precise spatial placement of PEN-2 subunit within the γ -secretase complex was unknown.

Most of our knowledge of the human PEN-2 topology and function comes from *in vivo* studies involving mutations. One such study involved the use of mutant hPEN-2 engineered with artificial glycosylation sites to study the topology within the ER membrane (Crystal, Morais et al. 2003). The study supported a model of hPEN-2 consisting of two membrane spanning helices with the N and C termini protruding into the ER lumen and the loop (connecting the two transmembrane domains) in the cytosolic region.

Another group studied the role of hPEN-2 *in vivo* and *in vitro* confirming the vital role of the subunit in the catalytic activity of the γ -secretase complex (Bammens, Chavez-Gutierrez et al. 2011). They were also the first to suggest the location of the first predicted transmembrane domain and the connecting loop to be inside a water-cavity. This model was a clear deviation from the hair-pin model suggested previously by Crystal et al. in 2003.

The then accepted model of the γ -secretase complex suggested a 1:1:1:1 complex (PS1: PEN-2: APh-1: NCSTN) but due to the heterogeneous glycosylation state of Nicastrin, an accurate determination of the molecular weight of the complex was not possible. The best estimate of the MW of the complex is 230 +/- 33 kDa which was determined by scanning transmission electron microscopy (Osenkowski, Li et al. 2009).

3.4.1 NMR characterization of the hPEN-2 in solution

To better understand the structure and function of PEN-2 subunit of the γ -secretase complex, we decided to proceed with the structural characterization of the full length human PEN-2 using NMR spectroscopy as the key technique.

One of the most challenging aspects in NMR studies of integral membrane proteins (IMPs) is reconstitution of the protein's native topology yet maintaining physical properties such as short rotational correlation time which makes NMR studies feasible. The other important aspect of IMP sample preparation is finding specific conditions from which resonance assignments are possible. With both of these criteria in mind, the recombinant protein also needs to be expressed to levels which make incorporation of isotopic labels (^{15}N , ^{13}C , ^2H) cost effective. Our initial studies with full length hPEN-2 were directed at finding suitable membrane mimetic conditions which would satisfy the above criteria. Many of the solvent systems used for NMR studies of membrane proteins were tried including detergents such as SDS, DPC, DHPC and bicelles (DMPC:DHPC). Samples dissolved in each condition were assessed for solubility and chemical shift dispersion using 2D Tr-HSQC experiments. Solubility, stability and the NMR spectral resolution of hPEN-2 was a major determinant in the elimination of particular solvent systems.

In our study, reconstitution of hPEN-2 into SDS micelles produced the best resolved spectra. Although we judged it to be the best, the spectrum had a mixture of sharp and broad peaks suggesting that hPEN-2 is in conformational exchange to some degree. The sharp peaks likely arise from the disordered extra-membrane loops while the broader peaks arise from the transmembrane segments. Using this sample, a set of 3D Heteronuclear experiments typically used for backbone resonance assignments of proteins were recorded. Resonance assignment of the full length hPEN-2 was approximately ~ 50% (66 spin systems) complete with a total of 121 spin systems, and the majority of the assignments belonging to the extra-membrane loops and part of the two putative transmembrane domains. The two major bottlenecks preventing completion of the chemical shift assignment of the full length protein was the chemical shift degeneracy of particularly the ^{13}C spins and incomplete observation of all expected resonances.

The structural information encoded in the chemical shifts, particularly backbone resonances was analysed. The chemical shift values of $^{13}\text{C}_\alpha$ resonances of assigned residues were tabulated and compared with database averages for secondary structural elements. Based on the secondary structure prediction, assigned residues in the loop showed propensities towards random-coil or beta-strand while assigned residues from within the putative transmembrane regions showed high alpha-helicity. The homology model, having highest confidence score, generated from I-TASSER (Figure 3.3.15) was used to map the assigned residues.

Challenging proteins such as hPEN-2 require novel strategies to glean insights into the structure and dynamics of these systems. One of these strategies was using specifically methyl labelled IVL sample for full length hPEN-2 in TROSY type NMR experiments. Since even very large proteins (>60 kDa) with well-defined tertiary structure have shown well resolved ^{13}C - ^1H TROSY spectrum previously, we expected the full length hPEN-2 (~ 14 kDa and ~ 35 kDa in DPC micelles) to also give a well-resolved spectrum. However, the Methyl-TROSY spectrum suffered from signal overlap problem and absence of some peaks (figure 3.16 and 3.17). The extra peaks in the Ile region also support our hypothesis that the protein is highly disposed to conformational exchange in solution.

Deuterium exchange experiments were also performed to understand the structural integrity of hPEN-2 in solution and the results showed complete disappearance of all resonances in 100% D_2O , during the course of data acquisition (figure 3.18). This was unexpected since the residues in the transmembrane portions of hPEN-2 should have been more protected from exchange with deuterons than the solvent exposed loops. However our results confirm low protection of the amide protons of full length protein in solution. When taken together with the lack of resonances coming from the transmembrane regions, these data suggest that there is considerable conformational exchange in the transmembrane regions of full length hPEN-2 in solution, occurring on an intermediate time scale which is not amenable to NMR spectroscopy. The net effect of this is pervasive line broadening of a majority of resonances beyond the limits of detection. The significance and precise source of these motions is not known and we hypothesize that these could be related to inherent properties of hPEN-2.

3.4.2 Topology of hPEN-2 in solution

For complete tertiary structure determination of hPEN-2 in solution, we performed PRE measurements with full length samples containing endogenous cysteine (C40) as the reporter. The comparison of the spectrum before and after the addition of the probe helped identify some residues which were undergoing sharp drop in signal intensity. These residues must be closer to the paramagnetic probe to undergo faster relaxation due to the unpaired electron on the paramagnetic ion. An analysis of the affected residues confirmed the dynamic N and C-termini which allows the flanking regions to be in close proximity to the C40 residue. Some of the residues affected by the paramagnetic ions are in the cytoplasmic loop and the proximal end of TM2, an indication that the two transmembrane helices are not parallel to each other but rather perpendicular. These findings are in good agreement with the recent Cryo-EM structure of the whole Gamma-secretase complex (figure 3.14B) which also shows similar topology for the PEN-2 subunit. In the absence of the partners, we propose that hPEN-2 TM helices are inclined to conformational dynamics.

The titration of urea in the hPEN-2 sample followed by acquisition of 2D NMR experiments was carried out to examine the stability of the protein. We observed an increase in the number of peaks in the Tr-HSQC spectrum upon increase of the denaturant concentration. This is due to non-cooperative unfolding of hPEN-2. The sharp resonances originate from the unstructured, dynamic part of the protein which were earlier broadened due to conformational fluctuations.

3.4.3 NMR characterization of hPEN-2 fragments

In order to maximize protein stability and improve spectral quality, we designed and expressed truncations of full length hPEN-2 with the flexible regions removed. NMR and CD results obtained from the S53-S93 fragment indicate that the putative transmembrane domain 2(TM2) is thermally stable but prone to self-association in solution, causing absence of a majority of resonances in the spectrum thereby preventing NMR assignment.

Mutation studies on the C-terminus of the hPEN-2 have shown that the sequence DYLSF (90-94) is needed for the interaction of PEN-2 with other subunits of

γ -secretase complex (Hasegawa, Sanjo et al. 2004). The recent Cryo-EM structure also reveals binding of this transmembrane domain of PEN-2 to the extracellular domain of Nicastrin. These studies indicate that the TM2 of PEN-2 needs its interacting partners for its stability in solution and prevent its aggregation (Prokop, Haass et al. 2005).

The R16-A61 fragment comprising the putative first transmembrane domain (TM1) was also used for NMR studies. This fragment was unstable and prone to aggregation and gave a poor spectrum with majority of resonances either absent or overlapping.

The structure of γ -secretase complex reveals the presence of a small hairpin within the putative TM1 of PEN-2, with N-terminus and loop in the cytoplasm and the C terminus extending into the extra-cellular region (Sun, Zhao et al. 2015). They also showed that the TM4 of PS1 binds to the TM1 and TM2 of PEN-2 in the complex. These studies are complementary to our findings which suggest that the TM1 fragment of hPEN-2 in the absence of its interacting partner is unstable in solution, even in the presence of detergent and is in considerable conformational exchange.

Overall, we hypothesize that hPEN-2 by itself in solution, is partially unstructured and disposed to a high degree of conformational dynamics. This is caused by alteration between ensembles of structures that are interconverting on a millisecond timescale. The conformational dynamics could be an inherent property of hPEN-2 in the absence of its interaction partners, or a result of hydrophobic mismatch. Thus we can conclude that our results are complimented by the structure of hPEN-2 bound to gamma secretase complex revealed by Cryo-EM studies (figure 3.3).

Chapter 4

4 Solid-state NMR study of Aquaporin Z

Collaboration with Prof. Jaume Torres (NTU) and Dr. Toshio Yamazaki.

Manuscript in preparation

4.1 Introduction

4.1.1 Physiological significance of Aquaporins

Koefoed-Johnsen *et. al.* suggested the existence of a water channel that allows the osmotic flow of water molecules across membranes in order to explain the high osmotic permeability shown by some lipid cell membranes ((Koefoed-Johnsen and Ussing 1953). Later, Preston *et. al* identified a protein, CHIP28, existing in several mammalian tissues possessing high water permeability. The protein was later termed aquaporin-1 (AQP1) and is found widely distributed in nature and contains hundreds of homologs shown to have multiple significant physiological functions (Agre, Bonhivers et al. 1998).

The two major subfamilies of the transmembrane water conducting channel proteins consist of the aquaporins (AQP) and aquaglyceroporins (glycerol and water conducting - AQGP), together forming a group of “highly selective passive transporters” (Heller, Lin et al. 1980; Preston, Carroll et al. 1992; Park and Saier 1996).

The *E. coli* glycerol facilitator aquaporin channel (GlpF) conducts not only glycerol but also linear polyalcohols, known as “alditols” across the cell membrane (Heller, Lin et al. 1980) and is highly selective as seen by the absence of backflow of radioactive substance counter to concentration gradient (Maurel, Reizer et al. 1993). The GlpF has wider diameter at the selectivity filter (3.3 Å) as compared to AqpZ (2.4 Å) (Wang, Schulten et al. 2005). The wider pore of GlpF allows passage of glycerol molecules but causes steric hindrances and prevents glycerol passage through the much narrower AqpZ channel (Wang, Schulten et al. 2005).

The human proteome consists of at least ten aquaporins (AQP – AQP9), with AQPs 1, 2, 4, 5 and 8 working mainly as water transporters, while AQPs 3, 7 and 9 being permeable to glycerol, small polar solutes and sugars (Tsukaguchi, Shayakul et al. 1998; Wu and Beitz 2007). Aquaporins have been demonstrated to play an important role in osmoregulation and are found to be present in different tissues throughout the body. Mutations in these genes lead to several disorders ranging from diabetes insipidus (mutations in AQP2) (Deen, Verdijk et al. 1994) to development of lens cataract (Shiels and Bassnett 1996).

Recent studies have shown an increase in the levels of AQP4 in brain injuries (Kapoor, Kim et al. 2013). Transgenic mouse models containing mutations in AQPs have been useful in studying the role of aquaporins in renal functions. Many AQP proteins are expressed in different epithelial and endothelial tissues of the body. The importance of AQP's in water reabsorption in the kidney can be gathered from naturally occurring mutations in the Aqp genes. Mutations in AQPs 1-4 reduce the urine concentrating ability of the collecting ducts by reducing the transcellular fluid permeability of the ducts (Verkman 2008). In contrast, the absence of AQP4 in mice has been shown to have adverse and favourable effects on the water balance in the brain and spinal cord under different conditions. Aquaporins also play an important role in the epithelial fluid transport (Tradtrantip, Tajima et al. 2009). There is evidence that at least seven of the aquaporins are expressed and involved in the functioning of the gastrointestinal tract (Ma and Verkman 1999). AQP3 is expressed in the cell membrane of the basal layer of the epidermis and is known to transport both water and glycerol. AQP3 knockout transgenic mice show resistance to skin tumorigenesis due to reduced glycerol uptake and reduced ATP energy consumption (Hara-Chikuma and Verkman 2008). Apart from these well studied roles, AQP protein has also been implicated in other physiological processes like neuroexcitation, skin homeostasis and pathologies like obesity. In cancer, AQPs have been shown to have tumour-promoting role. This has been shown to be due to their vital roles in cell migration and cellular proliferation. For example AQP3 null mice have been shown to be completely resistant to the formation of skin tumours. Owing to the redundancy in the function between the different AQP proteins, there are few human disorders which are caused by mutations or polymorphisms in an individual homolog of AQP. Mutations in AQP2 have been shown to lead to non-X-linked Diabetes insipidus. Owing to its expression in the brain, mutations in AQP4 has been shown to cause a disorder called neuromyelitis optica (NMO) which is characterized by blindness and paralysis.

Due to their abundant nature and high physiological importance, aquaporins are seen as key drug targets. There is ongoing research for both aquaporin inhibitors and activators to overcome and cure several aquaporin related diseases and disorders. Most of the research has focussed on developing antibody or small molecule based inhibitors of AQP function. These inhibitory molecules are being studied for their efficacy in the treatment of NMO, however the usefulness of these drugs in other pathologies like diabetes, glaucoma and cancer etc. should also be evaluated.

A comparison among the structures of the various aquaporins shows that the three dimensional fold of the aquaporins is highly conserved. They have a characteristic Asparagine-proline-Alanine (NPA) motif which is highly conserved and a selectivity filter region which forms a narrow constriction in the channel and forms the basis for the difference within the subfamilies of the aquaporins. For water selective aquaporins, the SF is narrower (2.4Å in AqpZ) and lined with polar residues, while the glycerol channels the SF is wider (3.3Å in GlpF) and lined with more hydrophobic residues. The *E.coli* AqpZ shares only about 36% sequence identity with the human aquaporin AQP1, yet the three dimensional structure of the two channels shows high similarity especially at the channel core. Figure 4.1 discusses the sequence and structural similarity between the AqpZ and AQP1. Since the most crucial region in the channel, the selectivity filter is also highly conserved among the subfamilies, the channel activity for AqpZ and AQP1 should be similar. This supports the use of AqpZ as a ‘model system’ for studying the physiology of water transport across membranes. It also serves as an excellent tool for understanding the structure and properties of the different integral membrane proteins that function as channel proteins. Another advantage of using the AqpZ model system is the ease of expression and purification of the aquaporin from *E.coli* which allows for both structural and functional studies.

