

Expression, Purification And Biophysical Characterization Of Large
Fragments Of The Apelin Receptor Enabling Delineation Of A
Juxtamembrane Helix With Amphipathicity Necessary For Plasma
Membrane Localization.

by

Aditya Pandey

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

at

Dalhousie University
Halifax, Nova Scotia
December 2015

© Copyright by Aditya Pandey, 2015

Dedication

To my parents and grandparents for enabling me to pursue education to the highest level.

To the teachers who have bestowed knowledge and skills upon me.

To my friends for their continuous support and encouragement.

Table of Contents

List of Tables	vi
List of Figures.....	viii
Abstract.....	xii
List of Abbreviations and Symbols Used.....	xiii
Acknowledgements	xvi
Chapter 1: Introduction.....	1
1.1. Membrane Proteins:	1
1.2. Membrane Protein Expression:	2
1.2.1. <i>Escherichia coli</i> (<i>E. coli</i>):	2
1.2.2. Yeast:	5
1.2.3. Baculovirus/Insect Cells:	7
1.2.4. Mammalian Cells:	8
1.2.5. Cell-Free Systems:	9
1.3. Membrane Protein Purification And Sample Preparation For NMR: ...	11
1.4. G-Protein Coupled Receptors (GPCRs):	12
1.4.1. History And Classification Of GPCRs:	13
1.4.2. Structural Architecture Of GPCRs:.....	14
1.4.3. Signaling Through GPCRs:	19
1.5. The Apelinergic System:.....	20
1.6. The Divide And Conquer Approach To Study Membrane Proteins:	22
1.7. Membrane-Mimetics For Membrane Protein Structural Characterization:	23
1.8. Introduction To Biophysical Techniques:.....	25
1.8.1. Circular Dichroism (CD) Spectroscopy:	26
1.8.2. NMR Spectroscopy For Structural Characterization Of Proteins:.....	27
1.8.3. General Method For Sequential Backbone Assignment Using 3D Data:	31
1.8.4. Study Of Protein Dynamics By NMR:	33
1.9. Overview:	34
Chapter 2: Use Of AT-Rich Gene Tags To Enhance Production Of AR TM1-3 .	42
2.1. Introduction:.....	42
2.2. Materials And Methods:.....	47
2.2.1. Materials:	47

2.2.2.	Cloning And Plasmid Preparation:	48
2.2.3.	Expression Of AR Fragments:	49
2.2.4.	SDS-PAGE And Western Blotting:	50
2.2.5.	Inclusion Body Preparation:.....	51
2.2.6.	Ni ²⁺ Affinity Chromatography:.....	52
2.2.7.	Purification And Mass Confirmation:	52
2.3.	Results	53
2.3.1.	Difficulties In Production Of The Full-Length AR And Its Fragments:.....	53
2.3.2.	Expression Optimization Using AT-Rich Tags:	55
2.3.3.	Purification Of AR TM1-3:.....	56
2.4.	Discussion:	56
Chapter 3: Optimization Of Conditions For High Resolution Structural Characterization Of AR TM1-3 In Various Membrane-Mimetics		75
3.1.	Introduction:.....	75
3.2.	Materials And Methods:.....	78
3.2.1.	Materials:	78
3.2.2.	CD Spectropolarimetry:	78
3.2.3.	Solution-State NMR Spectroscopy:.....	79
3.3.	Results:	80
3.3.1.	CD Spectropolarimetry:	80
3.3.2.	Solution-State NMR Spectroscopy:.....	82
3.4.	Discussion:	84
Chapter 4: Use Of Split-Inteins To Obtain Full-Length AR.....		104
4.1.	Introduction:.....	104
4.2.	Material And Methods:	108
4.2.1.	Materials:	108
4.2.2.	Construction Of Fusion Genes And Expression Plasmids:.....	108
4.2.3.	Expression Of AR TM1-3-I _N And I _C -AR TM4-7 Constructs In <i>E. Coli</i> Cells:	109
4.2.4.	Cell-Free Expression Of AR TM1-3-I _N And I _C -AR TM4-7 Constructs:.....	110
4.2.5.	Condition Optimization For Splicing Of The AR Fragments:.....	110
4.3.	Results:	111
4.3.1.	Expression In <i>E. coli</i> Cells:.....	111
4.3.2.	Cell-Free Expression:.....	111

4.3.3.	Intein-Mediated Splicing Of The Expressed AR Fragments:	112
4.4.	Discussion:	112
Chapter 5:	Structural And Functional Characterization Of The	127
	Juxtamembrane 8th Helix Of The AR	127
5.1.	Introduction:.....	127
5.2.	Materials And Methods:.....	129
5.2.1.	Materials:	129
5.2.2.	Cloning Of AR C-Tail And Its Mutants For Recombinant Expression:	130
5.2.3.	Mutagenesis Of The AR C-Tail Of HA-Tagged AR For	131
	Immunofluorescence Experiments:.....	
5.2.4.	Expression Of The WT AR C-Tail And Its Mutants:	131
5.2.5.	Purification Of The WT AR C-Tail And Its Mutants:	131
5.2.6.	CD Spectroscopy:	132
5.2.7.	Solution-State NMR Spectroscopy:	133
5.2.8.	Cell Culture:.....	134
5.2.9.	Transient Transfections:.....	134
5.2.10.	Cell Surface Immunofluorescence:.....	135
5.2.11.	Intracellular Immunofluorescence Microscopy:	135
5.3.	Results:	136
5.3.1.	Expression And Purification Of WT AR C-Tail:.....	136
5.3.2.	CD Spectroscopy Indicates Helix Induction In The Presence Of	137
	Membrane Mimetics:	
5.3.3.	NMR Spectroscopy Demonstrates Formation Of The Helix-8 In Presence	139
	Of LPPG Micelles:.....	
5.3.4.	Mutations In Helix-8 Reduce Plasma Membrane Localization Of The AR:	140
5.4.	Discussion:	142
Chapter 6:	Conclusions And Future Directions	162
6.1.	AT-Rich Gene Tags As A Tool For Membrane Protein Expression:....	162
6.2.	Possible Future Studies With AR TM1-3:	163
6.3.	Using Intein-Mediated Splicing To Obtain Full-Length AR:	164
6.4.	Role Of AR C-Tail In Plasma Membrane Localization Of AR And	
	Possible Future Studies:	164
6.5.	Significance Of This Work:.....	165
References:	168

LIST OF TABLES

Table 2.1 Primers and constructs for AR TM1-3	58
Table 2.2 Primers and constructs for AR TM4-7	59
Table 2.3 Standard thermocycling conditions and reaction composition for	
amplification of DNA using PCR.....	60
Table 2.4 Summary of constructs tested for expression of AR TM2-3	61
Table 2.5 Nucleotide and amino acid sequence of the expression enhancement tags	62
Table 2.6 Difference in the level of AR TM1-3 in BL21(DE3) and C41(DE3) strains	
of <i>E. coli</i>	63
Table 3.1 List of 2D experiments used to access the effect of temprature on spectral	
quality of AR TM1-3.	88
Table 3.2 List of 2D and 3D experiments acquired for AR TM1-3 in 50% HFIP	89
Table 3.3 List of 2D and 3D experiments acquired for AR TM1-3 in SDS micelles.....	90
Table 3.4 List of 2D and 3D experiments acquired for AR TM1-3 in 50% TFE.....	91
Table 3.5 Deconvolution of CD data using the CDSSTR algorithm and SMP180	
reference dataset.....	92
Table 4.1 Sequence of split-inteins used in this work	115
Table 4.2 Primers and constructs for AR TM1-3 and AR TM4-7 with different split-.....	
inteins.....	116

Table 5.1 Primers used in making plasmid constructs for the WT AR C-tail and its mutants for expression in <i>E. coli</i>	145
Table 5.2 List of 2D and 3D experiments acquired for WT AR C-tail in buffer with and without LPPG micelles	146
Table 5.3 Deconvolution of CD data using the BeStSel server for AR, AR-G, and AR-S in pH 6 condition	147
Table 5.4 Deconvolution of CD data using the BeStSel server for AR, AR-G, and AR-S in pH 7 condition.	148

LIST OF FIGURES

Figure 1.1 Flowchart outlining various optimization steps allowing for structural..... characterization of membrane proteins by NMR spectroscopy	36
Figure 1.2 Snake plot for the apelin receptor representing its sequence and topology	37
Figure 1.3 Timeline of solved GPCR structures	38
Figure 1.4 Cartoon representation of the current understanding of membrane protein..... folding	39
Figure 1.5 Illustration of 1D vs. 2D vs. 3D spectra for AR TM1-3 in 50% TFE	40
Figure 1.6 A general scheme for the “backbone walk”	41
Figure 2.1 The snake plot representing the “divide and conquer” approach based..... division of the receptor into four parts.....	64
Figure 2.2 The snake plot representing the “divide and conquer” approach based..... division of the receptor into two parts	65
Figure 2.3 AR TM2-3 expression using MBP and SUMO as fusion proteins	66
Figure 2.4 Factor Xa mediated cleavage of MBP from AR TM2-3 construct	67
Figure 2.5 Evaluation of AR TM1-3 expression in BL21(DE3) strain of <i>E. coli</i>	68
Figure 2.6 Evaluation of AR TM1-3 expression in C41(DE3) strain of <i>E. coli</i>	69
Figure 2.7 Comparison of AR TM1-3 expression in using AT-rich gene tags.....	70

Figure 2.8 Evaluation of purification and expression yield in minimal medium for AR..... TM1-3	71
Figure 2.9 HPLC chromatogram for AR TM1-3 purification	72
Figure 2.10 Mass spectrum to confirm identity and evaluate the efficiency of isotope-..... labeling for AR TM1-3	73
Figure 2.11 HPLC chromatogram to evaluate the purity of AR TM1-3 used for NMR..... sample preparation	74
Figure 3.1 Representation of membrane-mimetics used in this work	93
Figure 3.2 Comparison of secondary structure composition of AR TM1-3 in..... membrane-mimetics using far-UV CD	94
Figure 3.3 Comparison of secondary structure composition of AR TM1-3 in SDS and..... LPPG micelles using far-UV CD.....	95
Figure 3.4 ^1H - ^{15}N HSQC spectra for AR TM1-3 in DMSO at 22°C.....	96
Figure 3.5 ^1H - ^{15}N HSQC spectra for AR TM1-3 in SDS micelles at 47°C	97
Figure 3.6 ^1H - ^{15}N HSQC spectra for AR TM1-3 in 50%HFIP at 47°C	98
Figure 3.7 Effect of temperature on 1D ^1H and 2D ^1H - ^{15}N HSQC spectra for	
AR TM1-3 in 50% HFIP.....	99
Figure 3.8 Effect of temperature on 1D ^1H and 2D ^1H - ^{15}N HSQC spectra for	
AR TM1-3 in SDS micelles.....	100

Figure 3.9 ^1H - ^{15}N HSQC spectra for AR TM1-3 in 50%HFIP at 47°C	101
Figure 3.10 Comparison of the effect of protein concentration on ^1H - ^{15}N HSQC spectra for AR TM1-3.....	102
Figure 3.11 Partially assigned ^1H - ^{15}N HSQC for AR TM1-3 in 50% TFE.....	103
Figure 4.1 Representation of an intein.....	117
Figure 4.2 Representation of a split-intein.....	118
Figure 4.3 Five-step split-intein splicing mechanism.....	120
Figure 4.4 Schematic for segmental isotope-labeling of AR.....	121
Figure 4.5 Schematic for <i>in vitro</i> expression and subsequent splicing to obtain the..... full-length segmentally labeled AR	122
Figure 4.6 SDS-PAGE representing expression of AR constructs with split-inteins.....	123
Figure 4.7 Western blot representing expression of AR constructs with split-inteins....	124
Figure 4.8 Western blot demonstrating results of co-translational splicing	125
Figure 4.9 Western blot demonstrating results of post-translational splicing	126
Figure 5.1 Sequence-based secondary structure prediction for the WT AR C-tail and..... its mutants.....	149
Figure 5.2 SDS-PAGE demonstrating successful expression and purification of the..... WT AR C-tail.....	150

Figure 5.3 HPLC chromatogram demonstrating purification of WT AR C-tail.....	151
Figure 5.4 Identity of WT AR C-tail confirmed by mass spectrometry	152
Figure 5.5 CD spectra for the WT AR C-tail and its mutants with and without	
Membrane-mimetics	153
Figure 5.6 Difference CD spectra for the WT AR C-tail and its mutants in various.....	
membrane-mimetics.....	154
Figure 5.7 Comparison of ^1H - ^{15}N HSQC spectra for the WT AR C-tail in buffer with.....	
and without LPPG micelles	155
Figure 5.8 Assigned ^1H - ^{15}N HSQC for WT AR C-tail at 37 °C in presence of 100 mM.....	
LPPG.....	156
Figure 5.9 Estimation of secondary structure and dynamics of WT AR C-tail using	
NMR spectroscopy.....	157
Figure 5.10 FACS analysis of plasma membrane localization of full-length AR	159
Figure 5.11 AR cell surface expression is reduced in AR helix-8 mutants	160
Figure 5.12 Cyclohexamide treatment confirms reduced cell surface expression for	
AR helix-8 mutants	161
Figure 6.1 Fragments of AR studied in this work.....	167

ABSTRACT

The apelin receptor (AR or APJ) is a class-A G-protein coupled receptor (GPCR) activated by various isoforms of the peptide hormones apelin and, during embryonic development, apela. The apelin-AR system has been proposed as a therapeutic target for cardiovascular disease, central nervous system disorders, and various cancers. Because GPCRs are both inherently dynamic and function in a heterogeneous lipid bilayer environment, nuclear magnetic resonance (NMR) spectroscopy is well-suited for their characterization. NMR spectroscopy, however, requires multi-milligram quantities of proteins enriched with ^{15}N , ^{13}C , and/or ^2H isotopes. Full-length isotope-enriched GPCRs are very challenging to produce; therefore, I have employed the “divide and conquer” approach to study the AR in two pieces. Here, for the first time, I have demonstrated the use of small adenine- and thymine-rich (AT-rich) gene tags to enhance the expression of the 137 amino acid N-terminal portion of AR including the first three transmembrane segments (TM1-3) in *Escherichia coli*. Condition optimization in pursuit of an ideal membrane-mimetic environment has also been carried out. *In vitro* translation of full-length AR divided into two pieces, each fused with a split-intein fragment, is also demonstrated using an *E. coli*-based cell-free expression system. I also expressed, isotope-enriched, and purified the 71 amino acid intracellular C-terminal tail of the AR to allow correlation of its structure, membrane-interactions, and function. Biophysical characterization indicate the presence of a juxtamembrane 8th helix upon exposure to membranous environments. This helix is, putatively, common to class-A GPCRs but poorly characterized from a structure-function perspective. Directly correlating to biophysical data showing differences in helicity and micelle binding, I used immunofluorescence experiments to unambiguously show the importance of amphipathy of this helix in plasma membrane localization of the receptor. Overall, this work details both the challenges faced and successes achieved in characterization of the AR using NMR spectroscopy.

LIST OF ABBREVIATIONS AND SYMBOLS USED

χ : Magnetic susceptibility

γ : Gyromagnetic ratio

σ : Chemical shielding tensor

ω_0 : Larmor frequency in rad s^{-1}

ν : Larmor frequency in Hz

$[\theta]$: Mean residue ellipticity

APJ: Apelin Receptor

AR: Apelin Receptor

AR55: First 55 residues of the AR fused to a polyhistidine tail

AT TM2-3: Fragment of the AR containing transmembrane segment 2 and 3.

AR TM1-3: Fragment of the AR containing transmembrane segment 1, 2, and 3.

AR TM1-4: Fragment of the AR containing transmembrane segment 1, 2, 3, and 4.

AR TM5-7: Fragment of the AR containing transmembrane segment 5, 6, and 7.

AR TM4-7: Fragment of the AR containing transmembrane segment 4, 5, 6, and 7.

AR C-tail: 71 amino acid long C-terminal tail of the AR

\mathbf{B}_0 : External magnetic field

Brij 30: Polyoxyethylene(4)lauryl ether

Brij 35: Polyoxyethylene-(23)-lauryl-ether

Brij 78: polyoxyethylene-(20)-stearyl-ether

BSA: Bovine serum albumin

CD: Circular dichroism

CSI: Chemical shift index

DDM: *n*-Dodecyl- β -d-maltoside

DPC: Dodecylphosphocholine

DSS: 2,2-dimethyl-2-silapentane-5-sulfonate

DTT: Dithiothreitol

ECL: Extracellular loop

g: gravitational force

GPCR: G-protein coupled receptor

h: Plank's constant

HFIP: 1,1,1,3,3,3-hexafluoroisopropanol

HPLC: High performance liquid chromatography

HSQC: Heteronuclear single quantum coherence

ICL: Intracellular loop

I_N: Intein-N

I_C: Intein-C

IPTG: Isopropyl β -D-1-thiogalactopyranoside

PCR: Polymerase chain reaction

LPPG: 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-RAC-(1-glycerol)]

LPPC: 1-palmitoyl-2-hydroxy-*sn*-glycero-3-phosphocholine

MFI: Mean fluorescence intensity

MNG: maltose neopentyl glycol-3

NOE: Nuclear Overhauser effect

NOESY: Nuclear Overhauser effect spectroscopy

NMR: Nuclear magnetic resonance

NMR-3: Nuclear Magnetic Resonance Research Resource

NRC-IMB: National Research Council Institute for Marine Biosciences

PBS: Phosphate buffered saline

PDB: Protein Data Bank

NRMSD: Normalized root mean square deviation

SDS: Sodium dodecylsulphate

TFA: Trifluoroacetic acid

TFE: 2,2,2-Trifluoroethanol

TM: Transmembrane

ACKNOWLEDGEMENTS

First and foremost I would like to express my sincere gratitude to Dr. Jan K. Rainey and Dr. Paul Xiang-Qin Liu for providing me with an opportunity to pursue higher education under their guidance. Both Paul and Jan have been very supportive throughout my time at Dalhousie. Paul and Jan have been an excellent mentors and have always acted in my best interest. I sincerely thank them for the independence they have provided me in designing my own experiments. Their valuable suggestions have always been very helpful. The opportunities that Jan has provided me to attend conferences and workshops have been very helpful in my growth as a researcher. Letting me mentor undergraduate students in the lab is another example of how Jan has augmented my graduate school experience by enabling me to gain some supervisory experience.

Besides my advisors, I would like to thank the members of my graduate advisory committee: Dr. Roger McLeod, Dr. Denis Dupré and Dr. Roy Duncan, for their helpful advice during committee meetings and the encouragement they provided throughout the degree program.

I also thank Dr. Roy Duncan for allowing me access to his laboratory to be trained to conduct eukaryotic cell culture work. I would especially like to thank Dr. Hiren Parmar for helping me with FACS and microscopy; Dr. Roberto de Antueno and Nicole McMullen for helping me with baculovirus expressions.

The past five years at Dalhousie have been very enriching and enjoyable because of the following people that I have worked with; Andy Song, Aaron Banks, Nigel Chapman, Danielle LeBlanc, Robin Patterson, Nathan Weatherbee-Martin, Ellie M. Scott, Sara Sparavalo, Sam Kerzner, Doug Miller, Matthew Speckert, Kate Huang,

Pascaline Ngweniform, Ben Morash, Emily Truong, Lesley Seto, David Langelaan, Kyungsoo Shin, Muzaddid Sarker, and Lingling Xu. I also thank David for making me feel welcome in the lab and getting me started with basic protein work, as at that time, running a SDS-PAGE gel was a completely new thing for me! I would especially like to acknowledge both Kyung and Nigel for making 10H an enjoyable place to work. You guys have not been anything less than brothers to me!

I would also like to thank Roisin for her constant support, help, and advice throughout the degree program. Mitesh, Hiren, Manoj, Umang, and Milind are the reason I never felt away from home and was able to still talk in hindi with them, which is still the language I am most comfortable in.

Last, but not the least, I would like to thank Bruce Stewart for all the help he has provided during my stay at the Rainey lab. It is difficult to imagine the lab without you Bruce!

Chapter 1: Introduction

Note: *Parts of this chapter have been submitted as a mini-review titled “Current strategies for protein production and purification enabling membrane protein structural biology.” to the journal **Biochemistry and Cell Biology**.*

Three-dimensional structural information for biomolecules has had a significant impact on the field of drug discovery (Russell & Eggleston, 2000). Structural characterization of membrane proteins is very challenging and involves many bottlenecks during the course of the process. This thesis details the challenges faced *en route* to structural studies of a membrane protein, the apelin receptor, a class A G-protein coupled receptor, using nuclear magnetic resonance (NMR) spectroscopy and the strategies I developed to overcome them.

1.1. Membrane Proteins:

Membrane proteins play crucial roles in a wide variety of cellular functions, including regulation of ion transport across the membrane; sensing and transmitting chemical and electrical signals; mediating cellular attachment; and, controlling membrane lipid composition (von Heijne, 2007). The importance of membrane proteins is also highlighted by the fact that they represent approximately 26% of the human proteome and that more than 50% of currently marketed drugs target this class of proteins (Overington *et al.*, 2006; Yildirim *et al.*, 2007; Arinaminpathy *et al.*, 2009; Fagerberg *et al.*, 2010). Membrane proteins can, arguably, be referred to as the “Holy Grail” in the field of structural biology. Correspondingly, despite their importance and relative prevalence, less than 2% of the experimentally determined structures in the protein data bank (PDB) are of membrane proteins (Kozma *et al.*, 2013). Characterization of the

structure and dynamics of these proteins not only provides insight into their mechanisms of function, but also aids in rational design of new drugs (Vinothkumar & Henderson, 2010; Lounnas *et al.*, 2013).

1.2. Membrane Protein Expression:

Structural studies typically require milligram quantities of protein. Given that membrane proteins are generally scarce in their natural environment, heterologous expression of these proteins is necessary to obtain enough protein for characterization (Bill *et al.*, 2011). However, heterologous expression in the host of choice is not always feasible due to the often toxic nature of a membrane protein upon overexpression in that host (Bernaudat *et al.*, 2011). In addition, the discrepancy in the level of expression from one membrane protein to the next is not yet completely understood (Grisshammer, 2006). Thus, it is not currently possible to predict the degree of toxicity of a target protein and its relative expression in a given host is still a matter of trial and error. Correspondingly, a variety of expression hosts have been used successfully for expression of target membrane proteins. These expression hosts differ in their ability to carry out post-translational modifications, in level of expression, and in sensitivity to modification of culture conditions, which can dictate the selection of one host over the other. In this section, I briefly discuss the advantages and limitations of the most commonly used expressions systems.

1.2.1. *Escherichia coli* (*E. coli*):

E. coli is the most widely used host system for expression of recombinant proteins for five main reasons. (i) *E. coli* has fast growth kinetics with a doubling time of 20

minutes in optimal conditions (Sezonov *et al.*, 2007). This reduces the time required for expression of the protein of interest (POI). (ii) Cells can be grown to high density to achieve very high levels of protein production (Shiloach & Fass, 2005). (iii) Growth medium is inexpensive and can be manipulated without a significant loss in yield, enabling enrichment with stable isotopes for NMR spectroscopy (Marley *et al.*, 2001; Tyler *et al.*, 2005) and selenomethionine for X-ray diffraction (Stols *et al.*, 2004; Sreenath *et al.*, 2005). (iv) Extensive knowledge of genetics, physiology, and metabolism has enabled intelligent genetic manipulations to produce strains with specific advantages (Studier, 1991; Sorensen & Mortensen, 2005; Wagner *et al.*, 2008; Andersen *et al.*, 2013). (v) An abundance of expression vectors (https://www.embl.de/pepcore/pepcore_services/cloning/choice_vector/ecoli/index.html) enables easy incorporation of desired properties in the plasmid vector.

T7 RNA polymerase (T7RNAP) from *Enterobacteria phage λ* is often used to drive recombinant protein production in *E. coli* (Studier, 1991). T7RNAP exclusively recognizes the T7 promoter and synthesizes RNA several times faster than *E. coli* RNA polymerase with fewer observed incomplete transcripts due to premature termination, allowing higher protein yield (Iost *et al.*, 1992). Typically, T7RNAP is employed under control of the *lac* operon, where addition of isopropyl β-D-thiogalactoside (IPTG) induces expression of T7RNAP and subsequently enables overexpression of the POI (Studier, 1991). Since *E. coli* can grow in minimal medium, rich medium can be replaced with a medium containing isotope-enriched precursors or amino acids before inducing expression, enabling cost-effective production of proteins amenable to heteronuclear NMR characterization (Marley *et al.*, 2001).

Recently, auto-induction of membrane protein expression has been suggested to be more effective than traditional T7RNAP IPTG-induced expression (Gordon *et al.*, 2008). In this method, protein production is induced by lactose in the medium, which only occurs upon the depletion of metabolites such as glucose. Although the auto-induction protocol seems to provide an edge over the traditional IPTG induction protocol, its cost-effectiveness is somewhat compromised in producing isotope-enriched proteins for NMR (Tyler *et al.*, 2005). For additional background, the reader is referred to detailed reviews of various aspects of recombinant protein expression using *E. coli* expression systems (Makrides, 1996; Baneyx, 1999; Stevens, 2000; Sorensen & Mortensen, 2005; Rosano & Ceccarelli, 2014).

The prokaryotic *E. coli* system also has many limitations. Specifically, this organism often lacks the essential lipids, molecular chaperones, and machinery for PTMs required for correct membrane insertion and eukaryotic protein folding. It is important to note that efforts have been made to overcome these barriers by methods such as codon-optimization (Burgess-Brown *et al.*, 2008); addition of fusion tags for expression enhancement (Hsu *et al.*, 2013) and/or membrane insertion (Neophytou *et al.*, 2007); and, co-translation of post-translational machineries to facilitate protein folding (Mijakovic *et al.*, 2006). However, the use of *E. coli* still has many challenges to be overcome before it can be considered an ideal expression system for a membrane proteins. Miroux and co-workers (Hattab *et al.*, 2015) have recently provided an extensive analysis of the impact of *E. coli* expression systems in the field of membrane protein structural characterization.

1.2.2. Yeast:

Yeast have many similar advantages to *E. coli* for heterologous expression. For example, yeast also have a rapid growth rate, ability to be grown to high density, well-studied genetics, and availability of advanced tools for genetic manipulation. However, yeast cells have the added advantage of being capable of performing various eukaryotic PTMs. These modifications include proteolytic processing of signal peptide sequences; disulfide-bond formation; prenylation; phosphorylation; acylation; and certain types of *O*- and *N*-linked glycosylation that may be essential for activity, correct folding, and membrane insertion of proteins (Boer *et al.*, 2007). It is important to note that glycosylation patterns vary between different yeast strains, and some have been engineered to have more uniform and human-like glycosylation patterns (Hamilton *et al.*, 2003; Hamilton *et al.*, 2006; Vervecken *et al.*, 2007). These factors make yeast an inexpensive and efficient alternative to prokaryotic expression systems for production of membrane proteins. *S. cerevisiae* and *Pichia Pastoris* (*Pichia*) are the most commonly used species for expression of membrane proteins (Joubert *et al.*, 2010; Bornert *et al.*, 2012). *Schizosaccharomyce pombe* is a less frequently used yeast species that sometimes outperforms other yeast species in expression of mammalian membrane proteins (Sander *et al.*, 1994; Takegawa *et al.*, 2009).

Using *S. cerevisiae*, a model eukaryote with a long history of study, comes with a few perks. These include a completely sequenced genome, a variety of genetic manipulation tools, and a large number of available strains for expression of eukaryotic membrane proteins (Goffeau *et al.*, 1996; Winzeler *et al.*, 1999; Drew *et al.*, 2008; Li *et al.*, 2009). In comparison to other strains of yeast, protocols are available for high-throughput expression of membrane proteins (Newstead *et al.*, 2007). The ability to carry

out *in vivo* homologous recombination also makes *S. cerevisiae* an attractive host to produce membrane protein mutants in a high-throughput fashion to optimize expression for downstream applications (Ito *et al.*, 2008).

Despite these advantages, the methylotrophic yeast *Pichia* has been the most successfully applied membrane protein expression host among all the yeast strains. A *Pichia* expression system was first released for academic use in 1993, after which there has been an exponential increase in the knowledge base and the number of membrane proteins expressed using this system (Cregg *et al.*, 2000; Ramon & Marin, 2011). A key physiological trait of *Pichia* is its strong preference for respiratory growth, facilitating culturing at higher densities than fermentative yeast. Its methylotrophic nature also means that it can derive energy by metabolism of methanol using enzymes such as alcohol oxidase (AOX) and dihydroxyacetone synthase (DHAS) (Stewart *et al.*, 2001). Since the promoters of methanol utilization pathways are very efficient and tightly regulated, they are most commonly used for expression of heterologous proteins (Hollenberg & Gellissen, 1997). Using this system, very high expression yields have been achieved for membrane proteins (*e.g.*, 90 mg·L⁻¹ for human aquaporin (Nyblom *et al.*, 2007)). Similar to *E. coli*, *Pichia* can grow on minimal medium that can be modified to produce isotope-enriched protein required for structural characterization using NMR (Denton *et al.*, 1998; Wood & Komives, 1999; van den Burg *et al.*, 2001). Although the cost of production of an isotope-enriched protein may be higher compared to *E. coli*, *Pichia* is still a very attractive alternate host for production of isotope-enriched proteins that cannot be efficiently produced in *E. coli* (Clark *et al.*, 2015).

1.2.3. Baculovirus/Insect Cells:

There are many advantages of using the baculovirus/insect cell expression system. In comparison to other virus-based methodologies, baculovirus is safe due to its inability to infect mammals. Similarly to other eukaryotic expression systems, baculovirus/insect cell expression of heterologous genes typically permits proper protein folding as a consequence of PTMs that are often identical to those that occur in higher eukaryotes (Shi & Jarvis, 2007). Cell cultures from the insects *Spodoptera frugiperda* (Sf9, Sf21) and *Trichoplusia ni* (Hi5) infected by the baculovirus *Autographa californica* multi-nucleopolyhedrovirus (AcMNPV) are the most commonly used systems for membrane protein expression (Hitchman *et al.*, 2009). Heterologous expression of proteins varies in yield when expressed in different types of insect cells (Unger & Peleg, 2012).

A typically employed commercially available system for production of recombinant baculovirus for expression of proteins in insect cells is the BAC-to-BAC[®] (Invitrogen by life technologies, Waltham, MA, USA) system. In this protocol, the target gene is sub-cloned into the pFastBac vector, which is then transformed into DH10Bac *E. coli* competent cells harboring a baculovirus shuttle vector (bacmid) with a transposon site and a helper plasmid. Transposition of the target gene onto the bacmid gives rise to a recombinant bacmid that can be purified and used for production of recombinant virus that can be later used for infection of insect cells for testing protein expression. This entire process takes approximately 3-4 week to reach the stage of testing protein expression.

Recently, a faster and more convenient method employing transient transfection has also been designed for rapid screening of membrane protein expression (Chen *et al.*, 2013). To date, baculovirus-insect cell expression has given the most number of

eukaryotic membrane proteins for structural characterization (He *et al.*, 2014). Although this seems to be the best expression system for membrane protein expression, it is relatively expensive to produce proteins by this system compared to the systems discussed earlier. Moreover, due to complex metabolism and inability to grow in minimal medium, it is prohibitively expensive to produce isotope-enriched proteins for structural characterization by NMR (Gossert & Jahnke, 2012; Sitarska *et al.*, 2015). Notably, an economical approach has recently been put forward for expression of isotope-enriched proteins using a homemade isotope-enriched yeast extract (Opitz *et al.*, 2015). Given the reported success of this system, it appears highly promising to devise strategies to reduce cost over currently available protocols.

1.2.4. Mammalian Cells:

Membrane proteins that require specific PTMs and subcellular environments for proper folding and activity are often impossible to produce in functional form by prokaryotic or lower eukaryotic systems. Interestingly, the difficulty in competent protein production has been stated to not be proportional to the number of transmembrane segments or to the size of the protein; rather, it appears to arise from the complexity of the folding process for the protein in question (Tate, 2001). It is likely that a combination of various factors decides the fate of protein expression in a specific expression host. Therefore, mammalian cell lines seem to be an obvious choice to obtain functional mammalian proteins. However, these systems also have drawbacks. One of the main drawbacks is that levels of expression relative to the systems already introduced are typically very low, even under conditions of induced overexpression, making it prohibitively expensive to produce sufficient protein for structural studies (Andrell &

Tate, 2013). Chinese hamster ovary cells (CHO), human embryonic kidney cells (HEK293), baby hamster kidney cells (BHK-21) and monkey kidney fibroblast cells (COS-7) are the most commonly used cell lines for expression (Andrell & Tate, 2013). These different cell lines can vary significantly in terms of expression level depending on the POI, requiring individual optimization for each POI. It should be noted that mammalian cell line-based expression can also be performed either by transient transfection or by making stable cell lines. Transient transfection can be either carried out using recombinant viruses (Hassaine *et al.*, 2006; Dukkupati *et al.*, 2008; Matrai *et al.*, 2010) or by using chemical agents (Geisse & Fux, 2009). Although more time consuming, stable transfection provides a means of producing heterologous proteins with high reproducibility and expression level (Chelikani *et al.*, 2006; Camponova *et al.*, 2007). To date, due to complex metabolism and prohibitively expensive expression medium, a very limited number of proteins have been expressed using mammalian expression systems with isotope-enrichment, illustrating the current challenge of applying this class of expression system for NMR (Sastry *et al.*, 2012).

1.2.5. Cell-Free Systems:

In all of the conventional cell-based expression systems, the focus is on overexpression, which often results in improper protein folding and membrane insertion. In the case of membrane proteins, overexpression also frequently leads to formation of protein aggregates, destabilization of cellular membranes, and interaction with host metabolic machinery leading to cytotoxicity (Klammt *et al.*, 2006). Therefore, cell-free expression systems (also referred to as *in vitro* translation systems) were developed to bypass the complexities and sensitivities inherent in dealing with living cells (Nirenberg

& Matthaiei, 1961). In recent years, cell-free expression has gained popularity for the expression of membrane proteins.

Typically, a cell-free expression system is an open system enabling *in vitro* protein translation with the help of translation machineries provided by the cell lysates from various organisms (Schwarz *et al.*, 2008). A major attractive feature of cell-free systems for membrane protein expression is the ability to implement a carefully chosen folding environment. This allows for better control over the driving forces of membrane protein folding such as lipid composition (Phillips *et al.*, 2009); hydrophobic mismatch (Cybulski & de Mendoza, 2011); membrane lateral pressure and curvature (Botelho *et al.*, 2006; Marsh, 2007); membrane elasticity (Lundbaek, 2006; Lundbaek *et al.*, 2010); and, cholesterol content (Pucadyil & Chattopadhyay, 2006; Chini & Parenti, 2009). Cell-free expression systems exhibit suppressed interconversion of amino acids, enabling a wide range of isotope-enrichment patterns that can be exploited in various NMR approaches to solve the signal overlap problem that is common in membrane protein structural characterization (Parker *et al.*, 2004; Ozawa *et al.*, 2006; Junge *et al.*, 2011; Reckel *et al.*, 2011). Cell-free systems also allow production of proteins with Stereo-Array Isotope Labeling (SAIL) (Kainosho *et al.*, 2006). SAIL is a method of producing proteins with amino acids having certain stereospecific patterns of isotopes optimal for protein NMR analysis. Using the SAIL labeling strategy makes large proteins amenable for structural characterization by NMR (Takeda *et al.*, 2008; Tonelli *et al.*, 2011).

Currently, several types of cell-free system are available, based upon extracts from different sources, such as: *E. coli* (Schwarz *et al.*, 2007a), wheat germ (Harbers, 2014), rabbit reticulocyte (Anastasina *et al.*, 2014), *Spodoptera frugiperda* (Ezure *et al.*, 2014), CHO cells (Brodell *et al.*, 2014), mouse embryonic fibroblasts (Zeenko *et al.*,

2008), and the HeLa cell line (Mikami *et al.*, 2006). Among these, *E. coli* lysate is currently the most widely, and successfully, applied cell-free expression medium. Variations have been developed, such as the Protein synthesis Using Recombinant Elements (PURE) system (Shimizu *et al.*, 2001) containing a minimal set of purified elements required for the translation reaction, and the Cytomim system (Jewett *et al.*, 2008) where, besides the *E. coli* cell extract, inverted inner membrane vesicles from *E. coli* are also added. Recently, the PURE system has been optimized for rapid production of high quantities of functional membrane proteins by adding liposomes to the reaction mixture (Kuruma & Ueda, 2015). Although relatively expensive in terms of the required isotope-enriched precursors, cell-free expression is a very attractive technique for production of isotope-enriched proteins for structural characterization by NMR. It should also be noted that although production of properly folded membrane proteins has not yet been streamlined in cell-free expression systems, this class of expression has a great deal of potential, especially when it comes to production of membrane proteins for NMR.

1.3. Membrane Protein Purification And Sample Preparation For NMR:

NMR is a very versatile technique for atomic-level studies of biomolecular structure and dynamics. Membrane proteins may be characterized in environments amenable to either solution- or solid-state NMR methods. These methods differ in their requirement for isotropic tumbling of the nuclei being probed, with large molecules or complexes (>100 kDa) being extremely challenging to characterize by, if at all amenable to, solution-state NMR (Nietlispach & Gautier, 2011). NMR allows study of membrane proteins in their native environment (Murray *et al.*, 2013; Brown & Ladizhansky, 2015) or, following reconstitution, in native-like environments, including membrane-mimetics

such as organic solvent mixtures (Schwaiger *et al.*, 1998), micelles (Arora *et al.*, 2001), bicelles (Prosser *et al.*, 2006), nanodiscs (Gluck *et al.*, 2009; Hagn *et al.*, 2013), and bilayers (Das *et al.*, 2013). Based on the size of the reconstituted assembly of the membrane protein, either solution-state or solid-state approaches are used for their characterization. For solution-state NMR, approaches such as transverse relaxation optimized spectroscopy (TROSY) (Pervushin *et al.*, 1997) and specific isotope-labeling schemes (Goto *et al.*, 1999; Goto & Kay, 2000; Tugarinov *et al.*, 2002) have been used to extend the use of solution-state NMR for larger protein assemblies. These approaches rely upon exploitation of differential/optimal relaxation properties, and have also been exploited to study both structure and dynamics of integral membrane proteins (Hwang & Kay, 2005; Gautier *et al.*, 2010b; Qureshi & Goto, 2012). Challenges involved in sample preparation have been discussed further in Chapter 3. Figure 1.1 outlines the steps involved in structural characterization of membrane proteins.

1.4. G-Protein Coupled Receptors (GPCRs):

GPCRs represent the largest family of integral membrane proteins, which interact intracellularly with heterotrimeric guanine nucleotide-binding protein (G-proteins) or β -arrestins for their function (Moller & Gmeiner, 2015). GPCRs serve as gatekeepers, modulating numerous intracellular signaling cascades in response to a wide variety of extracellular stimuli including small molecules, amino acids, peptides, proteins, lipids, nucleotides, and photons. Corresponding to their physiological importance, GPCRs are the targets of 50-60% of currently marketed drugs (Lundstrom, 2009). Elucidating the structure of GPCRs is key both in terms of understanding this important class of proteins

and because of the indispensable role this may play in the field of drug discovery (Allen & Roth, 2011).

1.4.1. History And Classification Of GPCRs:

Long before the first GPCR (β_2 -adrenoceptor) was first isolated in 1979 by the Lefkowitz group (Caron *et al.*, 1979), various aspects of the signaling cascade through these receptors were already being studied (Sutherland, 1972). It was in the 1980s when research on the molecular aspects of GPCRs gained momentum, with rhodopsin and β_2 -adrenoceptor (β_2 AR) being the first receptors to be successfully sequenced, cloned and isolated (Nathans & Hogness, 1984; Dixon *et al.*, 1986). In 1975, using electron microscopy, the seven transmembrane (7TM) topology was first identified for bacteriorhodopsin (Henderson & Unwin, 1975). Although not a GPCR, bacteriorhodopsin has a 7TM topology similar to GPCRs. In later years, hydrophobicity plots of other identified GPCRs confirmed the speculations for existence of a large class of receptors with 7TM architecture. Continuous sequencing efforts led to identification of a number of other GPCRs, some with known and others with unknown (“orphan”) functions. Although GPCRs have a similar 7TM structural architecture, they share very little sequence identity, making it difficult to classify these receptors from sequence alone. In 1994, the GPCRs were classified for the first time into six classes (A–F) based on sequence homology and functional similarity (Attwood & Findlay, 1994). Among these six classes, D and E represent fungal pheromone receptors and cAMP receptors, respectively, and are not found in humans. Based on comprehensive phylogenetic analysis of the human genome, the GRAFS classification was later introduced (Fredriksson *et al.*, 2003). This system grouped the mammalian GPCRs into five main

classes: *Glutamate* (Class C), *Rhodopsin* (Class A), *Adhesion* (Class B), *Frizzled* (Class F), and *Secretin* (Class B).

1.4.2. Structural Architecture Of GPCRs:

GPCRs share a common topology, having seven α -helical membrane spanning segments connected by alternating intracellular and extracellular loops (ECLs and ICLs) with an extracellular N-terminal tail and intracellular C-terminal tail (e.g. Figure 1.2). Despite having a similar topology, GPCRs share less than 20% sequence identity between classes in the TM domain (Fredriksson *et al.*, 2003). In humans, out of approximately 800 GPCRS, 683 belong to the *Rhodopsin* class. As the largest of the GPCR families, and because the GPCR targeted herein is a member of this family, the *Rhodopsin* class will be the focus henceforth. This family is divided into four subgroups, α - δ , that are further divided into sub-families. For the TM domain, sequence identity within each subgroup is $\geq 25\%$, and this increases to $\geq 30\%$ for sub-families; some sub-families are further subdivided into sub-types with sequence identity sometimes higher than 35% (Nordstrom *et al.*, 2011). Recent molecular modelling and docking efforts suggest that homology models based on structures of GPCRs having 35-40% sequence identity to the target may provide a sufficiently accurate structure for ligand docking studies. A minimum of 35% sequence identity has been suggested to be required to obtain a satisfactory homology model for the structure of a solved GPCR structure (Kufareva *et al.*, 2011). Based on these efforts, it is speculated that at least 100 carefully selected representative receptors (including at least one from each sub-type) will be needed to cover the entire family (Katritch *et al.*, 2013). Recent studies have highlighted new challenges in GPCR modelling arising from the involvement of highly non-

conserved ECLs in ligand binding (Kufareva *et al.*, 2014). In particular, the structure of nociceptin (opioid-like) receptor also highlights the fact that structural diversity despite sequence similarity may add significant complexity in the modelling of GPCRs. This is evident from the fact that, despite having 60% sequence identity with the μ -, δ - and κ -opioid receptors, the nociceptin receptor has very different ligand selectivity (Katritch *et al.*, 2013). Another example is the difference in the ligand-binding pocket and the mode of binding for the sphingosine 1-phosphate 1 (S1P₁) receptor and lysophosphatidic acid 1 (LPA₁) receptor, despite 41% sequence similarity (Chrencik *et al.*, 2015).

In the last 8 years, there has been an exponential growth in the number of solved human GPCR structures. Prior to this point, rhodopsin was the only GPCR for which 11 structures had been solved, with the first one solved in the year 2000. The first non-rhodopsin structure was solved in 2007 for β_2 -AR. To date, 30 X-ray crystal structures (Dore *et al.*, 2014; Srivastava *et al.*, 2014; Tautermann, 2014; Chrencik *et al.*, 2015; Yin *et al.*, 2015; Zhang *et al.*, 2015a; Zhang *et al.*, 2015b) from various classes have been solved and one solid-state NMR derived backbone structure of CXCR1 (class A) has been reported (Park *et al.*, 2012b) (Figure 1.3). The credit for such unprecedented success in solving GPCR structures is attributable to numerous achievements that have been made in the areas of protein expression, sample preparation and structural characterization techniques (Milic & Veprincev, 2015). The availability of numerous GPCR structures has made it possible to compare the common structural features and identify the hallmarks of this large family of potential drug targets.

Broadly speaking, GPCR structures comprise three elements: first, the extracellular domain comprising the N-terminal tail and three extracellular loops (ECL1-ECL3) connecting the TM segments and modulating ligand access in many cases;

second, the hallmark TM region with seven α -helical segments that commonly constitute the ligand binding pocket; and, third, the intracellular region comprising three intracellular loops (ICL1-ICL3) connecting the TM segments, an amphipathic eighth helix proximal to the seventh TM segment (TM7), and the disordered C-terminal tail. The distinctive features of each structural element will now be described in more detail.

With regard to sequence and length, the extracellular region is highly variable among different GPCRs. In the case of class A GPCRs, the extracellular loops are either arranged such that they occlude the ligand binding pocket or else they leave the ligand-binding pocket solvent accessible. The ECLs have presumably evolved structurally to modulate the entry of the ligand into the ligand-binding pocket in consideration of the nature of the ligand (Venkatakrisnan *et al.*, 2013). In comparison to the other ECLs, ECL2 is very interesting due to its long length and structural diversity (Wheatley *et al.*, 2012). ECL2 of rhodopsin, CXCR4, protease-activated receptor-1, and opioid receptors have a β -hairpin that is very differently positioned in each. Whereas, ECL2 of β_2 -AR, β_1 -AR have an α -helical region. In contrast, ECL2 of D₃ dopamine receptor, sphingosine-1-phosphate receptor-1, lysophosphatidic acid receptor-1, and histamine H₁ receptor lack any secondary structure. Interestingly, adenosine A_{2A} receptor contains both an α -helix and a short β -strand in ECL2. The presence of disulphide bridges between cysteines in ECL2 and the TM region (TM3) or other extracellular loops is another important structurally constraining feature of the extracellular region. These disulphide linkages provide receptor stability and limit conformational flexibility of a GPCR. Point mutations and deletions leading to disruption of these disulfide bonds often leads to misfolding of the receptor (Unal & Karnik, 2012).

The TM region of all GPCRs has a similar architecture arising from similar secondary structural elements (α -helices) bundled together. Despite the general similarity of 7-helix bundle structuring, variable tertiary contacts arise from the inherent sequence diversity from GPCR to GPCR. A recent analysis comparing the structural aspects of GPCRs has revealed key similarities and differences among diverse receptors (Venkatakrisnan *et al.*, 2013). A comparison of the structures of known GPCRs (the first 21 solved) revealed a consensus network of 24 noncovalent inter-TM contacts mediated by 36 topologically equivalent amino acids. 14 out of these 36 positions were found to be critical as mutating them was found to modulate receptor activity. These inter-TM tertiary contacts were largely located in the central and cytoplasmic end of the TM helical bundles. This analysis also suggests a central role for TM3 in maintaining tertiary structure, as it shares consensus helical packing interfaces with all other TM helices except TM1 and TM7.

Since TM1 and TM2 are the first TMs to be translated, consensus non-covalent contacts observed between them may be expected to have an important role in membrane insertion, folding, and topogenesis of GPCRs. This hypothesis has been at least partially supported by mutagenesis of the rat neurotensin receptor 1, where Schlinkmann and co-workers used a directed evolution method to study the effect of amino acid mutation on biosynthesis and stability of the receptor (Schlinkmann *et al.*, 2012). Their works convincingly showed how selective pressure can lead to shift in amino acid sequence and also sheds light on evolution of GPCR sequence in different organisms. Furthermore, under selective pressure they observed a shift of the most conserved residues of the N-terminal helices of the rat neurotensin receptor 1. Additional studies are required to fully assess the importance and functional roles of TM1 and TM2.

A highly conserved triplet of amino acids commonly referred to as the E/DRY or DRY motif is found near the cytoplasmic end of TM3 (Rovati *et al.*, 2007). This triplet plays a pivotal role in regulation of conformational states during GPCR activation. The arginine of the DRY motif has also been shown to be directly involved in two ionic interactions. The first of these is with the sidechain of the immediately N-terminal acidic residue, the second with a sidechain of the nearby residue from TM6 forming a salt bridge referred to as the “ionic lock”. Although the ionic lock is clearly observed in the inactive state of rhodopsin (Palczewski *et al.*, 2000), its presence is lacking in other solved GPCR structures. Molecular dynamic simulations have demonstrated that the ionic lock is dynamic in nature and may be disrupted by any alteration, suggesting that its absence in other GPCR structures may be due to the presence of T4-lysozyme fused into ICL3 to facilitate crystallization (Dror *et al.*, 2009; Miao *et al.*, 2014). Another conserved stretch of amino acids (NPxxY, where x is any amino acid) occurs near the cytoplasmic end of TM7. This is commonly referred to as the NPxxY motif. The tyrosine in this motif is very highly conserved (92%) among class A GPCRs and is known as a major “microswitch” due to the rotameric conformational changes that occur in its sidechain during GPCR activation (Katritch *et al.*, 2013).

The third structural element, the intracellular face, forms the interface for interaction with cytosolic signaling proteins such as G-proteins (Rasmussen *et al.*, 2011), GPCR kinases (GRKs) (Ribas *et al.*, 2007), and β -arrestins (Kang *et al.*, 2015). In currently available structures, ICL1 is very short and contains a helical turn, whereas the relatively longer ICL2 is either unstructured or has a short α -helical stretch. Notably, some residues from ICL2 have been found to be involved in non-covalent interactions that affect receptor stability (Warne *et al.*, 2008). In most of GPCRs, ICL3 and the C-

terminal tail are long and intrinsically disordered (Jaakola *et al.*, 2005; Venkatakrishnan *et al.*, 2014). These intrinsically disordered regions have been speculated to provide the receptor with the ability to interact favourably with a variety of intracellular proteins, enabling diverse downstream signaling (Babu *et al.*, 2012). Alongside this intrinsically disordered character, the C-terminal tails of almost all structurally characterized GPCRs have an α -helical domain proximal to the TM7, commonly referred to as the “eighth helix” (helix-8). The importance of the C-terminal tail has been highlighted in a recent structure of rhodopsin in complex with visual arrestin-1 (Kang *et al.*, 2015), where the N-terminus of helix-8 of rhodopsin has been shown to interact with the arrestin.

1.4.3. Signaling Through GPCRs:

Classically, ligand binding to a GPCR leads to conformational change, which enables the interaction with heterotrimeric G-proteins. Subsequent activation and dissociation of heterotrimeric G-proteins leads to the generation of second messengers and, correspondingly, the requisite cellular response to a given ligand. Ligand-induced conformational change in turn makes the receptor amenable to phosphorylation by G-protein coupled receptor kinases (GRKs) (Gainetdinov *et al.*, 2004). Phosphorylation of the GPCR leads to arrestin recruitment, which in turn leads to desensitization of G-protein-dependent signaling pathways (Gurevich & Gurevich, 2006). Arrestin recruitment, in turn, promotes receptor internalization by clathrin-dependent endocytosis (Kang *et al.*, 2014). Research in the last 15 years has demonstrated that GPCR activation may be generally considered to be bimodal, following either the G-protein-dependent pathway or the arrestin-mediated pathway for downstream signaling. Different ligands for the same receptor may vary in their ability to promote a specific signaling path, and

may offer bias for one path over the other (Luttrell & Kenakin, 2011). Correspondingly, allostery and biased-agonism have emerged as new areas in GPCR-mediated signaling (Gentry *et al.*, 2015; Luttrell *et al.*, 2015).

Another interesting aspect of GPCR signaling is the role of location in signal regulation via membrane trafficking. Following the ligand induced activation, many GPCRs internalize by Endocytic trafficking to end up in the lysosomes for degradation. Whereas, some GPCRs take a different route and get recycled back to the plasma membrane after the ligand-induced endocytosis. This recycling pathway can promote rapid recovery of cellular response by the GPCRs. In addition, some GPCRs remain intact during endocytosis and continue to signal, or initiate new signaling pathways, from the endosomal membrane. The non-covalent interactions that GPCRs experience during the receptor trafficking in the recycling pathways have been considered to be physiologically important and have been nicely reviewed in (Hanyaloglu & von Zastrow, 2008; West & Hanyaloglu, 2015).

1.5. The Apelinergic System:

The apelin receptor (AR), initially known as APJ, is a class A rhodopsin-like G-protein coupled receptor (GPCR) discovered in 1993 (O'Dowd *et al.*, 1993). It is a 380 residue protein (Figure 1.2) and was considered “orphan” until the ligand “apelin” (short for “APJ endogenous ligand”) was identified in 1998. Apelin is a hydrophilic and basic peptide present in various forms in the human body as a circulating hormone, an adipokine, and a pancreatic hormone (Tatemoto *et al.*, 1998). It is expressed as a 77-residue preproprotein which is converted into a 55 residue proprotein (proapelin) following removal of a 22-residue signal peptide from the N-terminus. Proapelin is

further cleaved into apelin-13, -17 and -36, bio-active peptides, all of which retain the C-terminus (Table 1.1) (Tatemoto *et al.*, 1998; Habata *et al.*, 1999; Hosoya *et al.*, 2000). It is not clear whether apelin acts on AR in an autocrine, paracrine, and/or hormonal manner, but most evidence supports the paracrine mode of action (Barnes *et al.*, 2010). The apelinergic system has been shown to have a role in regulation of fluid homeostasis (De Mota *et al.*, 2004; Roberts *et al.*, 2009), angiogenesis during tumour formation (Kunduzova *et al.*, 2008; Novakova *et al.*, 2015), adipoinular axis function (Kunduzova *et al.*, 2008; Dray *et al.*, 2010), regulation of the cardiovascular system, (Cao *et al.*, 2015; Folino *et al.*, 2015) and in the central nervous system (Cheng *et al.*, 2012). Human immunodeficiency virus type 1 (HIV-1) infects cells by binding of the viral envelope (Env) protein to CD4 and various co-receptors; AR has been identified as one of those receptors (Zhou *et al.*, 2003). All of these identified roles for the apelinergic system make it a very promising therapeutic target (Pitkin *et al.*, 2010).