AqpZ	-----MFRKLA <u>A</u> ECFGTFWLVFGGCGSAVLAAGFP-----ELGIGFAGV <u>A</u> LAFGLT
AQP1	MASEFKKKLFWRAVVA <u>E</u> FLATTLFV <u>F</u> ISIGSALG-FKYPVGNNQTAVQDNVK <u>V</u> SLAFGLS
AqpZ	VLTMFAVGHISGGHFN <u>P</u> AVTIGLWAGGRFPAKEVVG <u>Y</u> VIAQVVGGIVAA <u>L</u> LYLIASGK
AQP1	IATLAQSVGHISGAHLN <u>P</u> AVTLGLLLSQCISIFRALMYIIAOCVGAIVAT <u>A</u> ILSGITSSL
AqpZ	TGFDAAASGFASNGYGEHSPGGYSMLSALVVELVLSAGFL <u>L</u> VI--HGATDKFAPAGF <u>A</u> PI
AQP1	TGNSLGRNDLAD---GVNSGQGL-----GIEIIGTLQLVL <u>C</u> VLATTDRLRRDLGG <u>S</u> APL
AqpZ	AIGLALT <u>L</u> IHLISIPVTNTSVNP <u>A</u> RS <u>T</u> AVAI <u>F</u> QGGWALEQLWFFWVVP <u>I</u> VGGIIGGL <u>I</u> YR
AQP1	AIGLSVALGHLAIDYTGCGINP <u>A</u> RSFGSAVITHN--FSNH <u>W</u> IFWVGPFIGGALAV <u>L</u> IYD
AqpZ	TLLEKRD-----
AQP1	FILAPRSSDLTDRVKVWTSQGVEEYDLDDADDINSRVEMKPK

Figure 4.1: Comparison of E.coli Aquaporin Z (AqpZ) with human Aquaporin 1 (AQP1). On top is the sequence alignment of AqpZ with AQP1. The two aquaporins show only about 36% identity in their sequence, highlighted by underline (BLAST align). The highly conserved NPA motif is highlighted in red. Arg 189 of the selectivity filter is also conserved. Below, the structures of AqpZ (orange) and AQP1 (green) superimposed with the conserved selectivity filter region and the NPA motif highlighted. Helices M1-M8 are also indicated. Modified from (Savage and Stroud 2007).

4.1.2 Structure of Aquaporins

To date about 48 aquaporin structures have been solved using X-ray diffraction, electron crystallography and electron microscopy techniques ([PDB](#)). A sub angstrom resolution structure of Aqp1 belonging to *Pichia pastoris* was recently submitted to PDB, enabling better understanding of the water channel activity (Kosinska Eriksson, Fischer et al. 2013). Around 26 structures of Aquaporin proteins, having resolution between 3 to 1.5 Å have been submitted to PDB over a period of 12 years. (Sui, Han et al. 2001) (Kosinska Eriksson, Fischer et al. 2013). Table 4.1 gives a comprehensive account of the different types of high resolution, aquaporin structures, submitted to the Protein Data Bank.

All the members of the aquaporin family consist of tetramers made up of four functionally active monomeric channels. Aquaporin 1 previously known as CHIP, is the first of the aquaporins to be discovered by Peter Agre, for which he was awarded the Noble Prize in Chemistry in 2003 (Agre, Preston et al. 1993; Jonsson 2003). Walz *et. al* in 1994, published the first structure (6 Å resolution 3D structure) of aquaporin CHIP, purifying it from the red blood cells and reconstituting it into lipid bilayers resulting in the formation of two dimensional crystals which also exhibit biological activity (Walz, Smith et al. 1994). The structure revealed a six transmembrane helix monomer containing extended loops with a highly conserved NPA domain (Walz, Smith et al. 1994; Jap and Li 1995; Li, Lee et al. 1997). The monomers of AQP1 have been known to assemble as tetramers, each monomer forming a water channel (Smith and Agre 1991).

The crystal structure of the glycerol facilitator of *E.coli* (GlpF) was recorded at 2.2 Å resolution previously (Fu, Libson et al. 2000). The structure reveals the presence of four monomeric units arranged symmetrically each containing one glycerol molecule. Each GlpF monomer contains six transmembrane domains and two half helices spanning the membrane forming a right handed helical bundle for each water channel. The N-terminus of the channel lies in the cytoplasmic region while the C-terminus ends in the periplasmic vestibule. The two half helices M3 and M7 begin with the highly conserved NPA motif which provides the h-bonds to successive hydroxyl groups of the glycerol molecules. The approximately 28 Å long channels have a width of ~ 15 Å at the periplasmic entrance and forms a constriction point (~

3.8 Å) at the start. The water conductivity of GlpF is much lower as compared to other water channels; one reason could be attributed to the hydrophobic nature of the channel itself and the constrain making the channel too narrow to allow passage of hydrated ions.

The most significant difference between the AQP and AQGP lies in the narrowest constriction point of the channel, which is termed as the Selectivity Filter (SF). Typically the SF in AQPs is roughly about 1.1 Å which is slightly less than the radius of a water molecule, while in AQGPs the SF is ~ 1.7 Å. The selectivity filter region contains four amino acids including a highly conserved Arginine residue, with the remaining three amino acids varying depending on the subfamily. In AQPs the SF is more hydrophilic with the presence of polar residues such as histidine allowing for water-selectivity while the AQGPs SF is more hydrophobic due to the presence of aromatic residues such as tyrosine and tryptophan. The side-chains of these residues provide favourable hydrophobic interaction with the carbon backbone of the alditols.

Organism	Type of Aquaporin	PDB ID	Experimental Method	X-ray resolution	Reference
<i>Escherichia coli</i>	AqpZ	1RC2, 2ABM	X-ray diffraction	2.5Å, 3.2Å	(Savage, Egea et al. 2003) (Jiang, Daniels et al. 2006)
	GlpF	1FX8	X-ray diffraction	2.2 Å	(Fu, Libson et al. 2000)
<i>Methanothermobacter marburgensis</i>	AqpM	2F2B	X-ray diffraction	1.68 Å	(Lee, Kozono et al. 2005)
<i>Pichia pastoris</i>	Aqy1	3ZOJ	X-ray diffraction	0.88 Å	(Kosinska Eriksson, Fischer et al. 2013)
<i>Plasmodium falciparum</i>	PfAQP	3CO2	X-ray diffraction	2.05 Å	(Newby, O'Connell et al. 2008)
<i>Spinacia oleracea</i>	SoPIP2	2B5F	X-ray diffraction	3.90 Å	(Tornroth-Horsefield, Wang et al. 2006)
<i>Rattus norvegicus</i>	AQP0	2B6O	Electron Crystallography	1.9 Å	(Gonen, Cheng et al. 2005)
<i>Bos taurus</i>	AQP1	1J4N	X-ray diffraction	2.2 Å	(Sui, Han et al. 2001)
<i>Homo sapiens</i>	AQP4	3GD8	X-ray diffraction	1.8 Å	(Ho, Yeh et al. 2009)
	AQP5	3D9S	X-ray diffraction	2.0 Å	(Horsefield, Norden et al. 2008)

Table 4.1: Aquaporin structures belonging to different class of organisms. They are categorised according to the type of aquaporin and the type and quality of the experimental method.

4.1.3 Structure of *E.coli* Aquaporin Z (AqpZ)

To better understand the highly selective function of the aquaporins and perform drug binding interaction studies we need to perform both structural and functional studies on the wild-type as well mutants. These studies usually require high yields of the proteins, which might prove difficult specially when trying to express large quantities in mammalian cells. For such system, an effective working model would be the *E.coli* AqpZ which can be expressed in high amounts and is more stable during isolation and purification procedures. Agre *et. al* was the first to publish the expression and purification of his-tagged AqpZ in *E.coli* (Borgnia, Kozono et al. 1999). The recombinant protein was purified in much higher yields than previously attempted and was also functionally active.

The monomer

The 2.5 Å crystal structure of the AqpZ shows the presence of six transmembrane alpha-helices termed M1, M2, M4, M5, M6 and M8 while two half helices (M3 and M7) embedded in the membrane (Savage, Egea et al. 2003). Each monomer itself consists of two domains, first domain having M1-M4 helices and the second domain with M5-M8 helices (Figure 4.2). The protomer itself, oligomerize to form a homogenous tetramer.

Figure 4.2: Schematic representation of Aquaporin Z in the membrane: Eight transmembrane alpha helices connected together by five inter-helical loops. [Modified from (Verkman, Anderson et al. 2014)].

The first domain of AqpZ begins with the N terminus in the cytoplasm with M1 traversing the membrane followed by a loop in the periplasmic vestibule and then by M2 traversing through the membrane. M2 is connected to the half helix M3 by a loop consisting of residues 54 to 62. The helix M3 itself begins with the distinctive “NPA” motif (Asparagine-Proline-Alanine) projecting its positive dipole towards water molecule in the central cavity and ends with M4 helix ending in the periplasmic cavity. The helix is followed by a long loop consisting of residues 103 – 131, which end in the second segment of AqpZ. The C-terminus segment of AqpZ resembles the topology of the first domain by beginning at the periplasmic side and ending into the cytoplasm. This “pseudo-2 fold symmetry” results in the formation of the channel containing carbonyls from the main chain for water binding.

Aquaporin Z tetramer

AqpZ is a homo-tetramer with four water channels that can transport water up to six times faster than GlpF (Calamita, Bishai et al. 1995; Borgnia, Kozono et al.

1999). A 2.5 Å resolution crystal structure of AqpZ was solved by Savage *et al.* in 2003. In their study, they described that the AqpZ crystals contain two tetramers per unit; each unit is made up of four copies each of protomers A and B (Savage, Egea *et al.* 2003). They proposed quasi-I4 symmetry of the unit cell with a slight rotation of the tetramer around the four-fold axis. Protomers A and B exhibit somewhat different conformations especially near the periplasmic vestibule. The formation of the tetramer takes place by interaction between the M1, M2 helices of one monomer with the M5, M6 helices of the adjacent monomer, with M2 and M6 forming the pore in the tetramer (Figure 4.3 inset). This interface is about 3340 Å² mostly due to the presence of around 11 aromatic residues. They also suggested that in all structures, the channel appears ‘open’ towards the periplasm but ‘closed’ towards the cytoplasm, mostly due to electrochemical gradient with a cavity size of about 7-10 Å. Recently Jiang *et al.* have also reported crystal structure of the AqpZ tetramer at a 3.2 Å resolution (Jiang, Daniels *et al.* 2006). Their findings confirm that AqpZ exists in a tetrameric form (Figure 4.3) and proposed a model containing ~ 270 water molecules, two detergent molecules and six lipid fragments, elucidating the electron density observed on the peripheral surface as well as the central cavity. The high electron density inside the four monomeric channels is due to the presence of water cavity within each monomer.

Figure 4.3: Periplasmic view of the Aquaporin Z tetramer reproduced from the Crystal structure [modified from with monomers marked A, B, C and D along with the water channel]. Crystals were formed in the presence of detergents and lipids which can also be seen in the structure. The inset shows “protomer - protomer interface” formed between the helices

M1, M2 of one monomer with the M5, M6 helices of the other [modified from (Jiang, Daniels et al. 2006)].

The channel

The AqpZ channel is about 28 Å in length and less than 4 Å in diameter (Savage, Egea et al. 2003). It is formed by the helices M1-M3 and M5-M7 such that the channel allows the conduction of water molecule one at a time. In the N- terminal domain, the carbonyl groups of the residues G59, G60, H61 and F62 form the hydrophilic region of the channel along with the carbonyls of the residues N182, T183, S184 and V185 from the C-terminus. Presence of large amounts of leucine, isoleucine, valines and phenylalanine imparts hydrophobic nature to the channel making it amphipathic. Each channel in the tetramer has two extremely conserved regions namely the Asparagine-Proline-Alanine (NPA) motif and the Selectivity Filter (SF).

The Selectivity Filter (SF)

The selectivity filter is formed by the sidechains of the residues F43, H174 and R189 with the carbonyls of T183. The three amino acids H174, T183 and R189 form a “hydrophilic triangle”. The SF is located about 7 Å from the periplasmic opening and forms the constriction point of the channel. The conserved NPA motif of each of the two domains within a monomer is oriented in a way such that the side chain of the Alanine residue aligns with the proline ring in a head to tail” manner. The side chains of N63 is involved in hydrogen bonds, such that the O(δ 1) bonds with the H^N of A65 and another bond between NH (δ 2) and CO of V185. A similar contact is observed between N186 and nearby residues. This allows the N δ 2 of the N to point into the channel cavity thus allowing hydrogen bonds to form with the water molecule in the cavity. The crystal structure published by Savage *et.al* show the presence of water placed next to COs of T183, S184, H61 and G60. The role of the residues at and near the SF has been discussed in the next section.

4.1.4 The “Open” and “Closed” state of AqpZ water channel

Aquaporin Z is a tetrameric, water channel membrane protein found in the cell membrane of wild type (WT) *Escherichia coli*. Each AqpZ tetramer is made up of four monomers and each monomer itself contains a lumen which allows the bi-directional passive diffusion of water across the membrane. Published crystal structure of AqpZ tetramer shows that there are two conformation states of Arg 189 side chain (one of the residues in the selectivity filter) that are associated with its water channel activity (Jiang, Daniels et al. 2006). More recently, Savage *et al* have also explained the role of R189 side chain in the opening and closing of the AqpZ channel (Savage, O'Connell et al. 2010).

They suggest that the WT AqpZ has R189 in a conformation known as “**down**” which positions the guanidinium group of R189 towards the pore (C_{β} – C_{γ} – C_{δ} –N ϵ dihedral angle = -101.0°). This down conformation of R189 side chain is stabilized by a number of hydrogen bonds. One H-bond is to the carbonyl of **Ala 117**, where A117 resides in the periplasmic loop (M4 and M5 linker) and another H-bond is to the carbonyl groups of the residues present in the M6-M7 linker region, especially T183

which also binds water molecule at the C-terminus. This downward orientation of R189 **closes** the pore.