Recently, another endogenous peptide ligand was identified for AR. This peptide, referred to as ELABELA, Toddler, or apela, is involved in zebrafish (*Danio rerio*) embryonic development (Chng *et al.*, 2013; Pauli *et al.*, 2014). Interestingly, apela and apelin activate the same receptor, and apela has been suggested to play a compensatory role in absence of apelin. Their exact mode of action is not yet clear and it still remains a mystery how the two ligands may work spatiotemporally. With growing interest in the field of biased agonism at GPCRs, apela and apelin have been suggested to be excellent candidates to study the role of biased agonism in drug discovery (Yang *et al.*, 2015).

1.6. The Divide And Conquer Approach To Study Membrane Proteins:

Although NMR spectroscopy is a very powerful technique for characterization of protein structure and dynamics, membrane protein characterization remains challenging (Sanders & Sonnichsen, 2006). One way to overcome some of obstacles in obtaining structural information for proteins, whether soluble or membrane proteins, is to divide the protein into smaller, tractable fragments and study each of these independently. The so-called “divide and conquer” approach is able to provide accurate structural information in advance of the availability of a full-length protein compatible with the structural characterization technique of choice.

In the case of membrane proteins, “dissected” TM segments are typically more readily obtained than larger protein segments. These are studied under the assumption that their folding is not compromised. This is based on hypothetical multi-stage membrane protein folding cycles (figure 1.4), where each TM segment of a polytopic protein inserts into the membrane and folds as a stable isolated unit prior to association with others (Cymer *et al.*, 2015). The “divide and conquer” approach has been extensively used for structural characterization of various membrane proteins (Bordag & Keller, 2010; Ruiz-Gomez *et al.*, 2010). It has been extremely successful, as demonstrate by work on cannabinoid receptors, both CB₁ (Tyukhtenko *et al.*, 2009) and CB₂ (Zhao *et al.*, 2006), where small fragments have been studied successfully. The success and potential of a fragment-based approach in structural characterization of GPCRs has been nicely reviewed in (Tikhonova & Costanzi, 2009; Cohen *et al.*, 2014).

The work by Yeagle and co-workers on bacteriorhodopsin provides the most compelling evidence in support of the divide and conquer approach. A set of 13 peptides

with overlapping primary sequence were synthesised and their structures were solved by solution-state NMR spectroscopy. The structure for all the independent peptides were solved by solubilizing the peptides in either DMSO, chloroform/methanol mixture, or detergent micelles. The average backbone RMSD for most of these peptides was 2.5 Å, when superposed onto the crystal structure. A three-dimensional structure for full-length rhodopsin was obtained using the peptide-based NMR restraints alongside longer-range experimentally obtained distance and angle constraints. The inter-helical distance constraints used in obtaining this structure were obtained from intermediate-resolution electron diffraction studies. The RMSD of the superposition of the thus obtained structure and the crystal structure was 2.9 Å in the helical region (Katragadda *et al.*, 2001a). Overall this work clearly demonstrates that reliable and valuable structural information can be obtained by studying small fragments of membrane proteins.

1.7. Membrane-Mimetics For Membrane Protein Structural Characterization:

The lipid compositions of biological membranes are very complex and vary between cell types, organelles, and organisms (Warschawski *et al.*, 2011). These membranes are composed of two monolayers of amphipathic lipids with the hydrophobic faces stacked against each other to form a bilayer. Specialized nano-domains known as “lipid rafts” are also believed to be present; these can form and dissipate rather quickly, with half-lives in the range of 10–20 ms (Nicolson, 2014). Due to the structural complexity of the biological membrane, it is necessary to reconstitute membrane proteins in membrane-mimetic environments containing lipids and detergents for structural characterization using biophysical techniques. However, the composition of these

membrane-mimetics may have both structural and functional ramifications and therefore must be chosen carefully (Zhou & Cross, 2013).

NMR spectroscopy is a unique tool that enables the study of both structural and dynamic aspects of proteins. To study membrane proteins by NMR, membrane-mimetics should be chosen to meet two criteria: first, they should closely mimic the natural environment of the protein to enable necessary interactions leading to proper protein folding to achieve its native state; and, second, they must also simultaneously allow for well-resolved NMR spectra. Beyond these factors, there are both biological and technical requirements that dictate the choice of an appropriate membrane-mimetic for structural characterization of membrane proteins by NMR. Membrane-mimetics are generally prepared using amphiphiles that have a tendency to self-associate in aqueous environment. Amphiphiles can either be classified as micelle forming surfactants or bilayer forming lipids. Although lipid bilayers are much closer to the native environment, their large size and long rotational correlation times make them unsuitable for solution-state NMR. However, bilayer-incorporated protein complexes are amenable to solid-state NMR and may provide a “native-like” environment for protein structural characterization.

Fast-tumbling micellar protein-surfactant complexes are the very suitable for characterization using solution-state NMR spectroscopy. Correlation times of under 50 ns provide suppressed orientation-dependent interactions, enabling acquisition of high-resolution NMR spectra in solution. Structural characterization of membrane proteins using NMR spectroscopy typically requires 2-D and 3-D experiments that are usually acquired for long periods of time. Therefore, sample stability and low conformational flexibility over long periods of time are key to obtaining satisfactory solution state-NMR

spectrum. Krueger-Koplin and co-workers evaluated a set of 25 detergents with a series of model membrane proteins to determine relative amenability of each for structural characterization of membrane proteins using solution-state NMR (Krueger-Koplin *et al.*, 2004). Their work demonstrated that LPPG and SDS micellar conditions tended to provide consistently superior NMR spectral quality vs. other micellar conditions. However, conformational heterogeneity observed in SDS made it less suitable in comparison to LPPG for the model proteins used in that work.

Organic solvent mixtures are another example of a membrane-mimetic environment that have been used for biophysical analysis of hydrophobic peptides, due to their spectroscopic properties which are ideal for circular dichroism and solution-state NMR (Goodman *et al.*, 1971; Nelson & Kallenbach, 1989; Bruch & Gierasch, 1990; Koppke *et al.*, 1993; Buck, 1998). In comparison to a membrane protein fragment embedded in micelles or bicelles, an organic solvent mixture solubilized membrane protein sample has a superior spectrum due to smaller correlation. Furthermore, there is evidence in literature where organic solvent mixtures have been used to obtain structures that appear to assume folded structures that may be relevant to the biologically active state (Girvin *et al.*, 1998; Schwaiger *et al.*, 1998; Penin *et al.*, 2004; Estephan *et al.*, 2005; Ma *et al.*, 2005). The specific membrane-mimetics used study AR fragments in this work have been further discussed in Chapter 3.

1.8. Introduction To Biophysical Techniques:

The ultimate goal of this work is to characterize AR structure and dynamics. The following sections outline the two primary biophysical techniques that are used extensively in this work.

1.8.1. Circular Dichroism (CD) Spectroscopy:

CD spectroscopy is an excellent technique for the identification and quantification of protein secondary structure. Determination of secondary structure by this method is based on the fact that different secondary structures exhibit differences in the molar absorptivity for left-handed vs. right-handed circularly polarized light. In CD spectroscopy, the difference in absorbance of right-handed and left-handed circularly polarized light (evident through production of elliptical polarization, reported as “ellipticity”) by an optically active centre is measured (Greenfield, 2006). Different secondary structural elements in a protein give rise to characteristic far-ultraviolet (far-UV; 180-250 nm) CD spectral behaviour. This is due to the splitting or shifting of optical transitions as a result of “exciton” interactions between peptide bond chromophores (Sreerama & Woody, 2004). For an α -helix, exciton splitting gives rise to bands of negative ellipticity at 222 and 208 nm and positive ellipticity at 193 nm (Holzwarth & Doty, 1965). β -sheets give rise to diverse spectra which is a manifestation of variations in length, extent, direction, and degree of twist and distortion in the β -sheets. The parallel and anti-parallel nature of these β -sheets can also affect the spectrum. Lower spectral amplitudes also complicate the identification of the specific type of β -sheet present in the protein (Micsonai *et al.*, 2015). In general, β -sheets have a band of negative ellipticity at 218 nm and a band of positive ellipticity at 195 nm. Disordered protein segments give rise to minimal ellipticity above 210 nm and negative ellipticity near 195 nm (Greenfield & Fasman, 1969). Importantly, the CD phenomenon is additive, meaning that the total ellipticity at a given wavelength will be given by the linear combination of the relative contributions from each secondary structural element. CD is an excellent technique for good estimation of protein secondary structure content, both in protein concentration-

dependent (Whitmore & Wallace, 2008) and -independent manners (Raussens *et al.*, 2003; McPhie, 2008). CD also can be used to assess the effects of physical properties of the environment on the secondary structure.

1.8.2. NMR Spectroscopy For Structural Characterization Of Proteins:

NMR spectroscopy is a biophysical technique that can be used for protein structure elucidation by exploiting the magnetic properties of its nuclei. For a nucleus to be NMR-active (i.e. observable in NMR experiments), it must possess nuclear spin, a form of angular momentum denoted by the quantum number I . Similar to other quantum numbers, the spin quantum number is quantized, and exists as an integer (0, 1, 2, ...) or as a half-integer (1/2, 3/2, 5/2, ...). In presence of a magnetic field, the Zeeman-effect causes displacement from degeneracy leading to splitting of the nuclear spin state I into $2I + 1$ sub-states. Transitions between these sub-states are probed in NMR spectroscopy (Levitt, 2008). Nuclear isotopes without nuclear spin (i.e., with $I=0$) such as ^{12}C , ^{16}O or ^{32}S do not display Zeeman splitting, and therefore are not amenable to NMR. Nuclear isotopes with spin $I=1/2$ (^1H , ^{13}C , ^{15}N , ^{31}P , ^{19}F , ...) are of particular importance in solution-state NMR, due to their easy to understand behavior in presence of external magnetic field. Isotopes with spin $I>1/2$ (^2H , ^{14}N , ...) are known as quadrupolar nuclei, and are difficult to utilize for high resolution structure elucidation by NMR.

Nuclei with non-zero spin angular momentum in a magnetic field (\mathbf{B}_0 ; a vector quantity) have a magnetic moment (μ) proportional to the spin angular momentum quantum number, and the proportionality constant is called the gyromagnetic ratio (γ). This induced magnetic moment is also proportional to \mathbf{B}_0 in the following manner:

$$\mu = \mu_0^{-1} V \chi \mathbf{B}_0 \quad 1.1$$

where $\mu_0^{-1} = 4\pi * 10^{-7}$ is a magnetic constant or vacuum permeability, V is the volume of the object, and χ is the magnetic susceptibility. When an ensemble of nuclear spins is not in a magnetic field, it has no net magnetization due to random orientation of the magnetic moments of each individual nucleus. Conversely, when a sample is in a magnetic field, the interaction between the magnetic moment and the applied field (\mathbf{B}_0) leads to the alignment of the magnetic dipole of each nucleus of the sample in the direction of the magnetic field. This leads to development of bulk magnetization (net magnetization for the sample as a whole) along the direction of the \mathbf{B}_0 . Build-up of this bulk-magnetization vector is an equilibration process, and is parallel (anti-parallel for nucleus with negative γ) to the \mathbf{B}_0 . This bulk-magnetization vector can be tipped away from direction of the \mathbf{B}_0 field by application of a radiofrequency pulse, as is done in an NMR experiment, leading to precession of the bulk-magnetization vector about \mathbf{B}_0 . The frequency of this precession is known as the Larmor frequency, given as either:

$$\omega_0 = -\gamma\mathbf{B}_0 \text{ in rad s}^{-1} \quad 1.2$$

or

$$\vartheta = -\gamma\mathbf{B}_0/2\pi \text{ in Hz} \quad 1.3$$

Here, γ dictates the Larmor frequency of each nucleus in a given magnetic field. At this point, it is important to note that the actual magnetic field experienced by a nucleus may vary, depending upon the environment around it. For example, the local magnetic field is attenuated (shielded) by the presence of electrons around the nucleus. The degree of shielding or deshielding experienced gives rise to an effective magnetic field (\mathbf{B}_{eff}) for a given nucleus that can be represented by:

$$\mathbf{B}_{\text{eff}} = (1 - \sigma)\mathbf{B}_0 \quad 1.4$$

where σ represents the degree of shielding. For this reason, each nucleus in a sample will have a defining Larmor frequency that depends upon the chemical and magnetic environment on the magnetic field experienced by it. This makes NMR spectroscopy an excellent technique to gauge changes in environment around any NMR-active nucleus through modifications to its Larmor frequency. As is clear from equations 1.1 and 1.2, Larmor frequency is dependent on B_0 . Therefore, in practice, to eliminate the effect of B_0 , Larmor frequency is represented as chemical shift (δ). Chemical shift is defined as the frequency of the resonance of a nucleus expressed with reference to a standard compound. Since these frequencies are of the order of 10^{-6} , they are reported in parts per million (ppm).

Assignment of chemical shifts to each of the specific nuclei in a molecule is the first step in characterization of molecules by NMR. In the case of small molecules, the number of NMR-active nuclei is small, and therefore it becomes easy to assign the chemical shifts to the specific nucleus using one-dimensional (1D) and, in some instances, two-dimensional (2D) experiments. Conversely, in the case of proteins, there is a large number of a very sensitive nucleus, ^1H , which leads to overlap of proton chemical shifts. This overlapping of the chemical shifts leads to ambiguity during assignment. This problem gets worse with increasing size of the protein. One solution to this problem is to increase the dimensionality of the experiments, where each peak position is defined by two or more resonance frequencies or chemical shifts. Use of 2D and three-dimensional (3D) NMR experiments has become fairly common for structural characterization of proteins. Examples of 1D, 2D and 3D NMR spectra are shown in figure 1.5. In typical multi-dimensional NMR experiments, each cross-peak arises from correlations between multiple nuclei that are coupled to one another either by scalar (J) coupling (through

bond), by dipolar coupling (through space), or through a combination of J-coupling and dipolar coupling. In multi-dimensional NMR experiments, magnetization is transferred from one nucleus to the other based on these couplings; depending on the types of nuclei between which this transfer happens, experiments may be classified as either homonuclear (same atomic nuclei) or heteronuclear (different atomic nuclei) (Rule & Hitchens, 2006).

Proteins have a low natural abundance of NMR-active nuclei other than ^1H . Thus, enrichment of proteins with NMR-active isotopes (^{15}N and ^{13}C) is generally necessary for performing heteronuclear NMR experiments, particularly if correlations between more than one non- ^1H nucleus are targeted. For structural characterization of a protein by NMR, the first goal is to assign the resonance for main-chain atoms and side-chain atoms. This can be done using either homonuclear or heteronuclear NMR experiments. For proteins less than 10 kDa, homonuclear NMR experiments may be sufficient to assign all chemical shifts, but for larger proteins heteronuclear NMR experiments are generally required for structural studies (Rule & Hitchens, 2006).

In this work I have used triple-resonance experiments involving magnetization transfer between ^1H , ^{15}N and/or ^{13}C nuclei for assignment of main-chain (or polypeptide backbone) chemical shifts. In this “backbone walk” approach, the first step is to assign the ^1H , ^{13}C and ^{15}N resonances for all amino acid residues in the protein. This is based on correlation of multiple multi-dimensional experiments HSQC (Palmer *et al.*, 1991; Kay *et al.*, 1992; Schleucher *et al.*, 1994), HNC0 (Yang & Kay, 1999), HN(CA)CO (Yamazaki *et al.*, 1994a), HNCA (Yamazaki *et al.*, 1994b), HN(CO)CA (Yamazaki *et al.*, 1994a), HNCACB (Wittekind & Mueller, 1993) to each other, allowing the assignment of each chemical shift. These experiments all rely upon cross-peak correlations arising

from J-coupled nuclei in neighbouring residues, allowing one to “walk” along the backbone during the assignment process (described in next section).

The next step is to assign the sidechains, which can be done using experiments like ^{13}C -/ ^{15}N - filtered/edited TOCSY-HSQC (Cavanagh, 2007) or the HCCH-TOCSY (Bax *et al.*, 1990) which provide cross-peaks arising from J-coupled nuclei contained within a single amino acid (or “spin system”). The final step of data analysis is to develop the constraints for structural calculation. Distance constraints are determined using experiments like the 2-D ^1H - ^1H NOESY and the 3-D ^{15}N -edited NOESY-HSQC and ^{13}C -edited NOESY-HSQC (Cavanagh, 2007). In all of the NOESY experiments, cross-peaks arise only for ^1H nuclei in close spatial proximity ($< \sim 6 \text{ \AA}$). Chemical shift assignments may also be used to provide secondary structure constraints for structural calculation based upon the observation that chemical shifts are perturbed from values observed in a random coil upon formation of either a β -strand or α -helix (Wishart & Sykes, 1994a). The restraints thus obtained are used as inputs for simulated annealing calculations in programs such as XPLOR-NIH (Schwieters *et al.*, 2003; Schwieters *et al.*, 2006) and CYANA (Guntert, 2004) to generate an ensemble of 3-D structures of the protein. The following web sources are excellent references for the assignment process and the experiments used in assignment (<http://www.ccpn.ac.uk/software/tutorials/analysis-beginners>; <http://www.protein-nmr.org.uk/solution-nmr/assignment-theory/>).

1.8.3. General Method For Sequential Backbone Assignment Using 3D Data:

In this work, I have used a set of 6 triple-resonance experiments, namely HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCANH, and CBCA(CO)NH. These 6 experiments represent 3 pairs (HNCA with HN(CO)CA; HNCO with HN(CA)CO; CBCANH with

CBCA(CO)NH) of experiments, where each pair can be used to walk along the backbone of a polypeptide chain. For example, the HNCA and HN(CO)CA pair can be used to walk along the backbone while assigning the $C\alpha$ and the NH for each residue except proline. This pair of experiments links each NH(i) group with the $C\alpha$ (i-1) (HN(CO)CA) or with $C\alpha$ (i) and $C\alpha$ (i-1) (HNCA). The peak representing the $C\alpha$ (i-1) can be linked to the NH (i-1) or $C\alpha$ (i) can be linked to NH (i+1). This way it is possible to walk in both the non-proline containing directions on the polypeptide backbone. In a similar fashion, both the HNCO, HN(CA)CO and CBCANH, CBCA(CO)NH pairs can also be used to walk along the backbone while assigning C' or $C\alpha$ and $C\beta$, respectively. A typical backbone walk approach using the HNCA, HN(CO)CA, HNCO, and HN(CA)CO spectra is shown in Figure 1.6. It should be noted that, in comparison to C' , using $C\alpha$ and $C\beta$ chemical shifts during assignment is advantageous because of very distinct paired values for certain amino acids, such as alanine, serine, threonine and glycine. Valine, isoleucine and proline also exhibit unusual, lower than typical $C\alpha$ chemical shifts. The $C\beta$ chemical shift information available in CBCANH and CBCA(CO)NH pair makes them very helpful 3D experiments in terms of the straightforwardness they allow in distinguishing between amino acids while assigning. For all the work in this thesis, the CcpNmr Analysis suite (Vranken *et al.*, 2005) has been used for resonance assignments. A descriptive tutorial for chemical assignment of the protein backbone using different NMR spectra by CcpNmr analysis can be found at (<http://www.ccpn.ac.uk/software/tutorials/analysis-beginners>).

1.8.4. Study Of Protein Dynamics By NMR:

Time dependent fluctuation in protein structure referred to as protein dynamics is a very important phenomenon that enables proteins to perform a multitude of functions. These fluctuations may occur on variable time scales and can be effectively studied using NMR spectroscopy by studying relaxation of nuclear spins. Observables such as R_1 (longitudinal relaxation rate constant; $R_1=1/T_1$), R_2 (transverse relaxation rate constant; $R_2=1/T_2$), and heteronuclear NOE (hetNOE) can be effectively used to study motion on the ps-ns time scale (Farrow *et al.*, 1994). These observables can be measured site-specifically for ^{15}N , ^{13}C and ^2H on the protein backbone and/or side chains for analysis of motion. Briefly, the hetNOE is measured by taking the ratio of peak intensity or peak volume in the saturated spectrum to those in the unsaturated spectrum, whereas R_1 and R_2 measurements utilize a series of spectra that quantify the time dependent buildup or decay of magnetization through longitudinal and transverse relaxation phenomena, respectively.

Interpretation of R_1 , R_2 , and hetNOE relaxation data can be carried out in a number of different ways. In this work, I have used the simplest of all - the site specific phenomenological method to interpret hetNOE. This method does not require any further processing of the acquired data and stays closest to the observed data set. In general, the enhancement factor of the hetNOE is reduced for spin-pairs located in regions of internal flexibility, making its local variation convenient to assign flexibility or rigidity to a specific region of the protein. The value on ^1H - ^{15}N hetNOE can range from -4 to +1, with the negative sign being a manifestation of the negative γ for ^{15}N . Various aspects of protein dynamics characterization by NMR are nicely reviewed in (Jarymowycz & Stone, 2006; Kleckner & Foster, 2011).

1.9. Overview:

This thesis details documentation of the hardships faced and the success achieved in studying a class A GPCR known as the apelin receptor. Nevertheless, in September of 2010, when I started this work, although structures of a few relatively larger membrane proteins had been elucidated, characterization of only one (Chopra *et al.*, 2000; Xie *et al.*, 2000; Katragadda *et al.*, 2001b; Tiburu *et al.*, 2009) or two (Zheng *et al.*, 2006; Zou *et al.*, 2008; Neumoin *et al.*, 2009) TM domains of GPCRs had been done by NMR. This was due to both challenges in production of isotope-enriched larger fragments and technical challenges involved in characterization of larger fragments of membrane proteins with NMR. At this stage, the N-terminal tail and TM1 (AR55) had already been successfully cloned, expressed, and characterized by a former graduate student (now published in (Langelaan *et al.*, 2013)). The goal of my project was to produce the remaining receptor fragments and to use inteins to put them together for structural characterization. In this thesis, I have separated my efforts into chapters based on various aspects of the project; Chapter 2 covers the strategies tested and the success achieved in production of various fragments of AR in *E. coli*; Chapter 3 details my biophysical characterization of the first three TM domains of AR (AR TM1-3) using CD and NMR; Chapter 4 covers the progress made in expression of AR in two pieces with split-inteins using an *E. coli*-based cell-free expression system; Chapter 5 covers the identification and biophysical characterization of helix-8 in the C-terminal tail of AR, and its involvement in anterograde transport of the receptor. This thesis not only paves the path to characterization of the structure and dynamics of AR and of AR-ligand interactions at

an atomic-level using NMR, but also sheds light on a structural element involved in GPCR anterograde trafficking.

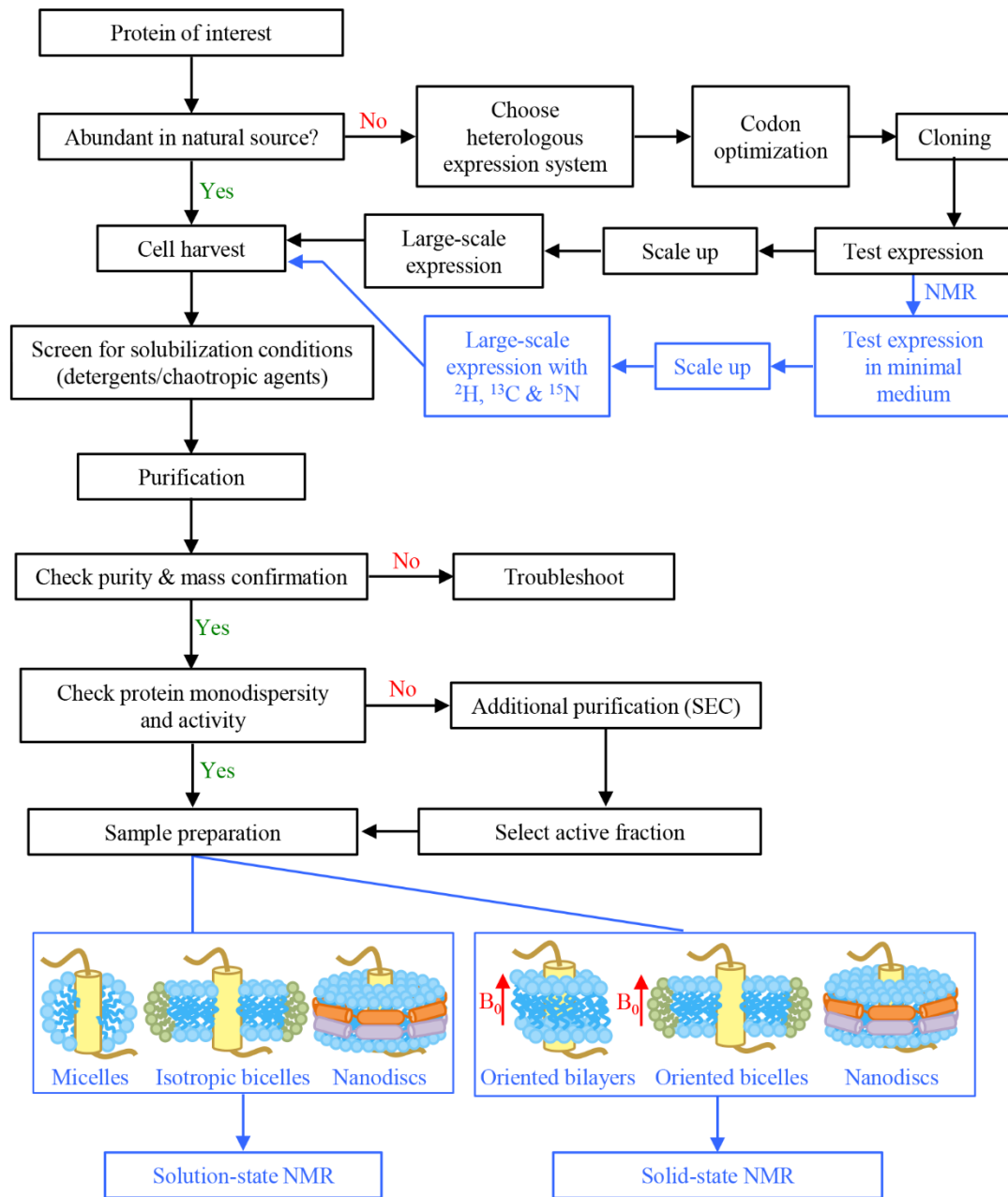


Figure 1.1: Flowchart outlining various optimization steps allowing for structural characterization of membrane proteins by NMR spectroscopy.

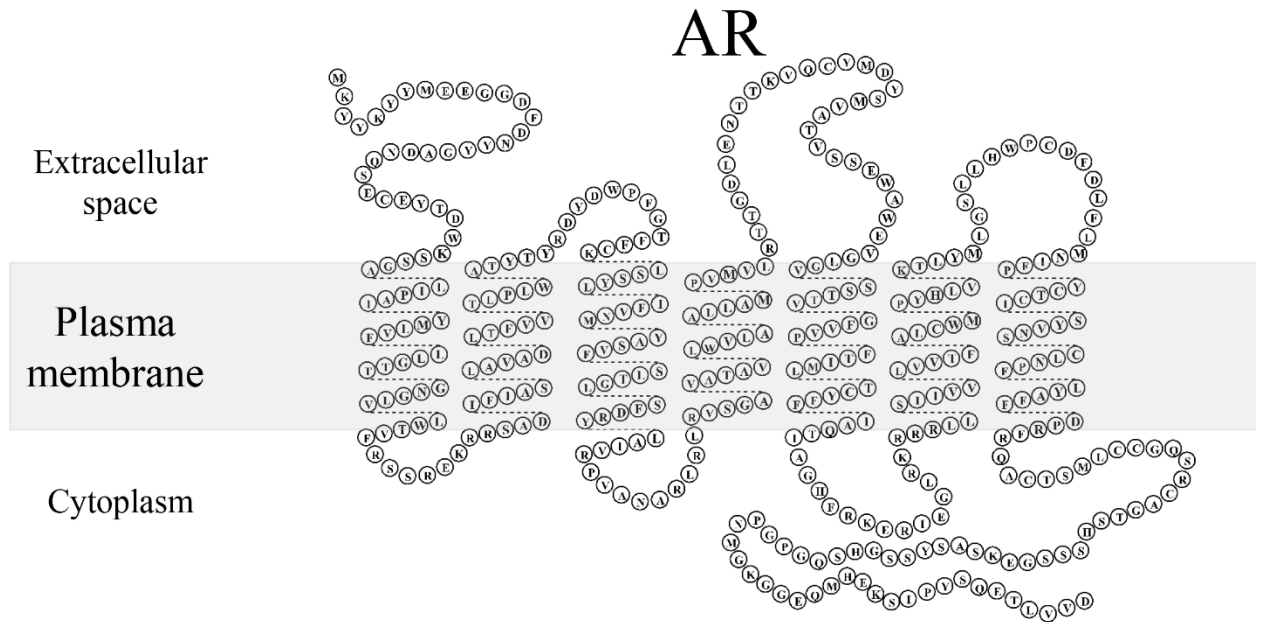


Figure 1.2 Snake plot for the apelin receptor representing its sequence and topology.

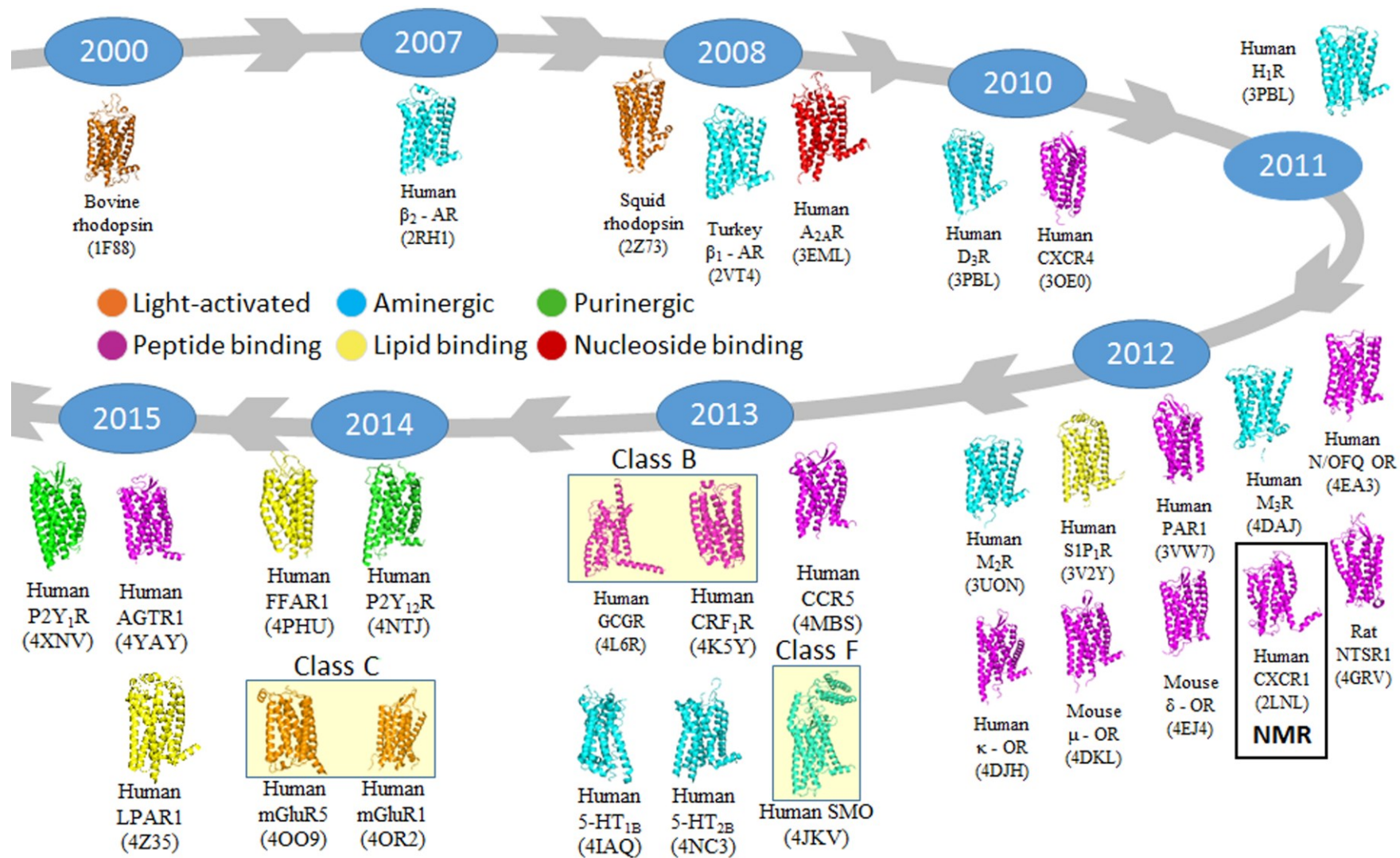


Figure 1.3: Timeline of solved GPCR structures.

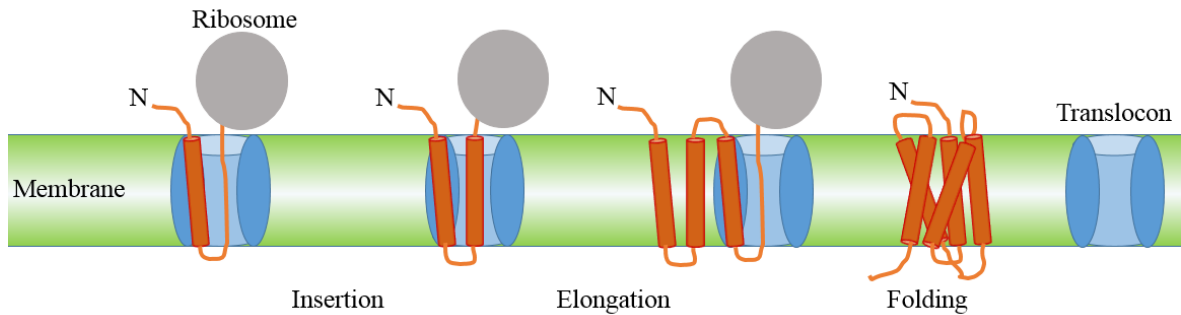


Figure 1.4: Cartoon representation of the current understanding of membrane protein folding. The figure represents the current understanding of the process of membrane protein folding, where the nascent protein from the ribosome is first delivered to the translocon on the ER membrane. The translocon then assists in insertion of the first transmembrane domain of the protein in the right topology. Subsequently, the remaining transmembrane domains are also inserted in a similar fashion. Finally, after the translation of the protein is complete, the transmembrane domains interact with each other to acquire an energetically favourable 3-dimensional geometry. (The figure is adapted from (Cymer *et al.*, 2015))

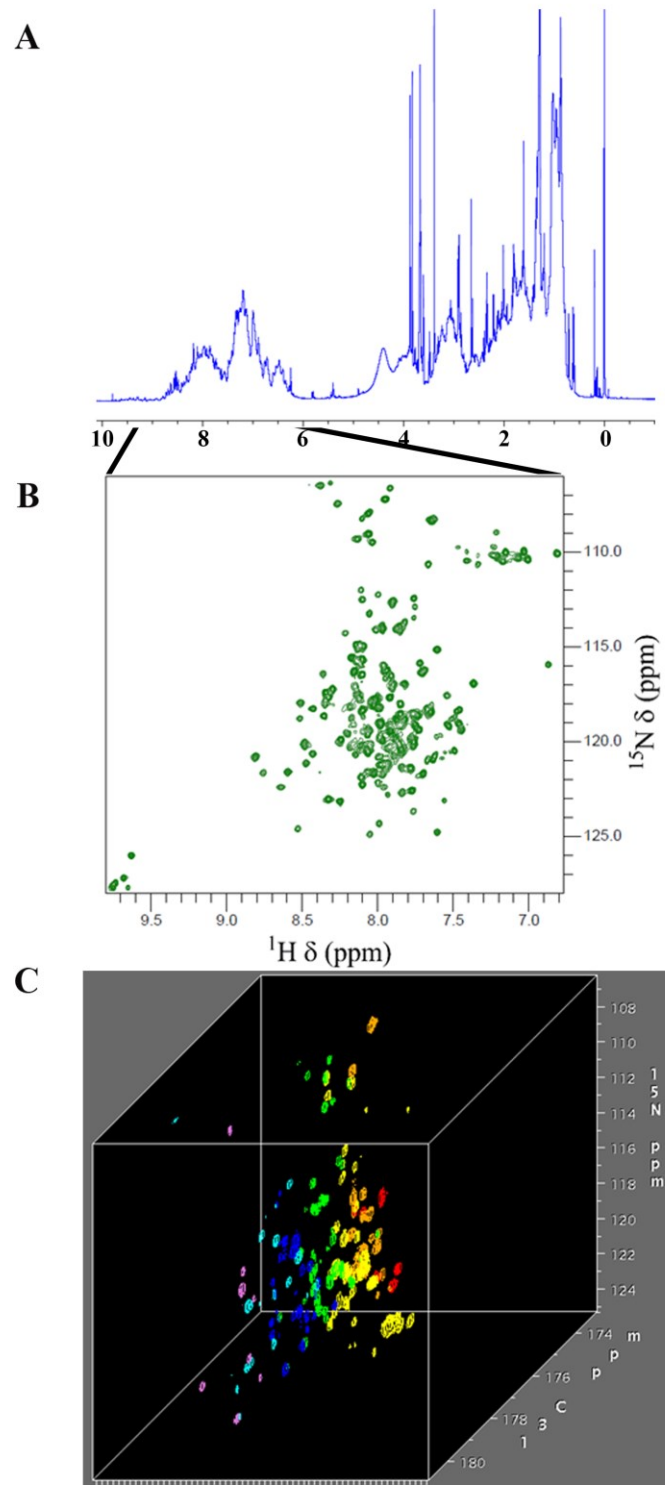
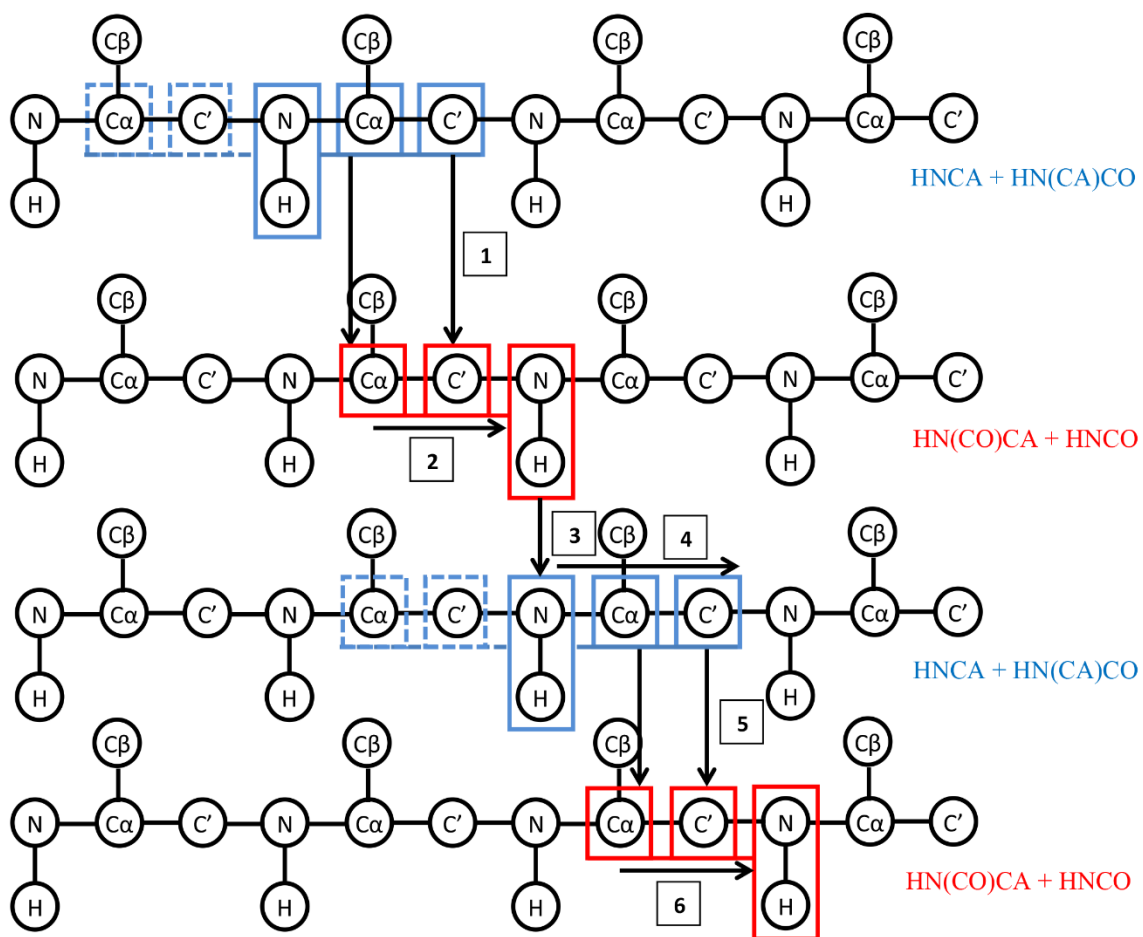


Figure 1.5: Illustration of 1D vs. 2D vs. 3D spectra for AR TM1-3 in 50% TFE. The figure shows a ^1H 1D (A), ^1H - ^{15}N (HSQC) 2D (B), and ^1H - ^{15}N - ^{13}C (HNCO) 3D (C) spectra for AR TM1-3 in 50% TFE condition discussed in chapter 3.



Using HNCA and HN(CA)CO experiments $C\alpha$ and C' (carbonyl carbon) are assigned for residue 'i'.

- 1- Identical $C\alpha$ and C' are found in HN(CO)CA and HNCO.
- 2- A new NH group is assigned for the residue 'i+1' using HN(CO)CA and HNCO.
- 3- New NH is matched in the HNCA and HN(CA)CO.
- 4- New $C\alpha$ and C' are assigned for 'i+1' residue corresponding to the new NH.
- 5- This step is the same as step 1 but for $C\alpha$ and C' of residue 'i+1'.
- 6- This step is same as step 2 and helps assigning NH group of residue 'i+2'.

Figure 1.6: A general scheme for the “backbone walk”.

Chapter 2: Use Of AT-Rich Gene Tags To Enhance Production Of AR TM1-3

Note: *Parts of this chapter have been published as a manuscript titled “Small expression tags enhance bacterial expression of the first three transmembrane segments of the apelin receptor” in **Biochemistry and Cell Biology** 92(4): 269-278.*

2.1. Introduction:

As mentioned earlier, multi-milligram quantities of a protein are typically required for structural characterization using experimental techniques such as X-ray crystallography and NMR spectroscopy. The amount of a GPCR that is produced in native tissues is very limited; therefore, methodologies to overexpress these proteins are typically required (Zhao & Wu, 2012). Furthermore, studies of proteins of ~100 amino acids or larger by NMR spectroscopy typically rely upon heteronuclear experiments relying upon enrichment in ^{15}N and ^{13}C nuclei at levels well above natural abundance (Cavanagh *et al.*, 2007). Among the five conventional expression systems (*Escherichia coli*, yeast, insect cells, mammalian cells and cell-free expression systems), expression in *E. coli* is the most economically feasible method to obtain an isotope-enriched protein suitable for structural characterization by NMR spectroscopy. However, obtaining a sufficient quantity of a target protein is typically a bottleneck in structural characterization of GPCRs, or any other class of membrane proteins, using NMR spectroscopy (Gautier *et al.*, 2010b). As a result of the difficulties in expression of isotope-labeled full-length GPCRs, fragments of receptors ranging in size from 20-120 amino acids have been studied (Zhao *et al.*, 2006; Zou *et al.*, 2008; Tikhonova & Costanzi, 2009; Tyukhtenko *et al.*, 2009; Tiburu *et al.*, 2011; Langelaan *et al.*, 2013).

Ste2p, a yeast GPCR, is the only example where larger fragments (4 and 5 transmembrane (TM) segments) have been reported for structural characterization by NMR spectroscopy (Potetinova *et al.*, 2012; Cohen *et al.*, 2014). Recently the structure of the first three transmembrane segments containing fragment (131 amino acids) has also been solved for this yeast GPCR (Fracchiolla *et al.*, 2015).

As discussed in Chapter 1, the so-called “divide and conquer” approach states that an isolated TM helix represents a distinct folding subunit that can be separated from its parent protein without losing its inherent structural characteristics (Bordag & Keller, 2010). Beyond general applicability to membrane proteins when judiciously applied (Bordag & Keller, 2010), this approach has been extensively used for studies of GPCRs (Cohen *et al.*, 2014). In accordance with this approach, the AR was divided into 4 fragments (AR N-terminal tail + TM1 (AR55), AR TM2-3, AR TM4-5, and AR TM6-7) (Figure 2.1).

Characterization of the structure and dynamics of AR55 (the 55 N-terminal amino acids of the AR) in a variety of membrane-mimetic environments were carried out previously in our lab. That set of studies has provided the only high-resolution structural information available for the AR (Langelaan *et al.*, 2013), shedding light on the positioning of N-terminal amino acids previously reported to be involved in apelin binding and activation. To supplement the existing structural information on the AR, I set out to produce the remaining fragments of the receptor. A few unsuccessful attempts to express the other fragments of the AR had already been made by previous students in the lab.

One of the most common strategies to solve the problem of protein expression is to fuse a proteinaceous tag to the protein of interest (POI) (Sorensen & Mortensen, 2005;

Walls & Loughran, 2011). Adding a tag not only has a potential to improve the yield of the target proteins in *E. coli* (Kefala *et al.*, 2007; Leviatan *et al.*, 2010), but can also increase solubility (Butt *et al.*, 2005; Marblestone *et al.*, 2006) to improve handling during subsequent purification and/or to enhance purification through affinity chromatography methods (Hochuli *et al.*, 1988). In the case of membrane proteins, targeted localization of an aggregate-prone and hydrophobic POI to the expression medium or periplasm has also been shown to improve yield (Better *et al.*, 1993). Another strategy to increase the yield of membrane proteins and other difficult-to-express proteins is to specifically target them to inclusion bodies using insoluble fusion partners. Targeting proteins to inclusion bodies shields them from proteolytic degradation, consequently increasing the yield for the POI (Hwang *et al.*, 2014). The benefits of fusion tags may potentially be both additive and complementary, meaning that they are often fused together in various arrangements to improve yield (Pryor & Leiting, 1997; Costa *et al.*, 2013; Liu *et al.*, 2015b), including for isotope-enriched membrane proteins for NMR (Kohno *et al.*, 1998; Leviatan *et al.*, 2010). Given the difficulties in expressing mammalian (or, indeed, non-mammalian) membrane proteins in heterologous systems such as *E. coli*, use of fusion tags is usually essential in producing sufficient quantities for structural studies.

Although fusion tags have many advantages, the design of fusion tag-POI constructs demand careful planning with consideration of the exact construct in question. One property of the tag itself that calls for careful selection is its size. Larger tags tend to impose a heavy metabolic burden on the host system, which increases with the complexity of the amino acid composition. Use of large tags can also lead to interactions between the fusion tag and the POI, which may also interfere with structure and/or

activity of the POI (Majtan & Kraus, 2012; Sabaty *et al.*, 2013; Singh *et al.*, 2013). Therefore, removal of these tags is also important. This leads to an additional layer of complexity in recombinant protein production and is often the bottleneck in such a procedure. The efficiency of tag removal can vary dramatically between different target proteins. For example, steric hindrance or unfavourable residues at cleavage sites can hamper protease-mediated tag removal. This may be circumvented by use of a different protease with different active site requirements; lengthening of the linker region; use of a smaller tag; or, through use of chemical cleavage methods involving reagents such as cyanogen bromide (CNBr) or 2-nitro-5-thiocyanatobenzoic acid (NTCB), which cleave C-terminal to Met and Cys, respectively (Crimmins *et al.*, 2005). In addition, different methods present different difficulties. For example, off-target cleavage can occur from proteases that recognize specific short amino acid motifs. Some proteases also leave residual amino acids on the target protein, thus increasing the protein size and, potentially, perturbing function. The removal of a tag can also yield improperly folded or disrupted target protein structures due to protein aggregation and precipitation under cleavage conditions. Thus, cleavage conditions require optimization through trial and error and can be highly time-consuming.

In this work, I have used Maltose-Binding Protein (MBP) (Hu *et al.*, 2011), Small Ubiquitin-like Modifier (SUMO) (Marblestone *et al.*, 2006), B1 domain of streptococcal protein G (GB1) (Kumar *et al.*, 2012), and Glutathione-S-Transferase (GST) (Park *et al.*, 2012a) as fusion partners to enable the expression of AR TM2-3 and AR TM4-5. The rationale behind using these tags was that they had been previously shown to facilitate membrane protein expression. None of these constructs successfully yielded appropriate quantity of pure protein. Issues with each construct have been discussed in the results

section.

At this point, the strategy of dividing the receptor into four parts was changed to dividing the receptor into two parts either (AR TM1-3 and AR TM4-7) or (AR TM1-4 and AR TM5-7). There were two reasons to do so; The first reason was the inefficient production of the smaller constructs. Since, AR TM1 had already been expressed by another graduate student without any tag, we thought adding it might facilitate the expression of of AR TM2-3 by restoring the native topology. This would not only provide us the means to express the protein but also enable us to study the interaction of AR TM1 with AR TM2 and AR TM3. The second reason was the incentive that we would gain in our efforts to produce a segmentally labeled full-length receptor. Expression of the receptor as two large fragments would eliminate the requirement for more than one split-intein pair to obtain full-length receptor (discussed in Chapter 4). Based on this, I sub-cloned AR TM1-3 and tested the expression. The result was rather disheartening as the protein did not express.

One day while going through the literature, I came across the work by Haberstock *et al.* (2012), which showed that having AT-rich codons next to the translation initiation site had a significant positive effect on translation efficiency in an *E. coli* extract-based cell-free expression system. It seemed reasonable to speculate a similar effect in *E. coli* cells. In *E. coli*, the nucleotide sequence around the ribosome binding site has been shown to play a critical role in the regulation of the amount of mRNA and protein (Kudla *et al.*, 2009). This may be related to the fact that an increase in translation efficiency has been shown to be directly related to better initiation of translation conferred by AT-rich codons (Qing *et al.*, 2003). Consequently, I made various constructs of AR TM1-3 with

the nucleotide tags rich in adenine and thymine base pairs (referred to hereafter as AT-rich gene tags).

In this chapter, I report the expression and purification of the first three TM segments (amino acids 1-137; Figure 2.2) of the AR. Expression was achieved in the C41(DE3) strain (Miroux & Walker, 1996) of *Escherichia coli* (*E. coli*) using AT-rich gene tags previously reported to enhance expression in a cell-free expression system (Haberstock *et al.*, 2012). Notably, inclusion of these tags extends the length of the protein by only 4-6 amino acids, a relatively minimal increase in length in comparison to most of the other tags tested. Through use of these expression tags, we achieved final purified protein yields of approximately 8 mg/L following expression in minimal growth medium suitable for stable NMR-active isotope enrichment.

2.2. Materials And Methods:

2.2.1. Materials:

The bacterial codon-optimized gene for the full-length AR (protein sequence in Figure 2.1) and DNA primers for cloning were obtained from BioBasic Inc. (Markham, ON). pMALc2x vector and amylose resin were purchased from New England Biolabs (Ipswich, MA). The pETHN vector was created by an earlier graduate student in the laboratory of Dr. Paul Liu by sub-cloning the 6xHis-SUMO protein sequence followed by a SapI restriction site into a pET-32 vector purchased from Novagen (Darmstadt, Germany). The pEGX vector was purchased from GE healthcare (Little Chalfont, UK). The pET-23b (+) vector was obtained from Novagen (Darmstadt, Germany). BL21(DE3) and C41(DE3) strain of *E. coli* (Miroux & Walker, 1996) was purchased from Lucigen

(Middletown, WI). Reagents for polymerase chain reaction (PCR) amplification of DNA were purchased from Invitrogen (Burlington, ON). Acetonitrile (high performance liquid chromatography (HPLC) grade), 2-propanol, trifluoroacetic acid (TFA), isopropyl β -D-1-thiogalactopyranoside (IPTG), dithiothreitol (DTT), ethylenediaminetetraacetic acid disodium salt dehydrate (EDTA- Na_2), ampicillin, and reagents for making lysogeny broth (LB) medium and 1x loading dye for SDS-PAGE were purchased from Fisher Scientific (Ottawa, ON). 15% Next gel solution for SDS-PAGE was purchased from Amresco (Solon, OH). Uniformly ^{15}N -labeled ammonium chloride and ^{13}C labeled D-glucose were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Columns for reverse phase (RP) HPLC were purchased from Agilent Technologies Canada (Mississauga, ON). Nitrocellulose membrane (0.2 μm) and Clarity Western ECL Substrate for western blotting were purchased from Bio-Rad (Mississauga, ON). Factor Xa protease and HisProbe-HRP was purchased from Thermo Fisher Scientific Inc. (Waltham, MA). **Tobacco etch virus (TEV) protease** was produced in-house, using the pRK793 plasmid obtained from Addgene (Cambridge, MA). All other materials were purchased from Sigma-Aldrich (Oakville, ON). Unless otherwise specified, reagents were used without further purification.