Figure 4.4: Closed state of AqpZ water channel. The orientation of side chains of residues R189 (in the selectivity filter), A117 (in the periplasmic loop) and T183 (in M6-M7 linker) which form a network of hydrogen bonds that “close” the water channel. The dashed lines indicate hydrogen bond between side chain of R189 and carbonyls of A117 and T183. [Modified and reproduced from the closed state monomer in crystal structure of AqpZ published by (Jiang, Daniels et al. 2006)].

In the “**open**” state, the side chain of R189 is oriented in an upward position and allows free passage of water molecules in a single file. Figure 4.5 below, depicts the single file flow of water molecules across the AqpZ channel and the role of the R189 side chain conformation in the gating mechanism.

Figure 4.5: Side chain conformation of R189 showing opening and closing of AqpZ water channel. A) “Open” state allowing water molecules to enter in a single file. B) “Down” state which closes the channel and water continuity is broken. [Figure modified and reproduced from (Wang, Schulten et al. 2005)].

Water is passively diffused through the channel due to the hydrophilic nature of the SF moieties and formation of a network of hydrogen bonds with surrounding residues. Though SF is crucial to water transport across the AQPZ channel, the role of

residues in the M6-M7 linker and periplasmic loop (i.e loop C) also play an important role in assisting AqpZ selectivity.

4.1.5 Aim of the project

In view of the importance of aquaporins in clinical medicines the development of small molecules acting as aquaporin modulators is highly promising. Aquaporin Z from *E.coli* serves as a good working model for human aquaporin-based diagnostics and therapeutics. The expression and purification of mammalian aquaporins is not easy, thus working with AqpZ which can be purified and expressed in good quantities (Borgnia, Kozono et al. 1999) can be used as a model to study the function of aquaporins.

Our collaborators have been able to successfully express and purify uniformly, isotopically labelled AqpZ from *E.coli* and reconstitute them into the lipids. These samples were used for recording solid-state NMR experiments. Our first aim in the AqpZ study was to identify the state of the channel, i.e. if the channel is in an ‘open’ or ‘closed’ state. This was achieved by comparing the chemical shifts predicted from the previously published ‘open’ state crystal structure with the chemical shifts obtained through NMR assignments. The two key residues under observation were Arg 189 (known selectivity filter gate) and Ala 117. Our results show an ‘open’ state for the AqpZ channel in the solid-state sample.

Based on the results of previously conducted biophysical and biochemical assays by our collaborators, we decided to proceed with the lead compound, (1-(4-methylphenyl) 1H-pyrrole-2, 5-dione) hypothesized to have inhibitory effect on the channel activity (unpublished data). The next aim of our study was to use solid state NMR to characterize the drug interaction site as well as to understand the inhibitory effect of the drug. Our results indicate that the drug binds to AqpZ in an ‘open’ state and does not convert it into a ‘closed’ state. Mapping of the chemical shift perturbations to the previously published crystal structure helped understand that the bound drug plugs the channel entrance or otherwise modifies the channel conductance in this open state. Since medium affinity drug cannot be seen by X-ray studies, NMR spectroscopy is a unique method to help identify such interactions and at the same time provide atomic-level resolution of the interaction site.

4.2 Materials and Methods

4.2.1 Cloning, Expression and Purification of Aquaporin Z from *E.coli*

The Aquaporin Z was expressed, purified and samples for MAS NMR were prepared in Prof. Jaume Torres' laboratory in School of Biological Sciences, NTU, using a previously described protocol (Junge, Schneider et al. 2008).

Reconstitution of AqpZ into liposomes

The film rehydration method was used for liposome preparation (Olson, Hunt et al. 1979). Briefly, a thin lipid film was formed by first dissolving 25 mg of total *Escherichia coli* lipid extract powder (Avanti Polar Lipids Inc.) in chloroform followed by drying under a nitrogen stream. The film was kept in a vacuum desiccator for at least 2 hours before the addition of purified AqpZ protein in OG to obtain a protein-to-lipid molar ratio of 1:85. The reconstitution mixture was incubated for 1 hour at room temperature and extruded through a 400 nm pore-size polycarbonate membrane using an Avestin extruder to facilitate the formation of proteoliposome. The sample was dialyzed in the presence of 20 mM Tris-HCl pH 8.0 for 3 days at 4°C. The compound NSC13691, obtained from the National Cancer Institute (NCI) compound library, was pre-dissolved in 100% dimethyl sulfoxide (DMSO). It was mixed with the dialyzed sample in a protein-to-drug molar ratio of 1:25, and the sample was freeze-thawed thrice before centrifugation at 20,000 g for 1 hour (the pellet shipped to Japan). An unbound AqpZ proteoliposome control was prepared in the same way without addition of drug. The final DMSO concentration in both samples was kept constant at 1.5%.

4.2.2 Nuclear Magnetic Resonance Spectroscopy sample preparation

Purified ^{15}N and ^{13}C labelled AqpZ was used for the solid state NMR experiments. For each sample, about 10 mg of AqpZ was added to about 25 mg of Escherichia coli lipid.

For the drug interaction studies, we used the drug ([1-\(4-methylphenyl\)1H-pyrrole-2,5-dione](#)) (ZINC ID 94134) having a molecular weight of 187.19466 g/mol with an octanol-water partition coefficient (XLogP3) value of 1.5 making it hydrophobic. It has three hydrogen bond acceptor sites consisting of two double bonded oxygen atoms and one tertiary amine group.

For NMR studies of drug-protein interaction, two samples were prepared. The sample without the drug contained a total of about 1.5% of DMSO and used as a control. The sample with the drug added to a ratio of 25:1 also contained a total of about 1.5% DMSO in the final concentration.

Both samples (with drug and the control) were pelleted with some liquid left on top and packed on ice and sent to Riken, Japan for the solid-state NMR data acquisition. Before the MAS NMR acquisition, the sample was placed inside a 3.2 mm rotor (ZrO₂: 30 μL) by setting a pipette man tip on the entry of the rotor as a funnel and scooping the sample pellet and putting it inside through the tip. Then the rotor along with the tip is placed in a 15mL Falcon tube and centrifuged with swing rotor (starting from slower rotation for some minutes) to allow the sample to go inside the rotor.

4.2.3 Solid-state MAS NMR spectroscopy

All solid-state MAS NMR experiments were performed in Riken, Japan by Dr. Nagashima, under the Supervision of Dr. Toshio Yamazaki.

For both the AqpZ control sample and AqpZ with drug, all the data was obtained on a Bruker 700 MHz AVANCE spectrometer having a 3.2 mm HCN magic angle spinning (MAS) probe. All experiments were recorded at 7°C and the spectrometer frequency was set at $^{13}\text{C} = 176.077\text{ MHz}$, $^{15}\text{N} = 70.957\text{ MHz}$ and having a MAS frequency set to 15750 Hz. Except otherwise mentioned, the magnetization transfer from ^{15}N to $^{13}\text{CA}/^{13}\text{CO}$ was achieved using a 6 ms cross-polarization and a recycle delay of 1.1 seconds.

The 1D CPMAS for sample with and without drug was acquired using 1024 data points, a recycle delay of 1 s and 16 scans. The homonuclear 2D ^{13}C - ^{13}C DARR

experiment for both samples, was acquired using a 20 ms mixing time with 1024 (F2) X 768 (F1) data points, 64 number of scans and a recycle delay of 1.5 s.

We recorded the 2D Heteronuclear ^{15}N - ^{13}C NCA and NCO of AqpZ control and with drug, using 384 scans each and recycle delay of 2 s. The data points were acquired as 1024 in F2 dimension and 96 in the F1 dimension. The spectral width was adjusted to 273 ppm and 74 ppm in ^{13}C and ^{15}N dimensions respectively.

The backbone assignment was achieved using the 3D NCACX and NCOCX experiments comprising a 20 ms DARR step for magnetization transfer from $^{13}\text{C}_\alpha$ to $^{13}\text{C}_\chi$ (NCACX) or from $^{13}\text{C}_\alpha$ to ^{13}CO (NCOCX) respectively with data points 1024X64X32 in F3XF2XF1 dimensions respectively. 96 scans for NCACX and 64 scans for NCOCX were chosen for good signal to noise ratio. 3D NCACX was acquired for the AqpZ-drug sample. The 3D CANCO experiment was set up with 128 scans and 20 ms DARR mixing time. For $^{13}\text{C}_\beta$ assignments we used the 3D NCACB experiments with a 4 ms C_α to C_β DREAM transfer. 96 scans with 1024X64X32 data points in F3XF2XF1 dimensions respectively were used.

The sample stability was checked by recording 1D CPMAS experiments before and after 3D experiments and if the sample was stored for long intervals of time.

All spectra were processed using TopSpin 2.1 and analyzed using CARA software package.

4.3 Results

4.3.1 Good Spectral Quality of solid-state NMR experiments

We used 1D ssNMR experiments to determine the quality of the AqpZ samples. These experiments were recorded before and after the 2D and 3D ssNMR experiments. The 1D, ^1H - ^{13}C CPMAS (Cross-polarization magic angle spinning) experiment recorded for the uniformly ^{15}N , ^{13}C labelled AqpZ in *E.coli* lipid is shown in Figure 4.6. Overall, the spectrum looks well dispersed and the carbonyl region $\sim 175\text{ppm}$ also shows a reasonably good dispersion of about 10 ppm. The presence of well-structured domains within the protein manifests in narrow line widths.

Figure 4.6: 1D, ^{13}C cross-polarization spectrum of ^{15}N , ^{13}C labelled AqpZ (in *E.coli* lipid). The spectrum was recorded at 900 MHz with spinning frequency of 14 kHz at 7°C. The insert displays the carbonyl region.

4.3.2 Solid-state NMR characterization of AqpZ reconstituted in *E.coli* lipids

The sequence specific assignment of AqpZ was achieved by using a combination of several 2D and 3D homo and Heteronuclear experiments. As the first step in sequence assignment, we used the ^{13}C - ^{13}C correlation, DARR (Dipolar Assisted Rotational Resonance) experiment, where magnetization is transferred from $^1\text{H}_\alpha$ to $^{13}\text{C}_\alpha$ nuclei of residue and from there to other ^{13}C nuclei close by in space (Takegoshi, Nakamura et al. 2001). With 25 ms mixing time, the spectrum was fairly well resolved and most of the characteristic peaks such as the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ of the threonine, serine and isoleucine could be identified (as seen in Figure 4.7). Since glycine lack the side chains they are absent in the aliphatic region of the spectra and appear around the C_α - C' region.

In order to proceed with backbone assignment, we need to know the N, C_α and CO chemical shifts of each residue. For this purpose, we used the 2D NCA experiment, in which the magnetization is transferred from hydrogen to ^{15}N nuclei through cross-polarization and then to $^{13}\text{C}_\alpha$ of the same residue via specific cross-polarization. Here the magnetization is allowed to evolve on the N nuclei and detected on the ^{13}C nuclei (Baldus, Petkova et al. 1998). Figure 4.8 shows the 2D NCA spectrum recorded for the AqpZ. Due to their characteristic chemical shift values, certain residues can be easily identified in the NCA spectrum. Example, glycine, with upfield shifted $^{13}\text{C}_\alpha$ and ^{15}N resonances. Similarly the sidechains of threonine and proline can also be identified based on their chemical shift values. Overall we could pick more than 200 spin systems. Further assignment would require other experiments.

Figure 4.7: 2D ^{13}C - ^{13}C DARR spectrum of ^{15}N , ^{13}C labelled AqpZ (in *E.coli* lipid). A) Full spectrum showing ^{13}C - ^{13}C correlation, recorded at 700 MHz at 15 kHz spinning frequency at 7°C using 20 ms mixing time. B) Aliphatic region of the spectrum

B

Figure 4.8: 2D ^{15}N - ^{13}C NCA spectrum showing correlation between ^{15}N and $^{13}\text{C}\alpha$ within the residues. A) Full spectrum showing assigned peaks. B) Central part of the spectrum zoomed for clarity purposes. The spectrum was recorded at 700 MHz at spinning frequency of ~ 15 kHz and at 7°C . Cross-polarization between ^{15}N and ^{13}C was achieved by using 4 ms adiabatic CP.

In the NCO experiment (figure 4.9), the magnetization transfer path is similar to NCA, except the nuclei now under detection is ^{13}CO (Baldus, Petkova et al. 1998). The experiment proved very useful in the sequential connection of the $\text{CO}_{(i-1)}$ with the $\text{N}_{(i)}$ residues. However in large proteins such as AqpZ, even the NCO suffered from severe signal overlaps, poor resolution and was used along with the 3D NCOCX experiment to remove ambiguity.

Figure 4.9: 2D ^{15}N - ^{13}CO , NCO spectrum showing correlation between ^{15}N of one residue with the ^{13}CO of the previous residue. Assigned peaks have been highlighted. The spectrum was recorded at 700 MHz at spinning frequency of ~ 15 kHz and at 7°C . Cross-polarization between ^{15}N and ^{13}CO was achieved by using 4 ms adiabatic CP.

Once the $\text{C}\alpha$ and N resonances from a spin system were identified, our next task was to extend the assignment to all ^{13}C and ^{15}N resonances within the same residue. This was achieved by using the 3D NCACX experiment. For NCACX, the magnetization begins on hydrogen and is propagated similar to the NCA experiment with the exception of introduction of a DARR step in between to transfer magnetization to other ^{13}C nuclei. The chemical shift is evolved on ^{15}N and $^{13}\text{C}\alpha$ nuclei and then detected on ^{13}C , resulting in a 3D spectrum (Pauli, Baldus et al. 2001). The NCACX spectra helps identify the $\text{C}\alpha$, $\text{C}\beta$ and C' of a single residue and are “dressed” with the N chemical shifts already assigned from the NCA experiment.

In large proteins, due to signal overlap and high proton density, the assignment of spin systems might become difficult. To overcome this problem, we use experiments such as NCACB, which help “de-crowd” the NCACX and help identify

the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ peaks from within the residue. Since the experiment utilizes a DREAM (transfer from the $^{13}\text{C}_\alpha$ to $^{13}\text{C}_\beta$, which is a double quantum step, the $^{13}\text{C}_\beta$ peaks appear negative and can be easily and unambiguously identified (Pauli, Baldus et al. 2001) (Figure 4.10).

Figure 4.10: Strips from systems A152 and T153 used to demonstrate the efficacy of 3D NCACB. The experiment was used to identify the $^{13}\text{C}_\beta$ resonances, observed as negative peaks, compared to $^{13}\text{C}_\alpha$ resonances. Strip A is from 3D NCACX spectrum and strip B is from 3D NCACB spectrum.