2.2.2. Cloning And Plasmid Preparation:

The plasmid for expression of AR TM2-3 with maltose-binding protein (MBP) was constructed by a previous graduate student by sub-cloning the gene encoding AR TM2-3 into the pMALc2x vector. Various mutants were made to enable cleavage of the fusion protein by adding a linker sequence using site-directed mutagenesis. A plasmid construct was also made with small ubiquitin-like modifier (SUMO) as the fusion protein

by sub-cloning the AR TM2-3 gene into the pETHN vector. Similarly, another construct was made with glutathione S-transferase (GST) as the fusion protein by sub-cloning the AR TM2-3 gene into the pEGX vector. Twenty different constructs having AT-rich gene tags for expression of the AR TM1-3, TM 1-4, TM 4-7, TM 5-7 and full-length AR were obtained by sub-cloning of the appropriate nucleotide sequence into the pET-23b (+) vector using the NdeI and XhoI restriction enzyme sites. Four open reading frames with different N-terminal tags were constructed and tested for expression of each AR construct (AR TM1-3, AR TM1-4, AR TM4-7, AR TM5-7 and full-length AR). Each plasmid construct thus obtained had an N-terminal expression enhancement tag and a hexahistidine-tag (6xHis-tag) at the C-terminus. Forward and reverse primers used to amplify AR TM1-3 and AR TM4-7 gene from a pUC 57 plasmid containing the sequence for the full-length AR are listed in Table 2.1 and 2.2 respectively. Forward primers for each construct had different overhangs at the 5' end for incorporation of different N-terminal tags. Standard thermocycling conditions and reaction composition for amplification of DNA using PCR are mentioned in Table 2.3.

2.2.3. Expression Of AR Fragments:

Expression of AR TM1-3 with and without N-terminal AT-rich tags was tested in *E. coli* BL21(DE3) and C41(DE3) strains. For small scale test expression of each construct, a single colony was grown in 4 mL LB medium with ampicillin (100 µg/mL) to an optical density at 600 nm (OD₆₀₀) of 0.6. 3 mL of this culture was then treated with IPTG at a final concentration of 0.5 mM to induce expression under the T7 promoter and allowed to grow at 37°C for 3 hours after induction. For optimization prior to large-scale protein expression, the effects of varying IPTG concentration (0.25 mM, 0.5 mM, 0.75

mM and 1 mM) and temperature (at 22°C for 14 hours, at 28°C for 14 hours and at 37°C for 2, 3 and 4 hours) were tested. For large-scale expression, the C41(DE3) strain of *E. coli* was transformed with pAR TM1-3_AT and grown for 14 hours at 37°C in a 250 mL baffled flask containing 15 mL of LB medium with ampicillin (100 µg/mL). This was used as a starter culture to inoculate 1L of LB medium with ampicillin (100 µg/mL). Expression was induced by addition of IPTG (to a final concentration of 0.5 mM) to the culture at OD₆₀₀ of 0.6. After induction, the cells were grown for 14 hours at 28°C and then harvested by centrifugation (6,000 x g at 4°C for 20 min).

For expression of ¹⁵N and ¹³C-enriched AR TM1-3, C41(DE3) *E. coli* cells were grown in LB medium with ampicillin (100 µg/mL) to an OD₆₀₀ of 0.6. Pelleted cells were re-suspended in a ½-equivalent volume of minimal medium (100 mM NaH₂PO₄, 40 mM K₂HPO₄, 4 mM MgSO₄, 1.8 µM FeSO₄, 2 g/L ¹³C₆ D-glucose, 1 g/L ¹⁵NH₄Cl and 100 mg/L ampicillin, titrated with NaOH to pH 7.3 (Marley *et al.*, 2001)) and grown for 14 hours at 28°C following induction.

2.2.4. SDS-PAGE And Western Blotting:

Based on observed OD₆₀₀ values, equal numbers of cells were pelleted from each sample and re-solubilized in 1x loading dye. Samples were heated at 90°C for 20 min before resolution with SDS-PAGE (15% Next gel; 180 V for 60 min). The gels were stained with Coomassie Brilliant Blue R-250, destained, and imaged using a VersaDoc 4000 (Bio-Rad). Expression was also analyzed by western blot, where the samples were first resolved by SDS-PAGE as described above and then transferred to a nitrocellulose membrane at 115 V for 60 min. Membranes were incubated in blocking buffer (Tris-buffered saline with 0.05% w/v tween 20 (TBST) and 2% w/v BSA) for 60 min at ~23°C.

Membranes were then washed twice with TBST for 10 min each prior to incubation with HisProbe-HRP (1:5000 dilution in blocking buffer) for 60 min at ~23°C. Membranes were washed again four times with TBST for 10 min each and then incubated for 5 min at ~23°C with Clarity Western ECL Substrate before imaging using a VersaDoc 4000.

2.2.5. Inclusion Body Preparation:

Inclusion bodies were purified as described in (Garboczi *et al.*, 1996), with minor changes as detailed by Björkman and co-workers (<http://www.its.caltech.edu/~bjorker/protocols.html>). Briefly, cell pellets were re-suspended in 12 mL of solution buffer (50 mM Tris-Cl, 25% sucrose, 1 mM EDTA-Na₂, 0.1% NaN₃, 10 mM DTT) and then mixed with 11 mL of lysis buffer (50 mM Tris-Cl, 1% Triton X-100, 1% sodium deoxycholate, 100 mM NaCl, 0.1% NaN₃, 10 mM DTT). The cells were lysed using a French pressure cell press (American Instrument Company; Silver Springs, MD). The resulting lysate was centrifuged at 11,000 x g for 20 min at 4°C. The supernatant was discarded and the pellet was re-suspended in 15 mL of wash buffer with Triton (50 mM Tris-Cl, 0.5% Triton X-100, 100 mM NaCl, 0.1% NaN₃, 10 mM DTT) using sonication. Centrifugation was again carried out at 11,000 x g for 20 min at 4°C. The pellet was again re-suspended in 15 mL of wash buffer without Triton (50 mM Tris-Cl, 1 mM EDTA-Na₂, 100 mM NaCl, 0.1% NaN₃, 10 mM DTT) using sonication. The inclusion body pellet containing partially pure AR TM1-3 was obtained by centrifuging again at 11,000 x g for 20 min at 4°C.

2.2.6. Ni²⁺ Affinity Chromatography:

Inclusion bodies were solubilized in a buffer (“binding buffer”) containing 10 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, and 1% SDS (w/v) at pH 8.0. A column (Econopac columns; Bio-Rad, Mississauga, ON) packed with Ni-NTA agarose resin (QIAGEN, Toronto, ON) was equilibrated with 2 column volumes of binding buffer. Solubilized inclusion bodies were applied to the column and allowed to flow through the column by gravity. The column was washed with 2 column-bed volumes of with a higher imidazole content (wash buffer: 10 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole and 1%SDS (w/v) at pH 8.0) before the bound proteins were eluted with 1 column-bed volume of a buffer with 250 mM imidazole (elution buffer: 10 mM NaH₂PO₄, 300 mM NaCl, 250 mM imidazole and 1%SDS (w/v) at pH 8.0). Constituents of fractions obtained through elution with loading, wash and elution buffers were collected and evaluated by SDS-PAGE.

2.2.7. Purification And Mass Confirmation:

The inclusion body pellet was re-suspended either in acetonitrile:isopropanol:water (2:1:2, by volume) or in TFA:water (4:1, by volume) at high protein concentration prior to RP-HPLC. Linear binary gradients (solvent A – H₂O and isopropanol; solvent B – acetonitrile and isopropanol; A and B both contain 0.1%TFA and 10% isopropanol for analytical separation, 20% isopropanol for preparative separation) were employed for both analytical and preparative RP-HPLC with the column temperature maintained at 50°C on a Prostar HPLC equipped with a diode array ultraviolet (UV)-visible detector (Varian Canada Inc.; Mississauga, ON). For analytical separation (ZORBAX Rapid resolution 300SB-C3 column, 4.6 mm x 150 mm,

3.5 μm), 100 μL of solubilized inclusion body preparation was subjected to a linear gradient of 30-90% solvent B over 40 min at a flow rate of 1 mL/min (chromatograms at 220 and 280 nm typically monitored). Each peak was collected and immediately lyophilized. Preparative RP-HPLC (ZORBAX PrepHT 300SB-C3 column, 21.2mm x 50mm, 5 μm) employed 1 mL of solubilized inclusion body preparation with a flow rate of 10 mL/min and a linear gradient of 35-50% solvent B over the first 8 min and 50-75% solvent B over the next 5 min. Due to the highly hydrophobic nature of the protein, the column had to be washed 3-4 times to recover all the injected protein. Each wash was performed by injecting TFA:water (4:1; volume equal to the volume of the sample injected) into the column and subjecting it to the same gradient as the sample. The purified protein fractions were pooled, lyophilized and stored at -20°C . Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry of lyophilized protein samples was performed either at the SPARC Biocentre (Hospital For Sick Children, Toronto, ON) or at the Genomics and Proteomics Facility (GaP, Memorial University of Newfoundland, St. John's, NL).

2.3. Results

2.3.1. Difficulties In Production Of The Full-Length AR And Its Fragments:

As mentioned in the introduction, fusion protein have been instrumental in production of membrane protein. Following the literature precedence of using MBP for membrane protein expression, plasmid constructs were made with MBP fused to the N-terminus of AR TM2-3, AR TM4-5, and full-length AR, with a Factor Xa cleavage site in between MBP and the POI. Among the three constructs tested, AR TM2-3 and AR TM4-

5 were expressed successfully in BL21(DE3) strain of *E. coli*. Despite the successful expression, pure AR TM2-3 or AR TM4-5 was not obtained, due to lack of cleavage activity by factor Xa protease. This construct also had a cysteine at the N-terminus of AR TM2-3, to enable cleavage using 2-nitro-5-thiocyanobenzoic acid (NTCB) (Smith, 2002). The NTCB cleavage worked, with the AR TM2-3 band visible by western blotting but not with Coomassie staining, implying that the amount of cleaved protein was low. In an attempt to solve the problem of fusion protein removal, a variety of constructs were made using various different fusion tags, cleavage sites and linkers (detailed in Table 2.4). Both Tev protease and Factor Xa protease were tested for their ability to cleave the MBP fusion protein. While following the manufacturers' protocol, neither protease was successful in removing the MBP from the AR TM segment. Various modifications to cleavage conditions, including the temperature, protease concentration, and incorporation of detergents into the buffer, were also tested. Serendipitously, the cleavage of MBP occurred with Factor Xa protease at very low protein concentrations (Figure 2.4). In comparison to GST and GB1 fusion constructs, MBP and SUMO fusion constructs yielded better expression levels (Figure 2.3). The SUMO fusion construct enhanced expression but was unfortunately not soluble, making the subsequent tag removal unsuccessful. These difficulties made it economically infeasible to obtain a suitable quantity of AR TM2-3 for structural characterization. All the attempts made to obtain pure AR TM2-3 have been listed in Table 2.4. Also, similar efforts were simultaneously made to obtain AR TM4-5 and full-length AR. No expression was ever observed for the full-length AR, whereas AR TM4-5 did express, but with same problems as with AR TM2-3. At this point, I started to look for tweaks and alternative strategies as mentioned in the introduction.

Consequently, the strategy to divide the receptor was changed and 20 different constructs were made with AT-rich gene tags to enhance the expression and bypass the tag removal problem. Among all the constructs that were tested, the AR TM1-3 constructs were the only ones that were successfully expressed. Therefore, AR TM1-3 will be in focus for the rest of the chapter. This also led to selection of an AR TM1-3 and AR TM4-7 pair for future work discussed in Chapter 4.

2.3.2. Expression Optimization Using AT-Rich Tags:

Four small AT-rich 5' gene tags (Table 2.5) were tested for the ability to enhance protein expression of recombinant AR TM1-3 in the BL21(DE3) and C41(DE3) strains of *E. coli*. Despite initial attempts, the BL21(DE3) strain was not suitable for expression as a large amount of cell death was observed after induction (Figure 2.5; Table 2.6). No target protein expression was observed for the AR TM1-3 construct without AT-rich tags (Figures 2.5 & 2.6). Conversely, SDS-PAGE (Figure 2.6) and western blot analysis (Figure 2.7) demonstrate that all four AT-rich tagged AR TM1-3 constructs led to AR TM1-3 expression in the C41(DE3) strain. Notably, although the G-tag was capable of enhancing protein expression over the untagged gene, the amount of protein produced with this tag was clearly lower than any of the other 3 tags (Figure 2.7). Western blot quantitation (Figure 2.7B) showed that the construct with the H-tag provided the best enhancement of AR TM1-3 expression. However, the comparative enrichment of the AR TM1-3 in comparison to background protein expression levels evident from SDS-PAGE analysis (Figure 2.6) led us to pursue the AR TM1-3 construct with the AT-tag for overexpression in both rich and minimal medium (relative levels illustrated in Figure 2.8A) and subsequent purification.

2.3.3. Purification Of AR TM1-3:

Overexpressed AR TM1-3 was found to form inclusion bodies; hence, the first step in purification was isolation and washing of the inclusion bodies. Very little to no loss of AR TM1-3 was observed while washing the inclusion bodies (Figure 2.8B). To further purify AR TM1-3, solubilized inclusion bodies were subjected to RP-HPLC (Figure 2.9). Despite the protein having a 6xHis-tag, RP-HPLC was chosen as the method of purification due to an inability to find conditions providing sufficient affinity of the 6xHis-tag to the Ni-NTA column (Figure 2.8B). The identity of the AR TM1-3 RP-HPLC eluent peak and the incorporation of ^{15}N and ^{13}C stable isotopes were confirmed by MALDI-TOF (Figure 2.10; expected mass of unlabeled AR TM1-3 17684.69 Da, observed mass 17684.52 Da; expected mass of 100% $^{15}\text{N}/^{13}\text{C}$ -enriched AR TM1-3 18690.55 Da, observed mass 18627.94 Da indicative of ~92% isotope incorporation efficiency.). The final yield of lyophilized AR TM1-3 powder following purification was ~ 8 mg/L of minimal medium, with a purity of >99% as confirmed by analytical HPLC (Figure 2.11).

2.4. Discussion:

Herein, I have shown that the use of AT-rich gene tags makes it possible to express a three TM long fragment (AR TM1-3) of a human GPCR in *E. coli*. Our findings demonstrate that AT-rich tags can enhance expression not only in an *E. coli*-based cell-free expression system, as previously reported, but also in *E. coli* cells. The degree of enhancement is extremely clear from the fact that, without the use of AT-rich tags, AR TM1-3 expression was undetectable even by western blotting (Figure 2.6;

probing for the C-terminal His-tag). It should also be noted that use of the C41(DE3) *E. coli* strain, which exhibits an increased tolerance to toxic protein expression relative to the more commonly employed BL21(DE3) strain (Miroux & Walker, 1996), was key for enhanced expression. In contrast, in BL21(DE3) *E. coli*, attempts at overexpression of AR TM 1-3 caused significant cell death, preventing production of sufficient quantities of the protein.

As mentioned in the introduction, fusion protein tags have proved to be instrumental in heterologous expression and purification of recombinant protein. Despite their applicability, their removal can be very challenging in some cases. In such cases, the post-translational tag removal can be avoided by the use of very small tags that can still enhance expression, such as the AT-rich tags (Haberstock *et al.*, 2012). The small size of these tags makes them very useful for structural studies using solution-state NMR spectroscopy (Zhou & Wagner, 2010).

In summary, I have successfully expressed and purified multimilligram quantities of a protein comprising the first three TM segments of the AR. This fragment is larger (by at least one TM segment) than other human GPCR fragments that have been produced and studied to date using an *E. coli* expression system (Cohen *et al.*, 2014). Biophysical characterization of AR TM1-3 will be presented in the next chapter.

Table 2.1: Primers and constructs for AR TM1-3.





Plasmid name	Construct	Forward primer (5' – 3')	Reverse primer (5' – 3')
pAR_TM1-3_AT		ATA CTA CAT ATG AAA TAT TAT AAA TAT TAT ATG GAG GAA GGC GGT GAT TTT	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG GTT AGC TAC AGG ACG AAC AAT A
pAR_TM1-3_G		ATA CTA CAT ATG AAA AGT AAA GGA GAA GAA ATG GAG GAA GGC GGT GAT TTT	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG GTT AGC TAC AGG ACG AAC AAT A
pAR_TM1-3_H		ATA CTA CAT ATG AAA CCA TAC GAT GGT CCA ATG GAG GAA GGC GGT GAT TTT	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG GTT AGC TAC AGG ACG AAC AAT A
pAR_TM1-3_AT4		ATA CTA CAT ATG AAA TAT TAT AAA ATG GAG GAA GGC GGT GAT TTT	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG GTT AGC TAC AGG ACG AAC AAT A

Table 2.2: Primers and constructs for AR TM4-7.

Plasmid name	Construct	Forward primer (5' – 3')	Reverse primer (5' – 3')
pAR_TM4-7_AT	AT-tag AR TM4-7 6xHis	ACT GTC CAT ATG AAA TAT TAT AAA TAT TAT GCA CGC CTG CGT CTG CGC	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG GTC CAC CAC CAG GGT TTC CT
pAR_TM4-7_G	G-tag AR TM4-7 6xHis	ACT GTC CAT ATG AAA AGT AAA GGA GAA GAA GCA CGC CTG CGT CTG CGC	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG GTC CAC CAC CAG GGT TTC CT
pAR_TM4-7_H	H-tag AR TM4-7 6xHis	ACT GTC CAT ATG AAA CCA TAC GAT GGT CCA GCA CGC CTG CGT CTG CGC	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG GTC CAC CAC CAG GGT TTC CT
pAR_TM4-7_AT4	AT4-tag AR TM4-7 6xHis	ATA CTA CAT ATG AAA TAT TAT AAA GCA CGC CTG CGT CTG CGC	ACT AGC CTC GAG TTA ATG ATG ATG ATG ATG ATG GTC CAC CAC CAG GGT TTC CT

Table 2.3: Standard thermocycling conditions and reaction composition for amplification of DNA using PCR.

Reaction composition	Stock concentrations	50 μL reaction
5X buffer	5x HF or GC buffer	10 μ L
Template DNA	50 ng/ μ L	10 ng (~100 ng needed when using stratagene protocol for site directed mutagenesis)
Primers	10 μ M	2.5 μ L
DNTPs	10 mM	1 μ L
Polymerase amount	2 U/ μ L	0.5 μ L
H ₂ O (μ L)		Add to 50 μ L

Cycling conditions		Number of cycles
Initial denaturation	98 °C, 30 sec	1
Melting step	98 °C, 10 sec	25
Annealing step	60-70 °C (temperature may differ between experiments)	
Extension step	72 °C, 30 sec/kb (time may differ between experiments)	
Final Extension	72 °C, 7 min	1
Hold	4 °C	

Note: Phusion polymerase has been used for all DNA amplification reactions in this thesis.

Table 2.4: Summary of constructs tested for expression of AR TM2-3.

Construct	Problem	Result
MBP-Factor Xa-AR TM 2-3	Difficult to cleave MBP	Worked at very low concentration
MBP-10XGly-Factor Xa-AR TM 2-3	MBP cleavage unsuccessful	X
MBP-Factor Xa-10XGly-AR TM 2-3	MBP cleavage unsuccessful	X
MBP-TEV-AR TM 2-3	MBP cleavage unsuccessful	X
GST-TEV-AR TM 2-3	Insoluble Protein Low expression level	X
GB1-TEV-AR TM 2-3	Low expression level	X
SUMO-AR TM 2-3	Insoluble Protein	X

Table 2.5: Nucleotide and amino acid sequence of the expression enhancement tags.

Tag name	Nucleotide sequence	Amino acid sequence	% of A and T
AT	AAATATTATAAATATTAT	KYYKYY	100
G	AAAAGTAAAGGAGAAGAA	KSKGEE	72
H	AAACCATACGATGGTCCA	KPYDGP	55
AT4	AAATATTATAAA	KYYK	100

Table 2.6: Difference in the level of AR TM1-3 in BL21(DE3) and C41(DE3) strains of *E. coli*. Comparison of volume (μL) of culture needed to have same number of cells (based on OD_{600}) in each sample for SDS-PAGE. Volumes are averaged from 2 and 3 experiments for BL21(DE3) and C41(DE3) respectively.

Strain → ↓ Tag	C41(DE3)		BL21(DE3)	
	Un-induced	Induced	Un-induced	Induced
No-tag	491	542	489	1552
AT-tag	512	979	478	2600
H-tag	516	773	486	1970
G-tag	526	520	483	2450
AT4-tag	527	808	458	2750

NOTE: Cell death in BL21(DE3) strain of *E. coli* is evident from the fact that much larger volume of culture was needed to have the same number of cells in the induced sample.

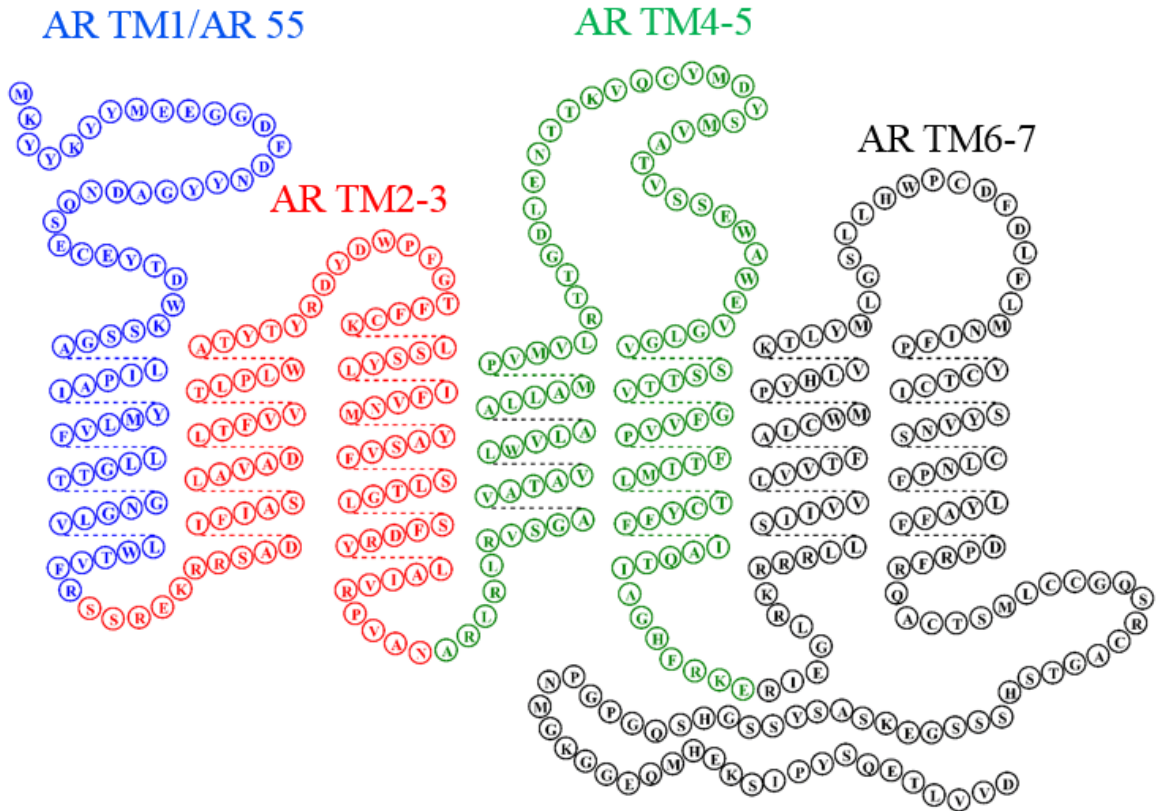


Figure 2.1: The snake plot representing the “divide and conquer” approach based division of the receptor into four parts.

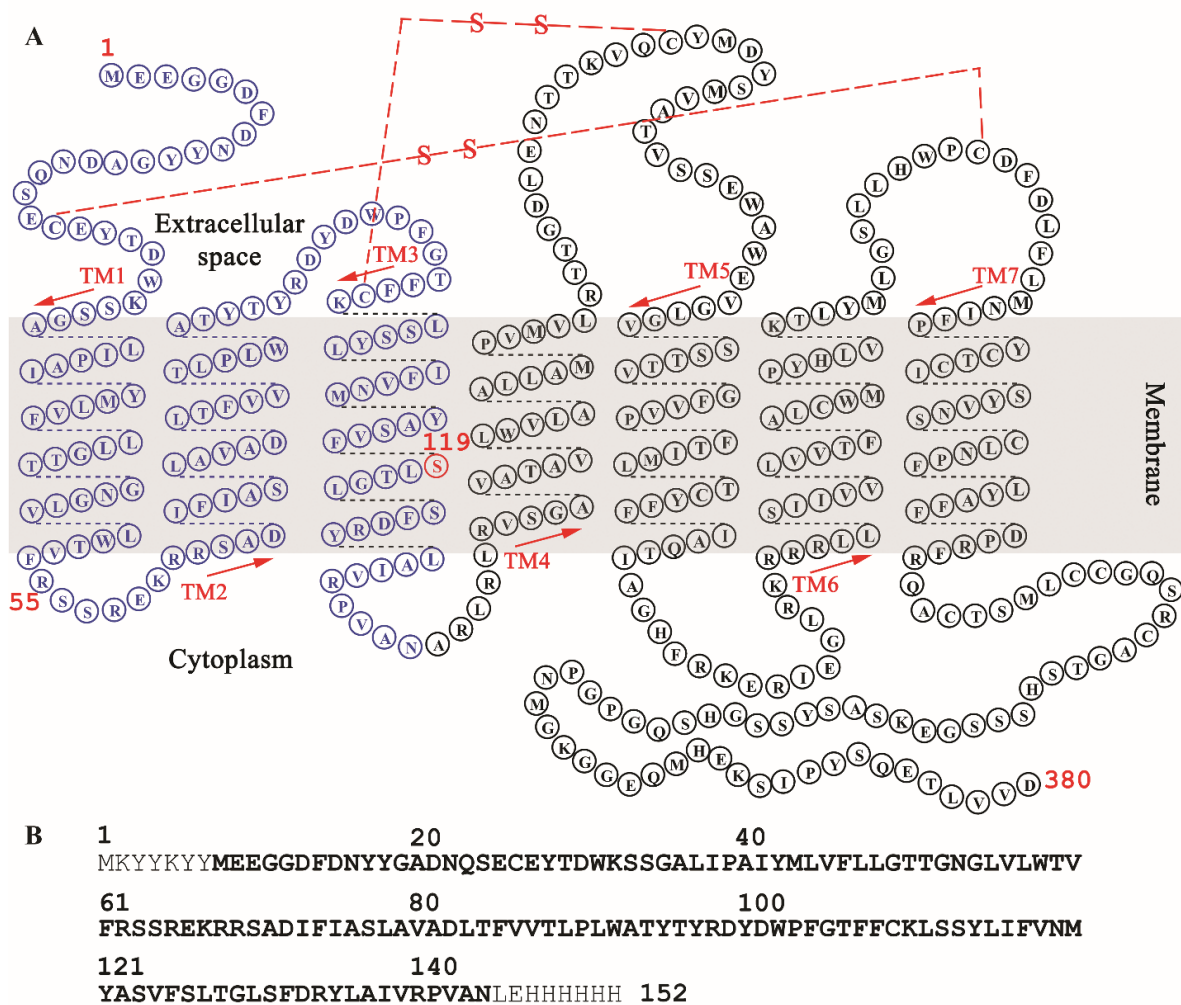


Figure 2.2 : The snake plot representing the “divide and conquer” approach based division of the receptor into two parts. (A) Sequence and topology of the apelin receptor (AR). The sequence coloured in blue represents AR TM1-3. Putative disulfide linkages are shown by red coloured dashed-lines (Alexandrov *et al.*, 2008). The Cys at position 119 (red) has been mutated to Ser to facilitate purification and handling. (B) Sequence of AR TM1-3 construct with N-terminal AT-tag and C-terminal 6xHis-tag.

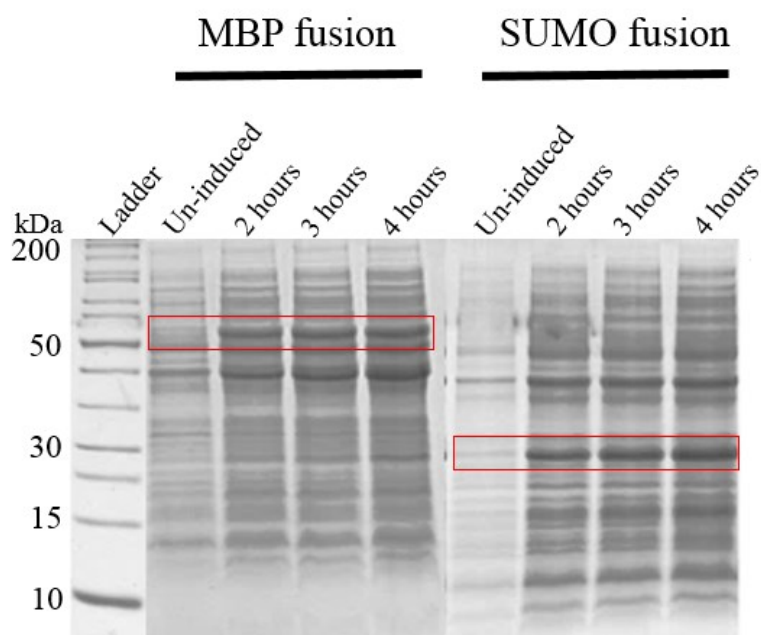


Figure 2.3: AR TM2-3 expression using MBP and SUMO as fusion proteins.

Coomassie blue stained gel showing test expression in C41(DE3) strains of *E. coli* for the AR TM2-3 constructs containing a either SUMO or MBP as a N-terminal fusion tag. After expression, equal volume of culture was pelleted from un-induced (UI) and at different time points after induction with IPTG. Samples were made by mixing the cell pellet from each culture with 100 μ L 1x loading dye and resolved using 15% SDS-PAGE. The red boxes indicate the position of the AR TM2-3 fusion protein.

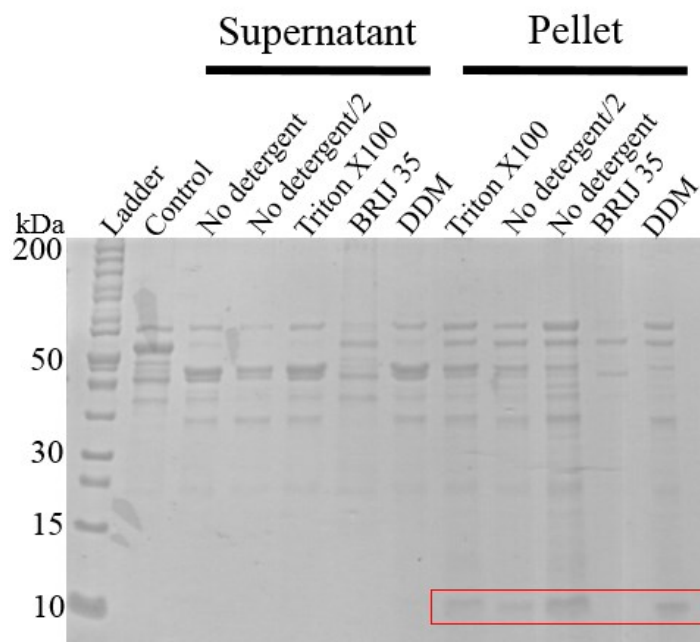


Figure 2.4: Factor Xa mediated cleavage of MBP from AR TM2-3 construct.

Coomassie blue stained 15% SDS-PAGE gel showing cleavage of MBP from the AR TM2-3 construct with MBP fusion. The fusion protein was purified on an amylose column using manufacturers' protocol. The cleavage of the MBP fusion tag was carried out at room temperature with and without (No detergent) different detergents. Cleavage of the fusion protein was carried out by factor Xa protease using the manufacturers' protocol. The cleavage occurred at low fusion protein concentration of ~500 ng/mL. No detergent/2 represents half the sample load in comparison to No detergent. The red box indicates the position of the cleaved AR TM2-3 protein.

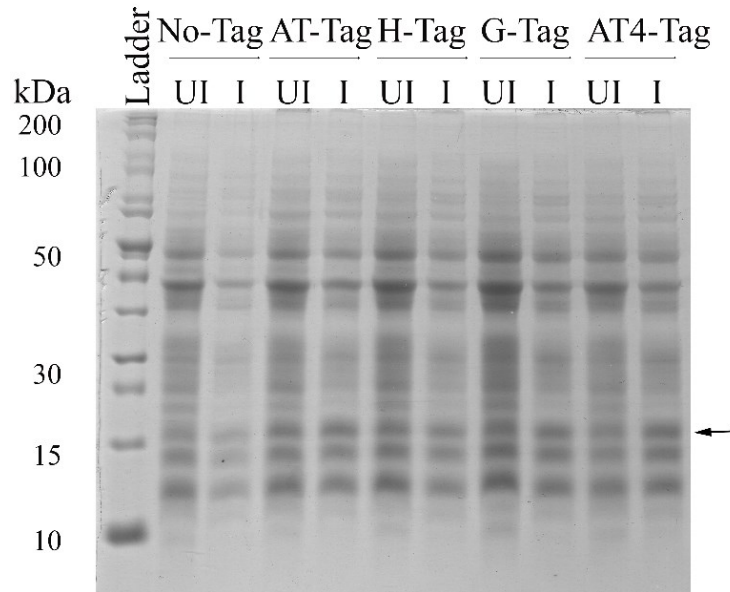


Figure 2.5: Evaluation of AR TM1-3 expression in BL21(DE3) strain of *E. coli*. Coomassie blue stained gel showing test expression in BL21(DE3) strains of *E. coli* for the AR TM1-3 constructs containing specified tags (sequences in Table 2.2) and a construct without any N-terminal tag. After expression, equal number of cells (using the volumes mentioned in Table 2.6) were pelleted from induced (I) and un-induced (UI) cultures of cells grown for 3 hours after induction with IPTG. Samples were made by mixing cell pellet from each culture with 100 μ L 1xloading dye and resolved using 15% SDS-PAGE. The arrow indicates the position of the AR TM1-3 band.

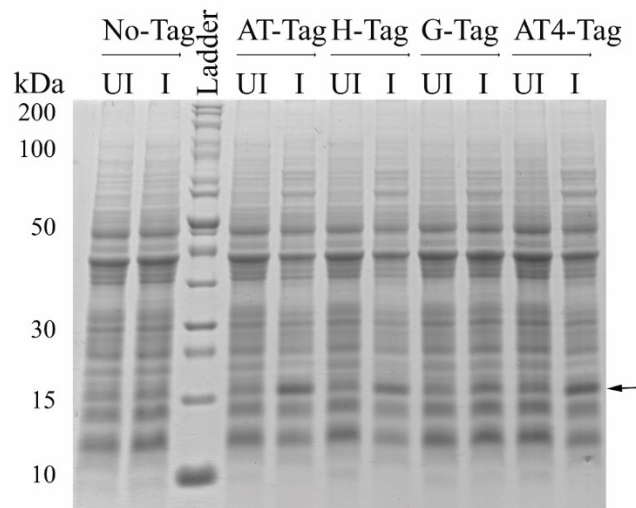


Figure 2.6 : Evaluation of AR TM1-3 expression in C41(DE3) strain of *E. coli*.

Coomassie blue stained gel showing test expression in C41(DE3) strains of *E. coli* for the AR TM1-3 constructs containing specified tags (sequences in Table 2.2) and a construct without any N-terminal tag. After expression, an equal number of cells (based on OD₆₀₀) were pelleted from induced (I) and un-induced (UI) cultures of cells grown for 3 hours after induction with IPTG. Samples were made by mixing the cell pellet from each culture with 100 μ L 1x loading dye and resolved using 15% SDS-PAGE. The arrow indicates the position of the AR TM1-3 band as confirmed by western blotting (Figure 2.7).

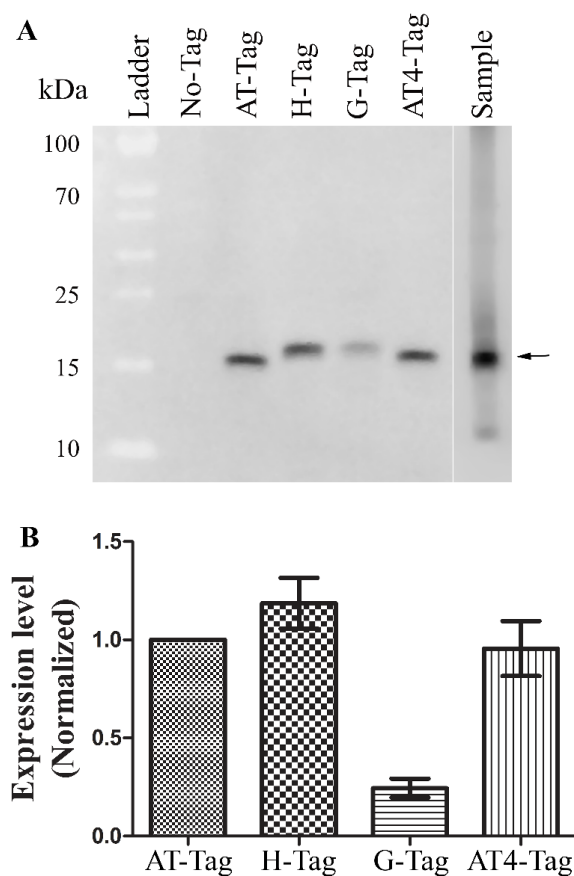


Figure 2.7: Comparison of AR TM1-3 expression in using AT-rich gene tags. (A) Western blot probed with HisProbe-HRP showing a comparison of expression in different constructs of AR TM1-3. Induced C41(DE3) *E. coli* for each construct were first resolved by SDS-PAGE (as in Figure 2.6) and then transferred onto a nitrocellulose membrane and probed (detailed in Materials and Methods). The AR TM1-3 NMR sample in SDS micelles (pH 6.00) was loaded in the lane labeled “sample”. (B) Quantitation (using ImageJ) of difference in expression observed by western blotting (n=3). Mean (bars) and standard error (error bars) for expression level are plotted, with the intensity of the AT-tag on each blot used as the normalization factor. Untagged AR TM1-3 is not plotted in (B) since no protein was observable by western blotting for constructs without an AT-rich tag.

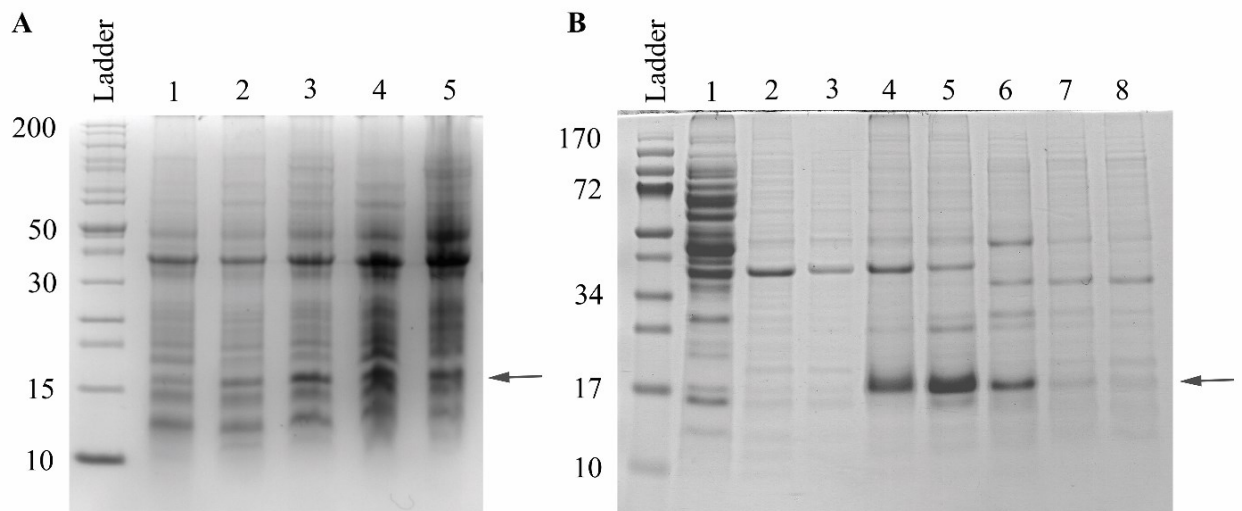


Figure 2.8 : Evaluation of purification and expression yield in minimal medium for AR TM1-3. Coomassie blue stained gels showing (A) Comparison of Expression of AR TM1-3 construct with AT-tag using C41(DE3) in M9 minimal medium (lane 3 and 5) versus in rich (lane 2 and 4) medium and (B) fractions from inclusion body preparation (lane 1, 2, and 3) and Ni²⁺ affinity chromatography of SDS solubilized 6xHis-tagged AR TM1-3 (lane 4, 5, 6, 7, and 8). In (A) lane 1 shows un-induced cells, lanes 2 and 3 show the expression after 3 hours after induction in rich and M9 media respectively at 37°C; lanes 4 and 5 show overnight expression in rich and M9 medium respectively at 28°C. In (B), lanes 1, 2, and 3 show supernatants from lysate, wash with buffer containing 1% Triton X-100, and wash with buffer without Triton X-100, respectively, during inclusion body preparation. Lanes 4-8 show flow-through, wash and three elution fractions, respectively, for inclusion bodies solubilized with 1% SDS and subjected to Ni²⁺-affinity purification. The arrow indicates the position of the AR TM1-3 band.

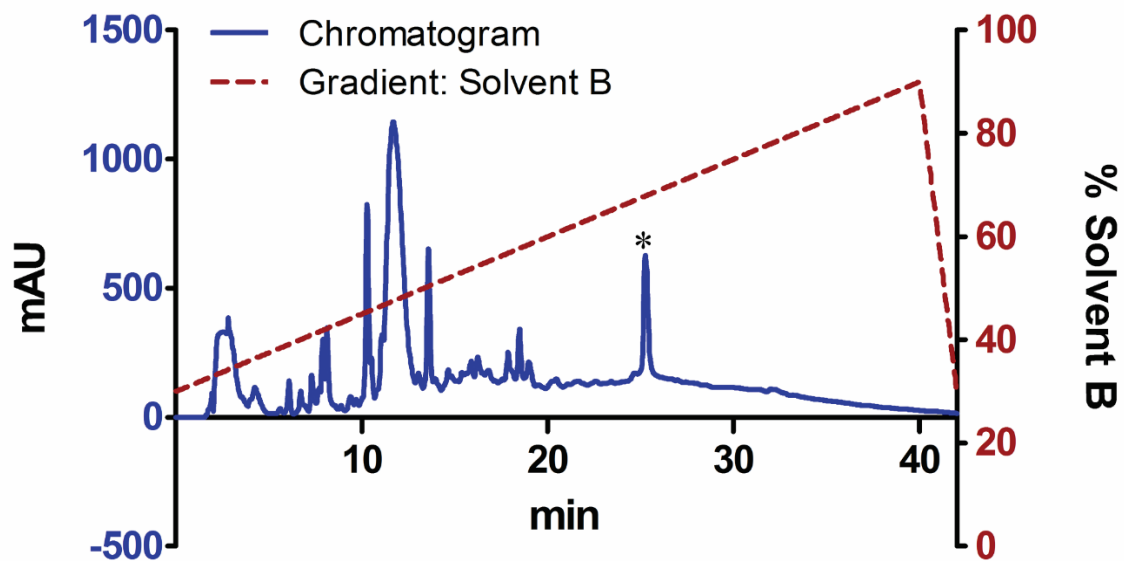


Figure 2.9: HPLC chromatogram for AR TM1-3 purification. (Solid blue line) Analytical RP-HPLC chromatogram (milliabsorbance units (mAU) at 220 nm) for the solubilized inclusion body containing AR TM1-3. (Dashed red line) Binary gradient of solvent A-B (A: 90% H₂O; B: 90% acetonitrile; both containing 10% isopropanol and 0.1%TFA). The * denotes the peak corresponding to AR TM1-3.

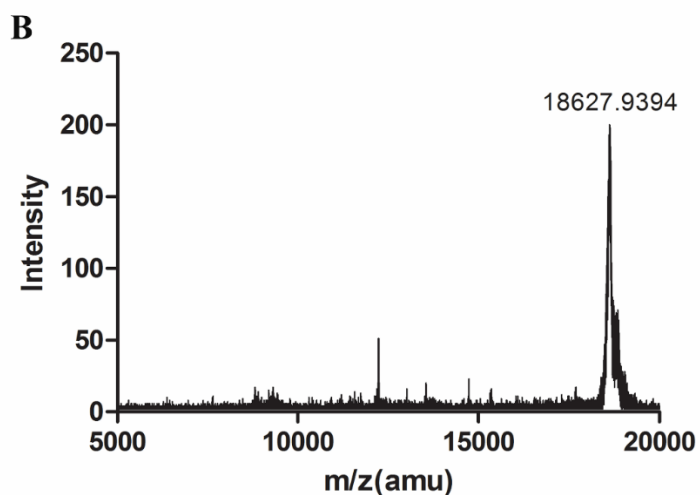
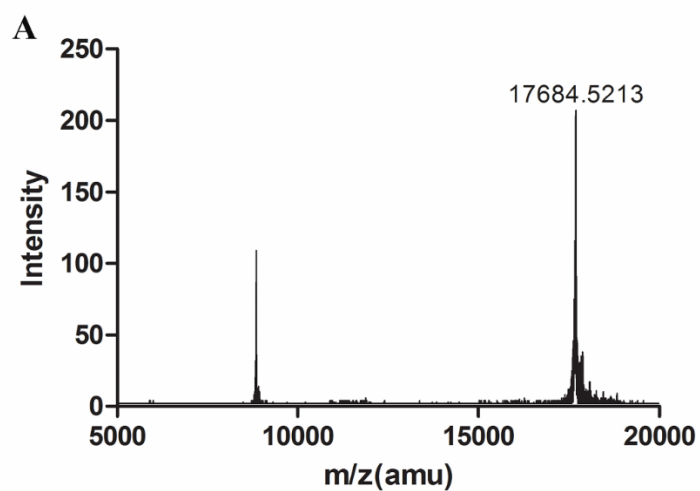


Figure 2.10: Mass spectrum to confirm identity and evaluate the efficiency of isotope-labeling for AR TM1-3. MALDI-TOF mass spectra of isolated AR TM1-3 expressed in (A) unlabeled rich medium (expected mass 17684.69 Da) and (B) $^{15}\text{N}/^{13}\text{C}$ isotope-enriched M9 minimal medium (expected mass with 100% isotope-enrichment 18690.55 Da).

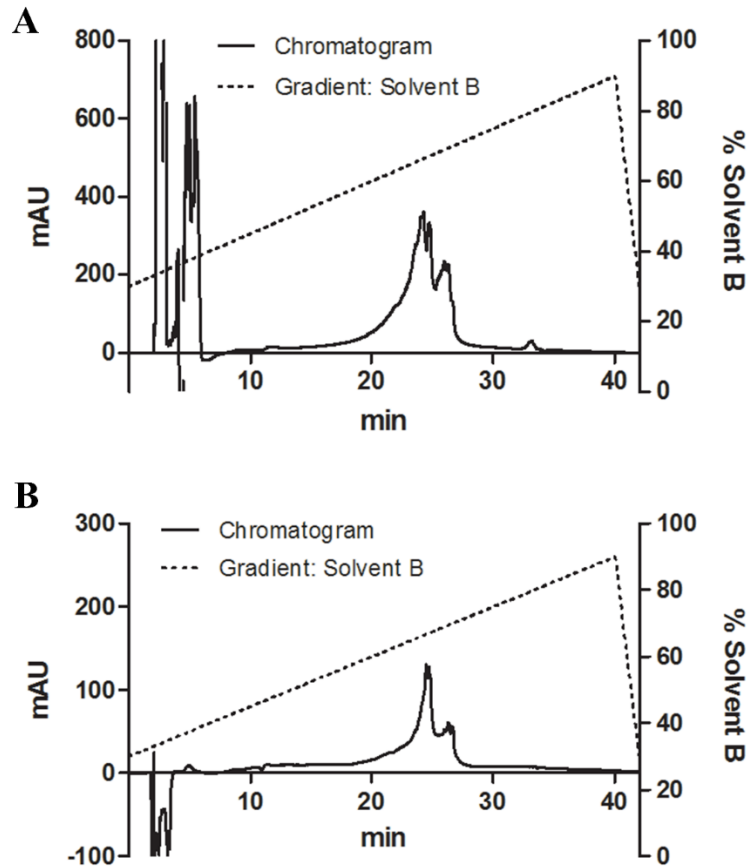


Figure 2.11: HPLC chromatogram to evaluate the purity of AR TM1-3 used for NMR sample preparation. (Solid line) Analytical RP-HPLC chromatogram (milliabsorbance units (mAU) at 220 nm) for purified AR TM1-3. (Dashed line) Binary gradient of solvent A-B (A: 90% H₂O; B: 90% acetonitrile; both containing 10% isopropanol and 0.1% TFA). (A) Purified AR TM1-3 solubilized in HFIP:water (1:1) was injected in the column. Eluent peaks in the first 7 mins of the chromatogram were from small molecules in the sample. AR TM1-3 elutes as two broad peaks from 20 to 26 min possibly due to conformational heterogeneity. (B) The column was washed by injecting TFA:water (4:1) and a chromatogram similar to (A) was observed, showing that a fraction of AR TM1-3 was retained on the column. As mentioned in the manuscript, 3-4 washes were required to recover all the injected protein. Both the peaks for AR TM1-3 in each chromatogram were confirmed by mass spectrometry.

Chapter 3: Optimization Of Conditions For High Resolution Structural Characterization Of AR TM1-3 In Various Membrane-Mimetics

Note: *Parts of this chapter have been published as a manuscript titled “Small expression tags enhance bacterial expression of the first three transmembrane segments of the apelin receptor” in **Biochemistry and Cell Biology** 92(4): 269-278.*

3.1. Introduction:

Following the work detailed in the previous chapter, the next objective is to characterize the atomic-level structure and dynamics AR TM1-3. The technique I have chosen for these studies is solution-state NMR, given its ability to shed light on both structure and dynamics of proteins. Solution-state NMR is a very powerful technique to study appropriately sized membrane protein samples in mimetics such as an organic solvent mixture, detergent micelles, isotropic bicelles and monodisperse nanodiscs. It is very difficult to reconstitute membrane proteins in their native environment in a way that it becomes amenable to characterization by biophysical techniques such as CD and NMR spectroscopy. The native biological membranes where GPCRs reside are a very complex mixture of many lipids that varies between species and cell types (Warschawski *et al.*, 2011). Further adding to this complexity, the inner and outer leaflets of the bilayer also differ substantially in their lipid compositions (Simons & Sampaio, 2011). The membrane protein folding theory states that it is the kinetics and thermodynamics of the peptide-bilayer interaction that drives the protein into its functional state within the membrane (Cymer *et al.*, 2015). Study of membrane proteins in a native environment

typically requires different types of lipids, leading to larger protein-lipid complexes such as non-isotropic bicelles, macrodiscs, and bilayers that are not compatible with solution-state NMR. In the case of these samples, where transverse relaxation is typically extremely rapid while longitudinal relaxation may be extremely slow, one must resort to solid-state approaches. Therefore, to study membrane proteins, it is essential to identify the optimal membrane-mimetic environment that is also compatible with solution-state NMR spectroscopy. Identification of an optimal membrane-mimetic environment is still a matter of trial and error, and is a bottleneck in structural characterization of membrane proteins.

Detergent micelles are one of the most frequently used membrane-mimetic environments to study membrane protein. Detergents are amphipathic molecules that self-assemble in aqueous solution to form aggregates known as micelles. These form in aqueous environment when the detergent concentration is above a specific threshold known as the critical micelle concentration (CMC). Micelles are small monolayer aggregates with a rough surface that differ in morphology, and may adopt shapes ranging from spherical to rod-like depending on the concentration and types of detergents used (Lipfert *et al.*, 2007). Beyond shape, micelles vary in their aggregation number (number of detergent molecules comprising a micelle), which is a characteristic of both detergent and solution conditions and dictates the size and weight of the micelle. Choosing detergents with small aggregation number is generally advantageous as they form fast-tumbling protein-detergent complexes that are desirable for solution-state NMR spectroscopy. Historically, detergents such as sodium dodecyl sulphate (SDS), dodecylphosphocholine (DPC), and 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-

(1-glycerol)] (LPPG) have been frequently used to study various membrane proteins (Sanders & Sonnichsen, 2006).

Mixtures of organic solvents such as such as methanol, chloroform, isopropanol, HFIP, TFE and DMSO are also commonly used as membrane-mimetic to study membrane protein using NMR spectroscopy. The exact mechanism of the solvent-solute interaction that allows these environments to mimic the anisotropic nature of lipid membranes is not completely understood. Molecular dynamics simulation-based analysis has shown that addition of water to organic solvent enhances the local phase separation of solvents, and thus provides a better membrane-mimetic effect (Mottamal *et al.*, 2007). Although the use of organic solvents as membrane-mimetics is more prevalent for small proteins/fragments (Boukadida *et al.*, 2014), a number of large proteins have also been characterized using these conditions (Fracchiolla *et al.*, 2015). Among organic solvent mixtures, TFE–water mixtures have become popular due to their ability to solubilize membrane proteins with minimal perturbations in the secondary structure. TFE-water mixtures also stabilize the secondary structure in proteins by forming a coat around the hydrophobic region of the peptide and reducing the local dielectric around it.

In this work, based upon my aim of characterizing AR TM1-3 by solution-state NMR spectroscopy, detergent micelles and organic solvent mixtures (Figure 3.1) were examined for their suitability. Correspondingly, a variety of micelle and solvent environments have been screened to identify the most suitable conditions for structural characterization of AR TM1-3 at the atomic level.

3.2. Materials And Methods:

3.2.1. Materials:

Sodium dodecyl sulfate (SDS), sodium acetate, sodium azide, 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), 2,2,2-trifluoroethanol (TFE) and reagents for making lysogeny broth (LB) medium were purchased from Fisher Scientific (Ottawa, ON). 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (LPPG) was purchased from Anatrace (Maumee, OH). Dimethylsulfoxide- d_6 (DMSO- d_6), HFIP- d_2 , TFE- d_2 , NaOD and sodium dodecyl sulfate- d_{25} (SDS- d_{25}) were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). DTT- d_6 , DPC- d_{38} , DCl, D_2O (99.8 atom % D) and D_2O containing 1% (w/w) 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) were obtained from C/D/N Isotopes (Pointe-Claire, QC). Quartz cuvettes for circular dichroism (CD) spectropolarimetry were purchased from Hellma (Vaughn, ON). 5 mm high quality NMR tube were purchased from Bruker Biospin (Fällanden, Switzerland). All other materials were purchased from Sigma-Aldrich (Oakville, ON). Unless otherwise specified, all reagents were used without further purification.

3.2.2. CD Spectropolarimetry:

Far-UV CD spectra for all samples of AR TM1-3 were recorded at 25 °C using a J-810 spectropolarimeter (Jasco, Easton, MD) at 20 nm/min with a data pitch of 0.1 nm from 260 nm to 190 nm. Cuvettes of path length (l) 0.1 cm and 0.01 cm were used for HFIP and detergent micelle samples, respectively. The HFIP sample contained 8.20 ± 0.05 μM protein in HFIP:H₂O (1:1 by volume); micelle samples contained 82.0 ± 0.5 μM protein, with 20 mM SDS and LPPG, in 20 mM sodium phosphate buffer at pH

6.00±0.05 and pH 7.00±0.05 or 10 mM DPC in 50 mM sodium acetate buffer at pH 4.00±0.05. All samples were prepared from the same protein stock solution, with concentration (c) determined using the Beer-Lambert law ($c=A \cdot \epsilon^{-1} \cdot l^{-1}$) by measuring absorbance (A) at 280 nm using a molar absorptivity (ϵ ; calculated following (Pace *et al.*, 1995)) of 42860 M⁻¹cm⁻¹. All data were collected in triplicate, averaged, blank subtracted, and are reported as mean residue ellipticity ([θ]). Data were deconvoluted using CDSSTR (Johnson, 1999) with the SMP180 dataset (Abdul-Gader *et al.*, 2011) as implemented in the DichroWeb server (Whitmore & Wallace, 2008).

3.2.3. Solution-State NMR Spectroscopy:

1D ¹H (Hwang & Shaka, 1995) and 2D ¹H-¹⁵N heteronuclear single quantum correlation (HSQC) (Palmer *et al.*, 1991; Kay *et al.*, 1992; Schleucher *et al.*, 1994) spectra of AR TM1-3 were acquired in two organic solvents and one detergent micelle membrane-mimetic conditions. Organic solvent samples were prepared by dissolving lyophilized protein to a final concentration of 0.6 mM in dimethyl sulfoxide (DMSO-d₆); 0.7 mM protein in 50% HFIP-d₂ plus 40% H₂O and 10% D₂O (by volume); or, 0.12 mM protein in 50% TFE-d₂ plus 40% H₂O and 10% D₂O (by volume). The HFIP and TFE samples also contained 1 mM DSS and 5 mM DTT-d₆. An SDS micelle sample was prepared at final concentrations of 0.6 mM protein and 150 mM SDS-d₂₅ in 95% H₂O and 5% D₂O. The SDS sample also contained 1 mM DSS, 10 mM DTT-d₆, 20 mM sodium phosphate and 1 mM NaN₃ and the pH of the sample was adjusted to 6.00±0.05 using DCl and NaOD. NMR spectra of the SDS sample were collected at 27, 37 and 47 °C using a Bruker Avance III 700 MHz spectrometer (Milton, ON) equipped with a cryogenically cooled 5 mm triple resonance inverse (TCI) probe and a z-axis gradient

(Biomolecular Magnetic Resonance Facility (BMRF) housed at the National Research Council, Halifax, NS). NMR spectra of the DMSO sample were collected at 22 and 37°C using a Bruker Avance 500 MHz spectrometer equipped with a room temperature 5 mm triple resonance inverse (TXI) probe and a z-axis gradient (NMR³ Facility at Dalhousie University). NMR spectra of the 50% HFIP sample were collected at 22, 37 and 47°C using the 700 MHz system and at 37 and 47°C using the 500 MHz system. NMR spectra of the 50% TFE sample were collected at 47°C using the 700 MHz system. NMR experimental details for all conditions are detailed in Tables 3.1, 3.2, 3.3 and 3.4. All NMR spectra were processed using Bruker Topspin 3.1 or NMRpipe 8.1 (Delaglio *et al.*, 1995). Spectra were plotted using CcpNmr Analysis 2.2.2 (Vranken *et al.*, 2005) for 2D experiments and 3D experiments, and Bruker Topspin 3.1 for 1D experiments. CCpNmr Analysis 2.2.2 was used for backbone walk (section 1.7.3) based assignment of C α , C β , C', and HN for the main chain of AR TM1-3.

3.3. Results:

3.3.1. CD Spectropolarimetry:

Far-UV CD spectropolarimetry was used to evaluate the secondary structure of purified AR TM1-3. This was carried out in 50% HFIP, in SDS and LPPG micelles at pH 6.00 \pm 0.05 and pH 7.00 \pm 0.05 and in DPC micelles at pH 4.00 \pm 0.05. In the case of DPC micelle solutions, AR TM1-3 was not soluble at higher pH than 4 – this is consistent with behaviour of the AR55 fragment studied previously (Langelaan *et al.*, 2013). Far-UV CD spectra for all of the samples except the DPC micelle sample showed characteristics of an α -helix-rich protein (Cantor & Schimmel, 1980), with a positive

band at ~192 nm ($\pi \rightarrow \pi^*$ transition) and negative bands at ~222 nm ($n \rightarrow \pi^*$ transition) and 208 nm ($\pi \rightarrow \pi^*$ transition) (Figure 3.2A). The signal intensity was higher in 50% HFIP. This could be indicative of a higher degree of structuring but could also be due to differences in $\Delta\epsilon$ (absorptivity) in HFIP/H₂O mixtures vs. aqueous solution. Therefore, the data are also shown in a normalized manner (with the minimum $[\theta]$ value for each spectrum used as the normalization factor; Figure 3.2B) to allow for evaluation of relative secondary structure content between samples. AR TM1-3 had slightly higher helical content in 50% HFIP, as is evident from higher signal intensity for the $\pi \rightarrow \pi^*$ transitions at 192 nm and 208 nm (Figure 3.2B).

Estimation of secondary structure composition was carried out through CD spectral deconvolution using the CDSSTR algorithm (Johnson, 1999) and employing the SMP180 (Abdul-Gader *et al.*, 2011) mixed soluble/membrane protein reference dataset (Table 3.5). Low normalized root mean square deviation (NRMSD) values demonstrate that reliable fits were obtained. The α -helical proportion predicted for AR TM1-3 by deconvolution in SDS micelles (~42-43%) is highly comparable to the ~40% expected from estimates based on the canonical GPCR topology and previous report from our lab (Langelaan *et al.*, 2013). Deconvolution of CD data for LPPG micelle samples also exhibited a comparable ~53% α -helical content (Table 3.5). It is also evident that the secondary structure composition of the protein in SDS and LPPG micelles is essentially the same at both pH 6.00 \pm 0.05 and 7.00 \pm 0.05 (Figure 3.3 and Table 3.5). Although, in comparison to SDS micelle condition, AR TM1-3 seems to have a higher helical content in LPPG micelle condition. This is also upheld by the CD data deconvolution results for both conditions (Table 3.5). LPPG micelle conditions were not pursued further because the samples were not stable for more than a day as evident from the increasing precipitate

observed in the sample with time. Despite the suitability of DPC for solubilization of AR55 in our previous structural study (Langelaan *et al.*, 2013), the CD spectrum for the DPC sample of AR TM1-3 resembles that of a protein rich in β -strand secondary structure, as upheld by deconvolution (only $\sim 3\%$ helical structure is estimated). This is indicative of improper folding and/or aggregation; hence, DPC micelle conditions were also not pursued further.

3.3.2. Solution-State NMR Spectroscopy:

Using uniformly ^{15}N and ^{13}C -enriched AR TM1-3, we performed 1D ^1H and 2D ^1H - ^{15}N HSQC NMR experiments in different solvent and micellar conditions and at a variety of temperatures in order to gauge feasibility of a future full-scale high-resolution NMR structural study. In particular, AR TM1-3 in DMSO showed much poorer peak dispersion (Figure 3.4) than in SDS micelles (figure 3.5) or 50% HFIP (Figure 3.6). On the basis of these results, DMSO was not pursued further. Consistent with the far-UV CD results, both 1D ^1H and 2D ^1H - ^{15}N HSQC spectra indicate that AR TM1-3 may be appropriately structured in 50% HFIP and SDS, as indicated by the dispersion of signals and the number of peaks in the HSQC spectra. Significant peak overlap is apparent in all conditions tested, as is typically observed in α -helical membrane proteins (Kim *et al.*, 2009). The global conformation and/or local dielectric constant experienced by a given region of the protein may change with the solvent system, as indicated by changes in the HSQC spectral peak patterns. The overall structuring under a particular solvent condition does not appear to have substantial temperature-dependent change, on the basis of monitoring HSQC peak positions as the temperature changes (HFIP, Figure 3.7; SDS, Figure 3.8). Unchanged spectral appearance even after weeks also demonstrates sample

stability over time, which is essential for pursuing triple-resonance NMR structural studies for full chemical shift assignment, structure determination and dynamics characterization.

Full resonance assignment has not yet been performed. However, distinct cross-peaks arising from the side chain H ϵ -N ϵ spin-pair of each of the four tryptophan residues in AR TM1-3 are clearly visible in all of the tested conditions (zooms in Figure 3.4-3.6). In the 50% HFIP, SDS micelle, four major Trp H ϵ -N ϵ peaks of ~equal intensity are clearly observable. Some additional peaks of much lower intensity are observed in HFIP (1 peak) and SDS (3 peaks) samples, indicating the likely presence of a minor conformation in a subset of tryptophan residues. As a whole, the number of cross-peaks (93 in SDS and 117 in HFIP; out of 147 expected) in HSQC spectra in SDS and HFIP samples at 47°C (Figure 3.5 and 3.6) seemed appropriate and these sample conditions appeared quite amenable for future high-resolution structural studies of AR TM1-3. Triple resonance NMR data have been collected in both conditions (Table 3.2-3.3).

Significant overlap was still apparent in the spectra arising from triple resonance experiments for both SDS micelle and 50% HFIP samples. For this reason, based on experience from HFIP samples and literature precedents (Potetinova *et al.*, 2012), another sample was prepared in 50% TFE at a lower protein concentration (0.12 mM). Similar to the 50% HFIP condition, four major Trp H ϵ -N ϵ peaks of ~equal intensity were clearly observable with an additional peak of much lower intensity (Figure 3.9). In comparison to the HFIP sample (at 0.7 mM), reduction in protein concentration for the TFE sample showed a significant improvement in peak dispersion, as represented in Figure 3.10. For the 50% TFE condition, the number of peaks observed in the HSQC were higher than that for the HFIP condition (138 in TFE vs 117 in HFIP). Using 3D backbone walk

experiments, I was only able to unambiguously assign C α , C β , C', and HN for 31 residues out of 152. Figure 3.11 shows a partially assigned ^1H - ^{15}N HSQC spectrum. Further assignment was difficult because of excessive overlap. Lack of some peaks in the 3D dataset also made it difficult to walk along the backbone. In my opinion, further condition optimization and higher number of scans for the 3D experiments may be required to obtain an easy to assign data set.

3.4. Discussion:

As discussed in the introduction, membrane proteins need to be in an environment closely mimicking the native environment, which requires judicious selection of a membrane-mimetic that is also amenable to technique of choice (Sanders & Mittendorf, 2011). There is a large number of membrane-mimetic systems (Warschawski *et al.*, 2011), varying in various physical aspects such as thickness, shape, curvature, lateral pressure, dielectric constant and hydration (Cross *et al.*, 2011). Out of various conditions tested, 50% TFE condition showed the best NMR spectral characteristics, particularly at low protein concentration.

The membrane-mimetic effect of TFE is caused by an increase in the order of the solvent, making solvent-peptide interactions energetically unfavourable and subsequently promoting interactions within the peptide backbone (i.e., driving secondary structure formation) (Buck, 1998). This effect is most likely due to the ability of TFE to coat the surface of the peptide with hydrophobic sidechains. Moreover, TFE does not severely disrupt hydrophobic interactions within peptides, due to its weak interaction with nonpolar residues. However, isotropic solvents such as TFE lack the chemical and structural heterogeneity of lipid bilayers, and may disrupt tertiary and quaternary contacts

(Roccatano *et al.*, 2002). Nevertheless, even in case of a lack of tertiary contacts, literature precedents show that structural information obtained from chemical shift assignments for AR TM1-3 in TFE conditions can be used to derive secondary structural information (Wishart *et al.*, 1992; Wishart & Sykes, 1994b).

In this work, biophysical studies in organic solvent/water mixtures (50% HFIP in water; 50% TFE in water) and in SDS micelles indicate that this GPCR fragment has appropriate secondary structuring and is suitable for long-term characterization by solution-state NMR spectroscopy. Although, complete assignment has not been performed yet in any of the conditions. The data acquired in 50% TFE condition may be revisited for further assignments. However, finding a more suitable condition may be a more practical thing to do. There is compelling evidence in the literature to suggest that the N-terminal tail and extracellular loops of class A GPCRs plays a crucial role in ligand binding and receptor function (Wheatley *et al.*, 2012). AR TM1-3 will therefore also provide a useful probe for the structural features and dynamics of the N-terminal tail and first extracellular loop of AR upon ligand binding. In future, NMR studies should also be carried out on AR TM1-3 to evaluate the effects of other membrane-mimetic environments, such as nanodiscs and bicelles, upon the structure of this protein. Solid-state NMR has strong potential for characterization of both structure and dynamics of membrane proteins, and is still being actively developed. Recent advances in solid-state NMR have enabled structural characterization of membrane proteins in environments very close to the native membranes, as very nicely reviewed in (Opella, 2015) and (Judge *et al.*, 2015). AR TM1-3 is also a suitable candidate that in future can be subjected to solid-state NMR experimentation and studied in a membrane environment that is arguably more similar to native membranes.

In terms of the challenges faced in characterization of AR TM1-3 using solution-state NMR, lessons can be learned from the works on structural characterization of sensory rhodopsin II (Nietlispach & Gautier, 2011) and proteorhodopsin (Reckel *et al.*, 2011). These represent the first solution-state NMR structures for membrane proteins containing seven transmembrane segments and, hence, are relevant to the apelin receptor. Despite the fact that both of these structures were solved in DHPC micelles at the same temperature, it is remarkable to see that the strategies used in their characterization were very different. These protein were expressed in different expression system yielding disparate labeling schemes, leading to different assignment methods for the backbone and sidechain resonances (Gautier *et al.*, 2008; Reckel *et al.*, 2011).

In both of these cases, the problem of line broadening due to unfavourable relaxation properties of nuclei in proteins of this size were addressed by deuteration of the protein. Since the magnetic moment of a deuteron is approximately one-sixth that of a proton, perdeuteration of a protein dramatically reduces the total dipolar contribution to transverse relaxation rates of any remaining protons and also the ^{13}C and ^{15}N heteronuclei. In addition to perdeuteration, transverse relaxation optimized spectroscopy (TROSY) based experiments at high magnetic fields were also used to obtain better resolved spectra (Gardner & Kay, 1998; Fernandez & Wider, 2003). Non-uniform sampling (NUS) (Barna *et al.*, 1987) schemes in combination with maximum entropy reconstruction (Rovnyak *et al.*, 2004) were also significant contributors enabling successful backbone assignment for sensory rhodopsin II. Selective labeling of sidechains either using oro-SAIL Leu- and Val for proteorhodopsin or selective methyl protonation of the hydrophobic amino acids (Ala β , Ile δ_1 , Leu δ , Met ϵ and Val γ) for sensory rhodopsin II also played an important role in assisting the assignment of sidechains

(Rosen *et al.*, 1996; Kainosho *et al.*, 2006). Paramagnetic relaxation enhancement (PRE) (Gottstein *et al.*, 2012) and residual dipolar coupling (RDC) (Lipsitz & Tjandra, 2004) derived restraints were also used in the final structure calculation of these proteins.

Based on the success achieved for these proteins, it is worthwhile to use some if not all of these approaches to obtain a more workable dataset. However, since AR TM1-3 is smaller than the proteins discussed above, there is definitely still merit in evaluating other membrane-mimetic conditions before investing the time and expense required to implement these approaches.

Table 3.1: List of 2D experiments used to access the effect of temperature on spectral quality of AR TM1-3.

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	¹ H frequency (MHz)	Facility
100% DMSO, 22 °C							
¹ H- ¹⁵ N HSQC	1	72	¹ H:2048 ¹⁵ N:320	¹ H:8012.8 ¹⁵ N:1520.5	¹ H:4.6 ¹⁵ N:111.6	500	NMR-3
50% HFIP, 22 °C							
¹ H- ¹⁵ N HSQC	1	64	¹ H:2048 ¹⁵ N:256	¹ H:7002 ¹⁵ N:1115.5	¹ H:4.53 ¹⁵ N:111.6	500	NMR-3
50% HFIP, 37 °C							
¹ H- ¹⁵ N HSQC	1	40	¹ H:2048 ¹⁵ N:256	¹ H:7002 ¹⁵ N:1115	¹ H:4.53 ¹⁵ N:111.6	500	NMR-3
50% HFIP, 47 °C							
¹ H- ¹⁵ N HSQC	1	48	¹ H:2048 ¹⁵ N:256	¹ H:7002 ¹⁵ N:1115	¹ H:4.69 ¹⁵ N:111.6	500	NMR-3
SDS, 47 °C							
¹ H- ¹⁵ N HSQC	1	16	¹ H:2048 ¹⁵ N:128	¹ H:9803.9 ¹⁵ N:2128.6	¹ H:4.7 ¹⁵ N:118	700	NRC-IMB

Table 3.2: List of 2D and 3D experiments acquired for AR TM1-3 in 50% HFIP.

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	¹ H frequency (MHz)	Facility
¹ H- ¹³ C HSQC	1	8	¹ H:2048 ¹³ C:256	¹ H:9804 ¹³ C:14084.7	¹ H:4.7 ¹³ C:40.0	700	NRC-IMB
¹ H- ¹⁵ N HSQC	1	8	¹ H:2048 ¹⁵ N:256	¹ H:8417.5 ¹⁵ N:2128.6	¹ H:4.7 ¹⁵ N:118	700	NRC-IMB
HNCO	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:9804 ¹³ C:1408.6 ¹⁵ N:1419	¹ H:4.7 ¹³ C:174 ¹⁵ N:115	700	NRC-IMB
HNcaCO	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:9804 ¹³ C:1408.6 ¹⁵ N:1419	¹ H:4.7 ¹³ C:174 ¹⁵ N:115	700	NRC-IMB
HNCA	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:9804 ¹³ C:4753.6 ¹⁵ N:1419	¹ H:4.7 ¹³ C:51.5 ¹⁵ N:115	700	NRC-IMB
HNcoCA	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:9804 ¹³ C:4753.6 ¹⁵ N:1419	¹ H:4.7 ¹³ C:51.5 ¹⁵ N:115	700	NRC-IMB
CBCANH	1	8	¹ H:2048 ¹³ C:128 ¹⁵ N:48	¹ H:9804 ¹³ C:10211.5 ¹⁵ N:1419	¹ H:4.7 ¹³ C:39 ¹⁵ N:115	700	NRC-IMB
CBCAcoNH	1	8	¹ H:2048 ¹³ C:128 ¹⁵ N:48	¹ H:9804 ¹³ C:10211.5 ¹⁵ N:1419	¹ H:4.7 ¹³ C:39 ¹⁵ N:115	700	NRC-IMB

Table 3.3: List of 2D and 3D experiments acquired for AR TM1-3 in SDS micelles.

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	¹ H frequency (MHz)	Facility
¹ H- ¹³ C HSQC	1	16	¹ H:2048 ¹³ C:512	¹ H:9804 ¹³ C:14084.8	¹ H:4.7 ¹³ C:40	700	NRC-IMB
¹ H- ¹⁵ N HSQC	1	16	¹ H:2048 ¹⁵ N:128	¹ H:9804 ¹⁵ N:2128	¹ H:4.7 ¹⁵ N:118	700	NRC-IMB
HNCO	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:8417.5 ¹³ C:1232.6 ¹⁵ N:1589	¹ H:4.7 ¹³ C:174 ¹⁵ N:115.8	700	NRC-IMB
HNcaCO	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:8417.5 ¹³ C:1232.6 ¹⁵ N:1589	¹ H:4.7 ¹³ C:174 ¹⁵ N:115.8	700	NRC-IMB
HNCA	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:8417.5 ¹³ C:4577.6 ¹⁵ N:1589	¹ H:4.7 ¹³ C:52 ¹⁵ N:115.8	700	NRC-IMB
HNcoCA	1	8	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:8417.5 ¹³ C:4577.6 ¹⁵ N:1589	¹ H:4.7 ¹³ C:52 ¹⁵ N:115.8	700	NRC-IMB
CACBNH	1	8	¹ H:2048 ¹³ C:128 ¹⁵ N:48	¹ H:8417.5 ¹³ C:10211.5 ¹⁵ N:1589	¹ H:4.7 ¹³ C:39 ¹⁵ N:115.8	700	NRC-IMB
CACBcoNH	1	16	¹ H:2048 ¹³ C:128 ¹⁵ N:48	¹ H:8417.5 ¹³ C:10211.5 ¹⁵ N:1589	¹ H:4.7 ¹³ C:39 ¹⁵ N:115.8	700	NRC-IMB

Table 3.4: List of 2D and 3D experiments acquired for AR TM1-3 in 50% TFE

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	¹ H frequency (MHz)	Facility
¹ H- ¹³ C HSQC	1	16	¹ H:2048 ¹³ C:256	¹ H:8418 ¹³ C:14085	¹ H:5.296 ¹³ C:40.0	700	NRC-IMB
¹ H- ¹⁵ N HSQC	1	16	¹ H:2048 ¹⁵ N:192	¹ H:8417.5 ¹⁵ N:1703	¹ H:5.296 ¹⁵ N:117.5	700	NRC-IMB
HNCO	1	16	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:8417.5 ¹³ C:1409 ¹⁵ N:1362	¹ H:5.296 ¹³ C:175 ¹⁵ N:116.4	700	NRC-IMB
HNcaCO	1	32	¹ H:2048 ¹³ C:48 ¹⁵ N:48	¹ H:8417.5 ¹³ C:1409 ¹⁵ N:1362	¹ H:5.296 ¹³ C:175 ¹⁵ N:116.4	700	NRC-IMB
HNCA	1	24	¹ H:2048 ¹³ C:56 ¹⁵ N:48	¹ H:8417.5 ¹³ C:4753.7 ¹⁵ N:1362	¹ H:5.296 ¹³ C:52.8 ¹⁵ N:116.4	700	NRC-IMB
HNcoCA	1	24	¹ H:2048 ¹³ C:56 ¹⁵ N:48	¹ H:8417.5 ¹³ C:4753.7 ¹⁵ N:1362	¹ H:5.296 ¹³ C:52.8 ¹⁵ N:116.4	700	NRC-IMB
CBCANH	1	40	¹ H:2048 ¹³ C:132 ¹⁵ N:48	¹ H:8417.5 ¹³ C:10563.6 ¹⁵ N:1561	¹ H:4.703 ¹³ C:39 ¹⁵ N:117	700	NRC-IMB

Table 3.5: Deconvolution of CD data using the CDSSTR algorithm and SMP180 reference dataset.

Conditions	Secondary structure ^a
AR TM1-3 in 50% HFIP	Helix 1- 0.55, Helix 2- 0.18 Strand 1- 0.01, Strand 2- 0.01 Turn – 0.09, Unordered- 0.14 NRMSD:0.010
AR TM1-3 in SDS micelles at pH 7.00 ±0.05	Helix 1- 0.23, Helix 2- 0.20 Strand 1- 0.06, Strand 2- 0.05 Turn – 0.17, Unordered- 0.30 NRMSD:0.022
AR TM1-3 in SDS micelles at pH 6.00 ±0.05	Helix 1- 0.24, Helix 2- 0.18 Strand 1- 0.08, Strand 2- 0.06 Turn – 0.14, Unordered- 0.29 NRMSD:0.019
AR TM1-3 in LPPG micelles at pH 7.00 ±0.05	Helix 1- 0.39, Helix 2- 0.14 Strand 1- 0.09, Strand 2- 0.05 Turn – 0.12, Unordered- 0.20 NRMSD:0.009
AR TM1-3 in LPPG micelles at pH 6.00 ±0.05	Helix 1- 0.38, Helix 2- 0.15 Strand 1- 0.09, Strand 2- 0.06 Turn – 0.12, Unordered- 0.22 NRMSD:0.007
AR TM1-3 in DPC micelles at pH 4.00 ±0.05	Helix 1- 0.02, Helix 2- 0.01 Strand 1- 0.43, Strand 2- 0.20 Turn – 0.03, Unordered- 0.29 NRMSD:0.012

^a Helix 1- regular α -helix; Helix 2 - distorted α -helix; Strand 1- regular β -strand; Strand 2- distorted β -strand; NRMSD- normalized root mean square deviation.

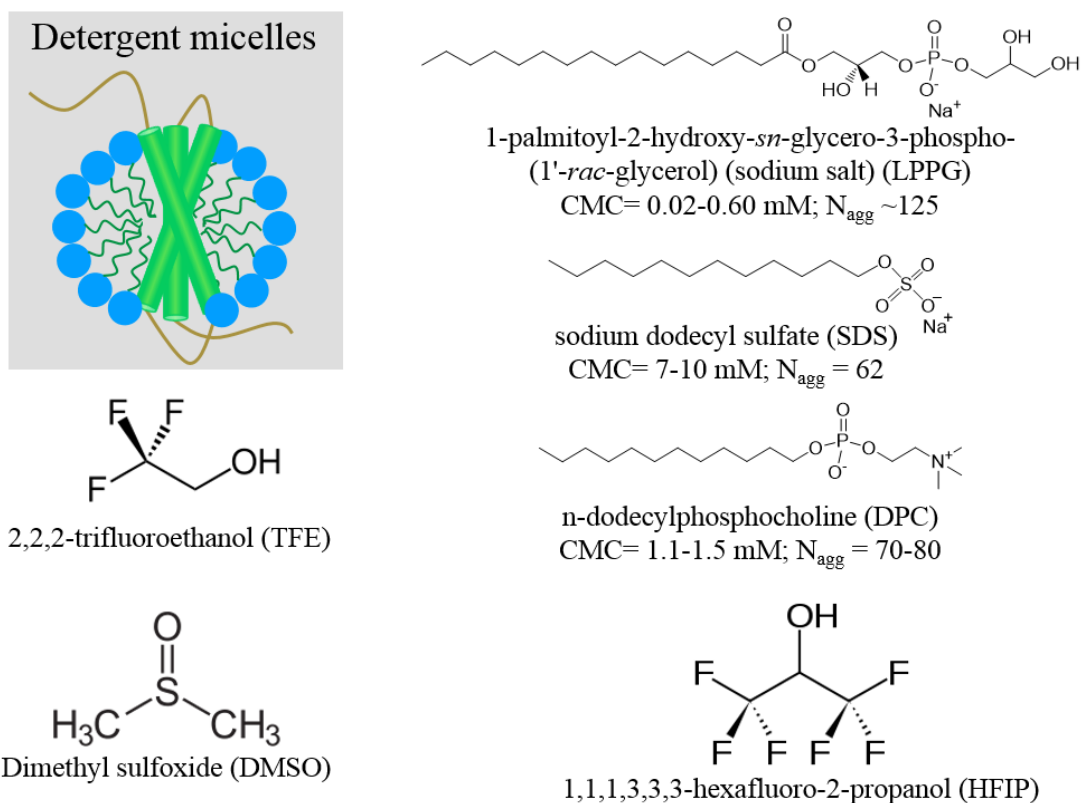


Figure 3.1: Representation of membrane-mimetics used in this work.

The figure shows a cartoon representation of a membrane protein in a micelle, which can be formed using detergents such as SDS, DPC and LPPG. The three organic solvents (TFE, DMSO and HFIP) used in this work are also shown. Here CMC stands for critical micelle concentration and N_{agg} stands for aggregation number (Lipfert *et al.*, 2007; Linke, 2009).

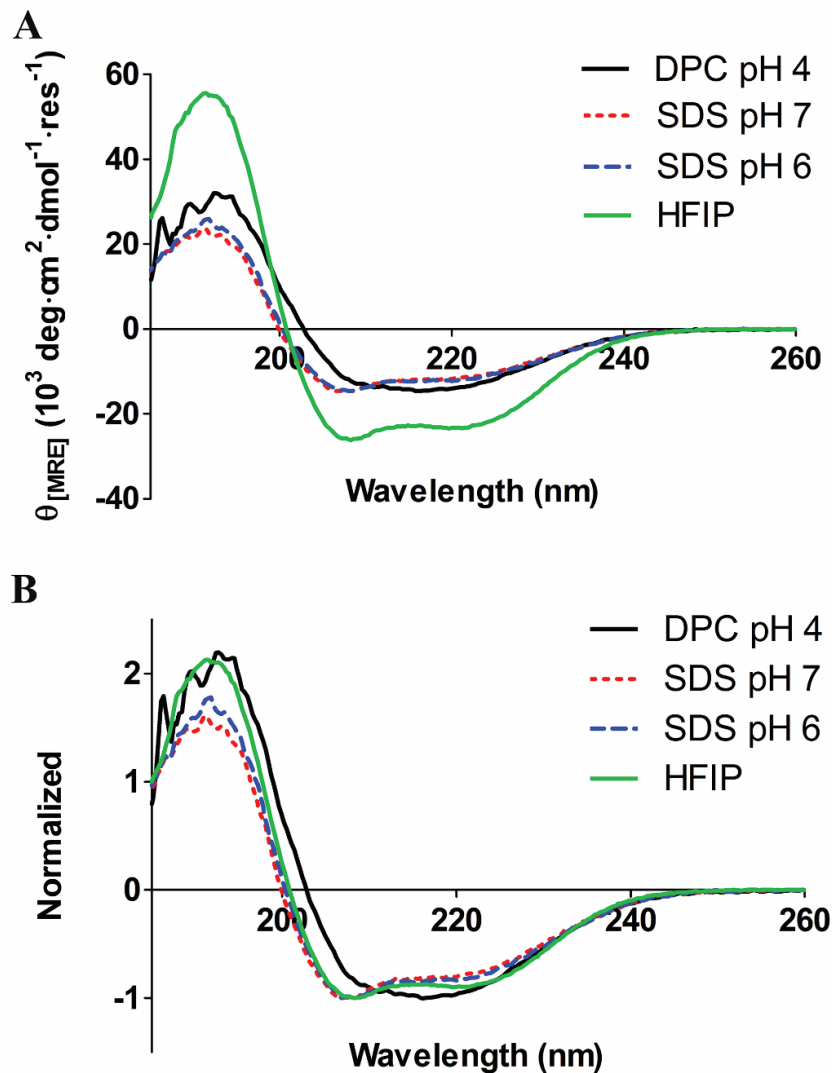


Figure 3.2: Comparison of secondary structure composition of AR TM1-3 in membrane-mimetics using far-UV CD. (A) Far-UV CD spectra of AR TM1-3 in indicated solvent or micellar environment, allowing evaluation of secondary structure composition. (B) Normalized representation (normalized to absolute minimum $[\theta]$), allowing direct comparison of relative helical content. HFIP refers to 50% HFIP in H₂O.

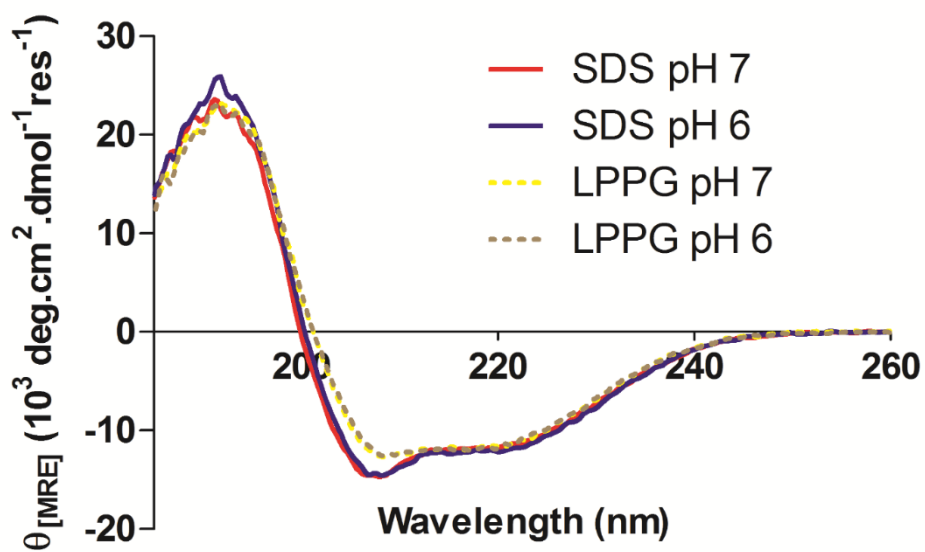


Figure 3.3: Comparison of secondary structure composition of AR TM1-3 in SDS and LPPG micelles using far-UV CD. Far-UV CD spectra of AR TM1-3 in indicated micellar environment, allowing evaluation of secondary structure composition.

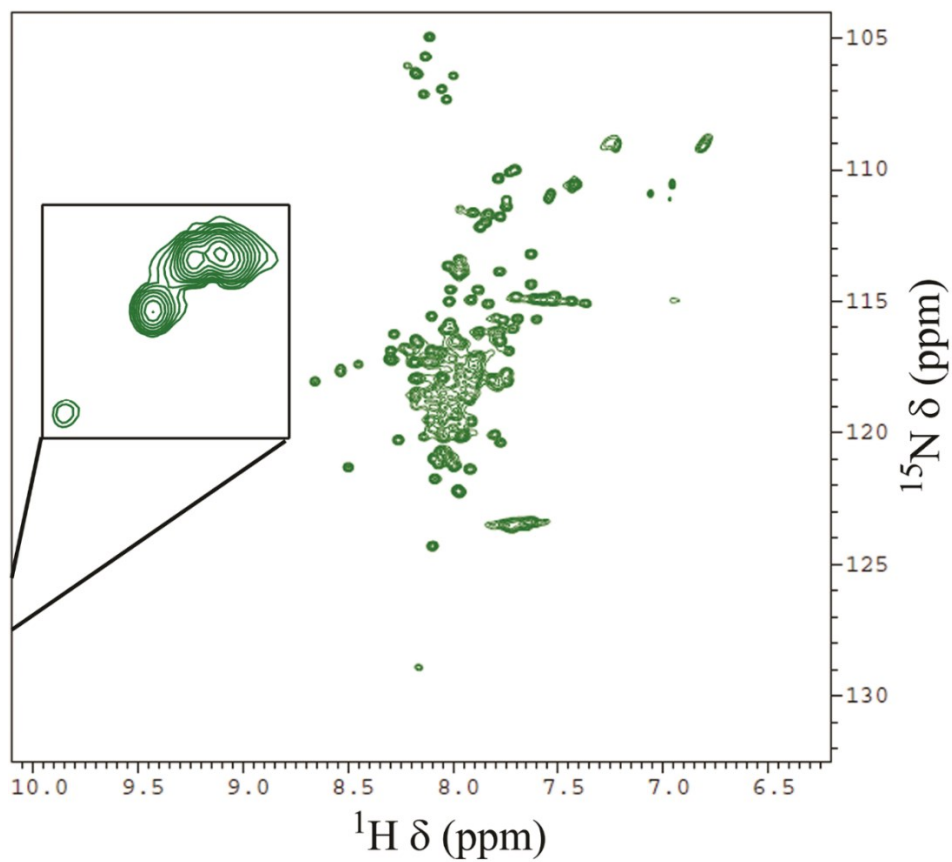


Figure 3.4: ^1H - ^{15}N HSQC spectra for AR TM1-3 in DMSO at 22°C. The zoomed region illustrates peaks arising from tryptophan side chain $\text{H}\epsilon$ - $\text{N}\epsilon$ and covers 10.69-10.93 ppm and 131.00-132.75 ppm in ^1H and ^{15}N dimensions, respectively.

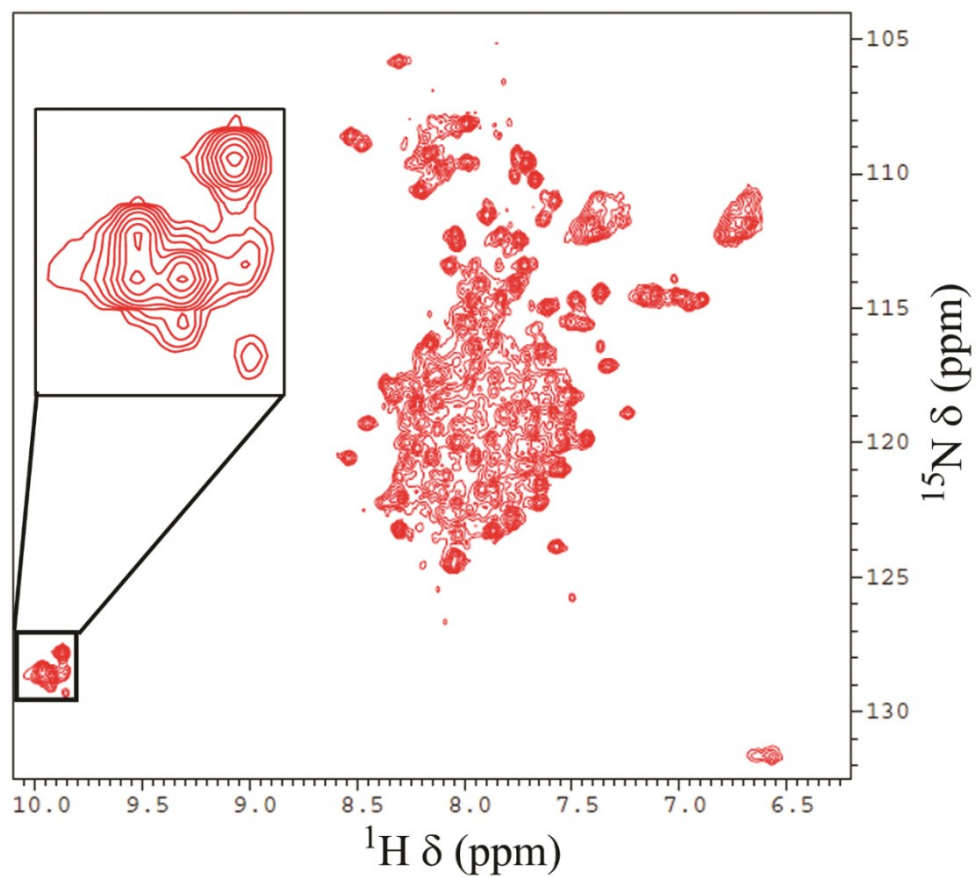


Figure 3.5: ^1H - ^{15}N HSQC spectra for AR TM1-3 in SDS micelles at 47°C. The zoomed region illustrates peaks arising from tryptophan side chain $\text{H}\epsilon$ - $\text{N}\epsilon$.

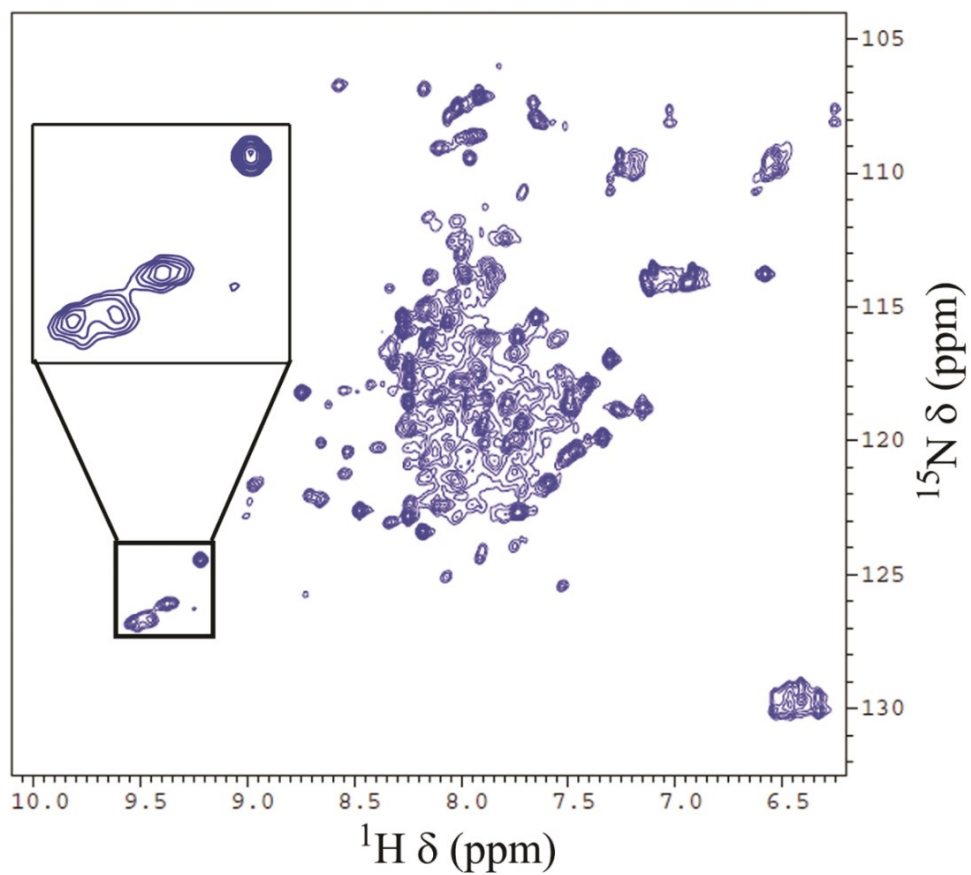


Figure 3.6: ^1H - ^{15}N HSQC spectra for AR TM1-3 in 50% HFIP at 47°C. Zoomed regions illustrates peaks arising from tryptophan side chain H ϵ -N ϵ .

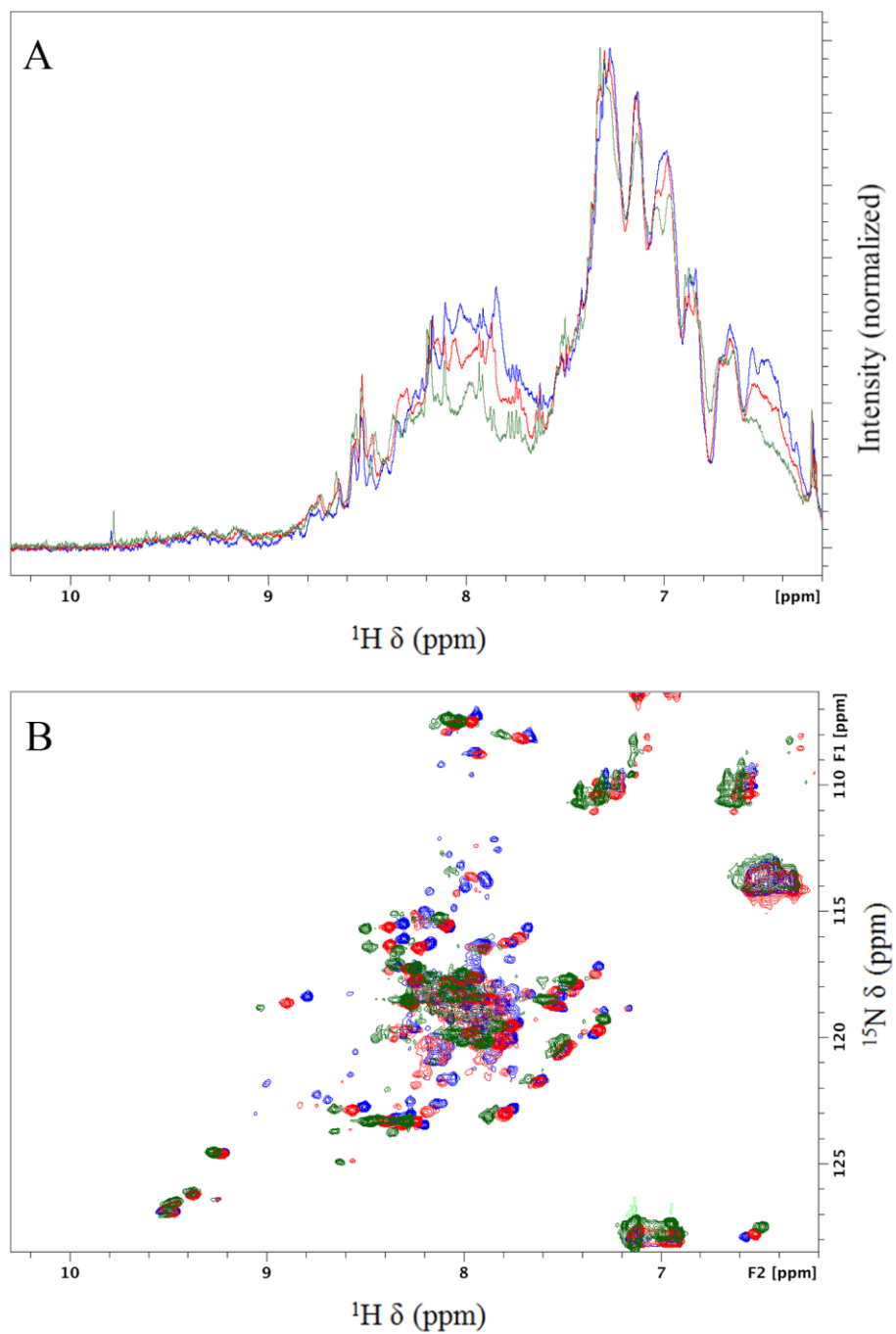


Figure 3.7: Effect of temperature on 1D ^1H and 2D ^1H - ^{15}N HSQC spectra for AR TM1-3 in 50% HFIP (A,B; red: 47°C, green 37°C, blue 22°C).

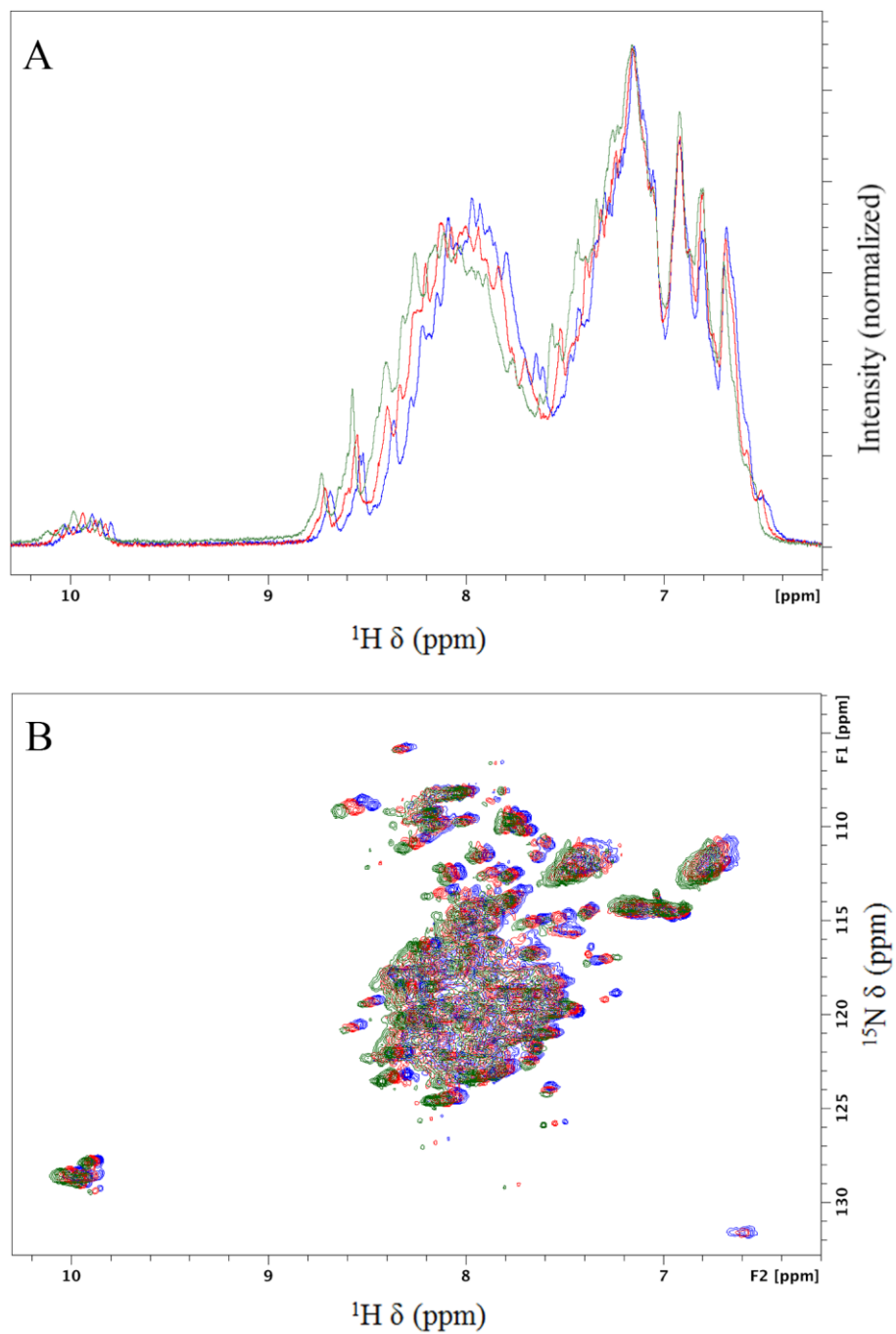


Figure 3.8: Effect of temperature on 1D ^1H and 2D ^1H - ^{15}N HSQC spectra for AR TM1-3 in SDS micelles (A,B; red: 47°C, green 37°C, blue 27°C).