To identify the $\text{C}_{\alpha-i}$ and the $\text{C}_{\beta-i}$ for identifying the (i-1) spin, the experiment NCOCX is used. These assignments will later help to link the predecessor fragment to the successor fragment and allow sequential walk. Like NCACX, in NCOCX, the

chemical shift is evolved on the $^{15}\text{N}_i$ and $^{13}\text{CO}_{(i-1)}$ nuclei and then detected on $^{13}\text{C}_{(i-1)}$, resulting in a 3D spectrum (Pauli, Baldus et al. 2001). The NCOCX usually gives a better dispersion than NCACX but in our hands, there was severe signal overlap, most probably due to large protein and presence of liposomes.

The stripscope in CARGA is then used to select the best candidates for the successor and predecessor of the reference strip. Once a good match is found the system is linked to the reference system and the walk continued. An example of the sequential walk for residues G105 to S114 is shown in figure 3.3.6. Both NCACX and NCOCX are used together to proceed with the backbone assignment.

Another useful addition to the NCACX and NCOCX is the 3D CANCO experiment. The spectrum provides sequential link between the $^{13}\text{C}_\alpha$ - ^{15}N of a residue to the CO of the previous residue (Li, Berthold et al. 2007). However, the spectrum suffers from poor signal to noise ratio due to the presence of three cross polarization transfers during the acquisition.

Figure 4.11: Sequential walk from residue G105 to F109 assigned with the help of 3D NCACX and 3D NCOCX. A) Strips from 3D NCACX highlighting the sequential walk. B) Strips from 3D NCOCX highlighting the sequential walk. Together, the walk establishes the $\text{C}^\alpha[i] - \text{C}^\alpha[i-1]$ connectivity.

With the help of both 2D and 3D solid-state MAS NMR experiments, we were able to complete about 30% of the backbone assignment (76/222 peaks expected). The

figure 4.12 below shows the assigned residues mapped onto the published crystal structure (Jiang *et. al.* 2006). Due to poor spectral resolution in NCO and NCOCX experiments and complete absence of signal from some regions, the assignment was not exhaustive. The assignments not only originate from the solvent accessible loops but also from residues located in helices M1, M2, M6, M7 and M8. The resonances coming from within the transmembrane domains M3 and M4 were not observable most probably due to interaction with the lipid molecules which causes line broadening. The assignment statistics are given in Appendix 5.

Figure 4.12: Assignment of AqpZ using MAS NMR experiments. Top view of the tetramer of AqpZ crystal structure (PDB code: 1RC2)(Jiang *et. al.* 2006) with the assigned residues highlighted in pink. Inset shows the assignment on a monomer of the AqpZ.

4.3.3 Secondary structure analysis of AqpZ

To verify the secondary structure of AqpZ in our study, we calculated the chemical Shift index for the C α resonances for each assigned residue based on the formula:

$$\Delta\delta \text{ of } C_{\alpha} = C_{\alpha} (\delta_{\text{observed}}) - C_{\alpha} (\delta_{\text{random coil}})$$

The upfield or downfield shifts of the $^{13}\text{C}_{\alpha}$ resonances helped us to identify the residues present in the alpha-helix or beta-strand of the protein (Wishart, Sykes et al. 1992). The “secondary chemical shift” of AqpZ of assigned residues is given below in figure 3.3.5.

As expected, AqpZ in *E.coli* lipid environment exhibits a high degree of alpha-helical content. Some residues in the loop region are prone to mobility and hence show little propensity towards alpha-helix or beta-strand. The secondary structure of almost all the assigned residues, predicted by the change in the $^{13}\text{C}_{\alpha}$ resonances agrees well with the previously published crystal structure of AqpZ (Savage, Egea et al. 2003; Jiang, Daniels et al. 2006).

Our results also show that solid-state NMR is more suitable for large membrane protein complexes since resonances from within the transmembrane domains can also be seen. These resonances are generally broadened beyond detection when acquiring solution-state NMR.

Figure 4.13: Chemical Shift Index analysis of $C\alpha$ resonances of AqpZ assigned residues. A) Residues 1-100 and B) Residues 101-231. All assigned residues show secondary structure similar to the published crystal structure (Savage, Egea et al. 2003).

4.3.4 Conformational Analysis of AqpZ by NMR data

When analysing the chemical shift values of assigned residues in AqpZ control sample, we observed large deviations in the backbone amide nitrogen (1H N) resonances of residues A117 and V193. When compared to their random coil values, these resonances were upfield shifted to > 10 ppm for A117 and > 15 ppm for V193.

From the crystal structure of AqpZ (PDB id: 1RC2) we know that an H-bond is formed between the guanidinium group of R189 and carbonyl of A117 (present in the M4-M5 loop) causing the channel to close (Savage, O'Connell et al. 2010). Previously

published literature suggests that the ^1H chemical shift of a residue in the protein backbone is strongly influenced by hydrogen bonding to the preceding carbonyl and can account for ~ 8 ppm shift (Williamson 2013). Thus we decided to compare the chemical shift values of the N resonances of A117 and V193 in the previously published “closed” state AqpZ crystal structure (PDB id: 1RC2) with the NMR observed chemical shifts for A117 and V193. We used the software ShiftX, to predict the ^1H , ^{13}C and ^{15}N chemical shifts of AqpZ by using its crystal structure pdb:1RC2 file as input (Neal, Nip et al. 2003). The difference in the observed and predicted chemical shift values ($\Delta\delta$) of the backbone amide N of A117 and V193 were then calculated (table 4.2). We observed a large upfield deviation of ^{15}N backbone amide resonance for both A117 and V193. To give an overview of the precision of the ShiftX software, a comparison of the predicted and calculated chemical shift values have been included in Appendix 6 for the fragment G105 – G120.

Residue	NMR δ (ppm)	Predicted δ (ppm)	$\Delta\delta$ (ppm)
A117 ^{15}N	108.4	114.86	-6.46
V193 ^{15}N	97.05	115.37	-18.32

Table 4.2: Comparison of the observed NMR, N chemical shift values of A117 and V193 with the “closed” state crystal structures of AqpZ. A large upfield shift of backbone amide N, for both residues was observed. The predicted chemical shift values of the crystal structure were obtained from ShiftX.

Based on our NMR data, we hypothesize that the hydrogen-bond between R189 and A117 is absent. This is shown by the large upfield deviation of the N chemical shift obtained from our NMR data from the closed state crystal structure of AqpZ. Hence we propose that the AqpZ sample in our study could possibly be in an “Open” state.

For V193, the much larger upfield shift (>15 ppm) could have resulted from a break in the alpha helix at the site of V193 residue, causing the half helix M7 to bend away from the selectivity filter region in the channel (R189). This conformational change could promote opening of the channel at the constriction point and allow the periplasmic loop (M4-M5 linker) to move away from the periplasmic channel entrance.

4.3.5 Characterization of the AqpZ - drug interaction site by solid-state NMR spectroscopy

In order to describe the binding site of the drug (shown in figure 4.14) on AqpZ, we performed the 2D ^{15}N - ^{13}C NCA and 3D ^{15}N - ^{13}C NCACX experiments on the labelled AqpZ sample, with and without drug to observe any changes in the protein backbone due to protein-drug interaction.

The spectra were analysed and the chemical shift perturbations calculated from the peak shifts of ^{15}N and $^{13}\text{C}_\alpha$ resonances. Upon drug addition, majority of the resonances suffered a slight drop in signal intensity. However there were some resonances which also experienced some chemical shift perturbations. These resonances belonged to the residues present in and around the loop region of AqpZ, earlier termed as loop C extending between M4 and M5 helices.

We compared the overlapped 2D NCA of AqpZ control and AqpZ in the presence of the drug (figure 4.15) to select resonances undergoing shifts in the $^{13}\text{C}_\alpha$ or ^{15}N planes. Since 3D experiments are more sensitive than 2D we used 3D NCACX spectrum of AqpZ with drug to calculate the shifts in the $^{13}\text{C}_\alpha$ and ^1H resonances of affected residues.

We found that several residues showed deviations of more than 0.2 ppm for $^{13}\text{C}_\alpha$ resonances (figure 4.16). $^{13}\text{C}_\beta$ chemical shift perturbations for residues T153, T183 and S184 were also observed. Some residues also showed > 0.4 ppm chemical shift deviation of ^1H resonances. As changes in ^1H resonances result from both hydrogen bonding as well as possible conformational changes, it is important to consider both $^{13}\text{C}_\alpha$ and ^1H resonances together when interpreting AqpZ-drug interaction data. To account for non-specific interactions, we have considered only chemical shift perturbation > 0.2 ppm for C_α resonances and > 0.4 ppm for backbone ^1H (figure 4.16 B and C respectively).

Certain residues such as L15, V16, F109 and V193 showed chemical shift perturbations of > 0.4 ppm for the $^{13}\text{C}_\alpha$ and ^1H resonances. The large perturbations caused by the addition of the drug could be an indication that these residues are possibly interacting with the drug molecule. As seen in figure 4.16D, these residues are localized near the periplasmic cavity, with V16 being near the selectivity filter region.

Figure 4.14: Drug molecule 1-(4-methylphenyl)1H-pyrrole-2,5-dione (ZINC id 94134) used in the AqpZ-drug interaction studies.

Figure 4.15: AqpZ-drug interaction studied by chemical shift perturbation of residues. Overlay of ^{15}N - ^{13}C , 2D spectra recorded of ^{13}C - ^{15}N labelled AqpZ control (blue) with drug (red) on 700 MHz at ~ 15 KHz spinning frequency. Assigned residues undergoing >0.5 ppm ^{15}N chemical shift perturbation have been highlighted.

Figure 4.16: Chemical shift perturbations after addition of the drug. A) Changes in the chemical shifts of $^{13}\text{C}\alpha$ (blue) and ^{15}N (red) resonances in the presence of the drug have been plotted for all assigned residues. The residues undergoing more than 0.4 ppm shift of both $^{13}\text{C}\alpha$ and ^{15}N resonances have been marked with an asterisk. These have been mapped onto the crystal structure (1RC2) of AqpZ monomer and highlighted in red in figure D. B represents $^{13}\text{C}\alpha$ deviations > 0.2 ppm highlighted in purple. C gives the ^{15}N perturbations > 0.4 ppm highlighted in blue. Due to absence of complete backbone assignment, the set of residues affected by the addition of drug is not exhaustive.

For these experiments, we hypothesize that the drug molecules could interact with the periplasmic entrance of the channel including the periplasmic loop. This interaction could lead to blocking of the channel thereby inhibiting the entry of the water molecules.

4.4 Discussion

In the present study we used solid-state MAS NMR to characterize Aquaporin Z protein, uniformly ^{15}N and ^{13}C labelled and reconstituted into *E.coli* liposomes. The backbone assignment was done using several 2D and 3D homonuclear and heteronuclear solid-state NMR experiments. We were able to unambiguously assign about 71 out of 231 (excluding 9 proline) residues in AqpZ monomer. Due to several, amino-acid repeats in the protein sequence and poor spectral resolution of crucial experiments like NCO and NCOCX, the complete backbone assignment was not possible. Certain residues such as Ala, Gly, Thr and Ser with their characteristic C_α and C_β chemical shifts were easy to assign and were helpful in linking together short fragments. Upon comparing the assigned residues to the previously published crystal structure of AqpZ, we observed that most of these residues mapped primarily to the M2, M4 helices, the periplasmic loop (M4-M5 linker) and the short M6-M7 linker. A tentative assignment of the SF residue R189 was also obtained, but due to poor fragment alignment, it was not included in the final set of assigned residues.

Despite the incomplete assignment of the protein we managed to identify some of the important residues involved in AqpZ selectivity and channel gating. With the help of these assigned residues we were able to monitor changes in the structure of AqpZ upon the addition of the drug and identify the possible AqpZ-drug interaction site.

4.4.1 Aquaporin Z reconstituted into liposomes exists in an “Open” state conformation

The “open” state of AqpZ water channel is characterized by an upward orientation of the guanidinium group of R189 while the “closed” state has the guanidinium group pointing towards the channel lumen. Hence in the closed state the passage of water molecules is blocked.

Our solid state NMR data suggests the presence of an “open” state conformation of the AqpZ protein when reconstituted into the *E.coli* liposomes.

Based on previous studies, we know that A117 and R189 are involved in the closing of the AqpZ channel (Savage, O'Connell et al. 2010) by the formation of a hydrogen bond between the two residues. Since hydrogen bond formation causes a downfield shift of NMR resonances of residues involved in the bonding, the “closed” state predictions should be similar to our NMR data. However our results show that there is an upfield shift (> 5 ppm) between the predicted and the observed backbone ^1H N amide resonance of A117. The large upfield shift of N atom could be an indication of the absence of H-bond between carbonyl of A117 and sidechain of R189. This could possibly result in a large conformational change in the loop (M6-M7 linker) and most likely causes it to “flip out”.

Another residue that shows an unusual chemical shift value of its ^1H N resonance is V193. We observed a shift of > 15 ppm of its backbone nitrogen. We know that in alpha-helices, hydrogen bonds are formed between the amide group of “i” residue and the carbonyl group of the “i+4th” residue to maintain the right handed helical structure. The large upfield shift of the ^{15}N resonance could be explained if there is a break in the H-bond between i (=CO of R189) and i+4th residue (-N_H of V193) which in turn would cause destabilization of the half helix M7.

Together the two residues A117 and V193 play a key role in describing the conformation of the AqpZ channel and based on their unique chemical shifts we postulate that the AqpZ sample in our study could possibly be in an open state. This hypothesis was further confirmed by the AqpZ-drug interaction studies where we found residues on the loop and near selectivity filter region perturbed in the presence of the drug.

4.4.2 The drug binds to AqpZ in an ‘open’ state and inhibits channel activity

The drug N-p-Tolylmaleimide (NSC 13691) was selected as the biochemical assays (unpublished data) have shown some inhibitory effect on AqpZ. In order to map the region where the drug binds to AqpZ and to understand the mechanism behind the inhibitory effect of the drug, we performed MAS NMR studies.

The NMR spectra analysis of AqpZ sample in the presence of the drug indicates that the drug possibly interacts with AqpZ near the periplasmic opening of

the channel. Widespread chemical shift perturbations in residues across the channel are indicators of probable allosteric changes within the protein in the presence of the drug. Large chemical shift deviations of residues residing in the M4-M5 linker, near the Selectivity Filter, and in the half helix M7 upon the addition of drug were observed and could possibly form the binding site of the drug to the protein. However due to excess of drug molecules added to the sample (ratio of 25:1, drug: protein), it is likely possible that there are more than one binding site for the drug. The high hydrophobicity of the drug itself could be the reason we observed perturbations of resonances coming from within the transmembrane domain.

The 2.5 Å crystal structure of AqpZ indicates that the channel width is ~ 2.4 Å around the selectivity filter and widens out into the periplasmic cavity (Savage, Egea et al. 2003). We hypothesize that the drug molecule, having dimensions of ~ 4 Å X 9 Å is not small enough to completely pass through the channel lumen and is very likely “trapped” near the SF which forms the constriction point (<4Å) in the channel cavity. Here, we speculate that there are ‘hydrophobic interactions’ occurring between the aromatic ring of the drug and the hydrophobic side chains of the four hydrophobic residues (L15, V16, F109 and V193) giving rise to the large chemical shift perturbations. Due to these interactions, the drug molecule ‘blocks’ the channel at the selectivity filter region (channel gate) and prevents the water molecules from passing through. The figure 4.17 below is a hypothetical model of the AqpZ-drug interaction site.

Figure 4.17: Hypothetic model of the AqpZ-drug interaction. We hypothesize that the drug interacts with the AqpZ near the periplasmic entrance of the channel. Residues L15, V16, F109 and V193 have been highlighted in red. The drug molecule is seen in blue. The selectivity filter residue involved in channel gating – R189 is also highlighted in purple. Left image is a view of the channel from the periplasmic side.

Even though the *E.coli* AqpZ shares only about 36% sequence similarity with human AQP1 (figure 4.1.1), the selectivity filter region (with R189) and the two NPA motif are highly conserved. This allows for high structural similarity of the channel core across the species within the same subfamily (figure 4.1.1). Since V16, which is highly conserved across the species, shows significant interactions with the inhibitor drug and is located near the selectivity filter region, we propose a similar ‘binding pocket’ for the human AQP1 where the drug interacts with V25 of human AQP1 with similar hydrophobic interactions. Our novel findings serve as a basis for further investigation of different inhibitory drugs for human aquaporins and will prove useful in aquaporin based therapeutics and diagnostics.

Chapter 5

Conclusion

Membrane proteins have long been recognized as important drug targets and a lot of effort has been directed towards solving the structure of vital membrane proteins in a native-membrane like environment. The basis of my study has been to explore the limits of solution state NMR and the use of solid-state NMR in characterizing dynamic membrane proteins. The three membrane proteins studied in this work are – human Stromal Interaction Molecule 1 (hSTIM1), human Presenilin Enhancer 2(hPEN-2) and the *E.coli* water channel Aquaporin Z (AqpZ).

In the hSTIM1 study, different fragments of the full length protein were designed and tested for expression. Fragments showing high stability and good yield were chosen for structural studies. The hSTIM1 fragments short-listed for NMR studies included the coiled-coil 1 domain (K246-K349 and K246-Q360), the calcium binding motif, the transmembrane domain (S58-Q360) and the C terminus (K246-S600). All four fragments showed thermal-stability and high alpha-helical content as determined from the circular dichroism data. The fragments containing only the first coiled-coil domain showed lack of tertiary structure in the absence of the other STIM1 domains. However the presence of an extension of the coiled coil 1 helps stabilize the tertiary fold of the fragment, as seen from the stability curve of the K246-S600 fragment. However, the fragment still suffered from aggregation and was disposed to degradation even at room temperatures. NMR data confirmed the presence of conformational exchange induced line-broadening in all fragments. The complete absence of signals and severe overlapping of most resonances confirmed the presence of intermediate exchange. Even in the absence of complete backbone assignment we observed structural changes in the hSTIM1 binding partner- ORAI1 by the addition of an ORAI1 C-terminus derived peptide, confirming the functional integrity of the protein in solution. Calcium binding was also monitored for the EF-hand motif comprising fragment, by using the translational diffusion experiment and a slight increase in the Diffusion Coefficient value (hence increase in the oligomeric state) was observed. We thus surmise that hSTIM1 is highly dynamic in nature, as is its inherent property and requires the presence of all its domains for stability and proper global fold. However, the functional integrity of the different domains is maintained even in the absence of other domains.

The second project involved working with human PEN-2, a transmembrane protein with two putative transmembrane helices. PEN-2 is a subunit of the gamma-secretase complex and is known to render the complex active upon binding. Our study

with PEN-2 involved structural characterization using novel methods in solution-state NMR. Secondary structure was confirmed using circular dichroism which revealed a high alpha-helical content and high thermal stability for PEN-2 in detergent micelles. With about 50% backbone assignment successful and coming from the extra-membrane loops, we were able to identify conformational dynamics within the PEN-2 protein in the micellar environment. The hydrogen-deuterium studies of PEN-2 in detergent micelles confirmed the ‘loose’ tertiary structure of the protein in solution displayed by the complete exchange of all resonances. The progressive unfolding of PEN-2 in micelles, in solution was monitored by adding increasing amounts of the denaturant and recording of 2D NMR experiments. The results from the denaturation study further prove that full length PEN-2 in detergent micelles is partially unfolded. NMR data obtained from selectively labelled samples support our hypothesis that the protein is disposed to conformational exchange. Judging from the poor NMR spectral quality, absence of signal from transmembrane domains and low protection of residues in the transmembrane domain we surmise that PEN-2 in detergent micelles is prone to high degree of conformational exchange at an intermediate timescale and in the absence of its binding partner Presenilin 1, lacks a rigid tertiary structure. Truncated fragments of PEN-2 were designed to overcome possible hydrophobic interactions between the two putative transmembrane domains. The fragment containing the first putative transmembrane domain (R16-A61) was less stable and gave a poorly dispersed NMR spectrum. The fragment S51-S93 comprising the second putative transmembrane domain was more thermally stable and gave a better resolved NMR spectrum but due to very low yields of the isotopically labelled S51-S93, triple resonance experiments for backbone assignments were not acquired. Our prediction of the PEN-2 topology verified by PRE data is in agreement with the recently published Cryo-EM structure of PEN-2 when bound to the subunits of active gamma-secretase complex.

My final project included solid-state NMR studies of the *E.coli* water channel, Aquaporin Z. AqpZ has a well-established tetrameric structure with each protomer having eight transmembrane domains forming the water channel. However, so far there has been no reported NMR structure and dynamics characterization of AqpZ. For NMR studies, the AqpZ was purified and reconstituted in *E.coli* liposomes and was used to identify the “open” or “closed” state of the channel in that state. Our NMR results suggested the existence of an “open” state of AqpZ channel in liposomes in

solid-state. These results were further confirmed by the AqpZ-drug interaction studies also monitored by solid-state NMR. The drug- '(1-(4-methylphenyl) 1H-pyrrole-2, 5-dione)' was proposed to have inhibitory effect on the channel activity. When excess of drug was added to the AqpZ sample, we observed significant chemical shift perturbation of certain residues located around the selectivity filter region of AqpZ channel. We propose that these residues stabilise the inhibitor inside the channel, by hydrophobic interactions between the aromatic side chain of the drug and the hydrophobic side-chains of the L15, V16, F109 and V193. Overall, our solid-state NMR findings suggest that AqpZ is in an 'open' conformation and this configuration allows for the drug binding. The drug possibly exerts an 'inhibitory' effect by blocking the channel in the open state and preventing the passage of water molecules. Medium affinity drug binding is not visible in X-ray and hence NMR serves as an excellent tool for drug-interaction studies in lipid environment.

One of the major problems when assigning the resonances to the membrane protein backbone was the conformational-induced line broadening. The exchange was significant in both full length proteins (PEN-2) and shorter fragments (of STIM1) suggesting that the individual subunits of a larger complex lack stability of tertiary structure in the absence of their natural binding partners. To circumvent the problem of conformational exchange we attempted different methods. These included acquiring almost all NMR spectra at high temperatures to induce faster exchange between the conformations resulting in a single, sharper peak. The samples were deuterated to overcome signal overlap problem and selectively labelled to make the less conspicuous signals more prominent. Even with better resolution, the lack of resonances from most transmembrane residues could not be overcome. To identify the global fold of the protein in such conditions, use of PRE measurements was very helpful. Another alternative to reduce conformational exchange was to design and use truncated fragments of full length proteins. The removal of certain dynamic regions from the amino-acid sequence allows for the production of much stable proteins in solution. These truncated regions may sometimes be more unstable than the full length protein, indicating the necessity of the other domains in maintaining the structural and functional integrity of the protein. The use of solid-state NMR allowed the study of large membrane proteins in its native-like environment. In the case of AqpZ, the good quality of NMR spectrum allowed for successful assignment of a majority of

resonances even in the tetrameric state of the membrane protein. This suggests that solid-state NMR is currently the best tool to study large membrane proteins structure and dynamics within the protein. The interaction studies of the drug with AqpZ followed by solid-state NMR serve as an excellent example of the uniqueness of NMR in identifying medium affinity drug binding.

Bibliography

- Abraham, S. J., R. C. Cheng, et al. (2015). "¹³C NMR detects conformational change in the 100-kD membrane transporter CIC-ec1." Journal of biomolecular NMR.
- Agre, P., M. Bonhivers, et al. (1998). "The aquaporins, blueprints for cellular plumbing systems." J Biol Chem **273**(24): 14659-14662.
- Agre, P., G. M. Preston, et al. (1993). "Aquaporin CHIP: the archetypal molecular water channel." The American journal of physiology **265**(4 Pt 2): F463-476.
- Ahmed, M., J. Davis, et al. (2010). "Structural conversion of neurotoxic amyloid-beta(1-42) oligomers to fibrils." Nat Struct Mol Biol **17**(5): 561-567.
- Almen, M. S., K. J. V. Nordstrom, et al. (2009). "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin." Bmc Biology **7**.
- Anderson, W. A. and R. Freeman (1962). "Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra." Journal of Chemical Physics **37**(1): 85-&.
- Angelucci, F., G. Spalletta, et al. (2010). "Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) patients are characterized by increased BDNF serum levels." Current Alzheimer research **7**(1): 15-20.
- Arora, A., F. Abildgaard, et al. (2001). "Structure of outer membrane protein A transmembrane domain by NMR spectroscopy." Nat Struct Biol **8**(4): 334-338.
- Baldus, M., A. T. Petkova, et al. (1998). "Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems." Molecular Physics **95**(6): 1197-1207.
- Bammens, L., L. Chavez-Gutierrez, et al. (2011). "Functional and topological analysis of Pen-2, the fourth subunit of the gamma-secretase complex." The Journal of biological chemistry **286**(14): 12271-12282.
- Battiste, J. L. and G. Wagner (2000). "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data." Biochemistry **39**(18): 5355-5365.
- Battiste, J. L. and G. Wagner (2000). "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data." Biochemistry **39**(18): 5355-5365.
- Baum, J., C. M. Dobson, et al. (1989). "Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin." Biochemistry **28**(1): 7-13.
- Bax, A. and N. Tjandra (1997). "High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium." Journal of Biomolecular Nmr **10**(3): 289-292.

- Bayburt, T. H. and S. G. Sligar (2010). "Membrane protein assembly into Nanodiscs." FEBS Lett **584**(9): 1721-1727.
- Bayrhuber, M., T. Meins, et al. (2008). "Structure of the human voltage-dependent anion channel." Proc Natl Acad Sci U S A **105**(40): 15370-15375.
- Berardi, M. J., W. M. Shih, et al. (2011). "Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching." Nature **476**(7358): 109-113.
- Berjanskii, M. V., S. Neal, et al. (2006). "PREDITOR: a web server for predicting protein torsion angle restraints." Nucleic acids research **34**(Web Server issue): W63-69.
- Berrier, C., K. H. Park, et al. (2004). "Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent." Biochemistry **43**(39): 12585-12591.
- Betz, H. and B. Laube (2006). "Glycine receptors: recent insights into their structural organization and functional diversity." J Neurochem **97**(6): 1600-1610.
- Borgnia, M. J., D. Kozono, et al. (1999). "Functional reconstitution and characterization of AqpZ, the E. coli water channel protein." Journal of molecular biology **291**(5): 1169-1179.
- Bowie, J. U. (2004). "Membrane proteins: a new method enters the fold." Proceedings of the National Academy of Sciences of the United States of America **101**(12): 3995-3996.
- Brustovetsky, N., J. M. Dubinsky, et al. (2003). "Two pathways for tBID-induced cytochrome c release from rat brain mitochondria: BAK- versus BAX-dependence." J Neurochem **84**(1): 196-207.
- Cady, S. D., K. Schmidt-Rohr, et al. (2010). "Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers." Nature **463**(7281): 689-692.
- Cai, S., C. Seu, et al. (2006). "Sensitivity enhancement of multidimensional NMR experiments by paramagnetic relaxation effects." Journal of the American Chemical Society **128**(41): 13474-13478.
- Calamita, G., W. R. Bishai, et al. (1995). "Molecular cloning and characterization of AqpZ, a water channel from Escherichia coli." The Journal of biological chemistry **270**(49): 29063-29066.
- Call, M. E., J. R. Schnell, et al. (2006). "The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor." Cell **127**(2): 355-368.
- Calloway, N., D. Holowka, et al. (2010). "A basic sequence in STIM1 promotes Ca²⁺ influx by interacting with the C-terminal acidic coiled coil of Orai1." Biochemistry **49**(6): 1067-1071.
- Carroll, H. R., B. B. A. McNaul, et al. (2006). "Immunodeficiency is a tough nut to CRAC: The importance of calcium flux in T cell activation." Molecular Interventions **6**(5): 253-+.
- Cavanagh, J., W. J. Fairbrother, et al. (1995). Protein NMR spectroscopy: principles and practice, Academic Press.
- Chill, J. H., J. M. Louis, et al. (2006). "NMR study of the tetrameric KcsA potassium channel in detergent micelles." Protein Sci **15**(4): 684-698.
- Cierpicki, T. and J. H. Bushweller (2004). "Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral