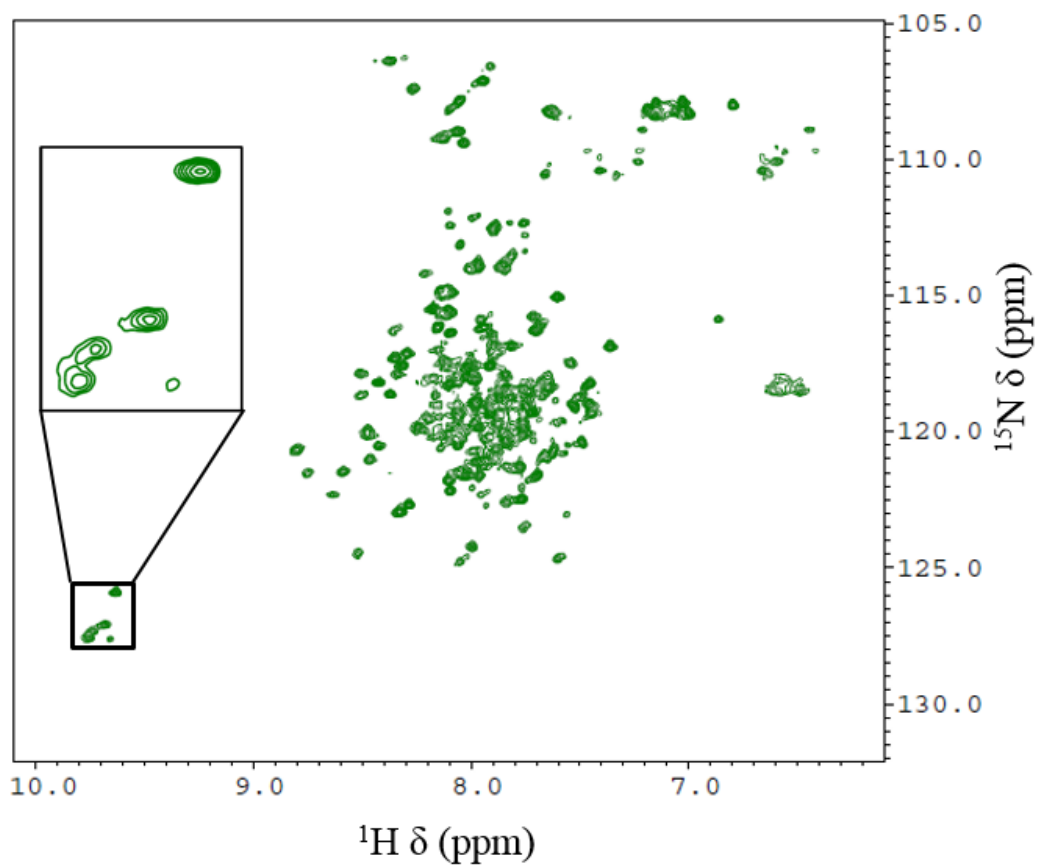


Figure 3.9: ^1H - ^{15}N HSQC spectra for AR TM1-3 in 50%HFIP at 47°C. Zoomed regions illustrates peaks arising from tryptophan side chain $\text{H}\epsilon$ - $\text{N}\epsilon$.

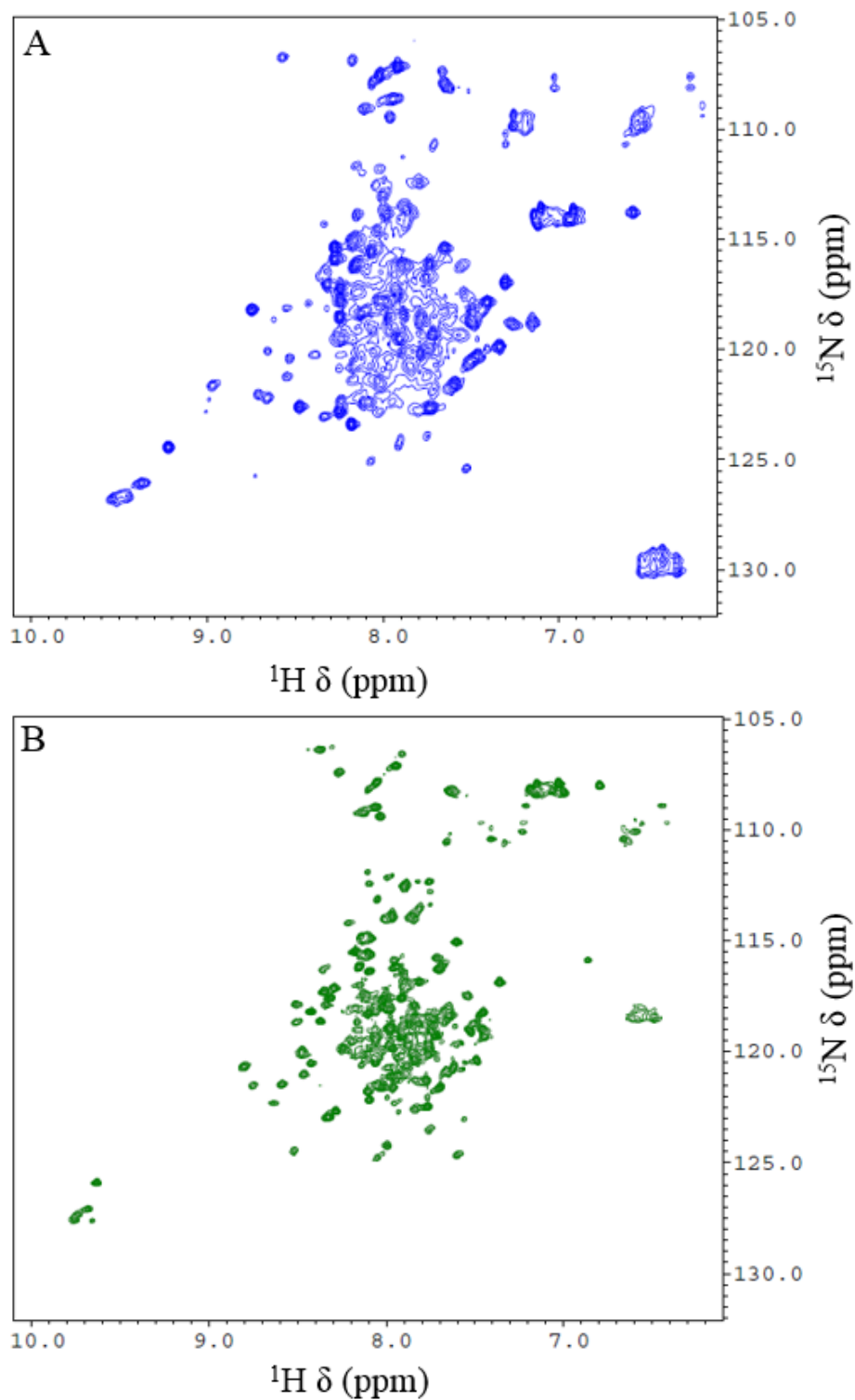
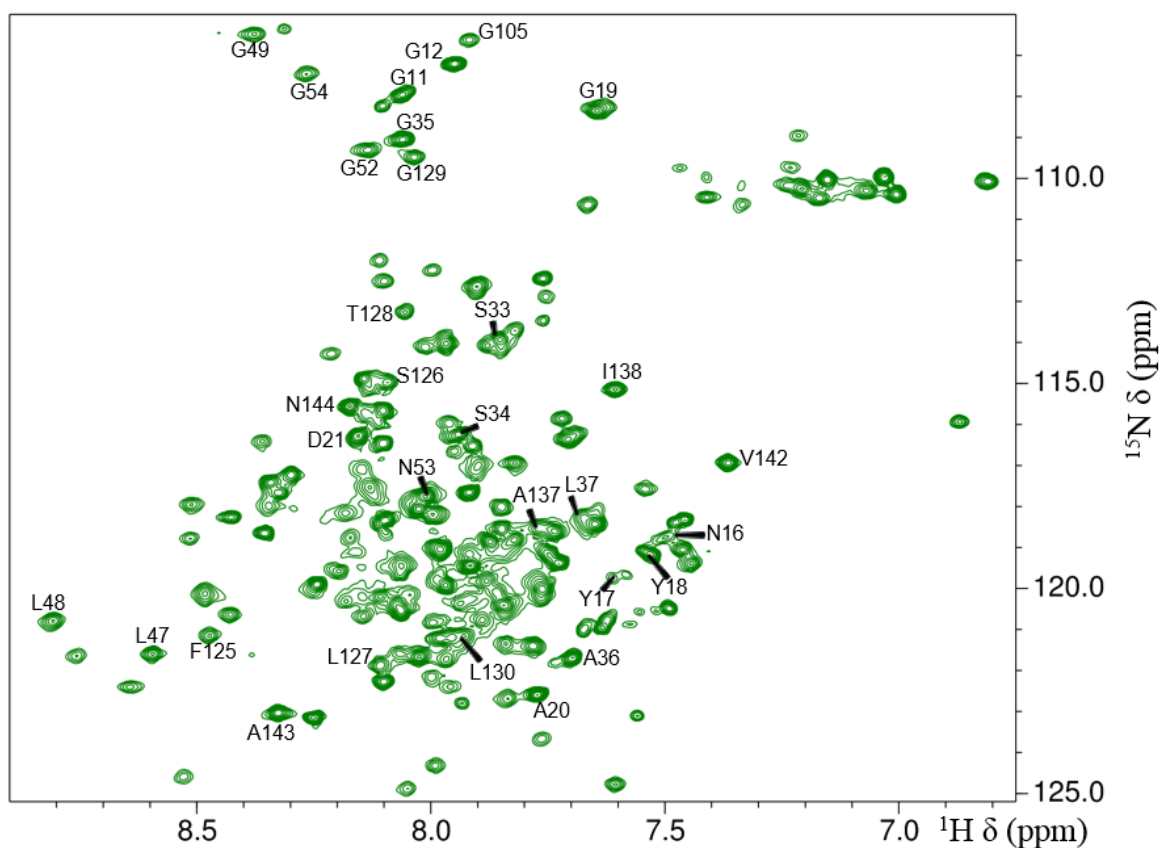


Figure 3.10: Comparison of the effect of protein concentration on ^1H - ^{15}N HSQC spectra for AR TM1-3 in (A) 50% HFIP (0.7 mM; blue) and (B) 50% TFE (0.12 mM; green).



Amino acid sequence of AR TM1-3

MKYYKYMEEGGDFDNYYGADNQSECEYTDWKSSGALIPAIYMLVF
 LLGTTGNGLVLWTVFRSSREKRRSADIFIASLAVADLTFVVTLPPLWAT
 YTYRDYDWPFGTFFCKLSSYLIFVNMYSVFSLTGLSFDRYLAIVRPV
 ANLEHHHHHH

Figure 3.11: Partially assigned ^1H - ^{15}N HSQC for AR TM1-3 in 50% TFE. ^1H - ^{15}N HSQC spectrum acquired at 47 °C on a 700 MHz spectrometer using a sample having 0.12 mM AR TM1-3 solubilized in 50% TFE and water mixture. Highlight represents the assigned amino acid.

Chapter 4: Use Of Split-Inteins To Obtain Full-Length AR

4.1. Introduction:

Advancements in the field of structural biology have led researchers to engage in characterization of larger and more complex target proteins. As mentioned earlier, to “divide and conquer” has become a very popular strategy to study such proteins. Although studying proteins as small fragments or domains provides important structural information, characterization of the full-length protein remains of paramount importance. When a full-length protein is produced through ligation of segments (small fragments or domains), one can carry out segmental isotope-labeling for NMR studies of full-length protein. The ligation reaction to join protein fragments may be catalyzed by split-inteins, which represent a subgroup of protein splicing elements referred to as inteins.

The term intein is derived from “internal protein” as inteins are embedded in a precursor protein in the same way as introns in a pre-mRNA. The host protein sequences flanking an intein are known as N-extein and C-extein (figure 4.1) (Volkman & Iwai, 2010). During the splicing process, the intein is excised, the flanking extein sequences are joined by a peptide bond, and the host protein assumes its normal folding and function (Gogarten *et al.*, 2002). It is well known that the intein splicing reaction does not require any co-factors; rather, it is a reaction that is self-catalyzed by the intein, which can be efficiently performed by optimizing the reaction conditions (Paulus, 2000). Split-inteins consists of two fragments (N-intein and C-intein) encoded by two separately transcribed and translated genes (figure 4.2) (Elleuche & Poggeler, 2010). The N-intein and C-intein of a split-intein can associate non-covalently and catalyze a protein *trans*-splicing

reaction (Elleuche & Poggeler, 2010). Protein ligation by intein-mediated *trans*-splicing does not require energy equivalents such as ATP, but it is dependent on the intein structure plus the first residue (called the +1 residue) of the C-extein which must be a cysteine, serine, or threonine.

The currently accepted canonical mechanism of protein *trans*-splicing proceeds via four concerted nucleophilic displacement reactions (Xu & Perler, 1996) (figure 4.3) preceded, in the case of split-inteins, by an association step between the N-intein (I_N) and C-intein (I_C) (Volkman & Iwai, 2010). All known inteins share a low degree of sequence similarity, with conserved residues at the N- and C-termini. Most inteins begin with Ser or Cys and end in His-Asn or His-Gln. The first amino acid of the C-extein is an invariant Ser, Thr, or Cys, but the residue preceding the intein at the C-terminus of the N-extein is not conserved (Elleuche & Poggeler, 2010) (figure 4.1). However, the positioning of certain residues proximal to the intein-splicing junction in both the N- and C-terminal exteins was recently found to accelerate or attenuate protein splicing (Amitai *et al.*, 2009).

A number of natural split-inteins have now been reported (Wu *et al.*, 1998; Aranko *et al.*, 2014). Split-inteins can also be artificially engineered from contiguous inteins by dividing the intein sequence into two parts through recombinant DNA technology (Perler, 2002; Sun *et al.*, 2004). Non-canonical split sites can be chosen, usually in the loop regions of the intein, to make I_N and I_C fragments of desirable sizes (Volkman & Iwai, 2010). For example, choosing a S1 site (one of the 14 split-sites shown in (Sun *et al.*, 2004)) makes the I_N shorter, whereas choosing a S11 site makes the I_C shorter. Although the ligation efficiency of non-canonical split-inteins is generally

lower than their canonically split variant (Volkman & Iwai, 2010), they are still an attractive tool for protein engineering.

Protein engineering using inteins has many applications in the field of structural biology. Proteins that are difficult to produce in full-length form can be produced as smaller fragments that can be co- or post-translationally spliced to produce the full-length protein (Figure 4.4). Intein-mediated protein *trans*-splicing can be carried out both *in vitro* and *in vivo* (Aranko *et al.*, 2009). The former is particularly valuable for protein fragments that are not soluble and are expressed in inclusion bodies, since these can be solubilized later for splicing *in vitro*. Use of *in vitro* splicing is also the most straightforward approach to allow intein-mediated segmental isotope-labeling for NMR spectroscopy (Zuger & Iwai, 2005; Tremblay *et al.*, 2015). Alternatively, *in vivo* splicing may be used by expressing the two split intein fused fragments of AR using different promoter systems in the same cell. The two promoters can be activated after exchanging the expression medium to enable isotope-labeling of one of the fragments (Aranko *et al.*, 2009). In the segmental labeling approach, a specific protein fragment, as opposed to the entire protein, is isotope-labeled, greatly reducing signal overlapping and improving spectral assignability. Alternatively, different fragments of the protein may be labeled in different manners, allowing isotopically-discriminated NMR experiments (Golovanov *et al.*, 2007) to be carried out in the same fashion as done in (Tremblay *et al.*, 2015).

Joining of more than two segments requires the selected intein pairs to be functionally orthogonal in order to prevent side products due to cross reactivity (Volkman & Iwai, 2010). Therefore, the aforementioned change in strategy to produce AR as two fragments instead of four (outlined in Chapter 2), significantly improves the

probability of being able to efficiently produce a full-length receptor by reducing the number of splicing reactions and split-intein pairs required (figure 4.4).

To date, intein-mediated splicing has been mainly used for soluble proteins. However, intein splicing can also be used to combine fragments of membrane proteins. Until recently, the only example of this was the β -barrel outer membrane protein from *E. coli* (Omp) (Brenzel *et al.*, 2009). Proteorhodopsin, a 7 TM protein was recently been spliced to obtain the functional full-length protein, serving as a proof of principle for this work (Mehler *et al.*, 2015). Thus, intein-mediated splicing for membrane proteins has not yet been exploited to its full potential, principally due to the difficulties in production and handling of these proteins.

Cell-free expression systems are a very attractive alternative for protein production. In the case of membrane proteins, these can be used to overcome some of the difficulties inherent in the traditional methods of protein production and handling, such as cytotoxicity, inefficient translocation, and proteolytic degradation. These systems also offer the possibility to very readily and controllably modulate the expression environment using a large number of membrane-mimetics (Schwarz *et al.*, 2008).

Herein, I report the expression of the AR in two fragments (AR TM1-3 and AR TM4-7) with three different split-inteins using an *E. coli*-based cell-free expression system and demonstrate a cell-free expression-based approach for splicing of membrane proteins (Figure 4.5). Three pairs of split-inteins (Table 4.1) have been tested for expression and subsequent splicing. To date, this represents the only example of expression of a human GPCR in two fragments which can potentially be ligated to obtain a full-length receptor. In future, this will provide the additional advantages of selective

labeling strategies that will be extremely beneficial, and potentially ground-breaking, for the study of GPCR structure, dynamics, and ligand binding by NMR spectroscopy.

4.2. Material And Methods:

4.2.1. Materials:

All strains of *E. coli* were purchased from Lucigen (Middleton, WI). 7.5% Next gel solution for SDS-PAGE was purchased from Amresco (Solon, OH). Nitrocellulose membrane (0.2 μm) and Clarity Western ECL Substrate for western blotting were purchased from Bio-Rad (Mississauga, ON). The Expressway™ Mini Cell-Free Expression kit, HisProbe-HRP, polyoxyethylene(4)lauryl ether (Brij-30), polyoxyethylene-(23)-lauryl-ether (Brij-35), polyoxyethylene-(20)-stearyl-ether (Brij-78) and polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether (Triton X-100) were purchased from Thermo Fisher Scientific Inc. (Waltham, MA). *n*-dodecyl- β -D-maltoside (DDM), maltose neopentyl glycol-3 (MNG), 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (LPPG), 1-palmitoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LPPC) were purchase from Anatrace (Maumee, OH). The Genopure Plasmid Midi Kit was purchased from Roche Life Sciences (Penzberg, Germany). All other materials have been listed in previous chapters. Unless otherwise specified, all reagents were used without further purification.

4.2.2. Construction Of Fusion Genes And Expression Plasmids:

Coding sequences of the split-inteins (listed in Table 4.1) were from existing plasmids in the Liu lab. The plasmids for expression of AR TM1-3 and AR TM4-7 were prepared as described in Chapter 2. A restriction free cloning approach (Unger *et al.*,

2010; Erijman *et al.*, 2011) was used to incorporate the gene encoding each split-intein from the plasmids with split-intein to the plasmids encoding the appropriate AR fragment. Briefly, the gene encoding each split-intein was PCR-amplified using two primers, each of which included a target-specific sequence (i.e split-intein sequence) and a 5' extension that overlapped the insertion site in the destination vector (i.e. the plasmid containing the AR fragment). Following purification, the double-stranded PCR product was used as a pair of mega-primers for the second PCR amplification. During the second PCR reaction, each of the mega-primers annealed to the destination vector at a pre-designed position and got extended in a linear-amplification reaction. The two new DNA strands thus formed gave rise to a plasmid with nicks on each strand of DNA. The parental methylated DNA was then digested by *DpnI* treatment and the newly synthesized plasmid was electroporated into DH5 α strain *E. coli* cells. All the constructs with forward and reverse primers for transfer-PCR are listed in Table 4.2.

4.2.3. Expression Of AR TM1-3-I_N And I_C-AR TM4-7 Constructs In *E. Coli* Cells:

Each construct detailed in Table 4.2 was tested for expression in BL21(DE3), C41(DE3) and C43(DE3) strains of *E. coli*. For test expression of each construct, a single colony was grown in 4 mL LB medium with ampicillin (100 μ g/mL) to an optical density at 600 nm (OD₆₀₀) of ~0.6. Then, 3 mL of this culture was induced with IPTG at a final concentration of 0.5 mM and grown at 37 °C for 3 hours after induction. For expression optimization, the effects of varying IPTG concentration (0.25 mM, 0.5 mM, 0.75 mM and 1 mM) and of varying temperature and expression time (at 22 °C for 14 hours, at 28 °C for 14 hours and at 37 °C for 2, 3 and 4 hours) were tested.

4.2.4. Cell-Free Expression Of AR TM1-3-I_N And I_C-AR TM4-7 Constructs:

The Expressway™ Cell-Free *E. coli* Expression System was used for *in vitro* transcription and translation of the split-intein-fused AR constructs using the plasmids listed in Table 4.2. The plasmid DNA used in the expression was purified from the DH5 α strain of *E. coli* using the Genopure Plasmid Midi Kit. The *in vitro* translation was carried out following manufacturers' protocol. Briefly, the reaction mixture contained *E. coli slyD*-Extract (20 μ L), 2.5X IVPS *E. coli* Reaction Buffer (-amino acids) (20 μ L), 50 mM Amino Acids (-Met), 75 mM Methionine, T7 Enzyme Mix (1 μ L) and 1 μ g DNA template. The final reaction volume was adjusted to 50 μ L, and the reaction was carried out at 30 °C. The expression was confirmed using SDS-PAGE and Western blotting using the protocol detailed in section 2.2.4.

4.2.5. Condition Optimization For Splicing Of The AR Fragments:

Splicing of AR fragments to obtain the full-length AR was tested using two approaches. *In vitro* co-translational splicing was the first approach, where two plasmids encoding the AR TM1-3-I_N and I_C-AR TM4-7 were expressed together in the same cell-free reaction mixture. This approach is likely to improve the probability of enabling assembly of a split-intein pair during translation and to promote subsequent splicing. Mild detergents such as polyoxyethylene(4)lauryl ether (Brij-30; 4 mM), polyoxyethylene-(23)-lauryl-ether (Brij-35; 4 mM), polyoxyethylene-(20)-stearyl-ether (Brij-78; 4 mM) and polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether (Triton X-100; 5 mM) were tested for their ability to enhance protein solubility and splicing without affecting the translation efficiency. Expression was carried out overnight at 30 °C in each of these detergent conditions.

The second approach employed post-translational splicing, where each construct was expressed separately and the splicing was tested by independently solubilizing each protein thus obtained, and then mixing the compatible split-intein partners. In these cases, the splicing reactions were carried out in 20 mM sodium phosphate buffer at pH 7 with 1 mM DTT. Various detergents such as *n*-dodecyl- β -D-maltoside (DDM; 5 mM), maltose neopentyl glycol-3 (MNG; 5 mM), 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-rac-(1-glycerol)] (LPPG; 2 mM), 1-palmitoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LPPC; 2mM) and sodium dodecyl sulfate (SDS; 2 mM) were added in the second splicing reaction in order to improve splicing. For the post-translational splicing approach, the reaction mixture was incubated at 37 °C for 24 hours. The success of each splicing reaction was evaluated by SDS-PAGE and western blotting.

4.3. Results:

4.3.1. Expression In *E. coli* Cells:

Expression of each construct in each strain of *E. coli* seemed to be toxic to the cells, as evident from very low cell density of the cultures where protein expression was induced. Expression could be observed only by western blot, not by SDS-PAGE, because of low level of expression while testing AR TM1-3-I_N constructs for expression. No expression was ever detected for any AR TM 4-7 construct.

4.3.2. Cell-Free Expression:

All the constructs were translated successfully using by cell-free expression (Figures 4.6 and 4.7). The AT-tagged-AR TM1-3-I_N-His construct was poorly expressed in comparison to all of the other constructs tested (Figure 4.7). However, it did express

well during co-translational splicing experiments (Figure 4.8). Expression levels were not quantified. To put these in perspective, translation products from a 20 μL reaction were solubilized in 100 μL 1x loading buffer (final volume); 2 μL of a given 120 μL sample was then loaded onto the SDS-PAGE gel and probed by western blotting (Figure 4.5).

4.3.3. Intein-Mediated Splicing Of The Expressed AR Fragments:

Splicing using the co-translational approach was unsuccessful with all of the conditions tested thus far (Figure 4.8). Using the post-translational splicing approach, the constructs with SG split-intein seemed to produce full-length AR (Figure 4.9; red box). Although the identity of the protein has not yet been confirmed and the amount of putative spliced produce is relatively low, the appearance on the Western blot looks promising. If this is correct, however, the yield is as yet very low and further optimization is required.

4.4. Discussion:

The ultimate goal of the AR project is to structurally and functionally characterize the full-length receptor. Following the “divide and conquer” approach, we have been able to express the AR in two fragments. As mentioned in the introduction, membrane proteins can be notoriously difficult to express in *E. coli* cells, and cell-free expression system has proven to be an effective alternative. Ligation of the AR fragments has been tested in a variety of conditions using co-translational splicing and post-translational splicing. Promising results were achieved using post-translational splicing in reactions containing LPPG, LPPC, SDS, MNG, and DDM. These splicing reaction have not yet been optimized to obtain sufficient quantity of full-length AR for structural biology. In

addition, there still remain many conditions that can be tested for a successful and efficient ligation to obtain full-length AR (Klammt *et al.*, 2005; Schwarz *et al.*, 2007b). Production of the AR fragments solubilized in membrane-mimetics will be the key to obtaining high splicing efficiency.

Cell-free expression has been previously employed for expression of a few GPCRs (Wang *et al.*, 2013; Orban *et al.*, 2015) but the signal overlap arising from seven similarly hydrophobic α -helices alongside the relatively large size of these proteins (~40 kDa for a typical class-A GPCR) makes it difficult to study them by solution-state NMR (Walser & Zerbe, 2012). The method of intein-mediated splicing of a GPCR expressed using cell-free system will facilitate characterization by NMR as it opens the possibility for more innovative labeling schemes, such as segmental isotope-labeling. This would inherently simplify spectral overlap through reduction of the number of observable nuclei. Furthermore, the chemical shift dispersion expected to arise from differences in chemical environment afforded by tertiary structuring would be more pronounced in a full-length folded 7-TM GPCR than in a shorter GPCR fragment (Gautier *et al.*, 2010a).

Very recently, intein-mediated ligation of two fragments has also been applied to a heptahelical membrane protein (proteorhodopsin) to obtain the functional full-length protein (Mehler *et al.*, 2015). This is the first reported example where inteins have been used to splice an α -helical membrane protein. In this case, the splicing was carried out *in vivo*, to minimize the risk of potential stabilization problems by separate solubilization and purification steps of both fragments. This study provides a proof of principle for the work presented in this chapter. Using inteins to enable structural characterization of membrane proteins is still a nascent field; therefore, there will be substantial amount of trial and error required to identify ideal conditions where splicing can be carried out in a

membrane-mimetic environment. Advancements in cell-free expression protocols and ability to co-translationally fold membrane proteins have potential to produce well folded split-inteins fused membrane proteins that is necessary for efficient splicing of membrane proteins using inteins. Using inteins with fast reaction kinetics such as *Npu* DnaE (Zettler *et al.*, 2009) also have a high potential of success in the context of co-translational splicing of membrane proteins.

Table 4.1: Sequence of split-inteins used in this work.

Split-intein description	Amino acid sequence
N-intein of <i>Ssp</i> GyrB split-intein (SGn)	CFSGDTLVALTDGRSVSFEQLVEEEKQGKQNFNCYTIRH DGSIGVEKIINARKTKTNAKVIKVTLDNGESIICTPDHKF MLRDGSYKCAMDLTLDLDDSLMPLHRKISTTEDSGLE
C-intein of <i>Ssp</i> GyrB split-intein (SGc)	SMEAVLNYNHRIVNIEAVSETIDVYDIEVPHTHNFALAS GVFVHN
N-intein of <i>Rma</i> DnaB split-intein (RBn)	CLAGDTLITLADGRRVPIRELVSQQNFVWALNPQTYR LERARVSRAFCTGIKPVYRLTTRLGRSIRATANHRFLTP QGWKRVDELQPGDYALPRRIPRVLASLE
C-intein of <i>Rma</i> DnaB split-intein (RBc)	SMAAQSDVYWDPIVSIIEPDGVVEEVFDLTVPGPHNFVAN DIIAHN
Improved N-intein of <i>Ssp</i> GyrB split-intein (SG20n)	CFSGDTLVALTDGRSVSFEQLVEEEKQGKQNFNCYTIRH DGSIGVEKIINARKTKTNAKVIKVTLDNGESIICTPDHKF MLRDGSYKCAMELTHDDSLMPFHRKISTTEDSGLE
Improved C-intein of <i>Ssp</i> GyrB split-intein (SG20c)	AGEAVLNYNRRIVNIEAVSETIDVYDIEVPHTHNFALAS GVFVHN

Table 4.2: Primers and constructs for AR TM1-3 and AR TM4-7 with different split-inteins.

Plasmid name	Construct	Forward primer (5' – 3')	Reverse primer (5' – 3')
pAR_TM1-3_AT-SGn	AT-tag-AR TM1-3-I _N -His	CTATTGTTTCGTCCT GTAGCTAACGGAT GTTTTTCTGGAGA TACATTAGTCG	GGGCTTTGTTAG CAGCCGGATCTT AGTGGTGGTGGT GGTGGTGCT
pSGc-AR_TM4-7	His-I _C -AR TM4-7-His	TACTTTAAGAAGG AGATATACATATG CATCACCACCATC ATCACTC	CGCGCAGACGCA GGCGTGCGCTGT TATGGACAAACA CTCCGC
pAR_TM1-3_AT-RBn	AT-tag-AR TM1-3-I _N -His	CTATTGTTTCGTCCT GTAGCTAACGGAT GCCTTGCCGGAGA CACCTT	GGGCTTTGTTAG CAGCCGGATCTT AGTGGTGGTGGT GGTGGTG
pRBc-AR_TM4-7	His-I _C -AR TM4-7-His	TACTTTAAGAAGG AGATATACATATG CATCACCACCATC ATCACTC	CGCGCAGACGCA GGCGTGCGCTGT TATGAGCGATAA TGTCGT
pAR_TM1-3_AT-SG20n	AT-tag-AR TM1-3-I _N -His	CTATTGTTTCGTCCT GTAGCTAACGGAT GTTTTTCTGGAGA TACATTAGTCG	GGGCTTTGTTAG CAGCCGGATCTT AGTGGTGGTGGT GGTGGTGCT
pSG20c-AR_TM4-7	His-I _C -AR TM4-7-His	TACTTTAAGAAGG AGATATACATATG GGTCATCATCATC ATCATCATGCTGG TGAAGCAGTATTA AATTACA	CGCGCAGACGCA GGCGTGCGCTGT TATGGACAAACA CTCCGC

NOTE: I_N stands for intein-N and I_C stands for intein-C; His stands for the hexa-histidine tag.

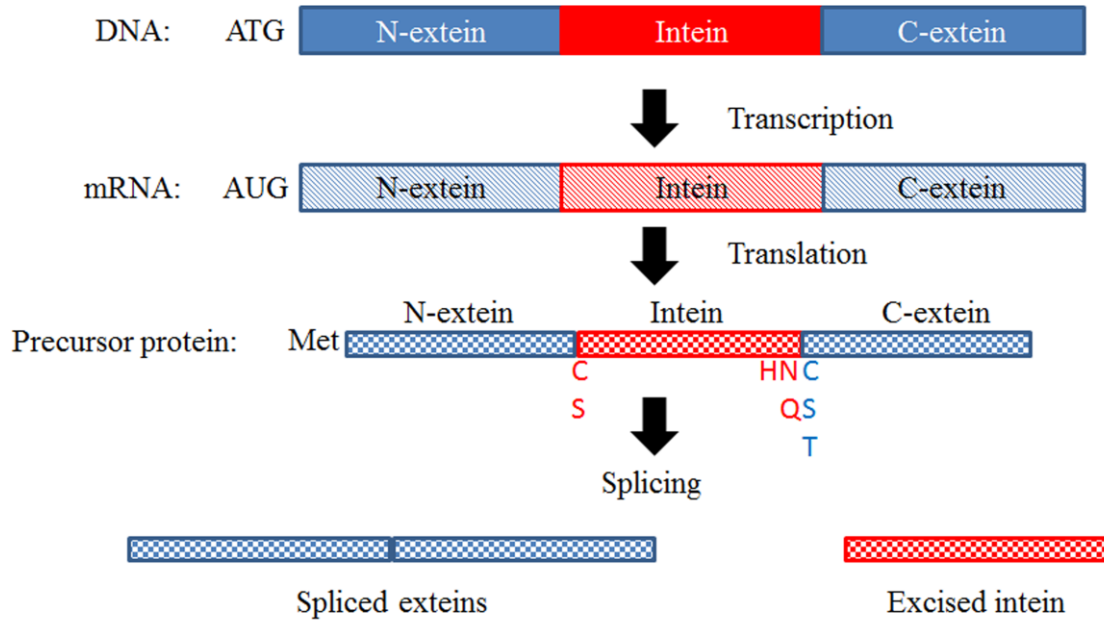


Figure 4.1: Representation of an intein. Conserved residues are displayed in red (intein) and blue (extein).

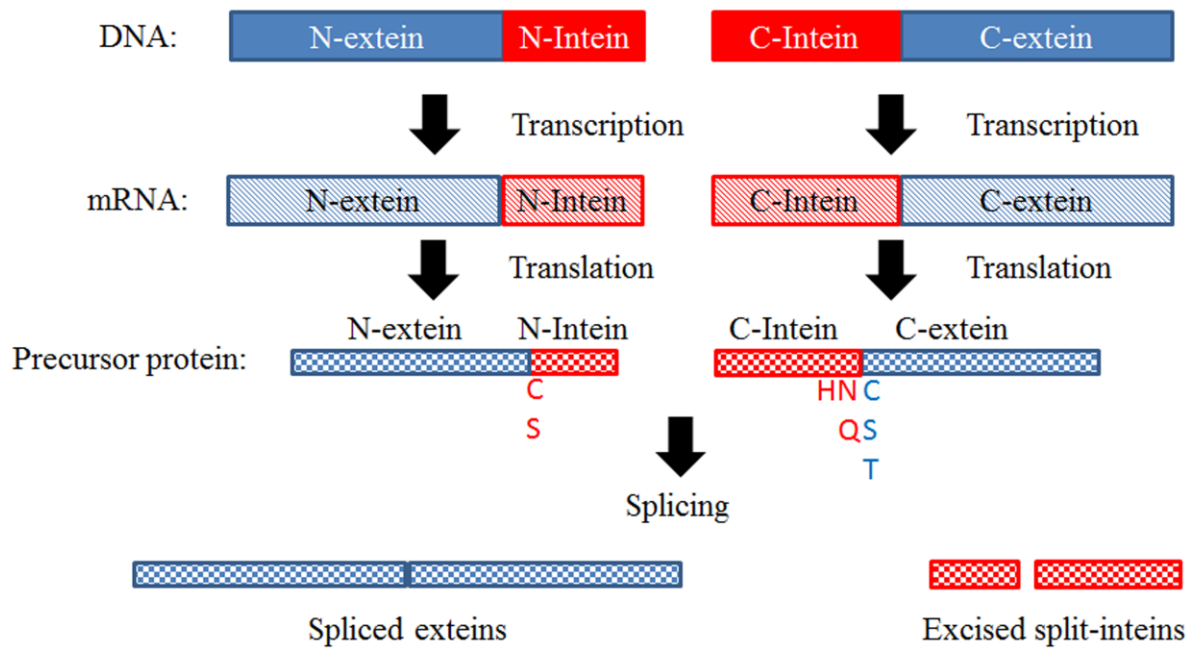


Figure 4.2: Representation of a split-intein. Conserved residues are displayed in red (intein) and blue (extein).

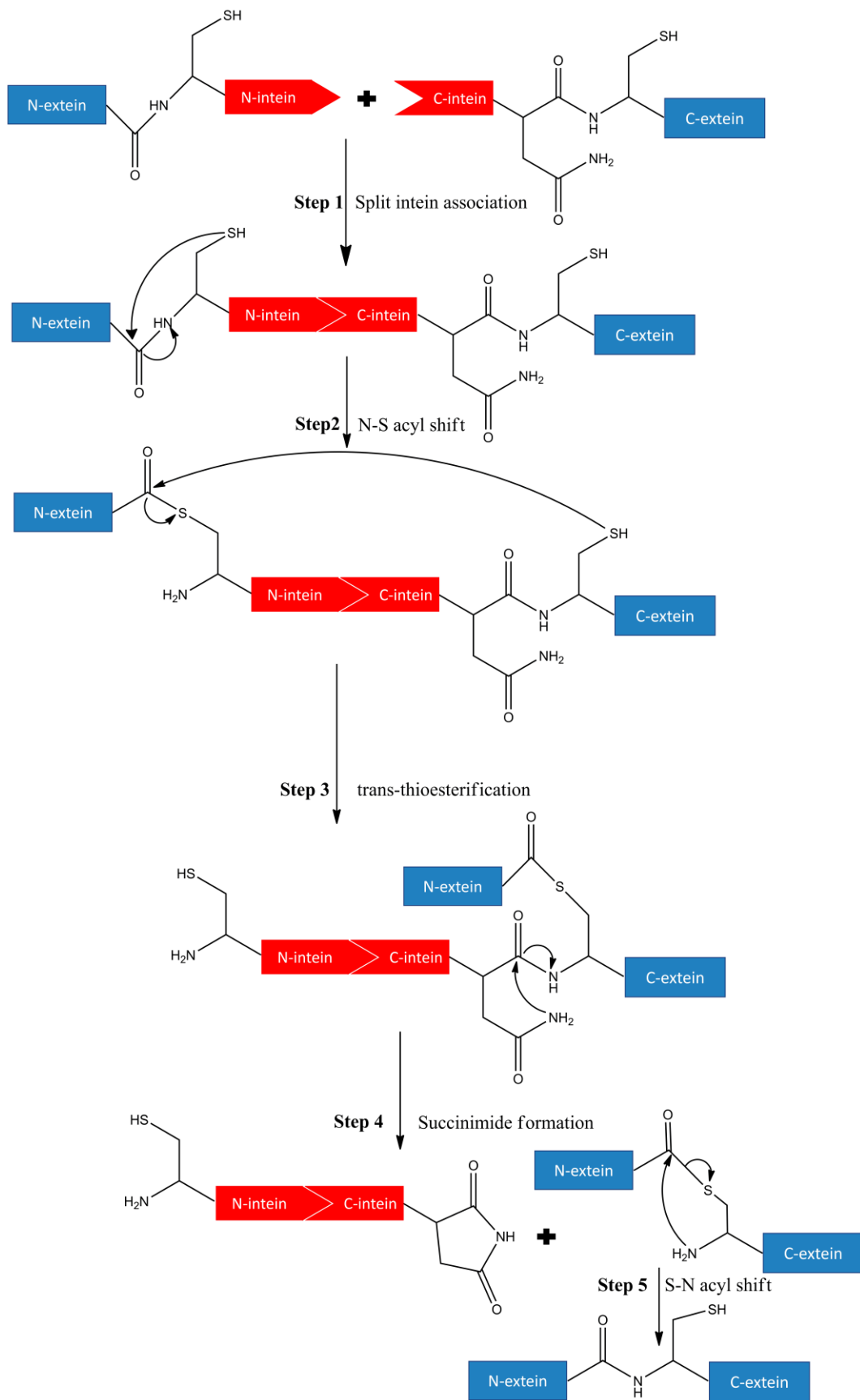
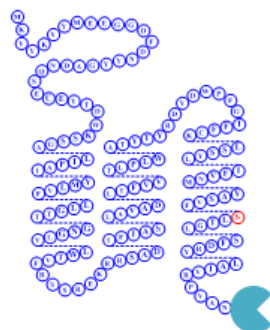


Figure 4.3. Five-step split-intein splicing mechanism. Step 1: Association of the two fragments (N-intein and C-intein) of the split-intein, which are located on two precursor proteins expressed from two different genes. Step 2: N-S acyl shift involving the thiol on the first residue of N-intein, leading to the replacement of the peptide bond with a thioester bond between N-extein and N-intein. Step 3: *Trans*-thioesterification where the thioester bond formed in step 2 is attacked by the thiol of the first residue of C-extein, leading to a branched intermediate. Step 4: Formation of a succinimide ring by the last residue (asparagine) of C-intein, separating the intein from the exteins. Step 5: S-N acyl shift leading to formation of a peptide bond between the two exteins. This figure has cysteine as the first residues of N-intein and N-extein, but the cysteine could be replaced by serine or threonine in different inteins.

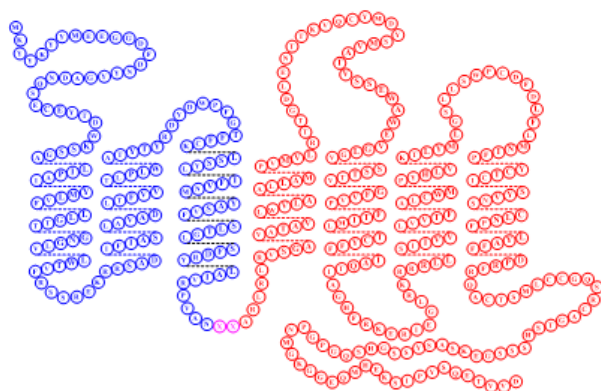
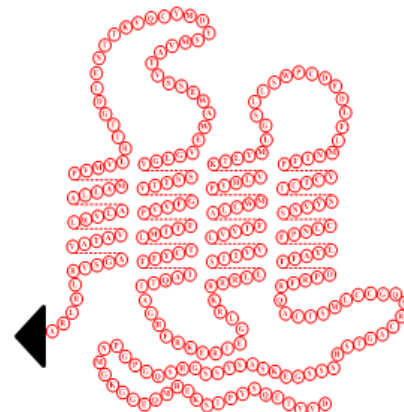
AR TM1-3 + N-Intein



Intein-mediated
trans-splicing

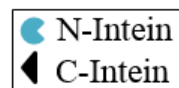


C-Intein + AR TM4-7



+

Excised
inteins



Segmentally isotope-labeled AR

Figure 4.4: Schematic for segmental isotope-labeling of AR.

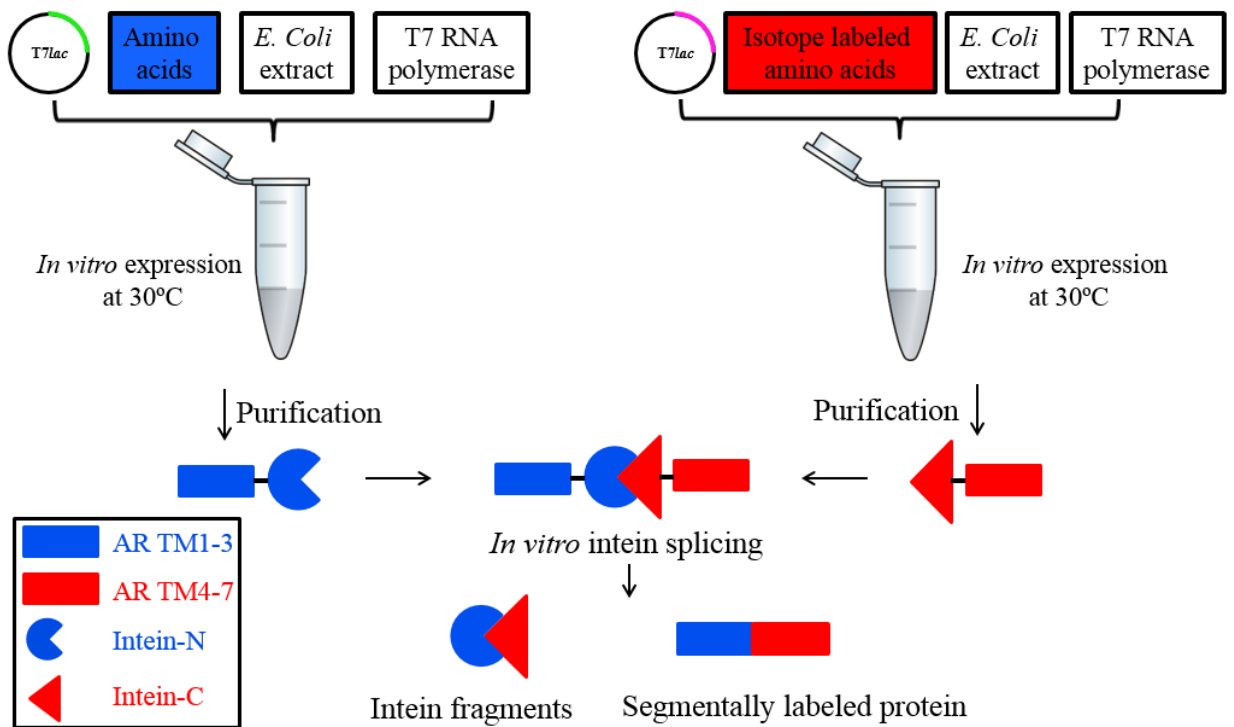


Figure 4.5: Schematic for *in vitro* expression and subsequent splicing to obtain the full-length segmentally labeled AR.

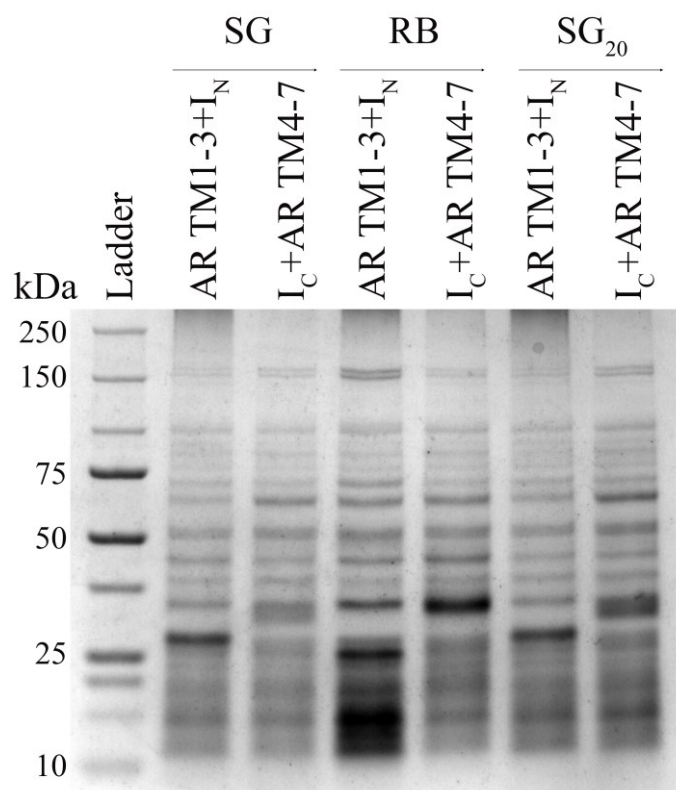


Figure 4.6: SDS-PAGE representing expression of AR constructs with split-inteins.

Coomassie blue stained SDS-PAGE gel showing expression of 3 pairs of different AR constructs (Table 4.2) using *E. coli* based cell-free expression system. Each pair of AR construct contain a split intein (Table 4.1) as indicated, with SG being *Ssp* GyrB split-intein, RB being *Rma* DnaB split-intien, SG₂₀ being improved *Ssp* GyrB split-intein, I_N being N-intein, and I_C being C-intein. After expression for 5 hours, the reaction mixture was centrifuged and the pellets were solubilized in 100 μ L of 1x loading dye and resolved using 7.5% SDS-PAGE.

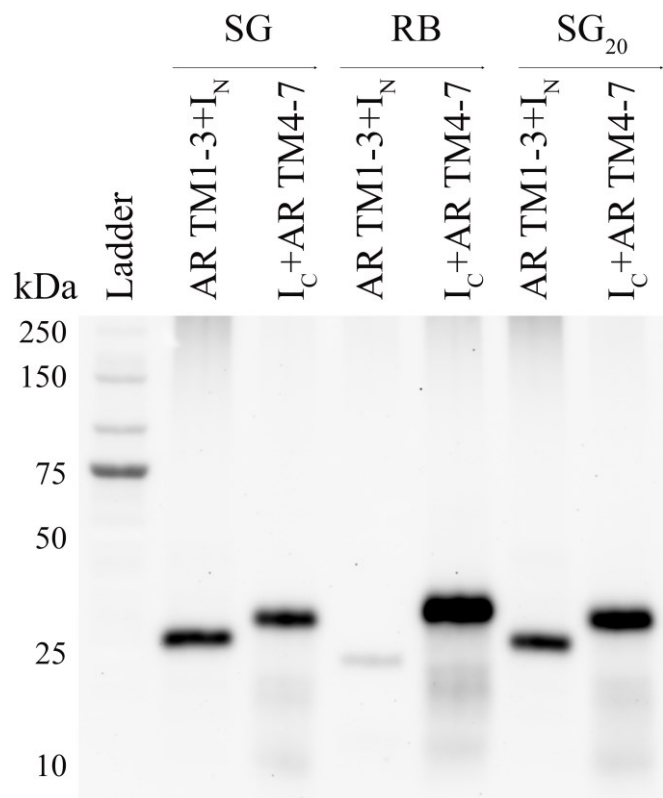


Figure 4.7: Western blot representing expression of AR constructs with split-inteins. Western blot probed with HisProbe-HRP to confirm the expression of different constructs of AR. The sample was first resolved by SDS-PAGE (as in Figure 4.4) and then transferred onto a nitrocellulose membrane and probed (detailed in Materials and Methods).

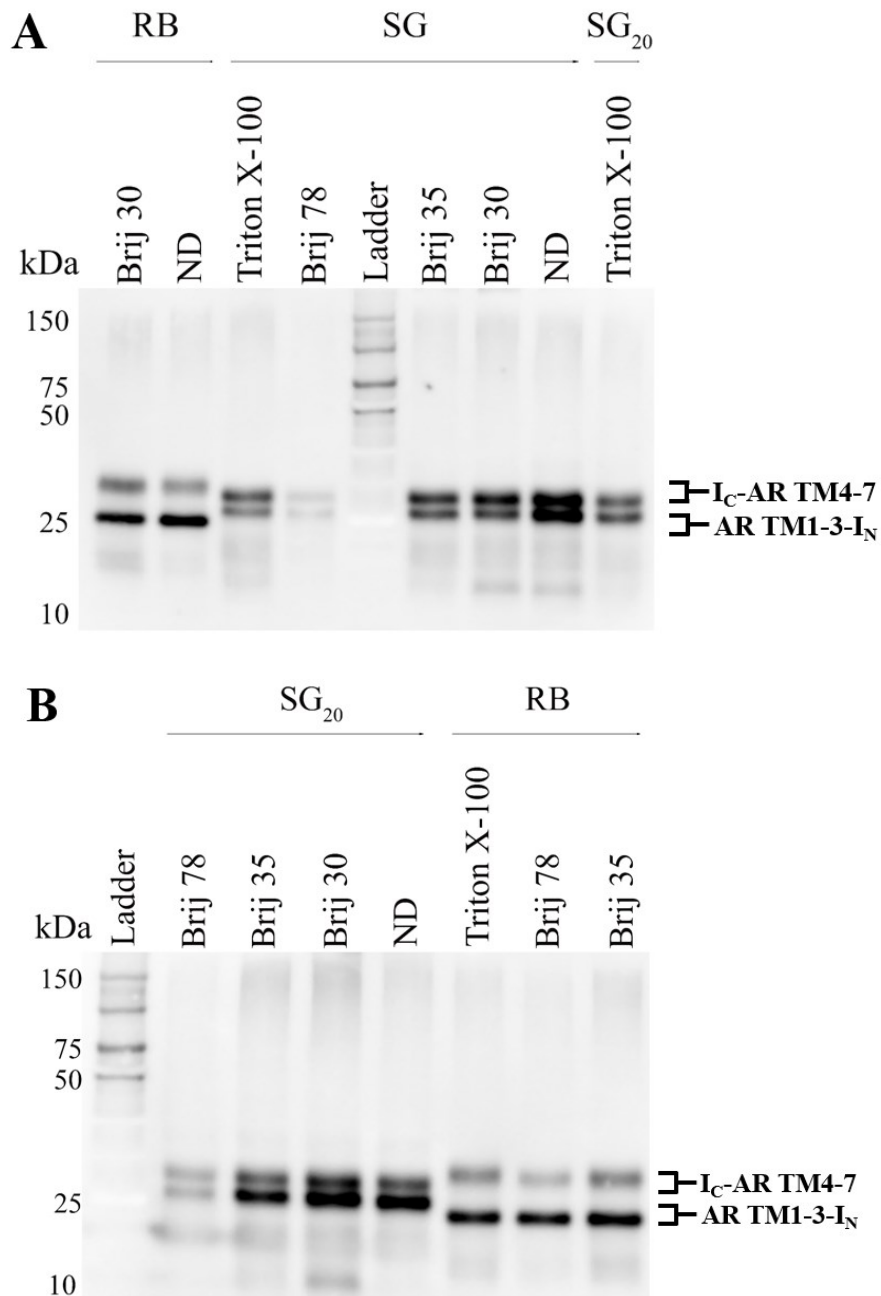


Figure 4.8: Western blot demonstrating results of co-translational splicing. Western blot probed with HisProbe-HRP showing a comparison of various detergent conditions used to optimize the co-translational splicing reaction. The pellets from the splicing reaction were first resolved by SDS-PAGE and then transferred onto a nitrocellulose membrane and probed (detailed in Materials and Methods). Collectively, both A and B represent all the splicing conditions tested to splice AR TM1-3 and AR TM4-7 using three pair of split inteins.