- membrane proteins." Journal of the American Chemical Society **126**(49): 16259-16266.
- Cierpicki, T., B. Y. Liang, et al. (2006). "Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-resolution structure of outer membrane protein A." Journal of the American Chemical Society **128**(21): 6947-6951.
- Crystal, A. S., V. A. Morais, et al. (2003). "Membrane topology of gamma-secretase component PEN-2." Journal of Biological Chemistry **278**(22): 20117-20123.
- Cui, B., X. Yang, et al. (2013). "The inhibitory helix controls the intramolecular conformational switching of the C-terminus of STIM1." PLoS One **8**(9): e74735.
- Dahmane, T., M. Damian, et al. (2009). "Amphipol-assisted in vitro folding of G protein-coupled receptors." Biochemistry **48**(27): 6516-6521.
- Davey, J. (2004). "G-protein-coupled receptors: new approaches to maximise the impact of GPCRS in drug discovery." Expert Opin Ther Targets **8**(2): 165-170.
- Deen, P. M. T., M. A. J. Verdijk, et al. (1994). "Requirement of Human Renal Water Channel Aquaporin-2 for Vasopressin-Dependent Concentration of Urine." Science **264**(5155): 92-95.
- Deisenhofer, J., O. Epp, et al. (1984). "X-ray structure analysis of a membrane protein complex. Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodospseudomonas viridis*." J Mol Biol **180**(2): 385-398.
- Di Capite, J. L., G. J. Bates, et al. (2011). "Mast cell CRAC channel as a novel therapeutic target in allergy." Curr Opin Allergy Clin Immunol **11**(1): 33-38.
- Doerr, A. (2006). "Together at last: crystallography and NMR." Nature Methods **3**(1): 6-6.
- Dowhan, W. (1997). "Molecular basis for membrane phospholipid diversity: why are there so many lipids?" Annu Rev Biochem **66**: 199-232.
- Dries, D. R. and G. Yu (2008). "Assembly, maturation, and trafficking of the gamma-secretase complex in Alzheimer's disease." Current Alzheimer Research **5**(2): 132-146.
- Duquesne, K. and J. N. Sturgis (2010). "Membrane protein solubilization." Methods in molecular biology **601**: 205-217.
- Dyson, H. J. and P. E. Wright (2005). "Intrinsically unstructured proteins and their functions." Nat Rev Mol Cell Biol **6**(3): 197-208.
- El Tahir, Y. and M. Skurnik (2001). "YadA, the multifaceted *Yersinia* adhesin." Int J Med Microbiol **291**(3): 209-218.
- Ellena, J. F., B. Liang, et al. (2009). "Dynamic structure of lipid-bound synaptobrevin suggests a nucleation-propagation mechanism for trans-SNARE complex formation." Proc Natl Acad Sci U S A **106**(48): 20306-20311.
- Fahrner, M., M. Muik, et al. (2009). "Mechanistic view on domains mediating STIM1-Orai coupling." Immunol Rev **231**(1): 99-112.
- Feske, S. (2007). "Calcium signalling in lymphocyte activation and disease." Nat Rev Immunol **7**(9): 690-702.

- Feske, S., Y. Gwack, et al. (2006). "A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function." Nature **441**(7090): 179-185.
- Fox, D. A., P. Larsson, et al. (2014). "Structure of the Neisserial outer membrane protein Opa(6)(0): loop flexibility essential to receptor recognition and bacterial engulfment." J Am Chem Soc **136**(28): 9938-9946.
- Franzin, C. M., X. M. Gong, et al. (2007). "NMR of membrane proteins in micelles and bilayers: the FXYD family proteins." Methods **41**(4): 398-408.
- Fu, D., A. Libson, et al. (2000). "Structure of a glycerol-conducting channel and the basis for its selectivity." Science **290**(5491): 481-486.
- Gardner, K. H. and L. E. Kay (1998). "The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins." Annual review of biophysics and biomolecular structure **27**: 357-406.
- Gardner, K. H., M. K. Rosen, et al. (1997). "Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR." Biochemistry **36**(6): 1389-1401.
- Gautier, A., H. R. Mott, et al. (2010). "Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy." Nat Struct Mol Biol **17**(6): 768-774.
- Gillespie, J. R. and D. Shortle (1997). "Characterization of long-range structure in the denatured state of staphylococcal nuclease .2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures." Journal of Molecular Biology **268**(1): 170-184.
- Girvin, M. E. and R. H. Fillingame (1995). "Determination of local protein structure by spin label difference 2D NMR: the region neighboring Asp61 of subunit c of the F1F0 ATP synthase." Biochemistry **34**(5): 1635-1645.
- Goedert, M. (2009). "Oskar Fischer and the study of dementia." Brain **132**: 1102-1111.
- Gonen, T., Y. Cheng, et al. (2005). "Lipid-protein interactions in double-layered two-dimensional AQP0 crystals." Nature **438**(7068): 633-638.
- Goto, N. K. and L. E. Kay (2000). "New developments in isotope labeling strategies for protein solution NMR spectroscopy." Curr Opin Struct Biol **10**(5): 585-592.
- Greenblatt, H. M., G. Kryger, et al. (1999). "Structure of acetylcholinesterase complexed with (-)-galanthamine at 2.3 angstrom resolution." Febs Letters **463**(3): 321-326.
- Haltia, T. and E. Freire (1995). "Forces and factors that contribute to the structural stability of membrane proteins." Biochimica et biophysica acta **1228**(1): 1-27.
- Hansen, M. R., P. Hanson, et al. (2000). "Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions." Methods in enzymology **317**: 220-240.
- Hara-Chikuma, M. and A. S. Verkman (2008). "Roles of aquaporin-3 in the epidermis." The Journal of investigative dermatology **128**(9): 2145-2151.

- Hardy, J. and D. J. Selkoe (2002). "Medicine - The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics." Science **297**(5580): 353-356.
- Hasegawa, H., N. Sanjo, et al. (2004). "Both the sequence and length of the C terminus of PEN-2 are critical for intermolecular interactions and function of presenilin complexes." Journal of Biological Chemistry **279**(45): 46455-46463.
- Heller, K. B., E. C. Lin, et al. (1980). "Substrate specificity and transport properties of the glycerol facilitator of Escherichia coli." J Bacteriol **144**(1): 274-278.
- Heller, K. B., E. C. Lin, et al. (1980). "Substrate specificity and transport properties of the glycerol facilitator of Escherichia coli." Journal of bacteriology **144**(1): 274-278.
- Henry, G. D. and B. D. Sykes (1994). "Methods to study membrane protein structure in solution." Methods in enzymology **239**: 515-535.
- Hiller, S., R. G. Garces, et al. (2008). "Solution structure of the integral human membrane protein VDAC-1 in detergent micelles." Science **321**(5893): 1206-1210.
- Hilty, C., C. Fernandez, et al. (2002). "Side chain NMR assignments in the membrane protein OmpX reconstituted in DHPC micelles." J Biomol NMR **23**(4): 289-301.
- Hilty, C., G. Wider, et al. (2004). "Membrane protein-lipid interactions in mixed micelles studied by NMR spectroscopy with the use of paramagnetic reagents." Chembiochem **5**(4): 467-473.
- Ho, J. D., R. Yeh, et al. (2009). "Crystal structure of human aquaporin 4 at 1.8 angstrom and its mechanism of conductance." Proceedings of the National Academy of Sciences of the United States of America **106**(18): 7437-7442.
- Hodge, T. and M. Colombini (1997). "Regulation of metabolite flux through voltage-gating of VDAC channels." J Membr Biol **157**(3): 271-279.
- Hong, M. (1999). "Resonance assignment of ¹³C/¹⁵N labeled solid proteins by two- and three-dimensional magic-angle-spinning NMR." J Biomol NMR **15**(1): 1-14.
- Horsefield, R., K. Norden, et al. (2008). "High-resolution x-ray structure of human aquaporin 5." Proceedings of the National Academy of Sciences of the United States of America **105**(36): 13327-13332.
- Hoth, M. and R. Penner (1992). "Depletion of intracellular calcium stores activates a calcium current in mast cells." Nature **355**(6358): 353-356.
- How, J., A. Zhang, et al. (2013). "Comprehensive Analysis and Identification of the Human STIM1 Domains for Structural and Functional Studies." PloS one **8**(1): e53979.
- Iwahara, J., C. D. Schwieters, et al. (2004). "Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule." Journal of the American Chemical Society **126**(18): 5879-5896.
- Jansen, M., M. Bali, et al. (2008). "Modular design of Cys-loop ligand-gated ion channels: functional 5-HT₃ and GABA_A rho1 receptors lacking the large cytoplasmic M3M4 loop." J Gen Physiol **131**(2): 137-146.

- Jap, B. K. and H. Li (1995). "Structure of the osmo-regulated H₂O-channel, AQP-CHIP, in projection at 3.5 Å resolution." Journal of molecular biology **251**(3): 413-420.
- Jehle, S., P. Rajagopal, et al. (2010). "Solid-state NMR and SAXS studies provide a structural basis for the activation of alphaB-crystallin oligomers." Nature structural & molecular biology **17**(9): 1037-1042.
- Jiang, J., B. V. Daniels, et al. (2006). "Crystal structure of AqpZ tetramer reveals two distinct Arg-189 conformations associated with water permeation through the narrowest constriction of the water-conducting channel." The Journal of biological chemistry **281**(1): 454-460.
- Jones, D. H. and S. J. Opella (2004). "Weak alignment of membrane proteins in stressed polyacrylamide gels." Journal of magnetic resonance **171**(2): 258-269.
- Jonic, S. and C. Venien-Bryan (2009). "Protein structure determination by electron cryo-microscopy." Current Opinion in Pharmacology **9**(5): 636-642.
- Jonsson, R. (2003). "The Nobel Prize in Chemistry for 2003." Scandinavian journal of immunology **58**(6): 587.
- Junge, F., B. Schneider, et al. (2008). "Large-scale production of functional membrane proteins." Cell Mol Life Sci **65**(11): 1729-1755.
- Kaiser, R. (1963). "Use of Nuclear Overhauser Effect in Analysis of High-Resolution Nuclear Magnetic Resonance Spectra." Journal of Chemical Physics **39**(10): 2435-&.
- Kapoor, S., S. M. Kim, et al. (2013). "Foxo3a transcriptionally upregulates AQP4 and induces cerebral edema following traumatic brain injury." The Journal of neuroscience : the official journal of the Society for Neuroscience **33**(44): 17398-17403.
- Kayed, R., E. Head, et al. (2003). "Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis." Science **300**(5618): 486-489.
- Keeler, J. (2005). Understanding NMR spectroscopy. Chichester, England ; Hoboken, NJ, Wiley.
- Kendrew, J. C., R. E. Dickerson, et al. (1960). "Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å. resolution." Nature **185**(4711): 422-427.
- Kerfah, R., M. J. Plevin, et al. (2015). "Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins." Curr Opin Struct Biol **32**: 113-122.
- Kidd, M. (1963). "Paired helical filaments in electron microscopy of Alzheimer's disease." Nature **197**: 192-193.
- Kielec, J. M., K. G. Valentine, et al. (2009). "Reverse micelles in integral membrane protein structural biology by solution NMR spectroscopy." Structure **17**(3): 345-351.
- Kim, K. K., R. Kim, et al. (1998). "Crystal structure of a small heat-shock protein." Nature **394**(6693): 595-599.
- Kim, S. H. and S. S. Sisodia (2005). "Evidence that the "NF" motif in transmembrane domain 4 of presenilin 1 is critical for binding with PEN-2." Journal of Biological Chemistry **280**(51): 41953-41966.

- Klammt, C., I. Maslennikov, et al. (2012). "Facile backbone structure determination of human membrane proteins by NMR spectroscopy." Nature methods **9**(8): 834-839.
- Knapp, K. G. and J. R. Swartz (2004). "Cell-free production of active E. coli thioredoxin reductase and glutathione reductase." FEBS Lett **559**(1-3): 66-70.
- Koefoed-Johnsen, V. and H. H. Ussing (1953). "The contributions of diffusion and flow to the passage of D₂O through living membranes; effect of neurohypophyseal hormone on isolated anuran skin." Acta Physiol Scand **28**(1): 60-76.
- Korukottu, J., R. Schneider, et al. (2008). "High-resolution 3D structure determination of kalitoxin by solid-state NMR spectroscopy." PLoS One **3**(6): e2359.
- Korzeniowski, M. K., I. M. Manjarres, et al. (2011). "Activation of STIM1-Orai1 involves an intramolecular switching mechanism." Faseb Journal **25**.
- Kosen, P. A. (1989). "Spin labeling of proteins." Methods in enzymology **177**: 86-121.
- Kosinska Eriksson, U., G. Fischer, et al. (2013). "Subangstrom resolution X-ray structure details aquaporin-water interactions." Science **340**(6138): 1346-1349.
- Kreft, A. F., R. Martone, et al. (2009). "Recent Advances in the Identification of gamma-Secretase Inhibitors To Clinically Test the A beta Oligomer Hypothesis of Alzheimer's Disease." Journal of Medicinal Chemistry **52**(20): 6169-6188.
- Krogh, A., B. Larsson, et al. (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." J Mol Biol **305**(3): 567-580.
- Lange, A., K. Giller, et al. (2006). "Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR." Nature **440**(7086): 959-962.
- Lange, A., S. Luca, et al. (2002). "Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids." J Am Chem Soc **124**(33): 9704-9705.
- Lau, S. Y., A. K. Taneja, et al. (1984). "Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils." J Biol Chem **259**(21): 13253-13261.
- Lau, T. L., A. W. Partridge, et al. (2008). "Structure of the integrin beta3 transmembrane segment in phospholipid bicelles and detergent micelles." Biochemistry **47**(13): 4008-4016.
- Lee, J. K., D. Kozono, et al. (2005). "Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 angstrom." Proceedings of the National Academy of Sciences of the United States of America **102**(52): 18932-18937.
- Lewis, R. S. (2007). "The molecular choreography of a store-operated calcium channel." Nature **446**(7133): 284-287.
- Li, H. L., S. Lee, et al. (1997). "Molecular design of aquaporin-1 water channel as revealed by electron crystallography." Nature Structural Biology **4**(4): 263-265.