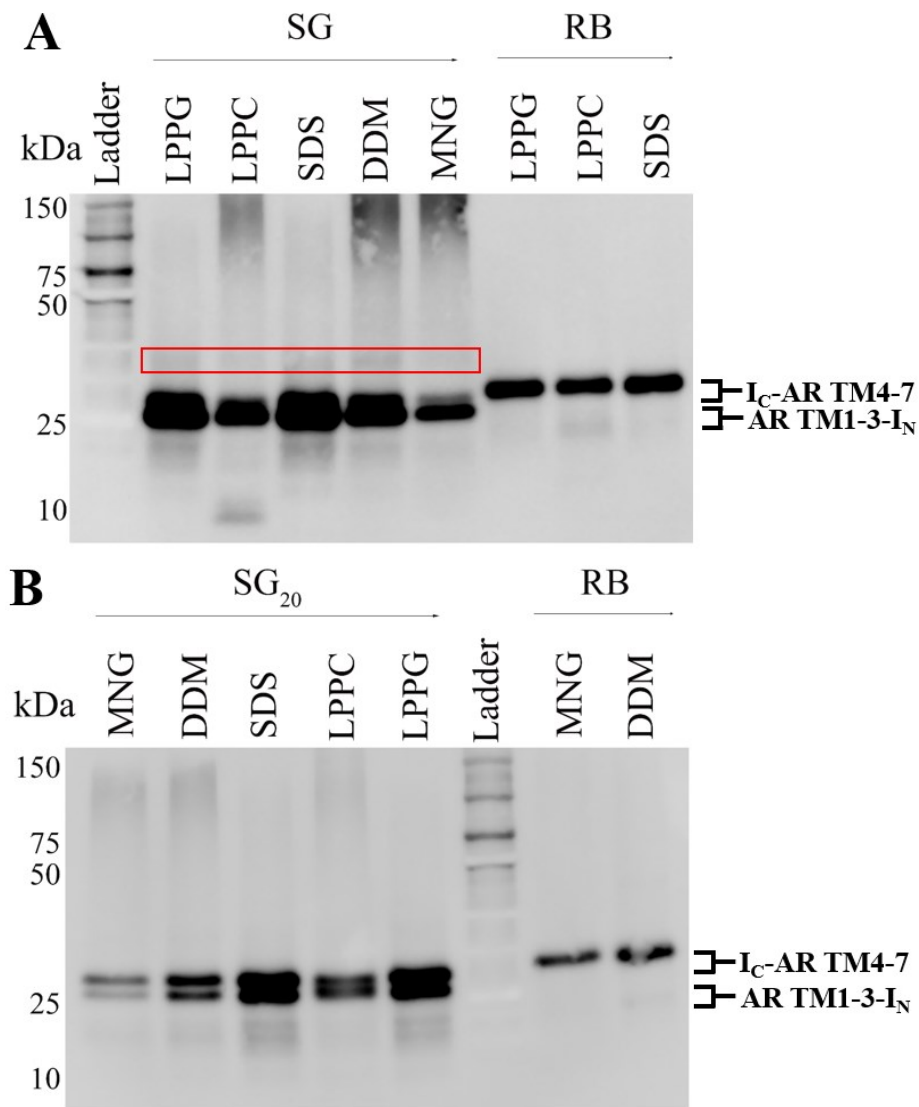


Figure 4.9: Western blot demonstrating results of post-translational splicing.

Western blot probed with HisProbe-HRP showing a comparison of various detergent conditions used to optimize the post-translational splicing reaction. The pellets from the splicing reaction were first resolved by SDS-PAGE and then transferred onto a nitrocellulose membrane and probed (detailed in Materials and Methods). Collectively, both A and B represent all the splicing conditions tested to splice AR TM1-3 and AR TM4-7 using three pair of split inteins. The red box shows very faint bands that may represent spliced full-length AR.

Chapter 5: Structural And Functional Characterization Of The Juxtamembrane 8th Helix Of The AR

5.1. Introduction:

GPCR-mediated signaling in the cell is dependent upon trafficking of properly folded GPCRs to the cell surface. Therefore, changes in trafficking of GPCRs can play very important roles in pathophysiology of various diseases (Conn *et al.*, 2007). Similarly to other membrane proteins, GPCRs are synthesized and folded in the ER, where chaperones may mediate the process of proper folding and maturation (Dong *et al.*, 2007; Young *et al.*, 2015). The properly folded receptors are then recruited to ER-derived COPII-coated transport vesicles (Tang *et al.*, 2005). These transport vesicles carrying cargo receptors then traverse from the ER to the plasma membrane through the ER–Golgi intermediate complex (ERGIC), the Golgi apparatus, and the trans-Golgi network (TGN) in sequence (Dong *et al.*, 2007). Post-translational modifications also occur *en route* to the plasma membrane.

Research in the last few decades has made it clear that the anterograde transport of GPCRs is a very complicated process involving various different motifs and regulators (Young *et al.*, 2015). The C-terminal tail has also been implicated in cell surface expression of receptors such as rhodopsin (Heymann & Subramaniam, 1997), luteinizing hormone/chorionic gonadotropin receptor (Rodriguez *et al.*, 1992), A1 adenosine receptor (Pankevych *et al.*, 2003), vasopressin 2 receptor (Oksche *et al.*, 1998), AT1a receptor (Gaborik *et al.*, 1998), melanin-concentrating hormone receptor 1 (Tetsuka *et al.*, 2004), and cannabinoid receptor 1 (Stadel *et al.*, 2011). Beyond the C-terminal tail in general, a number of conserved amino acid sequence motifs have now been identified in

GPCRs with important roles in correct plasma membrane localization of these receptors. For example, C-terminal motifs such as E(X)₃LL (Schulein *et al.*, 1998), F(X)₃F(X)₃F (Bermak *et al.*, 2001), FN(X)₂LL(X)₃L (Robert *et al.*, 2005), and F(X)₆LL (Duvernay *et al.*, 2004) have all been implicated in regulating GPCR export. However, the complete molecular mechanisms involving these motifs and, potentially, interacting protein partners in the transport of GPCRs from the ER to the cell surface are not yet understood.

Based on the structures of GPCRs solved to date, an amphipathic juxtamembrane helix-8 that follows TM7 appears to be a common structural feature of the C-terminal tail of Class A GPCRs. In general, the putative helix-8 long-axis lies parallel to the cell membrane surface. This means that it is suitably positioned to interact directly with the membrane and with intracellular signaling machinery (Huynh *et al.*, 2009). Interestingly, many functional studies suggest the possible involvement of this helix in various aspects of GPCR biology, such as trafficking (Ahn *et al.*, 2010), G-protein coupling and activation (Delos Santos *et al.*, 2006; Liu *et al.*, 2015a), receptor internalization (Aratake *et al.*, 2012), and interaction with arrestin (Kirchberg *et al.*, 2011; Kang *et al.*, 2015). Recently, work with the GPCR TGR5 also showed the importance of the helix-8 in plasma membrane localization of the receptor (Spomer *et al.*, 2014). In addition, helix-8 of rhodopsin has been shown to be a dynamic membrane-dependent conformational switch, possibly giving rise to functional ramifications in downstream signaling by GPCRs (Krishna *et al.*, 2002).

Despite the established importance of helix-8, structural and functional aspects of helix-8 have not been very extensively studied in comparison to the TM helices. In the case of the AR, there is no evidence showing either the presence of a helix-8 structure or involvement of this element in plasma membrane localization. In this chapter, I set out to

investigate the presence of the putative juxtamembrane helix-8 and, if present, its role in plasma membrane localization of the AR. The scarcity of comprehensive structural and functional analysis of this region of GPCRs in general provided further incentive to pursue these studies.

In this chapter, I report the expression, purification, and structural characterization of the C-terminal 71 amino acids of the AR (referred to as “AR C-tail”). Both CD and NMR spectroscopy were used in my structural characterization of the AR C-tail. Based upon the amphipathic helix-8 that is identified in structural studies, immunofluorescence experiments were used to examine its role in plasma membrane localization of the AR. Specifically, mutations were made in the helix-8 region of the AR C-tail (wild-type sequence referred to simply as AR) to replace the residues on the hydrophobic face by glycine (AR-G), serine (AR-S), or proline (AR-P); or, to flip the amphipathy (AR-FLIP) of the helix. These mutant versions of the AR are summarized in Figure 5.1. The effects of each set of mutations on the structure and function of the AR C-tail were evaluated by both CD spectroscopy and immunofluorescence experiments.

5.2. Materials And Methods:

5.2.1. Materials:

A pcDNA 3.1 plasmid DNA containing 3x haemagglutinin (HA) tag on the N-terminus of the AR gene was purchased from the Missouri University of Science and Technology cDNA Resource Centre (Rola, MO). Primary antibodies against the HA-tag were produce in-house in Dr. Roy Duncan’s laboratory, and protein disulfide isomerase (PDI) was purchased from Enzo Life Sciences (Farmingdale, NY). Alexa Fluor 488–

conjugated goat anti-mouse and Alexa Fluor 647–conjugated goat anti-rabbit secondary antibodies, trypsin-EDTA, OptiMEM medium, ProLong gold antifade reagent and Dulbecco's Modified Eagle Medium (DMEM) were purchased from Life Technologies (Carlsbad, CA). Polyethyleneimine (PEI) was purchased from Polysciences Inc. (Warrington, PA). 10% fetal bovine serum (FBS) and cyclohexamide were purchased from Sigma-Aldrich (St. Louis, MO). Glass coverslips and SnakeSkin™ Dialysis Tubing (3.5K MWCO, 22 mm) were purchased from Thermo Scientific (Hampton, NH). All other materials used in this work have been mentioned in previous chapters.

5.2.2. Cloning Of AR C-Tail And Its Mutants For Recombinant Expression:

The plasmid for expression of the AR C-tail (Complete amino acid sequence: A F F D P R F R Q A S T S M L C C G Q S R S A G T S H S S S G E K S A S Y S S G H S Q G P G P N M G K G G E Q M H E K S I P Y S Q E T L V V D) was obtained by sub-cloning the appropriate nucleotides into the pETHN vector (Chapter 2) at the *SapI* and *BamHI* restriction enzyme sites. The resulting plasmid was used as a template for producing the mutants using site-directed mutagenesis involving blunt-end ligations. Primers used to amplify the AR C-tail gene from a pUC 57 plasmid containing the sequence for the full-length AR contained the *SapI* restriction site as an overhang on the forward primer and the *BamHI* restriction site as an overhang on the reverse primer. All primers for the cloning experiments are listed in Table 5.1. The sub-cloned plasmids were electroporated into DH5 α strain of *E. coli* for storage at -80 °C.

5.2.3. Mutagenesis Of The AR C-Tail Of HA-Tagged AR For Immunofluorescence Experiments:

AR C-tail mutants were created by performing site-directed mutagenesis on a pcDNA 3.1 plasmid DNA containing the N-terminally 3xHA-tagged AR gene. The plasmid containing mutated gene was stored in DH5 α *E.coli*, and was extracted from it using the Genopure Plasmid Midi Kit for further use in immunofluorescence experiments.

5.2.4. Expression Of The WT AR C-Tail And Its Mutants:

Each plasmid containing the WT AR C-tail or its mutants were electroporated into BL21 (DE3) *E. coli* for expression. Protein expression was carried out for unlabeled and uniformly-labeled (enriched with ^{15}N isotopes and/or ^{13}C isotopes) AR C-tail proteins using the protocol described in section 2.2.3.

5.2.5. Purification Of The WT AR C-Tail And Its Mutants:

Econopac columns packed with Ni-NTA agarose resin were used to purify the WT AR C-tail and its mutants under native conditions using the protocol detailed in section 2.2.6 (without SDS). SUMO protease (produced in-house by expression in BL21(DE3) strain of *E. coli* using protocol detailed in section 2.2.3) was used at a protein:protease ratio of 1:1 (stoichiometric) to cleave off the SUMO tag. The cleavage was carried out in elution buffer for 12 hours at room temperature. The cleaved protein was dialysed against distilled water using 3.5 kDa molecular weight cut-off dialysis tubing, and then reverse purified using Ni-NTA chromatography. The protein of interest was collected as a flow-through fraction during reverse purification. The proteins were further purified using HPLC; a linear binary gradient (solvent A – H $_2$ O; solvent B –

acetonitrile; A and B, both containing 0.1%TFA) was employed for semi-preparative RP-HPLC on a Prostar HPLC equipped with a diode array ultraviolet (UV)-visible detector (Varian Canada Inc.; Mississauga, ON). Semi-preparative RP-HPLC (Cosmosil packed 5C18-AR-II column, 10 mm x 250 mm, 5 μ m) employed 1 mL of protein solubilized in 50% acetonitrile with a flow rate of 3 mL/min and a linear gradient of 2-50% solvent B over 20 min. Fractions containing purified protein were pooled, lyophilized, and stored at -20 °C. Positive-ion mode electrospray ionization (ESI) mass spectrometry of lyophilized protein samples was performed at the Mass Spectrometry Laboratory at the Department of Chemistry (Dalhousie University).

5.2.6. CD Spectroscopy:

Far-UV CD spectra for all samples of AR C-tail were recorded at 25 °C using a J-810 spectropolarimeter (Jasco, Easton, MD) at 20 nm/min with a data pitch of 1 nm from 260 nm to 190 nm. Cuvettes of path length (l) 0.01 cm were used for all micelle samples. All micelle samples contained 150.0 \pm 5.0 μ M protein, with either 40 mM SDS, 40 mM LPPG, or 30 mM DPC, in 20 mM sodium phosphate buffer at pH 6.00 \pm 0.05 and pH 7.00 \pm 0.05. CD data were also collected at the same protein concentration and with the same buffers without any membrane-mimetics. All samples were prepared from the same protein stock solution, with concentration (c) determined using the Beer-Lambert law ($c=A\cdot\epsilon^{-1}\cdot l^{-1}$) by measuring absorbance (A) at 280 nm using a molar absorptivity (ϵ ; calculated following (Pace *et al.*, 1995)) of 2980 M⁻¹cm⁻¹. All data were collected in triplicate, averaged, blank subtracted, and reported as mean residue ellipticity ($[\theta]$). The data were also smoothed by taking a sliding-window average with a window size of 3 nm. Difference spectra were calculated by subtracting the $[\theta]$ values for the sample

without micelles from that of the sample with micelles. CD data deconvolution was carried out using the BeStSel server (<http://bestsel.elte.hu/>) (Micsonai *et al.*, 2015) to estimate the secondary structure content in different conditions.

5.2.7. Solution-State NMR Spectroscopy:

1D ^1H (Hwang & Shaka, 1995) and 2D ^1H - ^{15}N HSQC (Palmer *et al.*, 1991; Kay *et al.*, 1992; Schleucher *et al.*, 1994) spectra of the WT AR C-tail were acquired in buffer and in LPPG micelles. The buffer sample was prepared by dissolving lyophilized protein to a final concentration of 0.92 mM in 20 mM sodium phosphate buffer. Micellar samples were prepared with protein at a final concentration of 0.72 mM or 0.5 mM was solubilized in 20 mM sodium phosphate with 120 mM or 100 mM LPPG, respectively. All the samples also contained 1 mM DSS, 10 mM DTT- d_6 , and 1 mM NaN_3 , 100 mM NaCl, with the pH of each sample was adjusted to 6.00 ± 0.05 using DCl and NaOD. All triple-resonance experiments and the ^1H - ^{15}N heteronuclear NOE experiments were conducted on ^{15}N - and ^{13}C -enriched protein sample. NMR spectra were acquired at 37 °C using either a Bruker Avance III 700 MHz spectrometer (Milton, ON) equipped with a cryogenically cooled 5 mm triple resonance inverse (TCI) probe and a z-axis gradient (Biomolecular Magnetic Resonance Facility (BMRF) housed at the National Research Council, Halifax, NS), or a Bruker Avance 500 MHz spectrometer equipped with a BBFO SmartProbe (NMR³ Facility at Dalhousie University). 3D triple-resonance experiments (HNCA, HN(CO)CA, HNC(O), HN(CA)CO, CBCANH, and CBCA(CO)NH) were acquired using the sample containing 0.5 mM WT AR C-tail in 100 mM LPPG using the 700 MHz instrument at 37 °C. ^1H - ^{15}N heteronuclear NOE data were also acquired for this sample at 16.4 T with a recycle delay of 5 s (Farrow *et al.*, 1994). The

NOE enhancement factor for a given position is reported as the ratio of cross-peak intensities in spectra collected with and without proton saturation. NMR experimental parameters used for study of each set of conditions are detailed in Tables 3.1, 3.2, 3.3 and 3.4. NMR spectra were processed using Bruker TopSpin 3.1 or NMRPipe 8.1 (Delaglio *et al.*, 1995). Spectra were plotted and analyzed using CcpNmr Analysis 2.2.2 (Vranken *et al.*, 2005) for 2D experiments and 3D experiments or Topspin 3.1 for 1D experiments. CCPNmr Analysis 2.2.2 was used for backbone walk (section 1.7.3) based assignment of C α , C β , C' and HN for the main chain of AR C-tail.

5.2.8. Cell Culture:

Human embryonic kidney (HEK) 293T cells were grown in DMEM supplemented with 10% FBS and 25 mM HEPES at 37 °C with 5% CO₂. Cells were maintained in 175 cm² flasks and sub-cultured using Trypsin-EDTA every 2-3 days to prevent over-confluency and cell death. Cells were maintained without the use of antibiotics.

5.2.9. Transient Transfections:

HEK 293T cells grown on culture plates to ~60% confluency were transfected with PEI. Transfection mixes were prepared in OptiMEM medium with 0.5 μ g of DNA and 3 μ g of PEI for each well of a 12-well culture plate. Cells were washed and supplemented with fresh serum-free growth medium before adding transfection mixes. After incubation for 4-6 h, the transfection mix was removed and cells were supplemented with fetal bovine serum for another 18-20 h for analysis by respective assays.

5.2.10. Cell Surface Immunofluorescence:

HEK 293T cells transfected with the HA-tagged AR, AR-G, AR-S, AR-P and AR-FLIP genes were cultured for 24 h in growth medium. Live cells were then incubated at 4 °C for 30 min in blocking buffer (5% normal goat serum, 1% bovine serum albumin (BSA), 0.02% NaN₃ in Hank's Balanced Salt Solution) and stained with a 1:500 dilution of anti-HA antibody and a 1:1000 dilution of Alexa 647–conjugated goat anti-mouse secondary antibody, each for 1 h at 4°C. Cells were washed 6 times with blocking buffer after each antibody incubation. Cells were rinsed quickly with PBS and then incubated in PBS containing 10 mM EDTA for one minute at room temperature. Following this, cells were resuspended in PBS and fixed with 3.7% formaldehyde. For each transfect, 10,000 cells were analyzed by flow cytometry (FACSCalibur; BD Biosciences) with data acquisition using the De Novo software package. Cells transfected with empty vector were used as negative controls. Mean fluorescence intensity (MFI) was calculated for each sample and background fluorescence intensity from empty vector-transfected cells was subtracted before mean percent surface fluorescence was calculated. Statistical analysis and sample comparison were performed using Prism (GraphPad, San Diego, CA). Standard error of the mean (SEM) values were calculated on the basis of sample values averaged between experiments. Samples were analyzed using ANOVA with a Tukey post-test to evaluate significance of the difference.

5.2.11. Intracellular Immunofluorescence Microscopy:

At 24 hours post-transfection, transfected HEK 293T cells cultured on glass coverslips were fixed with 3.7% formaldehyde for 20 min at room temperature and permeabilized for 20 min at room temperature with 0.1% Triton X-100. Alternatively,

cells were treated with 50 µg/ml cyclohexamide for 2 h prior to fixation and permeabilization. Cells were washed 3 times with PBS and blocked for 30 min in blocking buffer (1% BSA in PBS) and co-stained with mouse anti-HA antibody (1:200) and with 1:1000 dilutions of mouse monoclonal antibodies against an ER marker (rabbit anti-PDI), and subsequently with 1:1000 dilutions of Alexa 488–conjugated goat anti-mouse and Alexa 647–conjugated goat anti-rabbit secondary antibodies. Cells were washed 3 times with PBS after each antibody incubation. Coverslips were mounted on glass slides using ProLong Gold Antifade reagent. Microscopy was carried out using a Zeiss LSM 510 META confocal microscope, with micrographs acquired with 63X objective lens. The laser intensity was set to minimize saturated pixels, and detector gain was set to have minimal background in control samples using the ZEN 2009 image acquisition software. The same detector gain and laser intensity were used for all subsequent image acquisitions within a given experiment.

5.3. Results:

5.3.1. Expression And Purification Of WT AR C-Tail:

Expression of AR C-tail and its mutants was carried out at 37 °C in BL21(DE3) strain of *E. coli*. Initial purification was carried out using Ni²⁺ affinity chromatography. SUMO tag was cleaved off by overnight incubating the protein in the elution buffer (Ni²⁺ affinity) with SUMO protease. Subsequent purification was carried out using Ni²⁺ affinity based reverse purification. Samples from each step were evaluated by SDS-PAGE (Figure 5.2). However, it was observed that the purified protein was susceptible to degradation. Consequently, the protein was further purified using RP-HPLC to prevent

degradation, which may have been due to impurity not visible by SDS-PAGE. Following HPLC purification, the protein yield was approximately 10 mg/L. The chromatogram for the RP-HPLC purification step is shown in Figure 5.3. The identity of the WT AR C-tail RP-HPLC eluent peak was confirmed using positive mode ESI-MS (Figure 5.4). Purification of all the other mutants (AR-G, AR-S, AR-P, and AR-FLIP) was also successful (LeBlanc and Rainey, unpublished work). Interestingly, for the mutant AR-FLIP, where the hydrophathy has been flipped, cleavage of the SUMO tag was inefficient and half of the expressed protein remained un-cleaved.

5.3.2. CD Spectroscopy Indicates Helix Induction In The Presence Of Membrane Mimetics:

Far-UV CD spectra in buffer at pH 6 and pH 7 for AR, AR-G, and AR-S imply a predominantly disordered conformation, with an intense negative band at ~195 nm and a lack of bands in the 208-230 nm region (Figure 5.5). In each micellar condition, the band at 195 nm becomes red-shifted toward 200 nm and convolution with positive band at ~190 nm is apparent; additional negative ellipticity is also apparent over ~210-220 nm (Figure 5.5). While the micellar-state spectra do not resemble a canonical, helix-rich protein, the CD difference spectra (Figure 5.6) show features consistent with induction of α -helical character in the presence of detergent micelles, with a positive band at ~192 nm ($\pi \rightarrow \pi^*$ transition) and negative bands at ~222 nm ($n \rightarrow \pi^*$ transition) and 208 nm ($\pi \rightarrow \pi^*$ transition). This effect is more pronounced in LPPG and SDS micelles than in DPC micelles, suggesting a dependence of secondary structure induction upon the presence of an anionic membrane-mimetic environment. Although the overall appearances of the spectra at pH 6 are in general agreement with those at pH 7, the

variation in spectral characteristics for the difference spectra of AR-G and AR-S relative to WT AR C-tail point towards a difference in secondary structure content (Figure 5.5).

To better define these qualitative differences in CD spectral properties, secondary structure composition was estimated for each condition through CD spectral deconvolution using a recently released CD data deconvolution server (BeStSel) reported to have a higher predictive accuracy than previously available deconvolution protocols (Micsonai *et al.*, 2015). Although this deconvolution protocol is of particular interest for proteins having a high β -sheet content, the results presented in the study by (Micsonai *et al.*, 2015) indicate superior performance for helical structures as well. The deconvolution data for AR, AR-G, and AR-S in absence (ND) and presence (SDS, DPC, and LPPG) of detergent micelles are shown in Tables 5.3 (pH 6) and 5.4 (pH 7). Low normalized root mean square deviation (NRMSD) values (Table 5.3 and 5.4; green highlight) demonstrate that reliable fits were obtained. In accordance with the visual spectral characteristics observed in the difference spectra, the deconvolution-based estimation shows an approximately 5.6-7.5% overall increase in α -helical character for AR in SDS and LPPG micelles (Table 5.3 and 5.4; yellow highlight). It is noteworthy that despite similar overall α -helical content for AR, AR-G and AR-S in buffer conditions, the percentage of regular helix content (red highlight) is higher for AR, or, in other words, that the mutations lead to reduced regular helical content in absence of membrane-mimetics. It is also interesting to note that in the case of AR-G and AR-S, the percentage of distorted helix content (blue highlight) for the ND condition is the major contributor to the overall helical content. Overall, the CD data indicate that the presence of anionic membrane-mimetics induce helical character in the WT AR C-tail and its mutants. However, the dynamics of helix formation may vary in the mutants, as is evident from the difference in

the regular helical content (red highlight) in the ND and anionic micelle conditions for the non-amphipathic mutants (AR-G and AR-S) relative to the wild-type AR sequence.

5.3.3. NMR Spectroscopy Demonstrates Formation Of The Helix-8 In Presence Of LPPG Micelles:

The ^1H - ^{15}N HSQC spectra for the AR C-tail in buffer with and without LPPG were compared. Overlaid ^1H - ^{15}N HSQC spectra are presented in Figure 5.7, one containing 0.72 mM ^{15}N -enriched AR with 120 mM LPPG (blue) and the other containing 0.92 mM ^{15}N -enriched AR in buffer only (red). Although the two spectra mostly overlap, the spectrum for the sample with LPPG micelles showed better peak dispersion, indicative of increased structuring arising from the presence of a membrane-mimetic environment. To characterize the AR C-tail structure in the presence of LPPG, another sample was made with 0.5 mM uniformly ^{15}N - and ^{13}C -enriched AR with 100 mM LPPG for 2D and 3D NMR data collection on a 700 MHz spectrometer. Chemical shift assignment followed the strategy described in section 3.2.4. For the AR C-tail, 3D triple-resonance experiments (HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCANH, and CBCA(CO)NH) allowed unambiguous assignment of all $\text{C}\alpha$, $\text{C}\beta$, C' and HN chemical shifts for 52 of 71 residues, including the complete region predicted to be the putative helix-8. Figure 5.8 shows a ^1H - ^{15}N HSQC with backbone amide peaks labeled with amino acid assignments. Chemical shifts thus assigned were used to predict the secondary structure.

Chemical shift index (CSI) (Wishart & Sykes, 1994b) and DANGLE (Cheung *et al.*, 2010) are the two commonly used methods to predict the secondary structure of a protein based on the chemical shifts of its nuclei. The CSI algorithm uses secondary chemical shifts, which are a measure of deviations of the experimentally determined

chemical shifts from those for random coil. Generally, residues located in α -helices have positive $^{13}\text{C}\alpha$ and negative $^{13}\text{C}\beta$ secondary shifts, whereas amino acids present in β -sheets have negative $^{13}\text{C}\alpha$ and positive $^{13}\text{C}\beta$ secondary shifts (Wishart & Sykes, 1994b). DANGLE on the other hand predicts the secondary structure from amino acid sequence information, experimental chemical shifts and a database of known protein structures and their associated chemical shifts. Both CSI and DANGLE predict an α -helix from residue numbers 5-20 (Figure 5.9). As discussed in section 1.7.4, the ^1H - ^{15}N heteronuclear NOE is a sensitive probe for motion in ps-ns time scale and, hence, can be used to evaluate the stability of structured regions in proteins. In agreement with the secondary structure prediction, residues 5-17 show hetNOE enhancement factors of greater than 0.5, which is indicative of reduced dynamics in this region of the WT AR C-tail (Figure 5.9). In combination with the secondary structure prediction based on ^{13}C chemical shifts, the hetNOE measurements demonstrate the presence of a 12-15 amino acid long α -helix, which is in agreement with the size of the 8th helix observed in other class A GPCRs (Krishna *et al.*, 2002; Huynh *et al.*, 2009; Ahn *et al.*, 2010; Aratake *et al.*, 2012).

5.3.4. Mutations In Helix-8 Reduce Plasma Membrane Localization Of The AR:

Following identification of a structured membrane-induced 8th helix in the AR, the effects of the each of the mutations changing the properties of this helix (Figure 5.1) upon plasma membrane localization of AR were tested. This was carried out through quantification of cell surface localization of HA-tagged AR by flow cytometry. HEK 293T cells were transfected with pcDNA3.1 sub-cloned with either wild-type or mutant HA-tagged full-length AR or with an empty pcDNA3.1 vector as a negative control. Histograms demonstrating the mean fluorescence intensity for the negative control

relative to each AR helix-8 construct are shown in Figure 5.10. Reduction in the mean fluorescence intensity for all of the mutants is very apparent from the histograms. Statistical analysis shows that these differences are significant. It is interesting to note that the plasma membrane localization of the AR-FLIP mutant (where the topology of amphipathy for the helix-8 is flipped) is significantly higher than that for the other mutants, but still significantly decreased relative to the wild-type. This indicates that both the secondary structure and the amphipathic topology of helix-8 play a role in plasma membrane localization of AR.

Confocal microscopy was also used to visually confirm the observations made by flow cytometry analysis and determine the localization of AR mutants not on the cell surface. AR (green) showed a reticular staining pattern throughout cells that extended out to the cell periphery, characteristic of a membrane protein trafficking from the ER to the plasma membrane (Figure 5.11; anti-HA). Reticular staining pattern throughout cells was also observed for PDI (an ER marker; red) that did not extend to the periphery (Figure 5.11; anti-PDI). Whereas, AR-G and AR-S showed similar reticular staining, but minimal staining in the cell periphery. In comparison to AR, higher colocalization of AR-G and AR-S with PDI was observed in the overlay, which is evident from the yellow color. For AR-G and AR-S reticular staining using PDI can be observed beyond HA staining, indicating reduced localization of the mutants on the cell surface. However, due to small size of the cells and low magnification it remains difficult to comment on specific localization of the AR mutants.

To reduce reticular staining, cyclohexamide treatment was done 2 hours prior to fixation and permeabilization to arrest protein synthesis and allow the protein already produced to reach its destination. As expected, cyclohexamide treated cell expressing AR

showed minimal reticular staining and maximal staining in the periphery (Figure 5.12). Whereas cell expressing AR-G and AR-S showed minimal staining in the periphery and maximal staining dispersed within the cell. Colocalization with ER was not observed in overlay with PDI for any of the constructs, indicating that the receptor is not retained in ER. However, microscopy data clearly demonstrates reduced plasma membrane localization for helix-8 mutants of AR, further investigation needs to be done to ascertain the specific localization of the mutants.

5.4. Discussion:

Prior to the atomic-level structural characterization of rhodopsin in 2000 (Palczewski *et al.*, 2000), the presence of helix-8 was suggested in various studies on Class A GPCRs (Huynh *et al.*, 2009). The first rhodopsin structure in 2000 was the first high resolution physical evidence showing the presence of helix-8. Since then, it has become clear that helix-8 is a consistent feature in most class A GPCRs. Congruous to this accepted notion for class A GPCRs (Huynh *et al.*, 2009), biophysical data presented in this work show the presence of a 12-15 amino acid long helix-8 in AR that is accentuated in the presence of membrane-mimetic micelles.

The specific role of this short, amphipathic juxtamembrane helix is not clear. Various aspects of this helix have been linked to function. One such aspect is the presence of cysteine residues at the C-terminus of helix-8. These Cys residues fall at positions 325 and 326 in AR, near the C-terminus of helix-8 according to NMR chemical shift data (Figure 5.9). Acylation of these cysteines has been shown to be involved in regulation of signaling, internalization and desensitization (Qanbar & Bouvier, 2003). However, there are GPCRs that either do not contain cysteine or acylation of these

cysteines is not required for their function (Huynh *et al.*, 2009). This points towards a probable difference in the role of helix-8 in these two categories of receptors.

The juxtamembrane region of many GPCRs also contains a F(X)₆LL motif, which has been shown to be important for anterograde transport (Duvernay *et al.*, 2009; Hammad *et al.*, 2012). However, recent work with TGR5 showed that it is the secondary structure imparted by the presence of this motif which facilitates the anterograde transport. Their analysis reveals that F(X)₆LL motif is commonly found in helical region of proteins. They also showed that the conservative replacement of phenylalanine by tyrosine in this motif does not change the α -helix content with respect to the WT and still showed a high level of membrane localization and retained functional activity. They conclude that the F(X)₆LL motif has an important role in plasma membrane localization of GPCRs, not as sequence-specific sorting motif, but due to the secondary structure imparted by it. Work on TGR5 also shows that the presence of β -strands or a shorter α -helical stretch in the same place leads to retention of the receptor in the ER (Spomer *et al.*, 2014). Interestingly, AR does not possess the F(X)₆LL motif, indicating towards a different mechanism that governs the plasma membrane localization of AR.

Adding to the understanding of importance of helix-8 in GPCR plasma localization, our data demonstrates that it is the amphipathy in combination with α -helical secondary structure that is required for efficient plasma membrane localization of AR. However, it remains uncertain if the reduction in plasma membrane localization for the AR C-tail mutants is due to defect in anterograde transport or is due to bias towards receptor internalization. This remains an interesting question that should be pursued further.

Regulation of downstream signaling is also an important role of helix-8 that has recently started gaining in popularity. The most recent structure of rhodopsin in complex with arrestin-1 shows the interaction of the N-terminus of helix-8 with the finger loop of arrestin-1 (Kang *et al.*, 2015). Previously, it was also shown that helix-8 dynamics are involved in the regulation of arrestin activation (Kirchberg *et al.*, 2011). The low hetNOE (relative to a rigid secondary structure) observed in the helix-8 of AR also indicate that it may still exhibit significant dynamics, potentially with similar implications to the case of rhodopsin. Another example highlighting the importance of helix-8 is the recent work by (Sensoy & Weinstein, 2015), where helix-8 of the dopamine D2 receptor has been shown to interact with the PDZ domain of GIPC1 having implications in downstream signaling.

In summary, I have successfully expressed, purified, and structurally characterized the 71 amino acid long C-terminal tail of the AR. I specifically identify the presence of a 12-15 amino acid long juxtamembrane helix, thus identifying helix-8 of the AR for the first time, and show that this helix influences the plasma membrane localization of the receptor. Importantly, both the helical character and the amphipathic topology of this helix-8 influence the plasma membrane localization of AR. To date, this represents the most detailed structure-function analysis of the 8th helix of a GPCR.

Table 5.1: Primers used in making plasmid constructs for the WT AR C-tail and its mutants for expression in *E. coli*.

Plasmid name	Forward primer (5' – 3')	Reverse primer (5' – 3')
pAR C-tail-WT	ATCCGTGCTCTTCAGGA GCATTTTTTCGACCCGCG TTTT	ATATGCGGATCCTTAG TCCACCACCAGGGTTT CCT
pAR C-tail-Gly	TCCACCTCTGGTGGTTG CTGCGGTCAA	ACCCTGACGACCACGC GGGTCGAAAA
pAR C-tail-Ser	TCCACCTCTTCTTCCTG CTGCGGTCAA	GGA CTGACGAGAACGC GGGTCGAAAA
pAR C-tail-Pro	TCCACCTCTCCACCATG CTGCGGTCAA	TGGCTGACGTGGACGC GGGTCGAAAA
pAR C-tail-FLIP	TCCACCTGCGCCCAGC GTTGCTGCGGTCAAAG CCGTTCTG	CATCAGCGGACGAAAG TCGAAAAATGCTCCAC CAATCT

Table 5.2: List of 2D and 3D experiments acquired for WT AR C-tail in buffer with and without LPPG micelles.

Experiment	Recovery delay (s)	# of scans		# of complex points	Sweep width (Hz)	Center position (ppm)	¹ H frequency (MHz)	Facility
Without LPPG								
¹ H- ¹⁵ N HSQC	1	192		¹ H:2048 ¹⁵ N:144	¹ H:6009.6 ¹⁵ N:1520.5	¹ H:4.7 ¹⁵ N:119.5	500	NMR-3
With LPPG								
¹ H- ¹⁵ N HSQC	1	32		¹ H:2048 ¹⁵ N:128	¹ H:6009.6 ¹⁵ N:1216.4	¹ H:4.7 ¹⁵ N:119.5	500	NMR-3
¹ H- ¹³ C HSQC	1	16		¹ H:2048 ¹³ C:512	¹ H:8415.5 ¹³ C:12324.2	¹ H:4.7 ¹³ C:39	700	NRC-IMB
¹ H- ¹⁵ N HSQC	1	16		¹ H:2048 ¹⁵ N:256	¹ H:8417.5 ¹⁵ N:1419	¹ H:4.7 ¹⁵ N:116.6	700	NRC-IMB
HNCO	1	8		¹ H:2048 ¹³ C:40 ¹⁵ N:48	¹ H:8417.5 ¹³ C:1479 ¹⁵ N:1362.3	¹ H:4.7 ¹³ C:174.1 ¹⁵ N:116.6	700	NRC-IMB
HNcaCO	1	16		¹ H:2048 ¹³ C:40 ¹⁵ N:48	¹ H:8417.5 ¹³ C:1479 ¹⁵ N:1362.3	¹ H:4.7 ¹³ C:174.1 ¹⁵ N:116.6	700	NRC-IMB
HNCA	1	16		¹ H:2048 ¹³ C:56 ¹⁵ N:48	¹ H:8417.5 ¹³ C:4296 ¹⁵ N:1362.3	¹ H:4.7 ¹³ C:51.4 ¹⁵ N:116.6	700	NRC-IMB
HNcoCA	1	16		¹ H:2048 ¹³ C:56 ¹⁵ N:48	¹ H:8417.5 ¹³ C:4296 ¹⁵ N:1362.3	¹ H:4.7 ¹³ C:51.4 ¹⁵ N:116.6	700	NRC-IMB
CACBNH	1	16		¹ H:2048 ¹³ C:128 ¹⁵ N:48	¹ H:8417.5 ¹³ C:9683.3 ¹⁵ N:1362.3	¹ H:4.7 ¹³ C:41.5 ¹⁵ N:116.6	700	NRC-IMB
CACBcoNH	1	16		¹ H:2048 ¹³ C:128 ¹⁵ N:48	¹ H:8417.5 ¹³ C:9683.3 ¹⁵ N:1362.3	¹ H:4.7 ¹³ C:41.5 ¹⁵ N:116.6	700	NRC-IMB
¹ H- ¹⁵ N heteronuclear NOE HSQC	5	32		¹ H:2048 ¹⁵ N:320	¹ H:8417.5 ¹⁵ N:1362.3	¹ H:4.7 ¹⁵ N:116.6	700	NRC-IMB

Table 5.3: Deconvolution of CD data using the BeStSel server for AR, AR-G, and AR-S in pH 6 condition.

Estimated secondary structure content (%) at pH 6												
	AR				AR-G				AR-S			
	ND	SDS	DPC	LPPG	ND	SDS	DPC	LPPG	ND	SDS	DPC	LPPG
Helix1 (regular):	4.6	9.9	5.6	9.3	0.7	10.8	1.2	6	0	11	0.7	8.5
Helix2 (distorted):	6.2	6.4	6.7	6.8	8.6	7.9	7	7.3	8.5	9.7	8.2	10.3
Anti1 (left-twisted):	0	0	0	0	0	0	0	0	0	0	0	0
Anti2 (relaxed):	2.8	1.7	2.4	1	0	0.5	0	0	0	0	0	0
Anti3 (right-twisted):	15.1	16.2	14.4	15.1	12	12.5	15	10.7	13.4	13	12.1	7.6
Parallel:	0	3.5	0	0.2	0	3.3	0	0	0	1.4	0	0
Turn:	19.7	16.1	18.8	18.7	19.2	16.2	18.1	17.5	17.7	16.6	19.5	20
Others:	51.6	46.4	52.1	49	59.5	48.8	58.6	58.3	60.3	48.3	59.5	53.5
Total estimated secondary structure content for each major secondary structure (%)												
Helix:	10.8	16.4	12.3	16	9.4	18.7	8.2	13.3	8.5	20.7	8.8	18.8
Antiparallel:	17.9	17.7	16.8	16.1	12	13	15	10.8	13.4	13	12.1	7.6
Parallel:	0	3.5	0	0.2	0	3.3	0	0	0	1.4	0	0
Turn:	19.7	16.1	18.8	18.7	19.2	16.2	18.1	17.5	17.7	16.6	19.5	20
Others:	51.6	46.4	52.1	49	59.5	48.8	58.6	58.3	60.3	48.3	59.5	53.5
Spectral deviation												
RMSD:	0.1719	0.1296	0.157	0.1608	0.2539	0.1427	0.2292	0.1843	0.3053	0.1853	0.272	0.3291
NRMSD:	0.03168	0.02266	0.02974	0.03084	0.03331	0.02128	0.0321	0.02843	0.03609	0.02259	0.3028	0.03448

Table 5.4: Deconvolution of CD data using the BeStSel server for AR, AR-G, and AR-S in pH 7 condition.

Estimated secondary structure content (%) at pH 7												
	AR				AR-G				AR-S			
	ND	SDS	DPC	LPPG	ND	SDS	DPC	LPPG	ND	SDS	DPC	LPPG
Helix1 (regular):	4.2	9.4	5.2	10	2.5	9.7	4.9	7.6	0	9.4	0.1	3.3
Helix2 (distorted):	4.5	6.8	6.4	6.7	4.6	5.3	4.9	6.1	7.2	9.8	8.6	6.7
Anti1 (left-twisted):	0	0	0	0	0	0	0	0	0	0	0	0
Anti2 (relaxed):	4	2.9	3.7	0.4	3	3.2	2.7	2	0	0	0	0
Anti3 (right-twisted):	17.4	12.6	12.8	15	19.2	12.4	15.5	13.9	15.3	11.5	12.9	14.9
Parallel:	0	2.7	0	1.7	0	3.6	0	0	0	0	0	0
Turn:	18.9	16.3	19.5	19.7	18.8	16.1	18.9	20.1	18.2	16.9	18.9	16.8
Others:	51	49.2	52.4	46.6	51.9	49.8	53.1	50.2	59.2	52.4	59.6	58.2
Total estimated secondary structure content for each major secondary structure (%)												
Helix:	8.7	16.2	11.6	16.6	7.1	15	9.8	13.8	7.2	19.2	8.7	10
Antiparallel:	21.4	15.5	16.5	15.4	22.2	15.6	18.2	15.9	15.3	11.5	12.9	14.9
Parallel:	0	2.7	0	1.7	0	3.6	0	0	0	0	0	0
Turn:	18.9	16.3	19.5	19.7	18.8	16.1	18.9	20.1	18.2	16.9	18.9	16.8
Others:	51	49.2	52.4	46.6	51.9	49.8	53.1	50.2	59.2	52.4	59.6	58.8
Spectral deviation												
RMSD:	0.1566	0.1301	0.1316	0.1657	0.1675	0.1338	0.1563	0.1302	0.2460	0.2110	0.2560	0.2062
NRMSD:	0.03247	0.02327	0.02860	0.02974	0.03011	0.02340	0.02694	0.02537	0.03006	0.02584	0.03016	0.02740

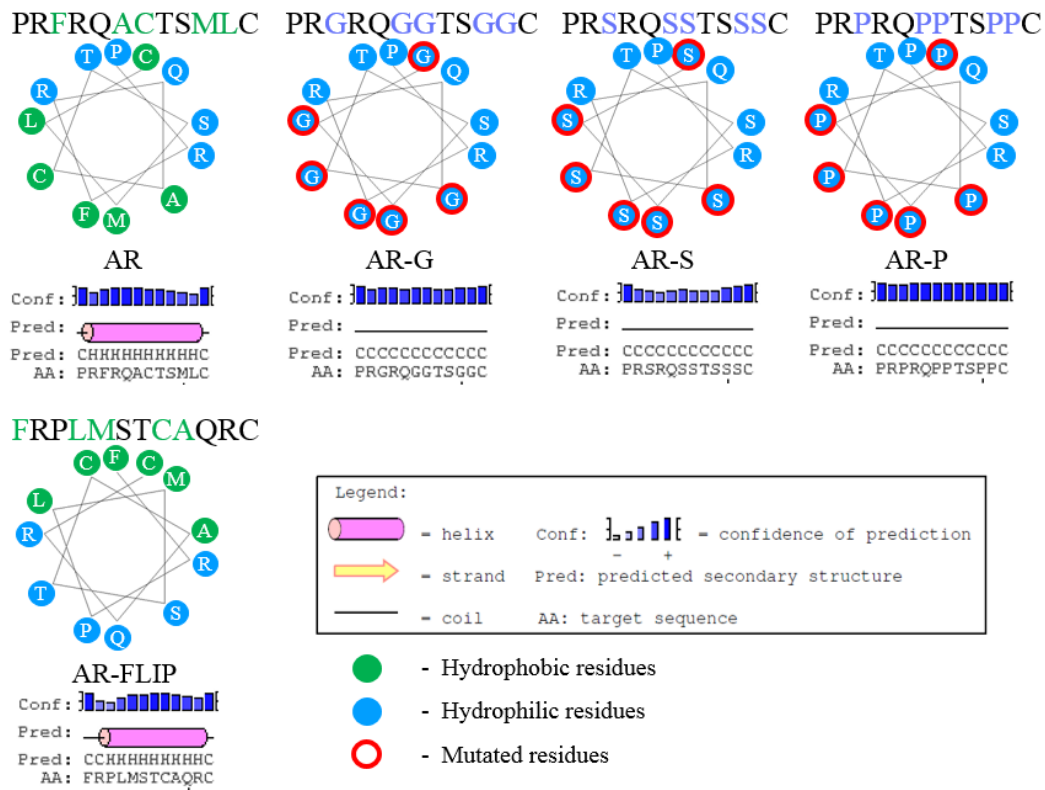


Figure 5.1: Sequence-based secondary structure prediction for the WT AR C-tail and its mutants. Bioinformatic analysis of the WT AR C-tail amino acid sequence representing an amphipathic helical region and the various mutants prepared to evaluate the structure and function of the predicted helix. Secondary structure predictions have been made using PSIPRED server v3.3 (Buchan *et al.*, 2013). Helical wheel has been plotted using the EMBOSS server (<http://embooss.bioinformatics.nl/cgi-bin/embooss/pepwheel>).

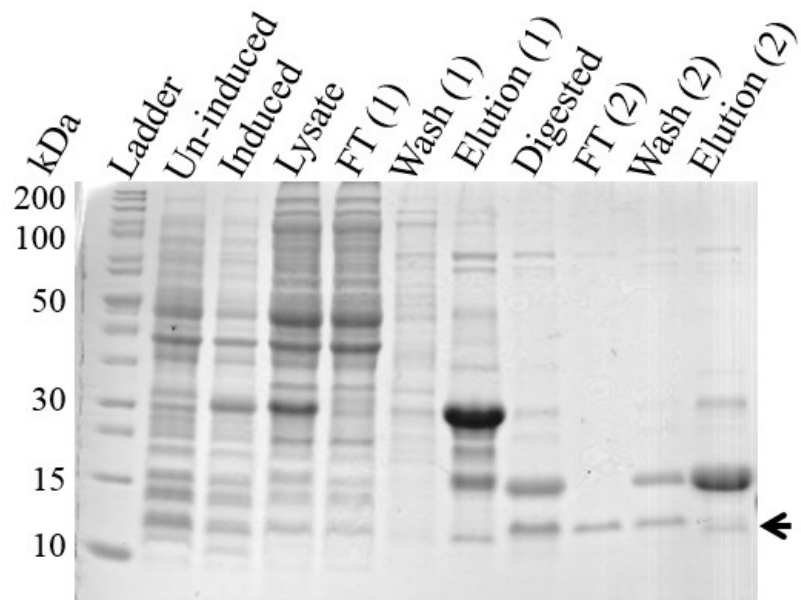


Figure 5.2: SDS-PAGE demonstrating successful expression and purification of the WT AR C-tail. 15% SDS-polyacrylamide gel showing AR C-tail expression, cleavage and Ni-NTA purification steps as described in Methods. The arrow denotes the position of AR C-tail on the gel following both purification steps and SUMO protease cleavage.

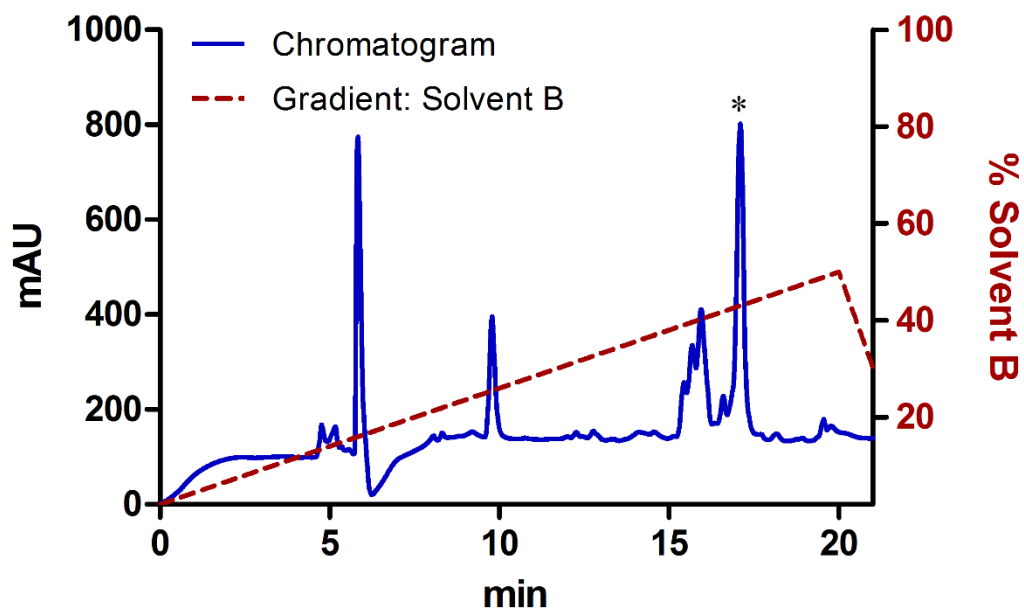


Figure 5.3: HPLC chromatogram demonstrating successful purification of WT AR C-tail. Analytical RP-HPLC chromatogram (milliabsorbance units (mAU) at 220 nm; solid blue line) for WT AR C-tail. The binary solvent gradient (A:B) is illustrated with the dashed red line (A: H₂O; B: acetonitrile; both containing 0.1%TFA). The * denotes the peak corresponding to WT AR C-tail.

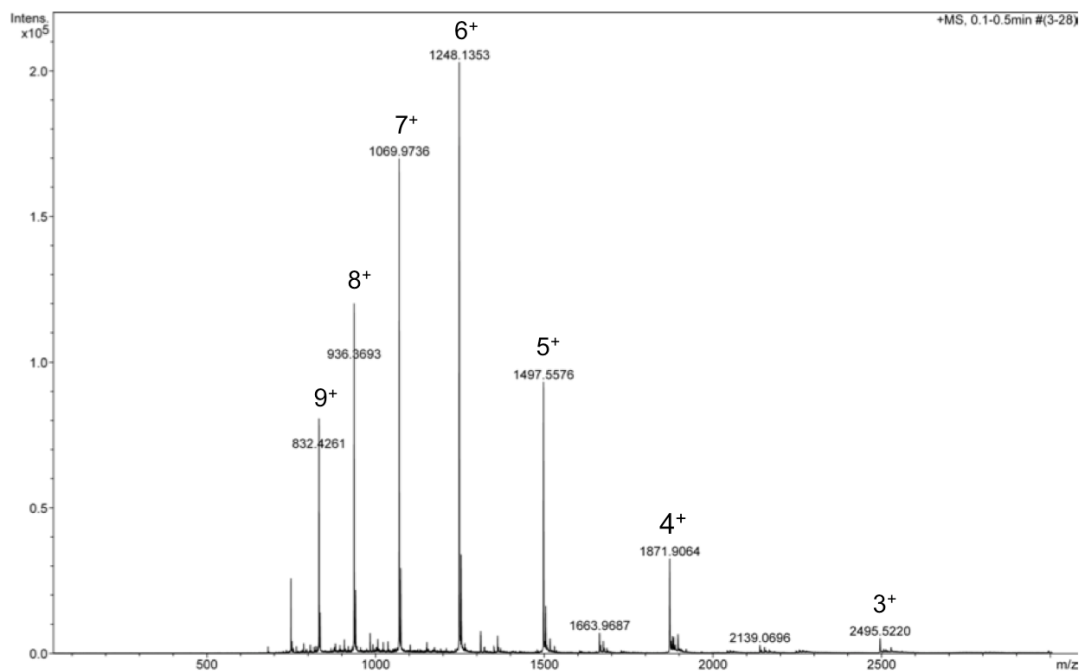


Figure 5.4: Identity of WT AR C-tail confirmed by mass spectrometry. Positive ion mode ESI-mass spectrum of HPLC purified WT AR C-tail. Corresponding charges are denoted by numbers above each charged species. Monoisotopic MW of AR C-tail = 7480.37; for the +4 charge in the mass spectrum $MW + 4 = 7484.37$ ($\sim 1871.9064 * 4$)

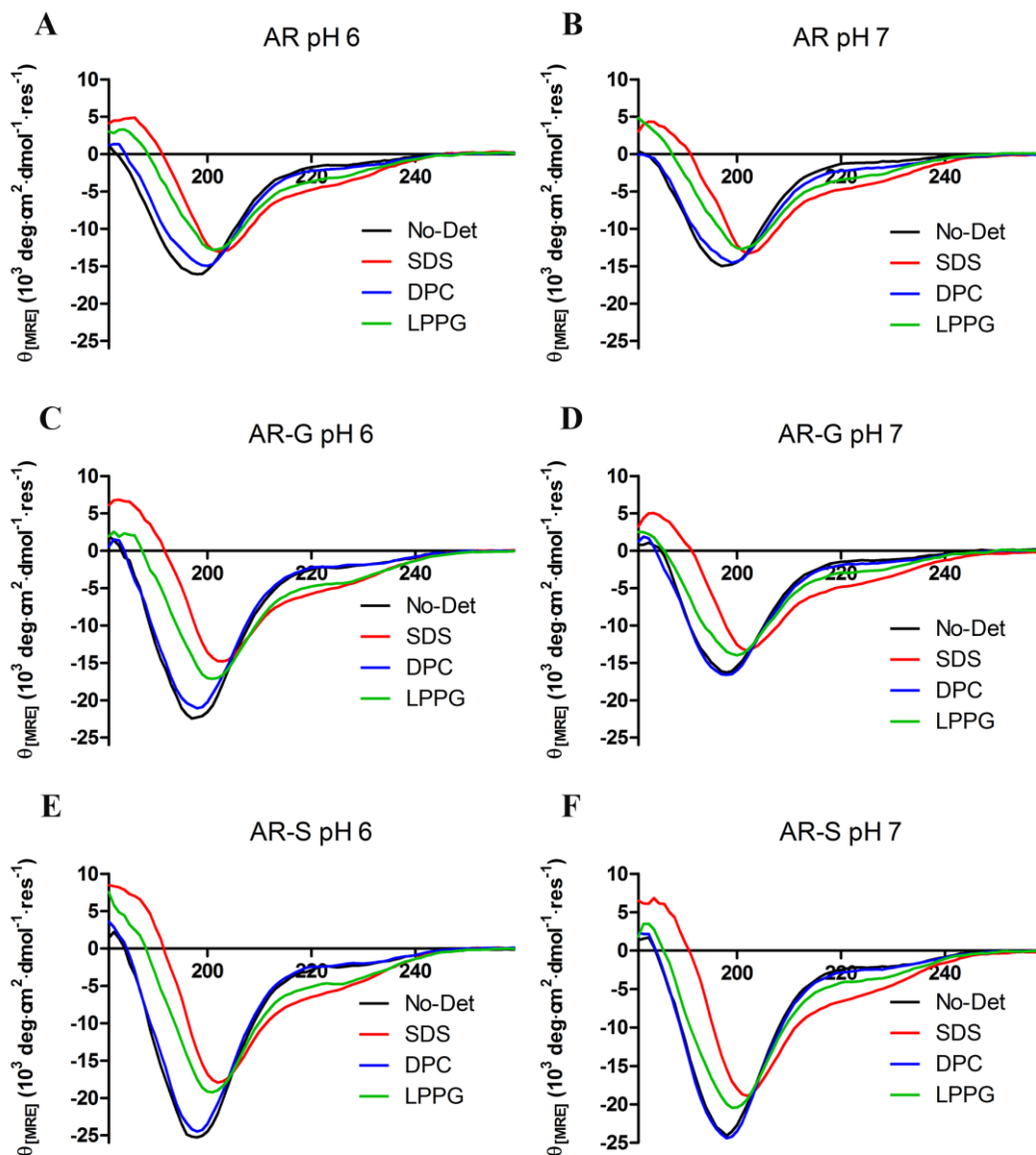


Figure 5.5: CD spectra for the WT AR C-tail and its mutants with and without membrane-mimetics. Far-UV CD spectra of WT AR C-tail (AR; A and B), glycine mutant (AR-G; C and D) and the serine mutant (AR-S; E and F) in buffer (No-Det) or indicated micellar environment, allowing evaluation of secondary structure composition at indicated pH. Protein concentration used in all the experiments was 150 μ M. All the data was collected at 25 $^{\circ}$ C.

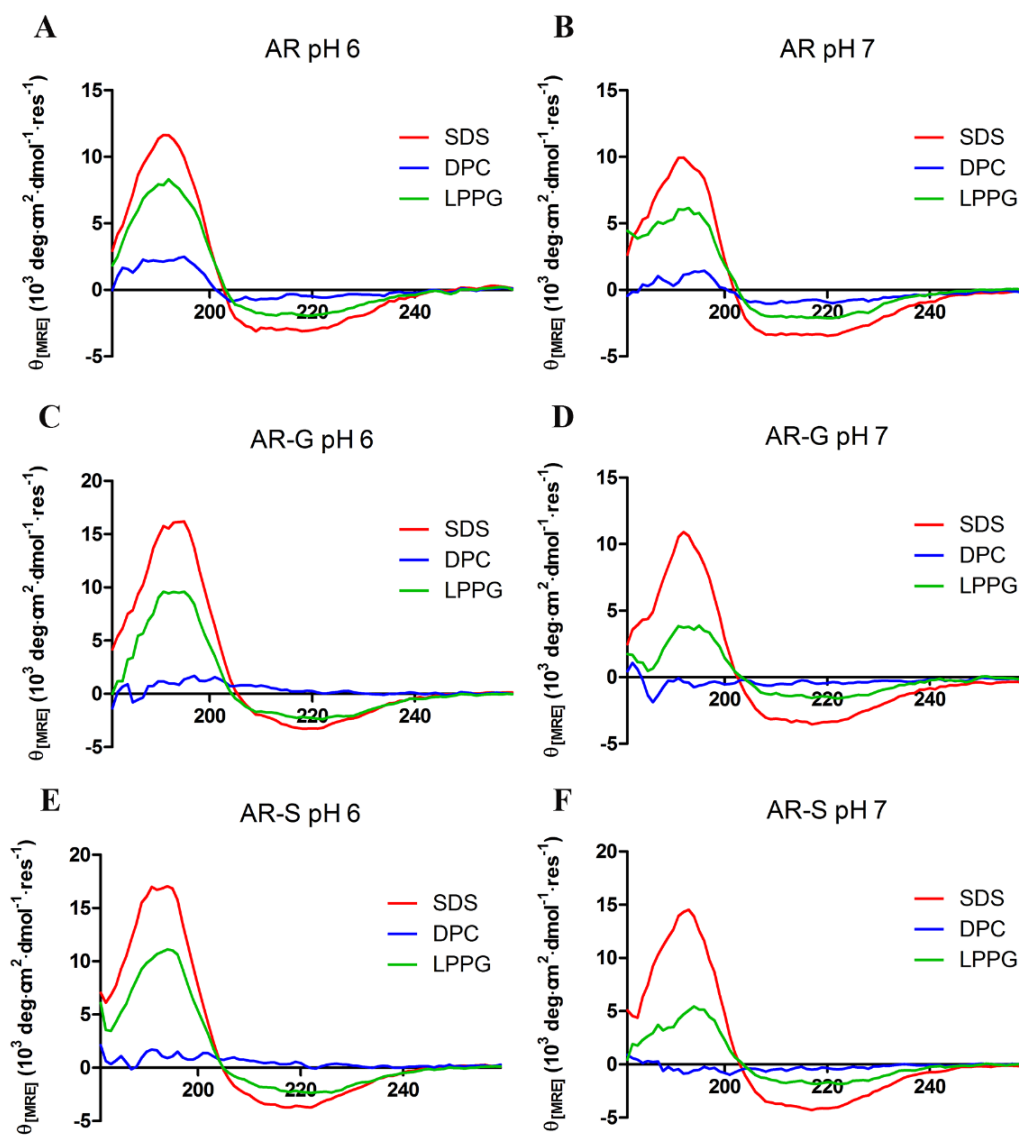


Figure 5.6: Difference CD spectra for the WT AR C-tail and its mutants in various membrane-mimetics. Far-UV CD difference spectra of WT AR C-tail (AR; A and B), glycine mutant (AR-G; C and D) and the serine mutant (AR-S; E and F) allowing evaluation of membrane-mimetic environment induced secondary structure composition at indicated pH. Protein concentration used in all the experiments was 150 μM . All the data was collected at 25 $^{\circ}\text{C}$.

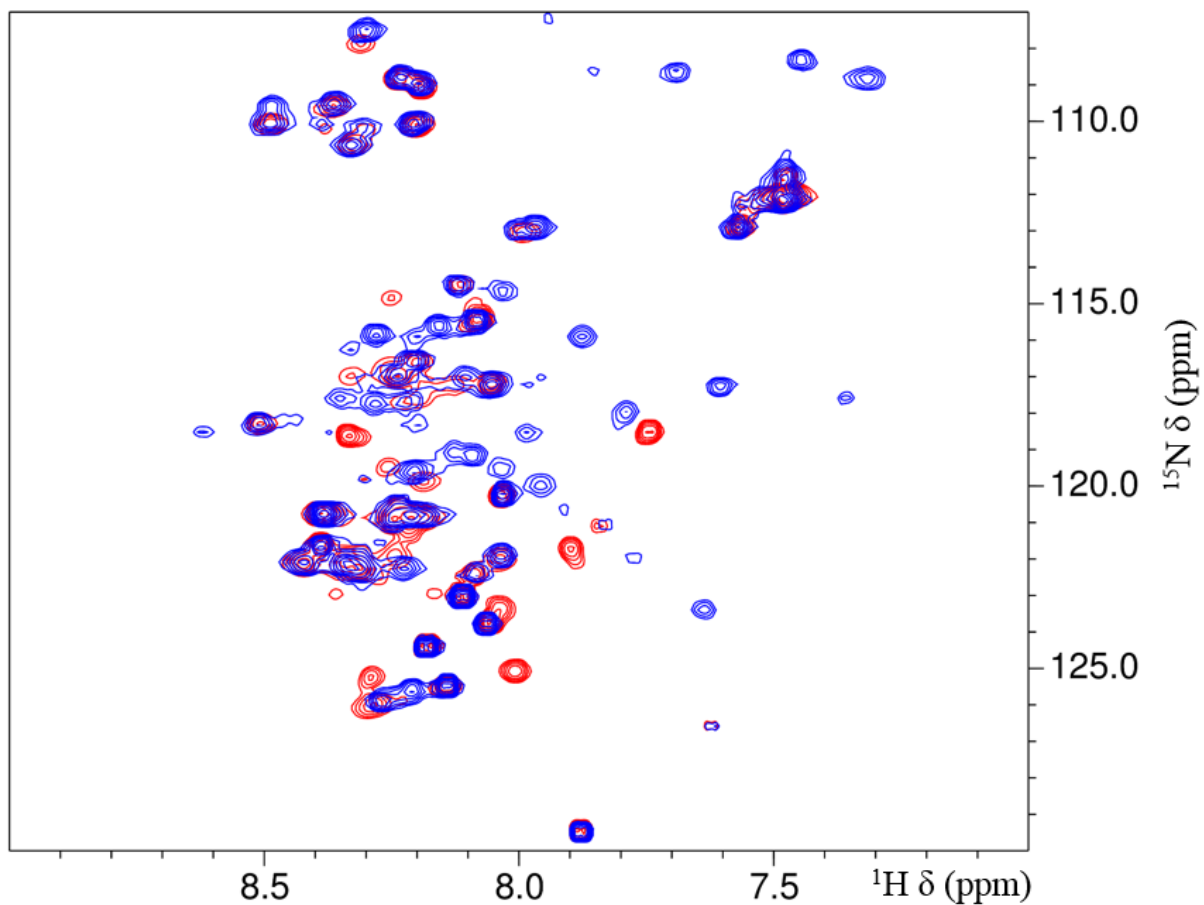


Figure 5.7: Comparison of ^1H - ^{15}N HSQC spectra for the WT AR C-tail in buffer with and without LPPG micelles. ^1H - ^{15}N HSQC spectra overlay for AR C-tail in the absence of detergent micelles (red) and the presence of 120 mM LPPG (blue), both collected at 37°C and pH 6, using a 500 MHz spectrometer.

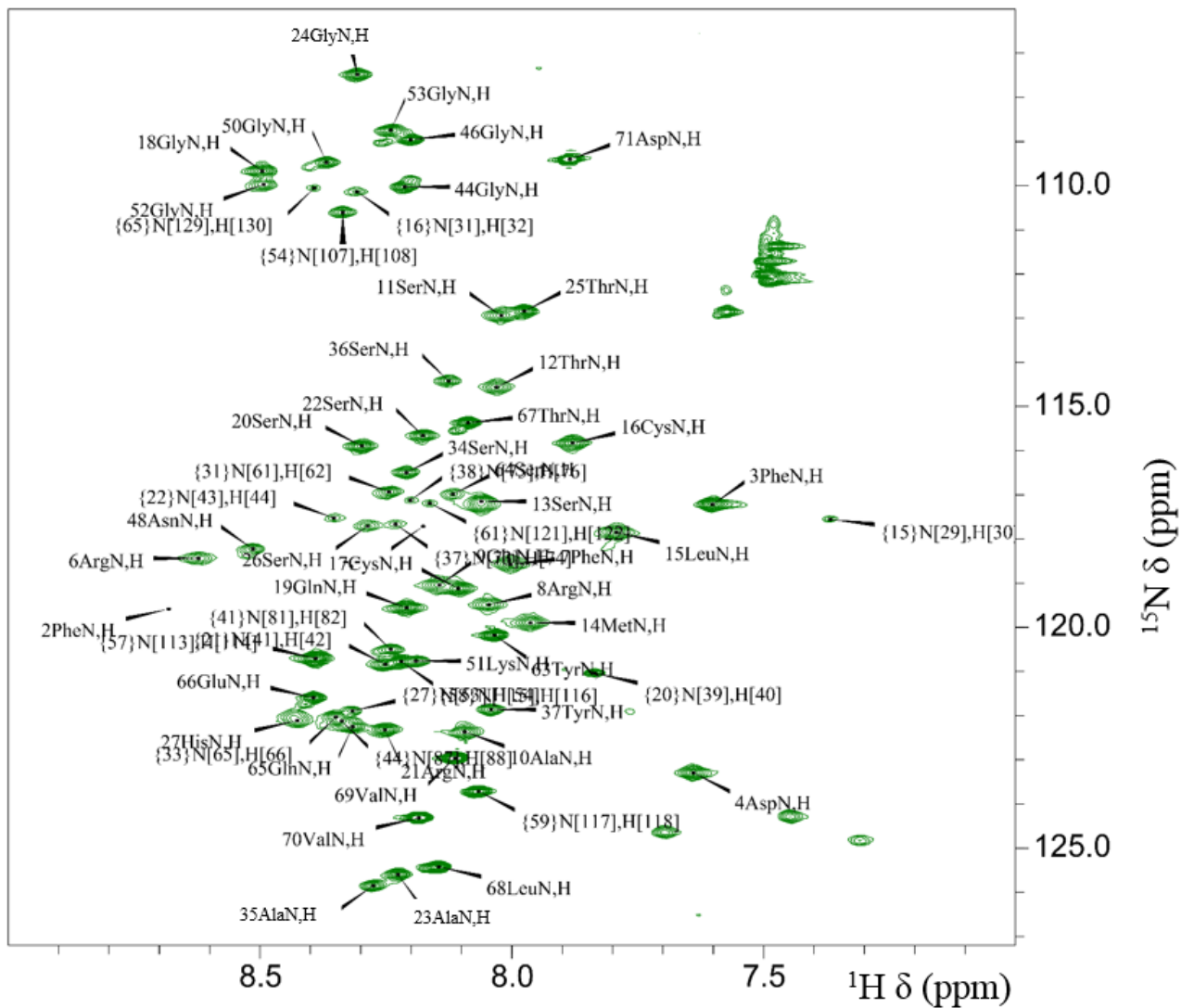


Figure 5.8: Assigned ^1H - ^{15}N HSQC for WT AR C-tail at 37 °C in presence of 100 mM LPPG. ^1H - ^{15}N HSQC spectrum for AR C-tail in presence of 100 mM LPPG. Data was collected at 37°C and pH 6, using a 700 MHz spectrometer.

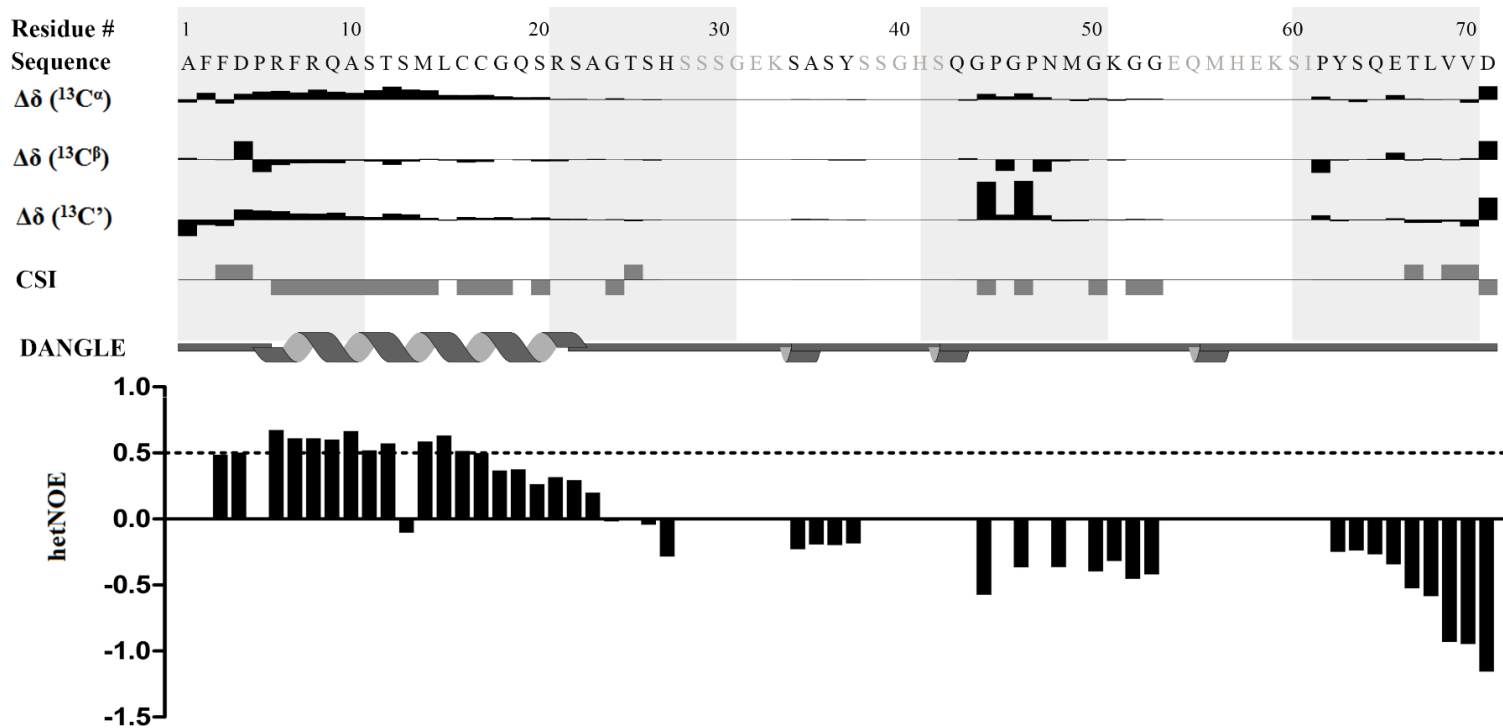


Figure 5.9: Estimation of secondary structure and dynamics of WT AR C-tail using NMR spectroscopy. (Top to bottom) Residue numbers are annotated above the amino acid sequence (assigned residues black, those remaining ambiguous grey) followed by the secondary chemical shifts for $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and the backbone carbonyl ($^{13}\text{C}'$) relative to the random coil chemical shifts reported by (Wishart *et al.*, 1995). Values of CSI are shown above the secondary structure predictions made by DANGLE. (The top part of the figure was modified from the output of CcpNmr Analysis 2.4.2.) Finally, ^1H - ^{15}N heteronuclear NOE enhancement values are plotted per residue, as calculated by taking the ratio of amide peak intensities in the saturated ^1H - ^{15}N HSQC spectrum to those in the unsaturated ^1H - ^{15}N HSQC spectrum.

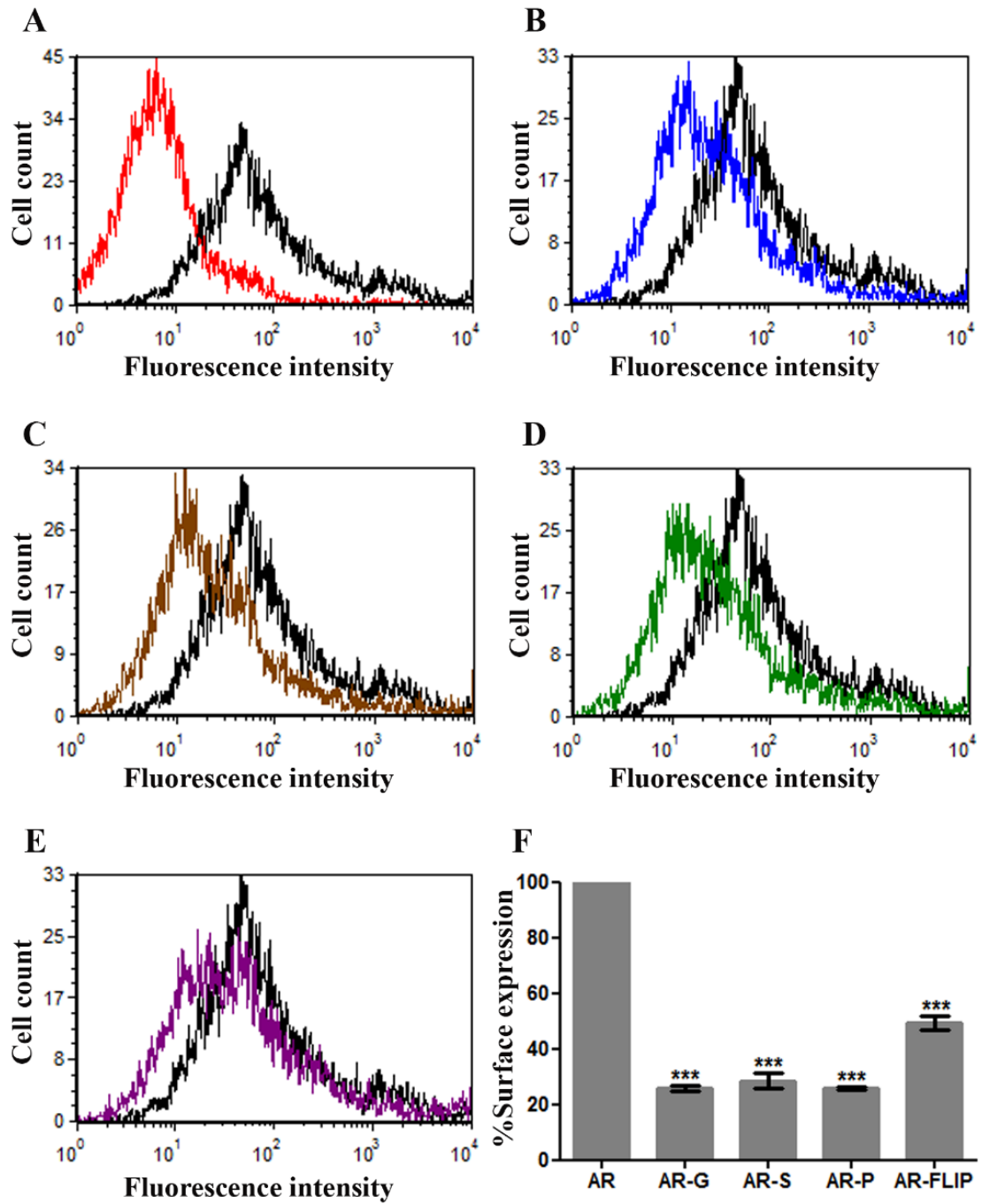


Figure 5.10: Mutations in helix-8 reduce plasma membrane localization of full-length AR. HEK 293T cells transfected with empty vector or with HA-tagged wild-type or mutant (Figure 5.1) AR were surface stained with anti-HA antibody and analyzed by flow cytometry. **A-E** show histograms with fluorescence intensity on the x-axis and cell count on the y-axis for empty vector (red, A), AR-G (blue, B), AR-S (brown, C), AR-P (green, D), and AR-FLIP (purple, E) overlaid with wild-type AR in each case (black). **(F)** Cell surface expression quantified using mean fluorescence intensity and reported as percent surface expression \pm SEM from three independent experiments performed in triplicate. Statistical significance compares AR to each helix-8 mutant (***) indicates $p < 0.005$).

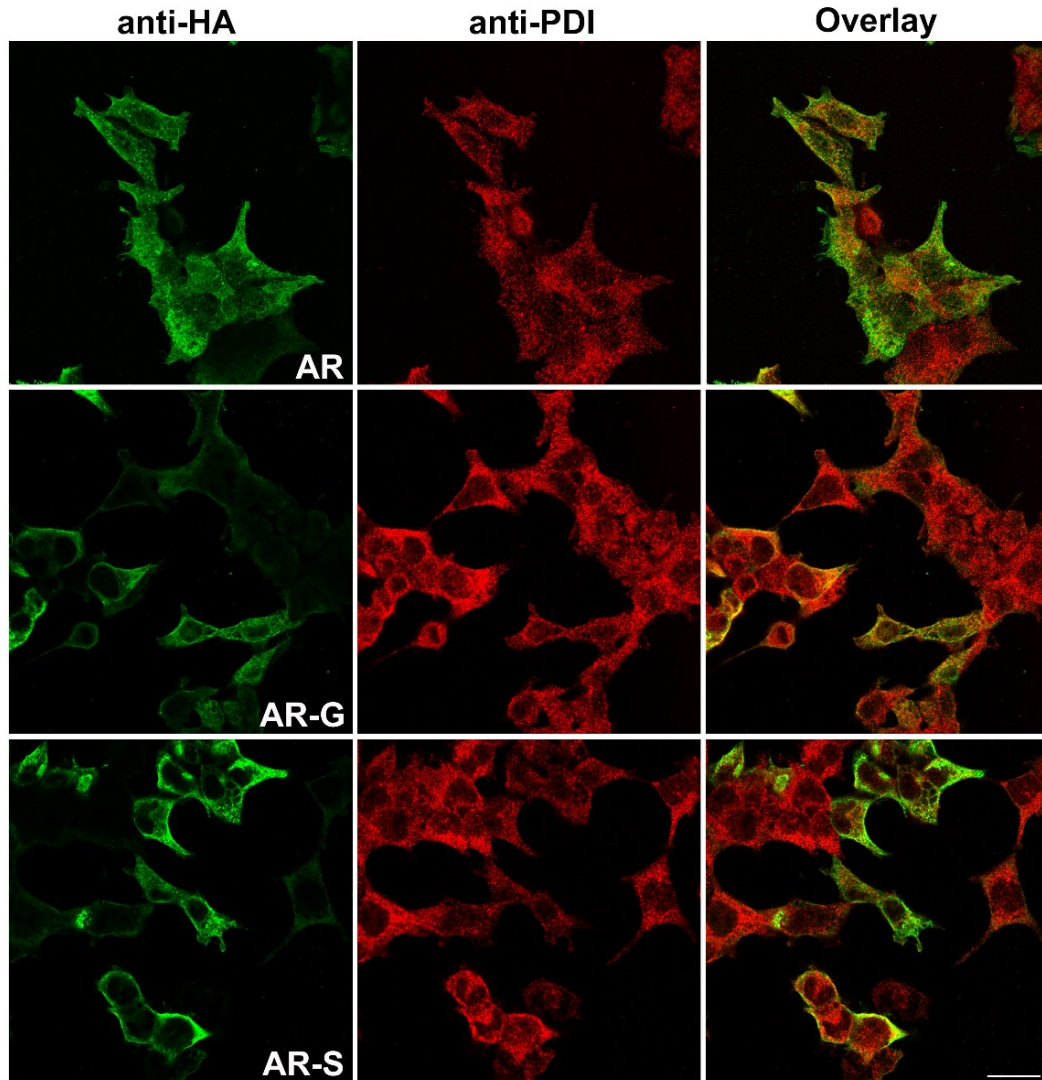


Figure 5.11: AR cell surface expression is reduced in AR helix-8 mutants. HEK 293T cells transfected with plasmids for the wild-type AR (AR), Glycine mutant (AR-G), and Serine mutant (AR-S) were fixed, permeabilized, and stained at 24 hours post-transfection with anti-HA antibody (green) and anti-PDI (red; ER marker). Right column shows merged images. Scale bar, 20 μ m.

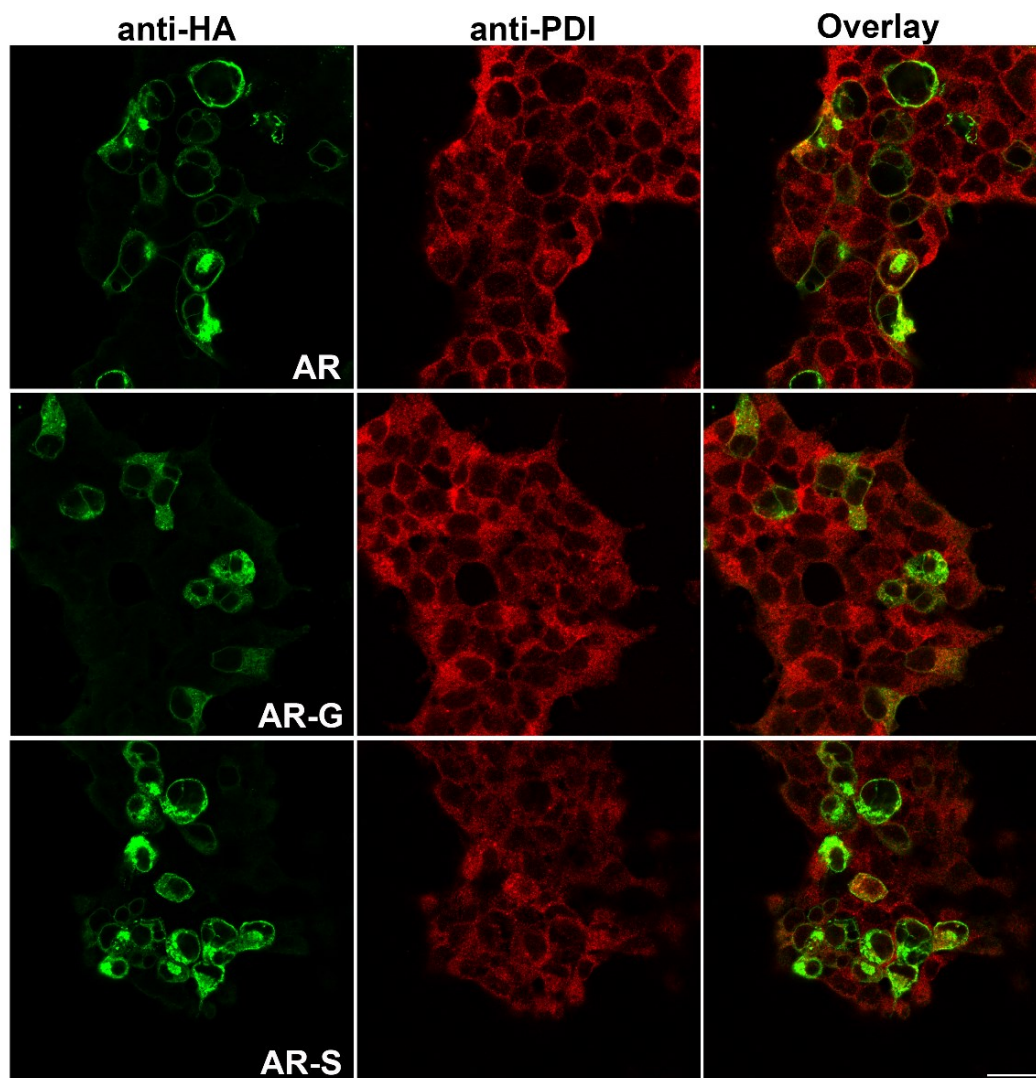


Figure 5.12: Cyclohexamide treatment confirms reduced cell surface expression for AR helix-8 mutants. HEK 293T cells transfected with plasmids for the wild-type AR (AR), Glycine mutant (AR-G), and Serine mutant (AR-S) were treated with cyclohexamide at 22 hours post-transfection and then fixed and stained at 2 hours post-treatment with anti-HA antibody (green) and anti-PDI (red; ER marker). Right column shows merged images. Scale bar, 20 μ m.

Chapter 6: Conclusions And Future Directions

In this thesis, I have provided a detailed description of the strategies I have developed and used to facilitate structural elucidation of the class A G-protein coupled apelin receptor. Herein, I have successfully demonstrated the production of large fragments of the AR, which will be used to gain structural understanding of the AR at an atomic level. I have also shown membrane-induced helix induction in a region of the AR C-terminal tail which would form a juxtamembrane helix, the so-called 8th helix, and have shown the functional significance of its helical structuring and amphiphathic topology in plasma membrane localization of AR.

6.1. AT-Rich Gene Tags As A Tool For Membrane Protein Expression:

A and T nucleotide-rich gene sequences near the ribosome binding site were shown herein to enhance the expression of a large fragment of the apelin receptor in *E. coli*. Four such AT-rich gene tags not only enhanced the expression of AR TM1-3 in *E. coli*, but made this possible; expression levels of this portion of the AR were not even observable by western blotting in the absence of an AT-tag. This represents the largest portion of a human GPCR produced in *E. coli* to date, outside of the full-length CXCR1 employed by Opella and co-workers in structural determination (Park *et al.*, 2012b). The small sizes (4-6 amino acids) of these tags provide an advantage over larger tags as the overall metabolic burden on the cell is reduced and removal may not be absolutely required. Based on the results shown in Chapter 2, these tags have a high potential to enhance expression of other so-called “difficult to-express” proteins. These tags may be of particular interest where the inefficiency in expression is due to inadequate initiation

of transcription or translation. However, it should be noted that the exact mechanism by which AT-rich sequences enhance expression is not fully understood; therefore, these tags may be even more generally applicable.

6.2. Possible Future Studies With AR TM1-3:

Following from the successful expression and purification of AR TM1-3, optimization of conditions was necessary to identify the membrane-mimetic environment best suited to characterization of this protein. My optimized conditions lay the ground work for future high-resolution structural studies using NMR spectroscopy. Although 1:1 TFE:H₂O conditions provided promising chemical shift dispersion and NMR spectroscopic behaviour, chemical shift assignment for AR TM1-3 in this condition has proven extremely challenging. It may be necessary to extend the suite of experiments beyond the standard set of triple-resonance backbone walk experiments in order to complete this assignment. As alluded to in Chapter 3, there is a possibility of lack of tertiary contacts in the 50% TFE condition. Completion of this assignment process would at least enable characterization of the secondary structure, allowing comparison to the topology predicted on the basis of conserved GPCR motifs. These chemical shift assignments could, in turn, potentially be used to assist in chemical shift assignment in other, more physiological conditions. Further screening of conditions is certainly required to identify a suitable membrane-mimetic for reconstitution of AR TM1-3. Following the work in our lab on the fragment of AR comprising residues 1-55 (Langelaan *et al.*, 2013), ongoing characterization of AR TM1-3 will also extend our understanding of structural and dynamic aspects of apelin binding to the apelin receptor.

6.3. Using Intein-Mediated Splicing To Obtain Full-Length AR:

As stated in Chapter 4, the ultimate goal of the AR project is to characterize the structure, dynamics, and ligand binding mechanism of the full-length receptor. I have used a “divide and conquer” approach to tackle the hurdle on sample preparation and facilitate structural elucidation by NMR spectroscopy. Following this approach I have shown the expression of the AR in two fragments fused to split-intein fragments. The potential of split-inteins to enable ligation of the two fragments of the AR makes them an attractive tool to achieve segmental-labeling that is desirable for characterization of challenging proteins by NMR. Although I observed evidence of splicing, for the conditions tested to date, I have not yet been achieved a highly efficient splicing reaction to allow preparation of a sufficient level of full-length AR for NMR studies. There still remain a number of conditions that can be modified in ongoing optimization of the splicing reaction. Successful splicing will provide a robust and generalizable method to enable structural characterization of large membrane proteins by NMR spectroscopy.

6.4. Role Of AR C-Tail In Plasma Membrane Localization Of AR And Possible Future Studies:

In the past few years, the C-terminal tail of GPCRs has been in focus for its role in modulation of GPCR-mediated signaling and for its role in regulation of GPCR trafficking. In light of the findings of Chapter 5, I have identified an α -helical region (commonly referred to as helix-8 in other GPCRs) in the C-terminal tail of AR. The degree of helicity is strikingly accentuated in the presence of membrane-mimetic micelles, giving credence to its annotation as a juxtamembrane helical segment. I have also successfully demonstrated the significance of the AR helix-8 in plasma membrane

localization of AR. Relative to the currently available literature, this work represents the most extensive study of a GPCR helix-8, with the most detailed correlation of structure to function to date. In future, based on the potential role that GPCR C-terminal tails are believed to play in interaction with intracellular proteins, this 71-residue protein may be extremely valuable to study the mode of interaction of AR with its intracellular partner proteins.

6.5. Significance Of This Work:

As described in the introduction, a large body of research relating to the adrenergic system has shown the role of this system in a wide variety of both physiological and pathological settings. This GPCR-ligand pair has been particularly highlighted for its involvement in cardiovascular system regulation, including the generation of new vasculature by tumours. The success achieved in my work has paved the way to high-resolution structural characterization of AR. With this work I have shown how AT-rich gene tags can be used for expression of notoriously difficult membrane proteins in *E. coli*. This work is the only example where a full-length GPCR has been expressed in two fragments, each attached to a split intein. The intein mediated splicing methodology has a very high potential to address the problem of signal overlap in characterization of full-length GPCRs using NMR spectroscopy. In this work, I have also identified a 12-15 amino acid long α -helix (helix-8) in the C-terminal tail of the AR. To date, this work represents the most extensive study of relating structure of helix-8 to its function for any GPCR. Due to involvement of helix-8 in plasma membrane localization, it has a high potential as a target for design of therapeutics to modulate signaling cascades regulated by the AR. Overall, this work is a significant step taken

towards complete elucidation of the AR structure and understanding the role of a structural motif conserved among class A GPCRs. Figure 6.1 demonstrates the different fragments that have been studied in this work.

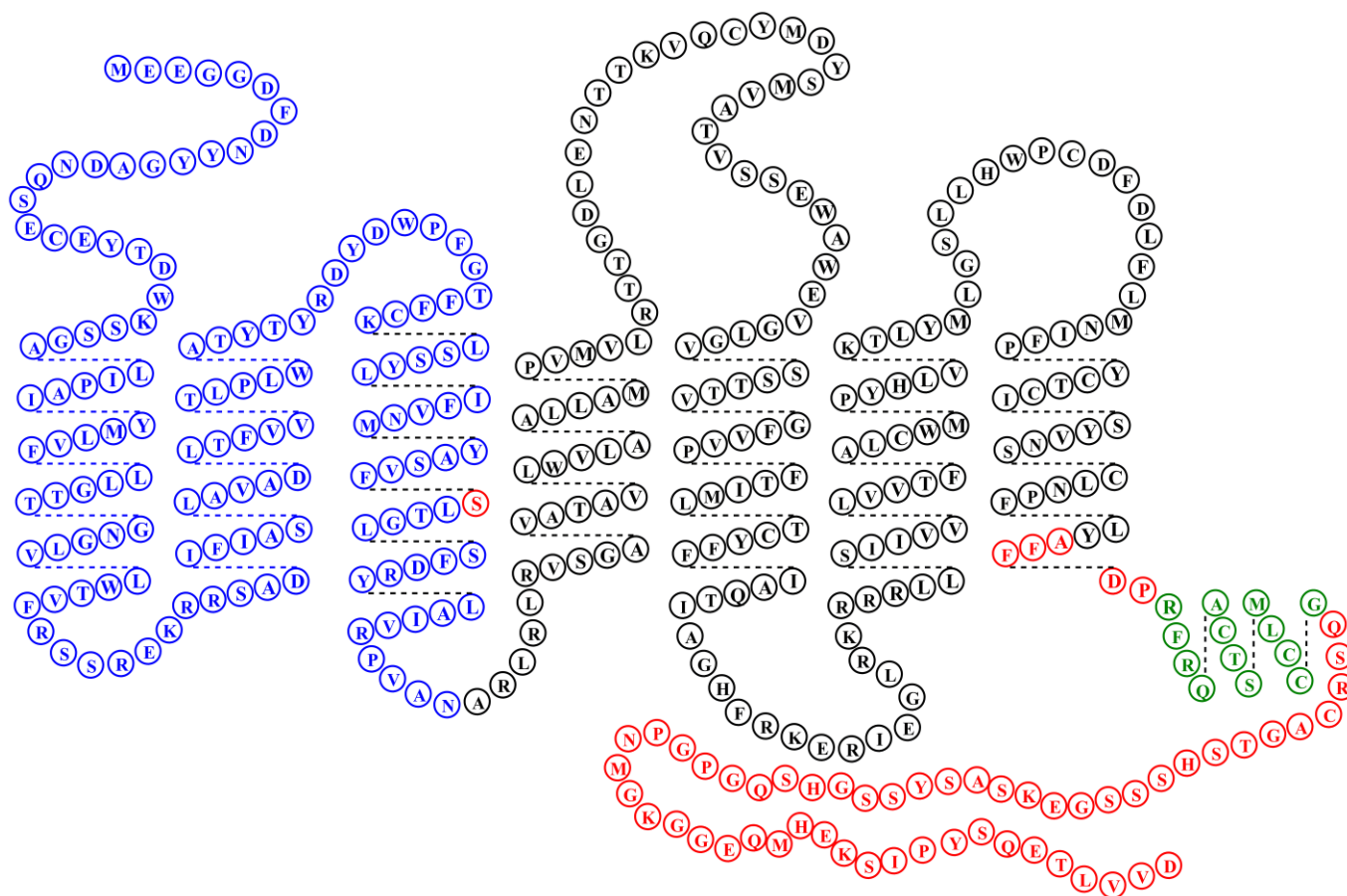


Figure 6.1: Fragments of AR studied in this work. AR TM1-3 is shown in blue, whereas AR TM4-7 is shown in of black, red, and green colour. AR C-tail is shown in red with the helix-8 represented in green.

REFERENCES:

1. Abdul-Gader A, Miles AJ, Wallace BA (2011). A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics* **27**(12): 1630-1636.
2. Ahn KH, Nishiyama A, Mierke DF, Kendall DA (2010). Hydrophobic residues in helix 8 of cannabinoid receptor 1 are critical for structural and functional properties. *Biochemistry* **49**(3): 502-511.
3. Alexandrov AI, Mileni M, Chien EYT, Hanson MA, Stevens RC (2008). Microscale Fluorescent Thermal Stability Assay for Membrane Proteins. *Structure* **16**(3): 351-359.
4. Allen JA, Roth BL (2011). Strategies to discover unexpected targets for drugs active at G protein-coupled receptors. *Annu Rev Pharmacol Toxicol* **51**: 117-144.
5. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M (2009). Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci U S A* **106**(27): 11005-11010.
6. Anastasina M, Terenin I, Butcher SJ, Kainov DE (2014). A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *Biotechniques* **56**(1): 36-39.
7. Andersen KR, Leksa NC, Schwartz TU (2013). Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **81**(11): 1857-1861.
8. Andrell J, Tate CG (2013). Overexpression of membrane proteins in mammalian cells for structural studies. *Mol Membr Biol* **30**(1): 52-63.
9. Aranko AS, Wlodawer A, Iwai H (2014). Nature's recipe for splitting inteins. *Protein Eng Des Sel* **27**(8): 263-271.
10. Aranko AS, Zuger S, Buchinger E, Iwai H (2009). In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS One* **4**(4): e5185.
11. Aratake Y, Okuno T, Matsunobu T, Saeki K, Takayanagi R, Furuya S, *et al.* (2012). Helix 8 of leukotriene B4 receptor 1 inhibits ligand-induced internalization. *Faseb j* **26**(10): 4068-4078.

12. Arinaminpathy Y, Khurana E, Engelman DM, Gerstein MB (2009). Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov Today* **14**(23-24): 1130-1135.
13. Arora A, Abildgaard F, Bushweller JH, Tamm LK (2001). Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat Struct Biol* **8**(4): 334-338.
14. Attwood TK, Findlay JB (1994). Fingerprinting G-protein-coupled receptors. *Protein Eng* **7**(2): 195-203.
15. Babu MM, Kriwacki RW, Pappu RV (2012). Structural biology. Versatility from protein disorder. *Science* **337**(6101): 1460-1461.
16. Baneyx F (1999). Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10**(5): 411-421.
17. Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP (1987). Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *Journal of Magnetic Resonance (1969)* **73**(1): 69-77.
18. Barnes G, Japp AG, Newby DE (2010). Translational promise of the apelin--APJ system. *Heart* **96**(13): 1011-1016.
19. Bax A, Clore GM, Gronenborn AM (1990). $^1\text{H} \cdot ^1\text{H}$ correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^1H and ^{13}C spectra of ^{13}C -enriched proteins. *Journal of Magnetic Resonance (1969)* **88**(2): 425-431.
20. Bermak JC, Li M, Bullock C, Zhou QY (2001). Regulation of transport of the dopamine D1 receptor by a new membrane-associated ER protein. *Nat Cell Biol* **3**(5): 492-498.
21. Bernaudat F, Frelet-Barrand A, Pochon N, Dementin S, Hivin P, Boutigny S, *et al.* (2011). Heterologous expression of membrane proteins: choosing the appropriate host. *PLoS One* **6**(12): e29191.
22. Better M, Bernhard SL, Lei SP, Fishwild DM, Lane JA, Carroll SF, *et al.* (1993). Potent anti-CD5 ricin A chain immunoconjugates from bacterially produced Fab' and F(ab')₂. *Proceedings of the National Academy of Sciences of the United States of America* **90**(2): 457-461.

- 23.** Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, *et al.* (2011). Overcoming barriers to membrane protein structure determination. *Nat Biotechnol* **29**(4): 335-340.
- 24.** Boer E, Steinborn G, Kunze G, Gellissen G (2007). Yeast expression platforms. *Appl Microbiol Biotechnol* **77**(3): 513-523.
- 25.** Bordag N, Keller S (2010). Alpha-helical transmembrane peptides: a "divide and conquer" approach to membrane proteins. *Chem Phys Lipids* **163**(1): 1-26.
- 26.** Bornert O, Alkhalfioui F, Logez C, Wagner R (2012). Overexpression of membrane proteins using *Pichia pastoris*. *Curr Protoc Protein Sci* **Chapter 29**: Unit 29.22.
- 27.** Botelho AV, Huber T, Sakmar TP, Brown MF (2006). Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophys J* **91**(12): 4464-4477.
- 28.** Boukadida C, Marnata C, Montserret R, Cohen L, Blumen B, Gouttenoire J, *et al.* (2014). NS2 proteins of GB virus B and hepatitis C virus share common protease activities and membrane topologies. *J Virol* **88**(13): 7426-7444.
- 29.** Brenzel S, Cebi M, Reiss P, Koert U, Mootz HD (2009). Expanding the scope of protein trans-splicing to fragment ligation of an integral membrane protein: towards modulation of porin-based ion channels by chemical modification. *Chembiochem* **10**(6): 983-986.
- 30.** Brodel AK, Sonnabend A, Kubick S (2014). Cell-free protein expression based on extracts from CHO cells. *Biotechnol Bioeng* **111**(1): 25-36.
- 31.** Brown LS, Ladizhansky V (2015). Membrane proteins in their native habitat as seen by solid-state NMR spectroscopy. *Protein Sci* **24**(9): 1333-1346.
- 32.** Bruch MD, Gierasch LM (1990). Comparison of helix stability in wild-type and mutant LamB signal sequences. *J Biol Chem* **265**(7): 3851-3858.
- 33.** Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**(Web Server issue): W349-357.

- 34.** Buck M (1998). Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys* **31**(3): 297-355.
- 35.** Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O (2008). Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expr Purif* **59**(1): 94-102.
- 36.** Butt TR, Edavettal SC, Hall JP, Mattern MR (2005). SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif* **43**(1): 1-9.
- 37.** Camponova P, Baud S, Matras H, Duroux-Richard I, Bonnafous JC, Marie J (2007). High-level expression and purification of the human bradykinin B(2) receptor in a tetracycline-inducible stable HEK293S cell line. *Protein Expr Purif* **55**(2): 300-311.
- 38.** Cantor CR, Schimmel PR (1980). *Biophysical Chemistry, Part 2: Techniques for the Study of Biological Structure and Function*. 2nd edn. W.H. Freeman and Company: San Francisco.
- 39.** Cao J, Li H, Chen L (2015). Targeting drugs to APJ receptor: the prospect of treatment of hypertension and other cardiovascular diseases. *Curr Drug Targets* **16**(2): 148-155.
- 40.** Caron MG, Srinivasan Y, Pitha J, Kociolek K, Lefkowitz RJ (1979). Affinity chromatography of the beta-adrenergic receptor. *J Biol Chem* **254**(8): 2923-2927.
- 41.** Cavanagh J (2007). *Protein NMR spectroscopy principles and practice*, 2nd edn, pp 1 online resource (xxv, 885 p.). Amsterdam ; Boston: Academic Press,.
- 42.** Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ (2007). *Protein NMR spectroscopy principles and practice.*, 2nd edn. Boston: Academic Press.
- 43.** Chelikani P, Reeves PJ, Rajbhandary UL, Khorana HG (2006). The synthesis and high-level expression of a beta2-adrenergic receptor gene in a tetracycline-inducible stable mammalian cell line. *Protein Sci* **15**(6): 1433-1440.
- 44.** Chen H, Shaffer PL, Huang X, Rose PE (2013). Rapid screening of membrane protein expression in transiently transfected insect cells. *Protein Expr Purif* **88**(1): 134-142.
- 45.** Cheng B, Chen J, Bai B, Xin Q (2012). Neuroprotection of apelin and its signaling pathway. *Peptides* **37**(1): 171-173.

- 46.** Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW (2010). DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance* **202**(2): 223-233.
- 47.** Chini B, Parenti M (2009). G-protein-coupled receptors, cholesterol and palmitoylation: facts about fats. *J Mol Endocrinol* **42**(5): 371-379.
- 48.** Chng SC, Ho L, Tian J, Reversade B (2013). ELABELA: a hormone essential for heart development signals via the apelin receptor. *Dev Cell* **27**(6): 672-680.
- 49.** Chopra A, Yeagle PL, Alderfer JA, Albert AD (2000). Solution structure of the sixth transmembrane helix of the G-protein-coupled receptor, rhodopsin. *Biochim Biophys Acta* **1463**(1): 1-5.
- 50.** Chrencik JE, Roth CB, Terakado M, Kurata H, Omi R, Kihara Y, *et al.* (2015). Crystal Structure of Antagonist Bound Human Lysophosphatidic Acid Receptor 1. *Cell* **161**(7): 1633-1643.
- 51.** Clark L, Zahm JA, Ali R, Kukula M, Bian L, Patrie SM, *et al.* (2015). Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*. *J Biomol NMR* **62**(3): 239-245.
- 52.** Cohen LS, Fracchiolla KE, Becker J, Naider F (2014). GPCR Structural Characterization: Using Fragments as Building Blocks to Determine a Complete Structure. *Biopolymers*.
- 53.** Conn PM, Ulloa-Aguirre A, Ito J, Janovick JA (2007). G protein-coupled receptor trafficking in health and disease: lessons learned to prepare for therapeutic mutant rescue in vivo. *Pharmacol Rev* **59**(3): 225-250.
- 54.** Costa SJ, Almeida A, Castro A, Domingues L, Besir H (2013). The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Appl Microbiol Biotechnol* **97**(15): 6779-6791.
- 55.** Cregg JM, Cereghino JL, Shi J, Higgins DR (2000). Recombinant protein expression in *Pichia pastoris*. *Mol Biotechnol* **16**(1): 23-52.
- 56.** Crimmins DL, Mische SM, Denslow ND (2005). Chemical cleavage of proteins in solution. *Curr Protoc Protein Sci* **Chapter 11**: Unit 11 14.