- Li, Y., D. A. Berthold, et al. (2007). "Partial (13)C and (15)N chemical-shift assignments of the disulfide-bond-forming enzyme DsbB by 3D magic-angle spinning NMR spectroscopy." Chembiochem : a European journal of chemical biology **8**(4): 434-442.
- Liang, B. and L. K. Tamm (2007). "Structure of outer membrane protein G by solution NMR spectroscopy." Proc Natl Acad Sci U S A **104**(41): 16140-16145.
- Linke, D., T. Riess, et al. (2006). "Trimeric autotransporter adhesins: variable structure, common function." Trends Microbiol **14**(6): 264-270.
- Liou, J., M. L. Kim, et al. (2005). "STIM is a Ca²⁺ sensor essential for Ca²⁺-store-depletion-triggered Ca²⁺ influx." Current Biology **15**(13): 1235-1241.
- Lipari, G., A. Szabo, et al. (1982). "Protein dynamics and NMR relaxation: comparison of simulations with experiment." Nature **300**(5888): 197-198.
- Luchette, P. A., T. N. Vetman, et al. (2001). "Morphology of fast-tumbling bicelles: a small angle neutron scattering and NMR study." Biochim Biophys Acta **1513**(2): 83-94.
- M, G., P. A., et al. (1974). Proton-enhanced nuclear induction spectroscopy, Google Patents.
- Ma, D. J., T. S. Tillman, et al. (2008). "NMR studies of a channel protein without membranes: Structure and dynamics of water-solubilized KcsA." Proceedings of the National Academy of Sciences of the United States of America **105**(43): 16537-16542.
- Ma, T. and A. S. Verkman (1999). "Aquaporin water channels in gastrointestinal physiology." The Journal of physiology **517** (Pt 2): 317-326.
- Mancia, F. and J. Love (2010). "High-throughput expression and purification of membrane proteins." Journal of structural biology **172**(1): 85-93.
- Mao, G. Z., M. Z. Cui, et al. (2012). "Pen-2 is dispensable for endoproteolysis of presenilin 1, and nicastrin-Aph subcomplex is important for both gamma-secretase assembly and substrate recruitment." Journal of Neurochemistry **123**(5): 837-844.
- Marsh, R. E., R. B. Corey, et al. (1955). "An investigation of the structure of silk fibroin." Biochim Biophys Acta **16**(1): 1-34.
- Maurel, C., J. Reizer, et al. (1993). "The Vacuolar Membrane-Protein Gamma-Tip Creates Water Specific Channels in Xenopus-Oocytes." Embo Journal **12**(6): 2241-2247.
- McDermott, A. E. (2004). "Structural and dynamic studies of proteins by solid-state NMR spectroscopy: rapid movement forward." Curr Opin Struct Biol **14**(5): 554-561.
- McGeer, P. L. and E. G. McGeer (1995). "The inflammatory response system of brain: Implications for therapy of Alzheimer and other neurodegenerative diseases." Brain Research Reviews **21**(2): 195-218.
- Mehring, M. (1983). Principles of high-resolution NMR in solids. Berlin ; New York, Springer-Verlag.
- Meier, S., D. Haussinger, et al. (2002). "Charged acrylamide copolymer gels as media for weak alignment." Journal of Biomolecular Nmr **24**(4): 351-356.

- Miroshnikov, K. A., E. I. Marusich, et al. (1998). "Engineering trimeric fibrous proteins based on bacteriophage T4 adhesins." Protein engineering **11**(4): 329-332.
- Miroux, B. and J. E. Walker (1996). "Over-production of proteins in Escherichia coli: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels." Journal of Molecular Biology **260**(3): 289-298.
- Mowrey, D. D., T. Cui, et al. (2013). "Open-channel structures of the human glycine receptor alpha1 full-length transmembrane domain." Structure **21**(10): 1897-1904.
- Murray, C. W. and T. L. Blundell (2010). "Structural biology in fragment-based drug design." Curr Opin Struct Biol **20**(4): 497-507.
- Nam, H. J., J. Jeon, et al. (2009). "Bioinformatic approaches for the structure and function of membrane proteins." BMB Rep **42**(11): 697-704.
- Neal, S., A. M. Nip, et al. (2003). "Rapid and accurate calculation of protein ¹H, ¹³C and ¹⁵N chemical shifts." J Biomol NMR **26**(3): 215-240.
- Newby, Z. E., J. O'Connell, 3rd, et al. (2008). "Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite Plasmodium falciparum." Nature structural & molecular biology **15**(6): 619-625.
- Ollerenshaw, J. E., V. Tugarinov, et al. (2003). "Methyl TROSY: explanation and experimental verification." Magnetic Resonance in Chemistry **41**(10): 843-852.
- Olson, F., C. A. Hunt, et al. (1979). "Preparation of Liposomes of Defined Size Distribution by Extrusion through Polycarbonate Membranes." Biochimica Et Biophysica Acta **557**(1): 9-23.
- Opella, S. J., A. Nevzorov, et al. (2002). "Structure determination of membrane proteins by NMR spectroscopy." Biochem Cell Biol **80**(5): 597-604.
- Osenkowski, P., H. Li, et al. (2009). "Cryoelectron microscopy structure of purified gamma-secretase at 12 Å resolution." Journal of molecular biology **385**(2): 642-652.
- Otzen, D. E. (2002). "Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature." Biophys J **83**(4): 2219-2230.
- Page, R. C., J. D. Moore, et al. (2006). "Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins." Journal of structural and functional genomics **7**(1): 51-64.
- Park, C. Y., P. J. Hoover, et al. (2009). "STIM1 clusters and activates CRAC channels via direct binding of a cytosolic domain to Orai1." Cell **136**(5): 876-890.
- Park, J. H. and M. H. Saier, Jr. (1996). "Phylogenetic characterization of the MIP family of transmembrane channel proteins." J Membr Biol **153**(3): 171-180.
- Parker, M. J., M. Aulton-Jones, et al. (2004). "A combinatorial selective labeling method for the assignment of backbone amide NMR resonances." Journal of the American Chemical Society **126**(16): 5020-5021.
- Pauli, J., M. Baldus, et al. (2001). "Backbone and side-chain ¹³C and ¹⁵N signal assignments of the alpha-spectrin SH3 domain by magic angle

- spinning solid-state NMR at 17.6 Tesla." Chembiochem : a European journal of chemical biology **2**(4): 272-281.
- Pervushin, K., R. Riek, et al. (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." Proceedings of the National Academy of Sciences of the United States of America **94**(23): 12366-12371.
- Phillips, K. and A. H. de la Pena (2011). "The combined use of the ThermoFluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization." Curr Protoc Mol Biol **Chapter 10**: Unit10 28.
- Pines, A., M. Gibby, et al. (1973). "Proton-enhanced NMR of dilute spins in solids." The Journal of Chemical Physics **59**(2): 569-590.
- Poget, S. F., S. M. Cahill, et al. (2007). "Isotropic bicelles stabilize the functional form of a small multidrug-resistance pump for NMR structural studies." J Am Chem Soc **129**(9): 2432-2433.
- Poget, S. F. and M. E. Girvin (2007). "Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better." Biochim Biophys Acta **1768**(12): 3098-3106.
- Popot, J. L., E. A. Berry, et al. (2003). "Amphipols: polymeric surfactants for membrane biology research." Cell Mol Life Sci **60**(8): 1559-1574.
- Prakriya, M., S. Feske, et al. (2006). "Orai1 is an essential pore subunit of the CRAC channel." Nature **443**(7108): 230-233.
- Prestegard, J. H., C. M. Bougault, et al. (2004). "Residual dipolar couplings in structure determination of biomolecules." Chemical reviews **104**(8): 3519-3540.
- Preston, G. M., T. P. Carroll, et al. (1992). "Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein." Science **256**(5055): 385-387.
- Prokop, S., C. Haass, et al. (2005). "Length and overall sequence of the PEN-2 C-terminal domain determines its function in the stabilization of presenilin fragments." Journal of Neurochemistry **94**(1): 57-62.
- Putney, J. W. (1986). "A Model for Receptor-Regulated Calcium Entry." Cell Calcium **7**(1): 1-12.
- Qureshi, T. and N. K. Goto (2012). "Contemporary methods in structure determination of membrane proteins by solution NMR." Topics in current chemistry **326**: 123-185.
- Renault, M., O. Saurel, et al. (2009). "Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications." J Mol Biol **385**(1): 117-130.
- Rienstra, C. M., M. Hohwy, et al. (2000). "2D and 3D ¹⁵N-¹³C-¹³C NMR chemical shift correlation spectroscopy of solids: assignment of MAS spectra of peptides." Journal of the American Chemical Society **122**(44): 10979-10990.
- Roos, J., P. J. DiGregorio, et al. (2005). "STIM1, an essential and conserved component of store-operated Ca²⁺ channel function." Journal of Cell Biology **169**(3): 435-445.

- Ruschak, A. M. and L. E. Kay (2010). "Methyl groups as probes of supra-molecular structure, dynamics and function." Journal of biomolecular NMR **46**(1): 75-87.
- Saier, M. H., Jr., V. S. Reddy, et al. (2014). "The transporter classification database." Nucleic Acids Res **42**(Database issue): D251-258.
- Salzmann, M., G. Wider, et al. (1999). "TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins." Journal of the American Chemical Society **121**(4): 844-848.
- Samudio, I., M. Fiegl, et al. (2009). "Mitochondrial uncoupling and the Warburg effect: molecular basis for the reprogramming of cancer cell metabolism." Cancer Res **69**(6): 2163-2166.
- Savage, D. F., P. F. Egea, et al. (2003). "Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z." PLoS biology **1**(3): E72.
- Savage, D. F., J. D. O'Connell, 3rd, et al. (2010). "Structural context shapes the aquaporin selectivity filter." Proceedings of the National Academy of Sciences of the United States of America **107**(40): 17164-17169.
- Savage, D. F. and R. M. Stroud (2007). "Structural basis of aquaporin inhibition by mercury." J Mol Biol **368**(3): 607-617.
- Scharenberg, A. M., L. A. Humphries, et al. (2007). "Calcium signalling and cell-fate choice in B cells." Nat Rev Immunol **7**(10): 778-789.
- Schnell, J. R. and J. J. Chou (2008). "Structure and mechanism of the M2 proton channel of influenza A virus." Nature **451**(7178): 591-595.
- Schultz, J., F. Milpetz, et al. (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proceedings of the National Academy of Sciences of the United States of America **95**(11): 5857-5864.
- Schumann, F. H., H. Riepl, et al. (2007). "Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions." J Biomol NMR **39**(4): 275-289.
- Shahid, S. A., B. Bardiaux, et al. (2012). "Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals." Nat Methods **9**(12): 1212-1217.
- Shen, Y., F. Delaglio, et al. (2009). "TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." Journal of Biomolecular Nmr **44**(4): 213-223.
- Shen, Y., R. Vernon, et al. (2009). "De novo protein structure generation from incomplete chemical shift assignments." Journal of biomolecular NMR **43**(2): 63-78.
- Shiels, A. and S. Bassnett (1996). "Mutations in the founder of the MIP gene family underlie cataract development in the mouse." Nat Genet **12**(2): 212-215.
- Smith, B. L. and P. Agre (1991). "Erythrocyte Mr-28,000 Transmembrane Protein Exists as a Multisubunit Oligomer Similar to Channel Proteins." Journal of Biological Chemistry **266**(10): 6407-6415.
- Smith, E. E. and S. M. Greenberg (2009). "Beta-amyloid, blood vessels, and brain function." Stroke; a journal of cerebral circulation **40**(7): 2601-2606.

- Stathopoulos, P. B., L. Zheng, et al. (2008). "Structural and mechanistic insights into STIM1-mediated initiation of store-operated calcium entry." Cell **135**(1): 110-122.
- Stouffer, A. L., R. Acharya, et al. (2008). "Structural basis for the function and inhibition of an influenza virus proton channel." Nature **451**(7178): 596-599.
- Struts, A. V., G. F. Salgado, et al. (2011). "Retinal dynamics underlie its switch from inverse agonist to agonist during rhodopsin activation." Nat Struct Mol Biol **18**(3): 392-394.
- Sui, H. X., B. G. Han, et al. (2001). "Structural basis of water-specific transport through the AQP1 water channel." Nature **414**(6866): 872-878.
- Sun, L., L. Zhao, et al. (2015). "Structural basis of human gamma-secretase assembly." Proc Natl Acad Sci U S A **112**(19): 6003-6008.
- Takegoshi, K., S. Nakamura, et al. (2001). "C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR." Chemical Physics Letters **344**(5-6): 631-637.
- Tate, C. G. (2001). "Overexpression of mammalian integral membrane proteins for structural studies." FEBS letters **504**(3): 94-98.
- Terstappen, G. C. and A. Reggiani (2001). "In silico research in drug discovery." Trends Pharmacol Sci **22**(1): 23-26.
- Tian, C., M. D. Karra, et al. (2005). "Membrane protein preparation for TROSY NMR screening." Methods in enzymology **394**: 321-334.
- Tjandra, N. and A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." Science **278**(5340): 1111-1114.
- Torizawa, T., M. Shimizu, et al. (2004). "Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol." J Biomol NMR **30**(3): 311-325.
- Tornroth-Horsefield, S., Y. Wang, et al. (2006). "Structural mechanism of plant aquaporin gating." Nature **439**(7077): 688-694.
- Tradtrantip, L., M. Tajima, et al. (2009). "Aquaporin water channels in transepithelial fluid transport." The journal of medical investigation : JMI **56 Suppl**: 179-184.
- Tsukaguchi, H., C. Shayakul, et al. (1998). "Molecular characterization of a broad selectivity neutral solute channel." The Journal of biological chemistry **273**(38): 24737-24743.
- Tugarinov, V., V. Kanelis, et al. (2006). "Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy." Nature Protocols **1**(2): 749-754.
- Tugarinov, V. and L. E. Kay (2003). "Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods." Journal of the American Chemical Society **125**(45): 13868-13878.
- Tugarinov, V. and L. E. Kay (2004). "An isotope labeling strategy for methyl TROSY spectroscopy." Journal of Biomolecular Nmr **28**(2): 165-172.
- Tycko, R., F. J. Blanco, et al. (2000). "Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR." Journal of the American Chemical Society **122**(38): 9340-9341.