- 57.** Cross TA, Sharma M, Yi M, Zhou HX (2011). Influence of solubilizing environments on membrane protein structures. *Trends Biochem Sci* **36**(2): 117-125.
- 58.** Cybulski LE, de Mendoza D (2011). Bilayer hydrophobic thickness and integral membrane protein function. *Curr Protein Pept Sci* **12**(8): 760-766.
- 59.** Cymer F, von Heijne G, White SH (2015). Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* **427**(5): 999-1022.
- 60.** Das N, Murray DT, Cross TA (2013). Lipid bilayer preparations of membrane proteins for oriented and magic-angle spinning solid-state NMR samples. *Nat Protoc* **8**(11): 2256-2270.
- 61.** De Mota N, Reaux-Le Goazigo A, El Messari S, Chartrel N, Roesch D, Dujardin C, *et al.* (2004). Apelin, a potent diuretic neuropeptide counteracting vasopressin actions through inhibition of vasopressin neuron activity and vasopressin release. *Proc Natl Acad Sci U S A* **101**(28): 10464-10469.
- 62.** Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**(3): 277-293.
- 63.** Delos Santos NM, Gardner LA, White SW, Bahouth SW (2006). Characterization of the residues in helix 8 of the human beta1-adrenergic receptor that are involved in coupling the receptor to G proteins. *J Biol Chem* **281**(18): 12896-12907.
- 64.** Denton H, Smith M, Husi H, Uhrin D, Barlow PN, Batt CA, *et al.* (1998). Isotopically labeled bovine beta-lactoglobulin for NMR studies expressed in *Pichia pastoris*. *Protein Expr Purif* **14**(1): 97-103.
- 65.** Dixon RA, Kobilka BK, Strader DJ, Benovic JL, Dohlman HG, Frielle T, *et al.* (1986). Cloning of the gene and cDNA for mammalian beta-adrenergic receptor and homology with rhodopsin. *Nature* **321**(6065): 75-79.
- 66.** Dong C, Filipeanu CM, Duvernay MT, Wu G (2007). Regulation of G protein-coupled receptor export trafficking. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1768**(4): 853-870.
- 67.** Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, *et al.* (2014). Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* **511**(7511): 557-562.

- 68.** Dray C, Debard C, Jager J, Disse E, Daviaud D, Martin P, *et al.* (2010). Apelin and APJ regulation in adipose tissue and skeletal muscle of type 2 diabetic mice and humans. *Am J Physiol Endocrinol Metab* **298**(6): E1161-1169.
- 69.** Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, Iwata S (2008). GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat Protoc* **3**(5): 784-798.
- 70.** Dror RO, Arlow DH, Borhani DW, Jensen MO, Piana S, Shaw DE (2009). Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc Natl Acad Sci U S A* **106**(12): 4689-4694.
- 71.** Dukkipati A, Park HH, Waghray D, Fischer S, Garcia KC (2008). BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expr Purif* **62**(2): 160-170.
- 72.** Duvernay MT, Dong C, Zhang X, Zhou F, Nichols CD, Wu G (2009). Anterograde trafficking of G protein-coupled receptors: function of the C-terminal F(X)6LL motif in export from the endoplasmic reticulum. *Mol Pharmacol* **75**(4): 751-761.
- 73.** Duvernay MT, Zhou F, Wu G (2004). A conserved motif for the transport of G protein-coupled receptors from the endoplasmic reticulum to the cell surface. *J Biol Chem* **279**(29): 30741-30750.
- 74.** Elleuche S, Poggeler S (2010). Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl Microbiol Biotechnol* **87**(2): 479-489.
- 75.** Erijman A, Dantes A, Bernheim R, Shifman JM, Peleg Y (2011). Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. *J Struct Biol* **175**(2): 171-177.
- 76.** Estephan R, Englander J, Arshava B, Samples KL, Becker JM, Naidler F (2005). Biosynthesis and NMR analysis of a 73-residue domain of a *Saccharomyces cerevisiae* G protein-coupled receptor. *Biochemistry* **44**(35): 11795-11810.
- 77.** Ezure T, Suzuki T, Ando E (2014). A cell-free protein synthesis system from insect cells. *Methods Mol Biol* **1118**: 285-296.
- 78.** Fagerberg L, Jonasson K, von Heijne G, Uhlen M, Berglund L (2010). Prediction of the human membrane proteome. *Proteomics* **10**(6): 1141-1149.

- 79.** Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, *et al.* (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* **33**(19): 5984-6003.
- 80.** Fernandez C, Wider G (2003). TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* **13**(5): 570-580.
- 81.** Folino A, Montarolo PG, Samaja M, Rastaldo R (2015). Effects of apelin on the cardiovascular system. *Heart Fail Rev* **20**(4): 505-518.
- 82.** Fracchiolla KE, Cohen LS, Arshava B, Poms M, Zerbe O, Becker JM, *et al.* (2015). Structural characterization of triple transmembrane domain containing fragments of a yeast G protein-coupled receptor in an organic : aqueous environment by solution-state NMR spectroscopy. *Journal of Peptide Science* **21**(3): 212-222.
- 83.** Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralagon groups, and fingerprints. *Mol Pharmacol* **63**(6): 1256-1272.
- 84.** Gaborik Z, Mihalik B, Jayadev S, Jagadeesh G, Catt KJ, Hunyady L (1998). Requirement of membrane-proximal amino acids in the carboxyl-terminal tail for expression of the rat AT1a angiotensin receptor. *FEBS Lett* **428**(3): 147-151.
- 85.** Gainetdinov RR, Premont RT, Bohn LM, Lefkowitz RJ, Caron MG (2004). Desensitization of G protein-coupled receptors and neuronal functions. *Annu Rev Neurosci* **27**: 107-144.
- 86.** Garboczi DN, Utz U, Ghosh P, Seth A, Kim J, VanTienhoven EA, *et al.* (1996). Assembly, specific binding, and crystallization of a human TCR-alpha-beta with an antigenic Tax peptide from human T lymphotropic virus type 1 and the class I MHC molecule HLA-A2. *J Immunol* **157**(12): 5403-5410.
- 87.** Gardner KH, Kay LE (1998). The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* **27**: 357-406.
- 88.** Gautier A, Kirkpatrick JP, Nietlispach D (2008). Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary structure, and dynamics. *Angew Chem Int Ed Engl* **47**(38): 7297-7300.

- 89.** Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D (2010a). Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol* **17**(6): 768-774.
- 90.** Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D (2010b). Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol* **17**(6): 768-774.
- 91.** Geisse S, Fux C (2009). Recombinant protein production by transient gene transfer into Mammalian cells. *Methods Enzymol* **463**: 223-238.
- 92.** Gentry PR, Sexton PM, Christopoulos A (2015). Novel Allosteric Modulators of G Protein-coupled Receptors. *J Biol Chem* **290**(32): 19478-19488.
- 93.** Girvin ME, Rastogi VK, Abildgaard F, Markley JL, Fillingame RH (1998). Solution structure of the transmembrane H⁺-transporting subunit c of the F1F0 ATP synthase. *Biochemistry* **37**(25): 8817-8824.
- 94.** Gluck JM, Wittlich M, Feuerstein S, Hoffmann S, Willbold D, Koenig BW (2009). Integral membrane proteins in nanodiscs can be studied by solution NMR spectroscopy. *J Am Chem Soc* **131**(34): 12060-12061.
- 95.** Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, *et al.* (1996). Life with 6000 genes. *Science* **274**(5287): 546, 563-547.
- 96.** Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E (2002). Inteins: structure, function, and evolution. *Annu Rev Microbiol* **56**: 263-287.
- 97.** Golovanov AP, Blankley RT, Avis JM, Bermel W (2007). Isotopically discriminated NMR spectroscopy: a tool for investigating complex protein interactions in vitro. *J Am Chem Soc* **129**(20): 6528-6535.
- 98.** Goodman M, Naider F, Rupp R (1971). Conformations of alanine oligopeptides in solution. *Bioorganic Chemistry* **1**(3): 310-328.
- 99.** Gordon E, Horsefield R, Swarts HG, de Pont JJ, Neutze R, Snijder A (2008). Effective high-throughput overproduction of membrane proteins in Escherichia coli. *Protein Expr Purif* **62**(1): 1-8.
- 100.** Gossert AD, Jahnke W (2012). Isotope labeling in insect cells. *Adv Exp Med Biol* **992**: 179-196.

- 101.** Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999). A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J Biomol NMR* **13**(4): 369-374.
- 102.** Goto NK, Kay LE (2000). New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* **10**(5): 585-592.
- 103.** Gottstein D, Reckel S, Dotsch V, Guntert P (2012). Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR. *Structure* **20**(6): 1019-1027.
- 104.** Greenfield N, Fasman GD (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**(10): 4108-4116.
- 105.** Greenfield NJ (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**(6): 2876-2890.
- 106.** Grisshammer R (2006). Understanding recombinant expression of membrane proteins. *Curr Opin Biotechnol* **17**(4): 337-340.
- 107.** Guntert P (2004). Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**: 353-378.
- 108.** Gurevich VV, Gurevich EV (2006). The structural basis of arrestin-mediated regulation of G-protein-coupled receptors. *Pharmacol Ther* **110**(3): 465-502.
- 109.** Habata Y, Fujii R, Hosoya M, Fukusumi S, Kawamata Y, Hinuma S, *et al.* (1999). Apelin, the natural ligand of the orphan receptor APJ, is abundantly secreted in the colostrum. *Biochim Biophys Acta* **1452**(1): 25-35.
- 110.** Haberstock S, Roos C, Hoevels Y, Dotsch V, Schnapp G, Pautsch A, *et al.* (2012). A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. *Protein Expr Purif* **82**(2): 308-316.
- 111.** Hagn F, Etzkorn M, Raschle T, Wagner G (2013). Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *J Am Chem Soc* **135**(5): 1919-1925.

- 112.** Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, *et al.* (2003). Production of complex human glycoproteins in yeast. *Science* **301**(5637): 1244-1246.
- 113.** Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, *et al.* (2006). Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313**(5792): 1441-1443.
- 114.** Hammad MM, Kuang YQ, Morse A, Dupre DJ (2012). Rab1 interacts directly with the beta2-adrenergic receptor to regulate receptor anterograde trafficking. *Biol Chem* **393**(6): 541-546.
- 115.** Hanyaloglu AC, von Zastrow M (2008). Regulation of GPCRs by endocytic membrane trafficking and its potential implications. *Annu Rev Pharmacol Toxicol* **48**: 537-568.
- 116.** Harbers M (2014). Wheat germ systems for cell-free protein expression. *FEBS Lett* **588**(17): 2762-2773.
- 117.** Hassaine G, Wagner R, Kempf J, Cherouati N, Hassaine N, Prual C, *et al.* (2006). Semliki Forest virus vectors for overexpression of 101 G protein-coupled receptors in mammalian host cells. *Protein Expr Purif* **45**(2): 343-351.
- 118.** Hattab G, Warschawski DE, Moncoq K, Miroux B (2015). Escherichia coli as host for membrane protein structure determination: a global analysis. *Sci Rep* **5**: 12097.
- 119.** He Y, Wang K, Yan N (2014). The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein Cell* **5**(9): 658-672.
- 120.** Henderson R, Unwin PN (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**(5521): 28-32.
- 121.** Heymann JA, Subramaniam S (1997). Expression, stability, and membrane integration of truncation mutants of bovine rhodopsin. *Proc Natl Acad Sci U S A* **94**(10): 4966-4971.
- 122.** Hitchman RB, Possee RD, King LA (2009). Baculovirus expression systems for recombinant protein production in insect cells. *Recent Pat Biotechnol* **3**(1): 46-54.

- 123.** Hochuli E, Bannwarth W, Dobeli H, Gentz R, Stuber D (1988). Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Nat Biotech* **6**(11): 1321-1325.
- 124.** Hollenberg CP, Gellissen G (1997). Production of recombinant proteins by methylotrophic yeasts. *Curr Opin Biotechnol* **8**(5): 554-560.
- 125.** Holzwarth G, Doty P (1965). THE ULTRAVIOLET CIRCULAR DICHROISM OF POLYPEPTIDES. *J Am Chem Soc* **87**: 218-228.
- 126.** Hosoya M, Kawamata Y, Fukusumi S, Fujii R, Habata Y, Hinuma S, *et al.* (2000). Molecular and functional characteristics of APJ. Tissue distribution of mRNA and interaction with the endogenous ligand apelin. *J Biol Chem* **275**(28): 21061-21067.
- 127.** Hsu MF, Yu TF, Chou CC, Fu HY, Yang CS, Wang AH (2013). Using Haloarcula marismortui bacteriorhodopsin as a fusion tag for enhancing and visible expression of integral membrane proteins in Escherichia coli. *PLoS One* **8**(2): e56363.
- 128.** Hu J, Qin H, Gao FP, Cross TA (2011). A systematic assessment of mature MBP in membrane protein production: overexpression, membrane targeting and purification. *Protein Expr Purif* **80**(1): 34-40.
- 129.** Huynh J, Thomas WG, Aguilar MI, Pattenden LK (2009). Role of helix 8 in G protein-coupled receptors based on structure-function studies on the type 1 angiotensin receptor. *Mol Cell Endocrinol* **302**(2): 118-127.
- 130.** Hwang PM, Kay LE (2005). Solution structure and dynamics of integral membrane proteins by NMR: a case study involving the enzyme PagP. *Methods Enzymol* **394**: 335-350.
- 131.** Hwang PM, Pan JS, Sykes BD (2014). Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in Escherichia coli. *FEBS Lett* **588**(2): 247-252.
- 132.** Hwang TL, Shaka AJ (1995). Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* **112**(2): 275-279.
- 133.** Iost I, Guillerez J, Dreyfus M (1992). Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J Bacteriol* **174**(2): 619-622.

- 134.** Ito K, Sugawara T, Shiroishi M, Tokuda N, Kurokawa A, Misaka T, *et al.* (2008). Advanced method for high-throughput expression of mutated eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* **371**(4): 841-845.
- 135.** Jaakola VP, Prilusky J, Sussman JL, Goldman A (2005). G protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Eng Des Sel* **18**(2): 103-110.
- 136.** Jarymowycz VA, Stone MJ (2006). Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev* **106**(5): 1624-1671.
- 137.** Jewett MC, Calhoun KA, Voloshin A, Wu JJ, Swartz JR (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol* **4**: 220.
- 138.** Johnson WC (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **35**(3): 307-312.
- 139.** Joubert O, Nehme R, Bidet M, Mus-Veteau I (2010). Heterologous expression of human membrane receptors in the yeast *Saccharomyces cerevisiae*. *Methods Mol Biol* **601**: 87-103.
- 140.** Judge PJ, Taylor GF, Dannatt HR, Watts A (2015). Solid-state nuclear magnetic resonance spectroscopy for membrane protein structure determination. *Methods Mol Biol* **1261**: 331-347.
- 141.** Junge F, Haberstock S, Roos C, Stefer S, Proverbio D, Dotsch V, *et al.* (2011). Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins. *N Biotechnol* **28**(3): 262-271.
- 142.** Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guntert P (2006). Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**(7080): 52-57.
- 143.** Kang DS, Tian X, Benovic JL (2014). Role of beta-arrestins and arrestin domain-containing proteins in G protein-coupled receptor trafficking. *Curr Opin Cell Biol* **27**: 63-71.
- 144.** Kang Y, Zhou XE, Gao X, He Y, Liu W, Ishchenko A, *et al.* (2015). Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**(7562): 561-567.

- 145.** Katragadda M, Alderfer JL, Yeagle PL (2001a). Assembly of a Polytopic Membrane Protein Structure from the Solution Structures of Overlapping Peptide Fragments of Bacteriorhodopsin. *Biophys J* **81**(2): 1029-1036.
- 146.** Katragadda M, Chopra A, Bennett M, Alderfer JL, Yeagle PL, Albert AD (2001b). Structures of the transmembrane helices of the G-protein coupled receptor, rhodopsin. *J Pept Res* **58**(1): 79-89.
- 147.** Katritch V, Cherezov V, Stevens RC (2013). Structure-function of the G protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol* **53**: 531-556.
- 148.** Kay L, Keifer P, Saarinen T (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* **114**(26): 10663-10665.
- 149.** Kefala G, Kwiatkowski W, Esquivies L, Maslennikov I, Choe S (2007). Application of Mystic to improving the expression and membrane integration of histidine kinase receptors from Escherichia coli. *J Struct Funct Genomics* **8**(4): 167-172.
- 150.** Kim HJ, Howell SC, Van Horn WD, Jeon YH, Sanders CR (2009). Recent Advances in the Application of Solution NMR Spectroscopy to Multi-Span Integral Membrane Proteins. *Prog Nucl Magn Reson Spectrosc* **55**(4): 335-360.
- 151.** Kirchberg K, Kim TY, Moller M, Skegro D, Dasara Raju G, Granzin J, *et al.* (2011). Conformational dynamics of helix 8 in the GPCR rhodopsin controls arrestin activation in the desensitization process. *Proc Natl Acad Sci U S A* **108**(46): 18690-18695.
- 152.** Klammt C, Schwarz D, Fendler K, Haase W, Dotsch V, Bernhard F (2005). Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *Febs j* **272**(23): 6024-6038.
- 153.** Klammt C, Schwarz D, Lohr F, Schneider B, Dotsch V, Bernhard F (2006). Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *Febs j* **273**(18): 4141-4153.
- 154.** Kleckner IR, Foster MP (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**(8): 942-968.
- 155.** Kohno T, Kusunoki H, Sato K, Wakamatsu K (1998). A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin

fusion system: an application to the production of mastoparan-X uniformly enriched with ¹⁵N and ¹⁵N/¹³C. *J Biomol NMR* **12**(1): 109-121.

156. Koppke KD, Bean JW, Bhandary KK, Briand J, D'Ambrosio CA, Peishoff CE (1993). Conformational mobility in cyclic oligopeptides. *Biopolymers* **33**(7): 1093-1099.

157. Kozma D, Simon I, Tusnady GE (2013). PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res* **41**(Database issue): D524-529.

158. Krishna AG, Menon ST, Terry TJ, Sakmar TP (2002). Evidence that helix 8 of rhodopsin acts as a membrane-dependent conformational switch. *Biochemistry* **41**(26): 8298-8309.

159. Krueger-Koplin R, Sorgen P, Krueger-Koplin S, Rivera-Torres I, Cahill S, Hicks D, *et al.* (2004). An evaluation of detergents for NMR structural studies of membrane proteins. *J Biomol NMR* **28**(1): 43-57.

160. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**(5924): 255-258.

161. Kufareva I, Katritch V, participants of GD, Stevens RC, Abagyan R (2014). Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* **22**(8): 1120-1139.

162. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R (2011). Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **19**(8): 1108-1126.

163. Kumar A, Ward P, Katre UV, Mohanty S (2012). A novel and simple method of production and biophysical characterization of a mini-membrane protein, Ost4p: a subunit of yeast oligosaccharyl transferase. *Biopolymers* **97**(7): 499-507.

164. Kunduzova O, Alet N, Delesque-Touchard N, Millet L, Castan-Laurell I, Muller C, *et al.* (2008). Apelin/APJ signaling system: a potential link between adipose tissue and endothelial angiogenic processes. *Faseb j* **22**(12): 4146-4153.

165. Kuruma Y, Ueda T (2015). The PURE system for the cell-free synthesis of membrane proteins. *Nat Protoc* **10**(9): 1328-1344.

166. Langelaan DN, Reddy T, Banks AW, Delleire G, Dupre DJ, Rainey JK (2013). Structural features of the apelin receptor N-terminal tail and first transmembrane segment

- implicated in ligand binding and receptor trafficking. *Biochim Biophys Acta* **1828**(6): 1471-1483.
- 167.** Leviatan S, Sawada K, Moriyama Y, Nelson N (2010). Combinatorial method for overexpression of membrane proteins in Escherichia coli. *J Biol Chem* **285**(31): 23548-23556.
- 168.** Levitt MH (2008). Spin dynamics basics of nuclear magnetic resonance, 2nd edn, pp 1 online resource (xxv, 714 p., 717 p. of plates). Chichester, England ; Hoboken, NJ: John Wiley & Sons,.
- 169.** Li M, Hays FA, Roe-Zurz Z, Vuong L, Kelly L, Ho CM, *et al.* (2009). Selecting optimum eukaryotic integral membrane proteins for structure determination by rapid expression and solubilization screening. *J Mol Biol* **385**(3): 820-830.
- 170.** Linke D (2009). Detergents: an overview. *Methods Enzymol* **463**: 603-617.
- 171.** Lipfert J, Columbus L, Chu VB, Lesley SA, Doniach S (2007). Size and shape of detergent micelles determined by small-angle X-ray scattering. *J Phys Chem B* **111**(43): 12427-12438.
- 172.** Lipsitz RS, Tjandra N (2004). Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct* **33**: 387-413.
- 173.** Liu R, Nahon D, le Roy B, Lenselink EB, AP IJ (2015a). Scanning mutagenesis in a yeast system delineates the role of the NPxxY(x)(5,6)F motif and helix 8 of the adenosine A(2B) receptor in G protein coupling. *Biochem Pharmacol* **95**(4): 290-300.
- 174.** Liu X, Huang A, Luo D, Liu H, Han H, Xu Y, *et al.* (2015b). Use of adenylate kinase as a solubility tag for high level expression of T4 DNA ligase in Escherichia coli. *Protein Expr Purif* **109**: 79-84.
- 175.** Lounnas V, Ritschel T, Kelder J, McGuire R, Bywater RP, Foloppe N (2013). Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* **5**: e201302011.
- 176.** Lundbaek JA (2006). Regulation of membrane protein function by lipid bilayer elasticity-a single molecule technology to measure the bilayer properties experienced by an embedded protein. *J Phys Condens Matter* **18**(28): S1305-1344.

- 177.** Lundbaek JA, Collingwood SA, Ingolfsson HI, Kapoor R, Andersen OS (2010). Lipid bilayer regulation of membrane protein function: gramicidin channels as molecular force probes. *J R Soc Interface* **7**(44): 373-395.
- 178.** Lundstrom K (2009). An overview on GPCRs and drug discovery: structure-based drug design and structural biology on GPCRs. *Methods Mol Biol* **552**: 51-66.
- 179.** Luttrell LM, Kenakin TP (2011). Refining efficacy: allostherism and bias in G protein-coupled receptor signaling. *Methods Mol Biol* **756**: 3-35.
- 180.** Luttrell LM, Maudsley S, Bohn LM (2015). Fulfilling the Promise of "Biased" G Protein-Coupled Receptor Agonism. *Mol Pharmacol* **88**(3): 579-588.
- 181.** Ma D, Liu Z, Li L, Tang P, Xu Y (2005). Structure and dynamics of the second and third transmembrane domains of human glycine receptor. *Biochemistry* **44**(24): 8790-8800.
- 182.** Majtan T, Kraus JP (2012). Folding and activity of mutant cystathionine beta-synthase depends on the position and nature of the purification tag: characterization of the R266K CBS mutant. *Protein Expr Purif* **82**(2): 317-324.
- 183.** Makrides SC (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev* **60**(3): 512-538.
- 184.** Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci* **15**(1): 182-189.
- 185.** Marley J, Lu M, Bracken C (2001). A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**(1): 71-75.
- 186.** Marsh D (2007). Lateral pressure profile, spontaneous curvature frustration, and the incorporation and conformation of proteins in membranes. *Biophys J* **93**(11): 3884-3899.
- 187.** Matrai J, Chuah MK, VandenDriessche T (2010). Recent advances in lentiviral vector development and applications. *Mol Ther* **18**(3): 477-490.
- 188.** McPhie P (2008). Concentration-independent estimation of protein secondary structure by circular dichroism: a comparison of methods. *Anal Biochem* **375**(2): 379-381.

- 189.** Mehler M, Eckert CE, Busche A, Kulhei J, Michaelis J, Becker-Baldus J, *et al.* (2015). Assembling a correctly folded and functional heptahelical membrane protein by protein trans-splicing. *J Biol Chem.*
- 190.** Miao Y, Nichols SE, McCammon JA (2014). Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. *Phys Chem Chem Phys* **16**(14): 6398-6406.
- 191.** Micsonai A, Wien F, Kernya L, Lee YH, Goto Y, Refregiers M, *et al.* (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci U S A* **112**(24): E3095-3103.
- 192.** Mijakovic I, Petranovic D, Macek B, Cepo T, Mann M, Davies J, *et al.* (2006). Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. *Nucleic Acids Res* **34**(5): 1588-1596.
- 193.** Mikami S, Masutani M, Sonenberg N, Yokoyama S, Imataka H (2006). An efficient mammalian cell-free translation system supplemented with translation factors. *Protein Expr Purif* **46**(2): 348-357.
- 194.** Milic D, Veprintsev DB (2015). Large-scale production and protein engineering of G protein-coupled receptors for structural studies. *Front Pharmacol* **6**: 66.
- 195.** Miroux B, Walker JE (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**(3): 289-298.
- 196.** Moller D, Gmeiner P (2015). Arrestin-Bound Rhodopsin: A Molecular Structure and its Impact on the Development of Biased GPCR Ligands. *Angew Chem Int Ed Engl.*
- 197.** Mottamal M, Shen S, Guembe C, Krilov G (2007). Solvation of transmembrane proteins by isotropic membrane mimetics: a molecular dynamics study. *J Phys Chem B* **111**(38): 11285-11296.
- 198.** Murray DT, Das N, Cross TA (2013). Solid state NMR strategy for characterizing native membrane protein structures. *Acc Chem Res* **46**(9): 2172-2181.
- 199.** Nathans J, Hogness DS (1984). Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc Natl Acad Sci U S A* **81**(15): 4851-4855.

- 200.** Nelson JW, Kallenbach NR (1989). Persistence of the alpha-helix stop signal in the S-peptide in trifluoroethanol solutions. *Biochemistry* **28**(12): 5256-5261.
- 201.** Neophytou I, Harvey R, Lawrence J, Marsh P, Panaretou B, Barlow D (2007). Eukaryotic integral membrane protein expression utilizing the Escherichia coli glycerol-conducting channel protein (GlpF). *Appl Microbiol Biotechnol* **77**(2): 375-381.
- 202.** Neumoin A, Cohen LS, Arshava B, Tantry S, Becker JM, Zerbe O, *et al.* (2009). Structure of a double transmembrane fragment of a G-protein-coupled receptor in micelles. *Biophys J* **96**(8): 3187-3196.
- 203.** Newstead S, Kim H, von Heijne G, Iwata S, Drew D (2007). High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **104**(35): 13936-13941.
- 204.** Nicolson GL (2014). The Fluid—Mosaic Model of Membrane Structure: Still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838**(6): 1451-1466.
- 205.** Nietlispach D, Gautier A (2011). Solution NMR studies of polytopic alpha-helical membrane proteins. *Curr Opin Struct Biol* **21**(4): 497-508.
- 206.** Nirenberg MW, Matthaei JH (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A* **47**: 1588-1602.
- 207.** Nordstrom KJ, Sallman Almen M, Edstam MM, Fredriksson R, Schioth HB (2011). Independent HHsearch, Needleman--Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. *Mol Biol Evol* **28**(9): 2471-2480.
- 208.** Novakova V, Sandhu GS, Dragomir-Daescu D, Klabusay M (2015). Apelinergic system in endothelial cells and its role in angiogenesis in myocardial ischemia. *Vascul Pharmacol*.
- 209.** Nyblom M, Oberg F, Lindkvist-Petersson K, Hallgren K, Findlay H, Wikstrom J, *et al.* (2007). Exceptional overproduction of a functional human membrane protein. *Protein Expr Purif* **56**(1): 110-120.

- 210.** O'Dowd BF, Heiber M, Chan A, Heng HH, Tsui LC, Kennedy JL, *et al.* (1993). A human gene that shows identity with the gene encoding the angiotensin receptor is located on chromosome 11. *Gene* **136**(1-2): 355-360.
- 211.** Oksche A, Dehe M, Schulein R, Wiesner B, Rosenthal W (1998). Folding and cell surface expression of the vasopressin V2 receptor: requirement of the intracellular C-terminus. *FEBS Lett* **424**(1-2): 57-62.
- 212.** Opella SJ (2015). Solid-state NMR and membrane proteins. *J Magn Reson* **253**: 129-137.
- 213.** Opitz C, Isogai S, Grzesiek S (2015). An economic approach to efficient isotope labeling in insect cells using homemade ¹⁵N-, ¹³C- and ²H-labeled yeast extracts. *J Biomol NMR* **62**(3): 373-385.
- 214.** Orban E, Proverbio D, Haberstock S, Dotsch V, Bernhard F (2015). Cell-free expression of G-protein-coupled receptors. *Methods Mol Biol* **1261**: 171-195.
- 215.** Overington JP, Al-Lazikani B, Hopkins AL (2006). How many drug targets are there? *Nat Rev Drug Discov* **5**(12): 993-996.
- 216.** Ozawa K, Wu PS, Dixon NE, Otting G (2006). N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. *Febs j* **273**(18): 4154-4159.
- 217.** Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**(11): 2411-2423.
- 218.** Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, *et al.* (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**(5480): 739-745.
- 219.** Palmer Iii AG, Cavanagh J, Wright PE, Rance M (1991). Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J Magn Reson (1969)* **93**(1): 151-170.
- 220.** Pankevych H, Korkhov V, Freissmuth M, Nanoff C (2003). Truncation of the A1 adenosine receptor reveals distinct roles of the membrane-proximal carboxyl terminus in receptor folding and G protein coupling. *J Biol Chem* **278**(32): 30283-30293.

- 221.** Park SH, Casagrande F, Chu M, Maier K, Kiefer H, Opella SJ (2012a). Optimization of purification and refolding of the human chemokine receptor CXCR1 improves the stability of proteoliposomes for structure determination. *Biochim Biophys Acta* **1818**(3): 584-591.
- 222.** Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, *et al.* (2012b). Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **491**(7426): 779-783.
- 223.** Parker MJ, Aulton-Jones M, Hounslow AM, Craven CJ (2004). A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. *J Am Chem Soc* **126**(16): 5020-5021.
- 224.** Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, *et al.* (2014). Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science* **343**(6172): 1248636.
- 225.** Paulus H (2000). Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* **69**: 447-496.
- 226.** Penin F, Brass V, Appel N, Ramboarina S, Montserret R, Ficheux D, *et al.* (2004). Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J Biol Chem* **279**(39): 40835-40843.
- 227.** Perler FB (2002). InBase: the Intein Database. *Nucleic Acids Res* **30**(1): 383-384.
- 228.** Pervushin K, Riek R, Wider G, Wuthrich K (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**(23): 12366-12371.
- 229.** Phillips R, Ursell T, Wiggins P, Sens P (2009). Emerging roles for lipids in shaping membrane-protein function. *Nature* **459**(7245): 379-385.
- 230.** Pitkin SL, Maguire JJ, Bonner TI, Davenport AP (2010). International Union of Basic and Clinical Pharmacology. LXXIV. Apelin receptor nomenclature, distribution, pharmacology, and function. *Pharmacol Rev* **62**(3): 331-342.
- 231.** Potetinova Z, Tantry S, Cohen LS, Carocchia KE, Arshava B, Becker JM, *et al.* (2012). Large multiple transmembrane domain fragments of a G protein-coupled receptor: biosynthesis, purification, and biophysical studies. *Biopolymers* **98**(5): 485-500.

- 232.** Prosser RS, Evanics F, Kitevski JL, Al-Abdul-Wahid MS (2006). Current applications of bicelles in NMR studies of membrane-associated amphiphiles and proteins. *Biochemistry* **45**(28): 8453-8465.
- 233.** Pryor KD, Leiting B (1997). High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr Purif* **10**(3): 309-319.
- 234.** Pucadyil TJ, Chattopadhyay A (2006). Role of cholesterol in the function and organization of G-protein coupled receptors. *Prog Lipid Res* **45**(4): 295-333.
- 235.** Qanbar R, Bouvier M (2003). Role of palmitoylation/depalmitoylation reactions in G-protein-coupled receptor function. *Pharmacol Ther* **97**(1): 1-33.
- 236.** Qing G, Xia B, Inouye M (2003). Enhancement of translation initiation by A/T-rich sequences downstream of the initiation codon in *Escherichia coli*. *J Mol Microbiol Biotechnol* **6**(3-4): 133-144.
- 237.** Qureshi T, Goto NK (2012). Contemporary methods in structure determination of membrane proteins by solution NMR. *Top Curr Chem* **326**: 123-185.
- 238.** Ramon A, Marin M (2011). Advances in the production of membrane proteins in *Pichia pastoris*. *Biotechnol J* **6**(6): 700-706.
- 239.** Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, *et al.* (2011). Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **477**(7366): 549-555.
- 240.** Raussens V, Ruysschaert JM, Goormaghtigh E (2003). Protein concentration is not an absolute prerequisite for the determination of secondary structure from circular dichroism spectra: a new scaling method. *Anal Biochem* **319**(1): 114-121.
- 241.** Reckel S, Gottstein D, Stehle J, Lohr F, Verhoeven MK, Takeda M, *et al.* (2011). Solution NMR structure of proteorhodopsin. *Angew Chem Int Ed Engl* **50**(50): 11942-11946.
- 242.** Ribas C, Penela P, Murga C, Salcedo A, Garcia-Hoz C, Jurado-Pueyo M, *et al.* (2007). The G protein-coupled receptor kinase (GRK) interactome: role of GRKs in GPCR regulation and signaling. *Biochim Biophys Acta* **1768**(4): 913-922.

- 243.** Robert J, Clauser E, Petit PX, Ventura MA (2005). A novel C-terminal motif is necessary for the export of the vasopressin V1b/V3 receptor to the plasma membrane. *J Biol Chem* **280**(3): 2300-2308.
- 244.** Roberts EM, Newson MJ, Pope GR, Landgraf R, Lolait SJ, O'Carroll AM (2009). Abnormal fluid homeostasis in apelin receptor knockout mice. *J Endocrinol* **202**(3): 453-462.
- 245.** Roccatano D, Colombo G, Fioroni M, Mark AE (2002). Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study. *Proc Natl Acad Sci U S A* **99**(19): 12179-12184.
- 246.** Rodriguez MC, Xie YB, Wang H, Collison K, Segaloff DL (1992). Effects of truncations of the cytoplasmic tail of the luteinizing hormone/chorionic gonadotropin receptor on receptor-mediated hormone internalization. *Mol Endocrinol* **6**(3): 327-336.
- 247.** Rosano GL, Ceccarelli EA (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**: 172.
- 248.** Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996). Selective Methyl Group Protonation of Perdeuterated Proteins. *J Mol Biol* **263**(5): 627-636.
- 249.** Rovati GE, Capra V, Neubig RR (2007). The highly conserved DRY motif of class A G protein-coupled receptors: beyond the ground state. *Mol Pharmacol* **71**(4): 959-964.
- 250.** Rovnyak D, Frueh DP, Sastry M, Sun Z-YJ, Stern AS, Hoch JC, *et al.* (2004). Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *Journal of Magnetic Resonance* **170**(1): 15-21.
- 251.** Ruiz-Gomez G, Tyndall JD, Pfeiffer B, Abbenante G, Fairlie DP (2010). Update 1 of: Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure. *Chem Rev* **110**(4): Pr1-41.
- 252.** Rule GS, Hitchens TK (2006). Fundamentals of protein NMR spectroscopy. In: *Focus on structural biology v 5*, pp 1 online resource (xxvi, 530 p.). Dordrecht: Springer,.
- 253.** Russell RB, Eggleston DS (2000). New roles for structure in biology and drug discovery. *Nat Struct Biol* **7 Suppl**: 928-930.

- 254.** Sabaty M, Grosse S, Adryanczyk G, Boiry S, Biaso F, Arnoux P, *et al.* (2013). Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis. *BMC Biochem* **14**: 28.
- 255.** Sander P, Grunewald S, Reilander H, Michel H (1994). Expression of the human D2S dopamine receptor in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*: a comparative study. *FEBS Lett* **344**(1): 41-46.
- 256.** Sanders CR, Mittendorf KF (2011). Tolerance to changes in membrane lipid composition as a selected trait of membrane proteins. *Biochemistry* **50**(37): 7858-7867.
- 257.** Sanders CR, Sonnichsen F (2006). Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No**: S24-40.
- 258.** Sastry M, Bewley CA, Kwong PD (2012). Mammalian expression of isotopically labeled proteins for NMR spectroscopy. *Adv Exp Med Biol* **992**: 197-211.
- 259.** Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletsky O, Glaser SJ, *et al.* (1994). A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR* **4**(2): 301-306.
- 260.** Schlinkmann KM, Honegger A, Tureci E, Robison KE, Lipovsek D, Pluckthun A (2012). Critical features for biosynthesis, stability, and functionality of a G protein-coupled receptor uncovered by all-versus-all mutations. *Proc Natl Acad Sci U S A* **109**(25): 9810-9815.
- 261.** Schulein R, Hermosilla R, Oksche A, Dehe M, Wiesner B, Krause G, *et al.* (1998). A dileucine sequence and an upstream glutamate residue in the intracellular carboxyl terminus of the vasopressin V2 receptor are essential for cell surface transport in COS.M6 cells. *Mol Pharmacol* **54**(3): 525-535.
- 262.** Schwaiger M, Lebendiker M, Yerushalmi H, Coles M, Groger A, Schwarz C, *et al.* (1998). NMR investigation of the multidrug transporter EmrE, an integral membrane protein. *Eur J Biochem* **254**(3): 610-619.
- 263.** Schwarz D, Dotsch V, Bernhard F (2008). Production of membrane proteins using cell-free expression systems. *Proteomics* **8**(19): 3933-3946.
- 264.** Schwarz D, Junge F, Durst F, Frolich N, Schneider B, Reckel S, *et al.* (2007a). Preparative scale expression of membrane proteins in *Escherichia coli*-based continuous exchange cell-free systems. *Nat Protoc* **2**(11): 2945-2957.

- 265.** Schwarz D, Klammt C, Koglin A, Lohr F, Schneider B, Dotsch V, *et al.* (2007b). Preparative scale cell-free expression systems: new tools for the large scale preparation of integral membrane proteins for functional and structural studies. *Methods* **41**(4): 355-369.
- 266.** Schwieters CD, Kuszewski JJ, Marius Clore G (2006). Using Xplor–NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc* **48**(1): 47-62.
- 267.** Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003). The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**(1): 65-73.
- 268.** Sensoy O, Weinstein H (2015). A mechanistic role of Helix 8 in GPCRs: Computational modeling of the dopamine D2 receptor interaction with the GIPC1–PDZ-domain. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1848**(4): 976-983.
- 269.** Sezonov G, Joseleau-Petit D, D'Ari R (2007). Escherichia coli physiology in Luria-Bertani broth. *J Bacteriol* **189**(23): 8746-8749.
- 270.** Shi X, Jarvis DL (2007). Protein N-glycosylation in the baculovirus-insect cell system. *Curr Drug Targets* **8**(10): 1116-1125.
- 271.** Shiloach J, Fass R (2005). Growing E. coli to high cell density--a historical perspective on method development. *Biotechnol Adv* **23**(5): 345-357.
- 272.** Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, *et al.* (2001). Cell-free translation reconstituted with purified components. *Nat Biotechnol* **19**(8): 751-755.
- 273.** Simons K, Sampaio JL (2011). Membrane organization and lipid rafts. *Cold Spring Harb Perspect Biol* **3**(10): a004697.
- 274.** Singh P, Sharma L, Kulothungan SR, Adkar BV, Prajapati RS, Ali PS, *et al.* (2013). Effect of signal peptide on stability and folding of Escherichia coli thioredoxin. *PLoS One* **8**(5): e63442.
- 275.** Sitarska A, Skora L, Klopp J, Roest S, Fernandez C, Shrestha B, *et al.* (2015). Affordable uniform isotope labeling with (2)H, (13)C and (15)N in insect cells. *J Biomol NMR* **62**(2): 191-197.
- 276.** Smith B (2002). Chemical Cleavage of Proteins at Cysteinyl-X Peptide Bonds. In: Walker J (ed). *The Protein Protocols Handbook*, edn: Humana Press. p[^]pp 503-506.

- 277.** Sorensen HP, Mortensen KK (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* **115**(2): 113-128.
- 278.** Spomer L, Gertzen CGW, Schmitz B, Häussinger D, Gohlke H, Keitel V (2014). A Membrane-proximal, C-terminal α -Helix Is Required for Plasma Membrane Localization and Function of the G Protein-coupled Receptor (GPCR) TGR5. *J Biol Chem* **289**(6): 3689-3702.
- 279.** Sreenath HK, Bingman CA, Buchan BW, Seder KD, Burns BT, Geetha HV, *et al.* (2005). Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr Purif* **40**(2): 256-267.
- 280.** Sreerama N, Woody RW (2004). Computation and analysis of protein circular dichroism spectra. *Methods Enzymol* **383**: 318-351.
- 281.** Srivastava A, Yano J, Hirozane Y, Kefala G, Gruswitz F, Snell G, *et al.* (2014). High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* **513**(7516): 124-127.
- 282.** Stadel R, Ahn KH, Kendall DA (2011). The cannabinoid type-1 receptor carboxyl-terminus, more than just a tail. *J Neurochem* **117**(1): 1-18.
- 283.** Stevens RC (2000). Design of high-throughput methods of protein production for structural biology. *Structure* **8**(9): R177-185.
- 284.** Stewart MQ, Esposito RD, Gowani J, Goodman JM (2001). Alcohol oxidase and dihydroxyacetone synthase, the abundant peroxisomal proteins of methylotrophic yeasts, assemble in different cellular compartments. *J Cell Sci* **114**(Pt 15): 2863-2868.
- 285.** Stols L, Millard CS, Dementieva I, Donnelly MI (2004). Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. *J Struct Funct Genomics* **5**(1-2): 95-102.
- 286.** Studier FW (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219**(1): 37-44.
- 287.** Sun W, Yang J, Liu XQ (2004). Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J Biol Chem* **279**(34): 35281-35286.

- 288.** Sutherland EW (1972). Studies on the mechanism of hormone action. *Science* **177**(4047): 401-408.
- 289.** Takeda M, Chang CK, Ikeya T, Guntert P, Chang YH, Hsu YL, *et al.* (2008). Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J Mol Biol* **380**(4): 608-622.
- 290.** Takegawa K, Tohda H, Sasaki M, Idris A, Ohashi T, Mukaiyama H, *et al.* (2009). Production of heterologous proteins using the fission-yeast (*Schizosaccharomyces pombe*) expression system. *Biotechnol Appl Biochem* **53**(Pt 4): 227-235.
- 291.** Tang BL, Wang Y, Ong YS, Hong W (2005). COPII and exit from the endoplasmic reticulum. *Biochim Biophys Acta* **1744**(3): 293-303.
- 292.** Tate CG (2001). Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett* **504**(3): 94-98.
- 293.** Tatemoto K, Hosoya M, Habata Y, Fujii R, Kakegawa T, Zou MX, *et al.* (1998). Isolation and characterization of a novel endogenous peptide ligand for the human APJ receptor. *Biochem Biophys Res Commun* **251**(2): 471-476.
- 294.** Tautermann CS (2014). GPCR structures in drug design, emerging opportunities with new structures. *Bioorg Med Chem Lett* **24**(17): 4073-4079.
- 295.** Tetsuka M, Saito Y, Imai K, Doi H, Maruyama K (2004). The basic residues in the membrane-proximal C-terminal tail of the rat melanin-concentrating hormone receptor 1 are required for receptor function. *Endocrinology* **145**(8): 3712-3723.
- 296.** Tiburu EK, Tyukhtenko S, Deshmukh L, Vinogradova O, Janero DR, Makriyannis A (2009). Structural biology of human cannabinoid receptor-2 helix 6 in membrane-mimetic environments. *Biochem Biophys Res Commun* **384**(2): 243-248.
- 297.** Tiburu EK, Tyukhtenko S, Zhou H, Janero DR, Struppe J, Makriyannis A (2011). Human cannabinoid 1 GPCR C-terminal domain interacts with bilayer phospholipids to modulate the structure of its membrane environment. *AAPS J* **13**(1): 92-98.
- 298.** Tikhonova IG, Costanzi S (2009). Unraveling the structure and function of G protein-coupled receptors through NMR spectroscopy. *Curr Pharm Des* **15**(35): 4003-4016.

- 299.** Tonelli M, Singarapu KK, Makino S, Sahu SC, Matsubara Y, Endo Y, *et al.* (2011). Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J Biomol NMR* **51**(4): 467-476.
- 300.** Tremblay ML, Xu L, Lefevre T, Sarker M, Orrell KE, Leclerc J, *et al.* (2015). Spider wrapping silk fibre architecture arising from its modular soluble protein precursor. *Sci Rep* **5**: 11502.
- 301.** Tugarinov V, Muhandiram R, Ayed A, Kay LE (2002). Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* **124**(34): 10025-10035.
- 302.** Tyler RC, Sreenath HK, Singh S, Aceti DJ, Bingman CA, Markley JL, *et al.* (2005). Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expr Purif* **40**(2): 268-278.
- 303.** Tyukhtenko S, Tiburu EK, Deshmukh L, Vinogradova O, Janero DR, Makriyannis A (2009). NMR solution structure of human cannabinoid receptor-1 helix 7/8 peptide: candidate electrostatic interactions and microdomain formation. *Biochem Biophys Res Commun* **390**(3): 441-446.
- 304.** Unal H, Karnik SS (2012). Domain coupling in GPCRs: the engine for induced conformational changes. *Trends in Pharmacological Sciences* **33**(2): 79-88.
- 305.** Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y (2010). Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J Struct Biol* **172**(1): 34-44.
- 306.** Unger T, Peleg Y (2012). Recombinant protein expression in the baculovirus-infected insect cell system. *Methods Mol Biol* **800**: 187-199.
- 307.** van den Burg HA, de Wit PJ, Vervoort J (2001). Efficient 13C/15N double labeling of the avirulence protein AVR4 in a methanol-utilizing strain (Mut+) of *Pichia pastoris*. *J Biomol NMR* **20**(3): 251-261.
- 308.** Venkatakrisnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM (2013). Molecular signatures of G-protein-coupled receptors. *Nature* **494**(7436): 185-194.
- 309.** Venkatakrisnan AJ, Flock T, Prado DE, Oates ME, Gough J, Madan Babu M (2014). Structured and disordered facets of the GPCR fold. *Curr Opin Struct Biol* **27**: 129-137.

- 310.** Vervecken W, Callewaert N, Kaigorodov V, Geysens S, Contreras R (2007). Modification of the N-glycosylation pathway to produce homogeneous, human-like glycans using GlycoSwitch plasmids. *Methods Mol Biol* **389**: 119-138.
- 311.** Vinothkumar KR, Henderson R (2010). Structures of membrane proteins. *Q Rev Biophys* **43**(1): 65-158.
- 312.** Volkmann G, Iwai H (2010). Protein trans-splicing and its use in structural biology: opportunities and limitations. *Mol Biosyst* **6**(11): 2110-2121.
- 313.** von Heijne G (2007). The membrane protein universe: what's out there and why bother? *J Intern Med* **261**(6): 543-557.
- 314.** Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, *et al.* (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**(4): 687-696.
- 315.** Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, *et al.* (2008). Tuning Escherichia coli for membrane protein overexpression. *Proc Natl Acad Sci U S A* **105**(38): 14371-14376.
- 316.** Walls D, Loughran ST (2011). Tagging recombinant proteins to enhance solubility and aid purification. *Methods Mol Biol* **681**: 151-175.
- 317.** Walser R, Zerbe O (2012). Approaches towards structures of Y receptors, examples of human G-protein coupled receptors, by solution NMR. *Chimia (Aarau)* **66**(10): 781-786.
- 318.** Wang X, Wang J, Ge B (2013). Evaluation of cell-free expression system for the production of soluble and functional human GPCR N-formyl peptide receptors. *Protein Pept Lett* **20**(11): 1272-1279.
- 319.** Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, *et al.* (2008). Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**(7203): 486-491.
- 320.** Warschawski DE, Arnold AA, Beaugrand M, Gravel A, Chartrand É, Marcotte I (2011). Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**(8): 1957-1974.

- 321.** West C, Hanyaloglu AC (2015). Minireview: Spatial Programming of G Protein-Coupled Receptor Activity: Decoding Signaling in Health and Disease. *Mol Endocrinol* **29**(8): 1095-1106.
- 322.** Wheatley M, Wootten D, Conner MT, Simms J, Kendrick R, Logan RT, *et al.* (2012). Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* **165**(6): 1688-1703.
- 323.** Whitmore L, Wallace BA (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**(5): 392-400.
- 324.** Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, *et al.* (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**(5429): 901-906.
- 325.** Wishart D, Sykes B (1994a). The ¹³C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* **4**(2): 171-180.
- 326.** Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995). ¹H, ¹³C and ¹⁵N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* **5**(1): 67-81.
- 327.** Wishart DS, Sykes BD (1994b). The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* **4**(2): 171-180.
- 328.** Wishart DS, Sykes BD, Richards FM (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.
- 329.** Wittekind M, Mueller L (1993). HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance, Series B* **101**(2): 201-205.
- 330.** Wood MJ, Komives EA (1999). Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation. *J Biomol NMR* **13**(2): 149-159.
- 331.** Wu H, Hu Z, Liu XQ (1998). Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A* **95**(16): 9226-9231.

- 332.** Xie H, Ding FX, Schreiber D, Eng G, Liu SF, Arshava B, *et al.* (2000). Synthesis and biophysical analysis of transmembrane domains of a *Saccharomyces cerevisiae* G protein-coupled receptor. *Biochemistry* **39**(50): 15462-15474.
- 333.** Xu MQ, Perler FB (1996). The mechanism of protein splicing and its modulation by mutation. *EMBO J* **15**(19): 5146-5153.
- 334.** Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE (1994a). A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ¹⁵N, ¹³C, ²H Labeled Proteins with High Sensitivity. *J Am Chem Soc* **116**(26): 11655-11666.
- 335.** Yamazaki T, Lee W, Revington M, Mattiello DL, Dahlquist FW, Arrowsmith CH, *et al.* (1994b). An HNCA Pulse Scheme for the Backbone Assignment of ¹⁵N,¹³C,²H-Labeled Proteins: Application to a 37-kDa Trp Repressor-DNA Complex. *J Am Chem Soc* **116**(14): 6464-6465.
- 336.** Yang D, Kay L (1999). Improved lineshape and sensitivity in the HNCQ-family of triple resonance experiments. *J Biomol NMR* **14**(3): 273-276.
- 337.** Yang P, Maguire JJ, Davenport AP (2015). Apelin, Elabela/Toddler, and biased agonists as novel therapeutic agents in the cardiovascular system. *Trends Pharmacol Sci* **36**(9): 560-567.
- 338.** Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M (2007). Drug-target network. *Nat Biotechnol* **25**(10): 1119-1126.
- 339.** Yin J, Mobarec JC, Kolb P, Rosenbaum DM (2015). Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. *Nature* **519**(7542): 247-250.
- 340.** Young B, Wertman J, Dupré DJ (2015). Chapter Thirteen - Regulation of GPCR Anterograde Trafficking by Molecular Chaperones and Motifs. In: Guangyu W (ed). *Prog Mol Biol Transl Sci*, edn, Vol. Volume 132: Academic Press. p[^]pp 289-305.
- 341.** Zeenko VV, Wang C, Majumder M, Komar AA, Snider MD, Merrick WC, *et al.* (2008). An efficient in vitro translation system from mammalian cells lacking the translational inhibition caused by eIF2 phosphorylation. *Rna* **14**(3): 593-602.
- 342.** Zettler J, Schütz V, Mootz HD (2009). The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* **583**(5): 909-914.

- 343.** Zhang D, Gao ZG, Zhang K, Kiselev E, Crane S, Wang J, *et al.* (2015a). Two disparate ligand-binding sites in the human P2Y1 receptor. *Nature* **520**(7547): 317-321.
- 344.** Zhang H, Unal H, Gati C, Han GW, Liu W, Zatsepin NA, *et al.* (2015b). Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. *Cell* **161**(4): 833-844.
- 345.** Zhao J, Zheng H, Xie XQ (2006). NMR characterization of recombinant transmembrane protein CB2 fragment CB2(180-233). *Protein Pept Lett* **13**(4): 335-342.
- 346.** Zhao Q, Wu BL (2012). Ice breaking in GPCR structural biology. *Acta Pharmacol Sin* **33**(3): 324-334.
- 347.** Zheng H, Zhao J, Sheng W, Xie XQ (2006). A transmembrane helix-bundle from G-protein coupled receptor CB2: biosynthesis, purification, and NMR characterization. *Biopolymers* **83**(1): 46-61.
- 348.** Zhou H-X, Cross TA (2013). Modeling the membrane environment has implications for membrane protein structure and function: Influenza A M2 protein. *Protein Sci* **22**(4): 381-394.
- 349.** Zhou N, Fan X, Mukhtar M, Fang J, Patel CA, DuBois GC, *et al.* (2003). Cell-cell fusion and internalization of the CNS-based, HIV-1 co-receptor, APJ. *Virology* **307**(1): 22-36.
- 350.** Zhou P, Wagner G (2010). Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. *J Biomol NMR* **46**(1): 23-31.
- 351.** Zou C, Naider F, Zerbe O (2008). Biosynthesis and NMR-studies of a double transmembrane domain from the Y4 receptor, a human GPCR. *J Biomol NMR* **42**(4): 257-269.
- 352.** Zuger S, Iwai H (2005). Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat Biotechnol* **23**(6): 736-740.