- Van Horn, W. D., H. J. Kim, et al. (2009). "Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase." Science **324**(5935): 1726-1729.
- van Montfort, R. L., E. Basha, et al. (2001). "Crystal structure and assembly of a eukaryotic small heat shock protein." Nat Struct Biol **8**(12): 1025-1030.
- Verkman, A. S. (2008). "Dissecting the roles of aquaporins in renal pathophysiology using transgenic mice." Seminars in Nephrology **28**(3): 217-226.
- Verkman, A. S., M. O. Anderson, et al. (2014). "Aquaporins: important but elusive drug targets." Nature Reviews Drug Discovery **13**(4): 259-277.
- Vinarov, D. A., B. L. Lytle, et al. (2004). "Cell-free protein production and labeling protocol for NMR-based structural proteomics." Nat Methods **1**(2): 149-153.
- Vold, R. R., R. S. Prosser, et al. (1997). "Isotropic solutions of phospholipid bicelles: a new membrane mimetic for high-resolution NMR studies of polypeptides." J Biomol NMR **9**(3): 329-335.
- Wagner, G. (1993). "Prospects for Nmr of Large Proteins." Journal of Biomolecular Nmr **3**(4): 375-385.
- Wagner, S., M. L. Bader, et al. (2006). "Rationalizing membrane protein overexpression." Trends in biotechnology **24**(8): 364-371.
- Wallin, E. and G. von Heijne (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms." Protein science : a publication of the Protein Society **7**(4): 1029-1038.
- Walz, T., B. L. Smith, et al. (1994). "Biologically active two-dimensional crystals of aquaporin CHIP." The Journal of biological chemistry **269**(3): 1583-1586.
- Wand, A. J., M. R. Ehrhardt, et al. (1998). "High-resolution NMR of encapsulated proteins dissolved in low-viscosity fluids." Proc Natl Acad Sci U S A **95**(26): 15299-15302.
- Wang, Y., K. Schulten, et al. (2005). "What makes an aquaporin a glycerol channel? A comparative study of AqpZ and GlpF." Structure **13**(8): 1107-1118.
- Williamson, M. P. (2013). "Using chemical shift perturbation to characterise ligand binding." Prog Nucl Magn Reson Spectrosc **73**: 1-16.
- Wishart, D. S. and B. D. Sykes (1994). "The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data." Journal of Biomolecular Nmr **4**(2): 171-180.
- Wishart, D. S., B. D. Sykes, et al. (1992). "The Chemical-Shift Index - a Fast and Simple Method for the Assignment of Protein Secondary Structure through Nmr-Spectroscopy." Biochemistry **31**(6): 1647-1651.
- Wollmann, P., K. Zeth, et al. (2006). "Purification of the YadA membrane anchor for secondary structure analysis and crystallization." International journal of biological macromolecules **39**(1-3): 3-9.
- Wu, B. and E. Beitz (2007). "Aquaporins with selectivity for unconventional permeants." Cellular and molecular life sciences : CMLS **64**(18): 2413-2421.

- Wu, M. M., J. Buchanan, et al. (2006). "Ca²⁺ store depletion causes STIM1 to accumulate in ER regions closely associated with the plasma membrane." Journal of Cell Biology **174**(6): 803-813.
- Wuthrich, K. (2003). "NMR studies of structure and function of biological macromolecules (Nobel Lecture)." Journal of Biomolecular Nmr **27**(1): 13-39.
- Xie, T., C. Yan, et al. (2014). "Crystal structure of the gamma-secretase component nicastrin." Proceedings of the National Academy of Sciences of the United States of America **111**(37): 13349-13354.
- Yang, D. and L. E. Kay (1999). "Improved 1HN-detected triple resonance TROSY-based experiments." Journal of biomolecular NMR **13**(1): 3-10.
- Yang, J., R. Yan, et al. (2015). "The I-TASSER Suite: protein structure and function prediction." Nat Methods **12**(1): 7-8.
- Yang, X., H. Jin, et al. (2012). "Structural and mechanistic insights into the activation of Stromal interaction molecule 1 (STIM1)." Proceedings of the National Academy of Sciences of the United States of America **109**(15): 5657-5662.
- Yee, A. A., A. Savchenko, et al. (2005). "NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins." Journal of the American Chemical Society **127**(47): 16512-16517.
- Yoshiura, C., Y. Kofuku, et al. (2010). "NMR analyses of the interaction between CCR5 and its ligand using functional reconstitution of CCR5 in lipid bilayers." J Am Chem Soc **132**(19): 6768-6777.
- Yu, L., C. Sun, et al. (2005). "Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex." Biochemistry **44**(48): 15834-15841.
- Zhou, Y., T. Cierpicki, et al. (2008). "NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation." Molecular cell **31**(6): 896-908.
- Zoonens, M., L. J. Catoire, et al. (2005). "NMR study of a membrane protein in detergent-free aqueous solution." Proceedings of the National Academy of Sciences of the United States of America **102**(25): 8893-8898.
- Zweifach, A. and R. S. Lewis (1993). "Mitogen-Regulated Ca²⁺ Current of T-Lymphocytes Is Activated by Depletion of Intracellular Ca²⁺ Stores." Proceedings of the National Academy of Sciences of the United States of America **90**(13): 6295-6299.

Appendix

Appendix 1: SEC profile of hSTIM1 (S58-Q360) in different buffers at pH 7.0 and 300 mM sodium chloride and 0.1% FC12. The buffers used were 20 mM Hepes, 50 mM Arginine-50 mM Glutamic acid, 20 mM Sodium phosphate and 20 mM Tris buffer.

Appendix 2: SEC profile of STIM1 (S58-Q360) in different sodium chloride concentrations at pH 7.0 and in 20 mM Tris and 0.1% FC12. The different salt concentrations tested were 250 mM, 200 mM, 150 mM and 100 mM NaCl.

Appendix 3: Size Exclusion Chromatogram profile of hSTIM1 (S58-Q360) at different pH values in 20 mM Tris, 150 mM NaCl and 0.1% FC12. The different pHs values tested were 7.0, 7.5, and 8.0.

Appendix 4: Size Exclusion Chromatogram profile of hSTIM1 S58-Q360 (^{15}N and ^2H labelled) using HiLoad 16/600 Superdex 200. The buffer contains 0.1% Fos-Choline 12 detergent. The elution volume of > 60 mL is an indication of MW ~ 66 kDa hence suggesting the presence of higher molecular weight oligomers.

Appendix 5: Size Exclusion Chromatogram profile of hSTIM1 S58-Q360 (^{13}C , ^{15}N and ^2H labelled) using HiLoad 16/600 Superdex 200. The buffer contains 0.1% Fos-Choline 12 detergent. The elution volume of > 60 mL is an indication of MW ~ 66 kDa hence suggesting the presence of higher molecular weight oligomers.

Cα	Cβ	Amino Acid
55.80	28.89	
45.33		GLY 8
58.77	63.95	SER 9
62.14	69.39	THR 10
58.61	63.74	SER 11
56.58	41.43	LEU 12
58.40	38.06	TYR 13
56.79	31.93	LYS 14
56.65	31.93	LYS 15
52.85	18.51	ALA 16
45.49		GLY 17
58.71	63.95	SER 18
56.72	29.16	GLU 19
53.35	38.87	ASN 20
55.20	41.50	LEU 21
58.26	38.46	TYR 22
57.97	38.76	PHE 23
56.34	28.83	GLN 24
45.60		GLY 25
53.76	38.81	MET 26
55.80	40.69	ASN 27
55.91	40.78	
57.45	28.76	GLU 29
57.05	29.71	ARG 30
62.74	31.66	VAL 31
58.44	63.88	SER 32

53.62	38.80	ASN	33
57.02	29.23	GLU	34
56.55	29.18	GLU	35
56.34	32.34	LYS	36
55.15	41.63	LEU	37
52.90	38.59	ASN	38
54.96	41.38	LEU	39
51.30	58.28	CYS	40
55.09		ARG	41
61.01	38.88		
57.75	40.94	TYR	43
61.59	38.40	TYR	44
57.54	31.93	LEU	45
46.05		GLY	46
45.84		GLY	47

63.81	30.65		
52.78	18.45	ALA	71
58.30	38.20	TYR	72
62.10	69.61	THR	73
56.79	29.30	GLU	74
56.32	28.49		
58.77	63.66	SER	76
56.24	28.81	GLN	77
61.97	37.69	ILE	78

60.59	29.25	TRP	92
66.79	30.65	VAL	93
65.37	36.31	ILE	94
67.15	30.71	VAL	95
58.29	40.89	LEU	96
67.03		THR	97

56.02	29.37		
57.86	29.48	TRP	110
45.90	69.98	GLY	111
52.84	18.59	ALA	112
57.14	41.03	LEU	113
46.47		GLY	114
55.67	39.88	ASP	115
59.35	38.87	TYR	116
54.80	41.63	LEU	117
58.37	63.54	SER	118
57.06	39.61	PHE	119
61.39	69.95	THR	120
58.66	37.86	ILE	121
62.48	34.15		
55.81	41.84	LEU	123
45.02		GLY	124
59.25	69.61	THR	125

fragment			
54.82	42.11		
45.46			
58.57	70.49	overlap	

fragment			
63.16	30.06		
54.73	42.31		
45.02		overlap	

fragment		
58.36	63.54	
56.82	39.61	
61.16	70.15	
57.80	41.45	overlap

fragment		
58.77	64.03	
56.75	29.10	
52.72		overlap

fragment		
64.47		
61.35	38.44	
52.21	19.03	
61.24	38.81	
60.55	38.53	overlap

fragment		
58.78		
57.42	28.74	overlap

Appendix 4: $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts obtained from sequential assignment (MAPPER) of the spin systems in full-length, his-tagged hPEN-2. Several fragments with $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts could not be included in the sequential walk due to low alignment. All assignments were made using solution-state NMR experiments.

Residue	Amino acid	C	C _α	C _β	N
13	F	178.866	57.595	39.472	114.479
14	W	117.139	61.951	28.753	120.501
15	L	176.695	54.182	42.697	121.156
16	V	176.99	58.583	31.465	0
17	F	176.587	62.249	39.275	119.068
18	G	175.115	45.76	0	101.142
19	G	175.287	44.697	2.66	110.728
20	C	176.217	62.121	24.61	118.38
21	G	174.972	47.439	2.66	106.89
22	S	176.232	62.664	61.184	117.709
23	A	2.66	54.603	18.166	125.299
24	V	178.915	57.324	38.365	115.474
41	L	2.66	65.881	37.92	117.723
42	A	179.6	55.855	17.956	121.138
43	F	180.372	64.005	38.235	112.363
44	G	176.471	48.077	2.66	104.356
45	L	181.494	57.235	41.091	116.883
46	T	175.548	65.845	68.003	116.355
47	V	176.526	66.03	32.56	123.155
48	L	2.66	57.896	42.925	119.753
49	T	177.019	65.635	69.746	105.219
50	M	178.287	56.45	31.236	119.483
56	H	177.33	66.005	29.323	117.979
57	I	176.556	63.209	38.069	119.818
58	S	176.026	58.331	66.668	107.188
105	G	172.03	45.051	2.66	108.121
106	K	175.134	53.892	34.527	121.868
107	T	175.387	64.277	68.787	121.869
108	G	174.532	44.818		
109	F	2.66	58.942	39.409	121.236
110	D	175.796	52.486	42.523	129.283
111	A	177.858	55.291	18.595	130.863
112	A	181.507	54.449	18.75	117.852
113	A	179.842	54.143	18.645	121.075
114	S	176.36	59.727	64.406	109.357
115	G	173.378	45.772	2.66	112.003
116	F	170.912	57.395	38.431	124.012
117	A	177.772	52.027	16.78	108.446
118	S	174.902	61.044	64.638	116.326
119	N	172.138	54.725	38.255	122.6
120	G	169.867	45.186	2.66	107.242
121	Y	177.238	54.529	43.366	111.525
122	G	176.463	46.544	2.66	112.34
123	E	175.955	57.845	28.83	128.457
151	G	176.069	48.797	2.66	107.34

152	A	177.72	53.341	17.686	119.06
153	T	173.718	60.103	69.116	101.852
154	D	177.894	54.192	42.625	126.257
175	L	175.722	57.505	31.085	121.222
176	I	175.189	63.489	40.019	109.939
177	S	176.273	59.378	67.429	99.643
178	I	175.112	67.154	35.005	128.832
180	V	179.06	63.8	31.457	110.798
181	T	177.363	60.198	71.267	104.544
182	N	171.837	54.033	39.9	123.929
183	T	172.365	66.288	64.729	108.144
184	S	173.905	58.372	63.896	128.835
185	V	174.861	62.199	33.185	111.48
186	N	173.765	51.821	39.226	110.777
189	R	178.753	57.391	28.117	120.46
190	S	175.713	62.416	62.845	114.901
191	T	173.538	66.898	69.614	118.262
192	A	175.845	55.732	19.232	118.517
193	V	178.374	62.125	30.297	97.053
194	A	179.246	54.457	19.477	124.972
197	Q	2.66	60.421	27.946	122.505
198	G	174.046	46.699	2.66	98.026
199	G	175.082	45.606	2.66	111.53
200	W	177.512	58.469	27.961	130.963
201	A	181.063	55.768	16.164	124.504
202	L	176.907	62.856	2.66	116.041
206	W	176.74	58.982	31.906	121.83
207	F	175.184	61.133	39.711	126.11
208	F	173.42	59.637	37.832	115.26
209	W	2.66	60.313	29.673	113.66
210	V	176.05	66.662	32.75	113.12

Appendix 5: Chemical shift values obtained from the sequential assignment of the spin systems in full length AqpZ (using CARA). Several fragments with C α and C β chemical shifts could not be included in the sequential walk due to low alignment.

Residue	Amino Acid	¹⁵N from ssNMR (δ ppm)	Predicted ¹⁵N from ShiftX (δ ppm)	Δδ (ppm)
105	G	108.121	109.6045	-1.4835
106	K	121.868	120.776	1.092
107	T	121.869	122.5643	-0.6953
108	G	114.328	113.3382	0.9898
109	F	121.236	122.4408	-1.2048
110	D	129.283	127.2078	2.0752
111	A	130.863	128.4753	2.3877
112	A	117.852	118.7405	-0.8885
113	A	121.075	117.6308	3.4442
114	S	109.357	112.1304	-2.7734
115	G	112.003	109.2154	2.7876
116	F	124.012	119.3721	4.6399
117	A	108.446	114.865	-6.419
118	S	116.326	113.5567	2.7693
119	N	122.6	121.0065	1.5935
120	G	107.242	107.0692	0.1728

Appendix 6: Table showing comparison of the calculated chemical shifts by NMR and predicted chemical shift values of ¹⁵N resonances by ShiftX. The fragment shown was assigned using CARA. Previously published crystal structure 1RC2 was used to obtain predicted chemical shift values using ShiftX software. The difference between calculated and predicted chemical shifts is mostly within 2 ppm.