Structural Studies Of Apelin And Its Receptor As Well As The Characteristics And Causes Of Membrane Protein Helix Kinks

by

David Nicholas Langelaan

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

at

Dalhousie University
Halifax, Nova Scotia
March 2012

DALHOUSIE UNIVERSITY

DEPARTMENT OF BIOCHEMISTRY & MOLECULAR BIOLOGY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Structural Studies Of Apelin And Its Receptor As Well As The Characteristics And Causes Of Membrane Protein Helix Kinks" by David Nicholas Langelaan in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Dated: March 26, 2012

External Examiner: _____

Research Supervisor: _____

Examining Committee: _____

_____

Departmental Representative: _____

DALHOUSIE UNIVERSITY

DATE: March 26, 2012

AUTHOR: David Nicholas Langelaan

TITLE: Structural Studies Of Apelin And Its Receptor As Well As The Characteristics And Causes Of Membrane Protein Helix Kinks

DEPARTMENT OR SCHOOL: Department of Biochemistry & Molecular Biology

DEGREE: PhD CONVOCATION: May YEAR: 2012

_____

Signature of Author

Dedication

For my parents (*Kees and Irma*),

my wife (*Robyn*),

and the great friends that I have met while at Dalhousie.

Table Of Contents

List Of Tables ..... **x**

List Of Figures ..... **xii**

Abstract ..... **xvi**

List Of Abbreviations And Symbols Used ..... **xvii**

Acknowledgements ..... **xxi**

Chapter 1: Introduction **1**

1.1. G-Protein Coupled Receptors 1

1.1.1. Physiological Importance 1

1.1.2. GPCR Signaling 2

1.2. Physiology Of AR And Apelin 4

1.3. Structural Studies Of GPCRs 6

1.4. Circular Dichroism (CD) Spectropolarimetry 9

1.5. Using NMR Spectroscopy To Study Proteins 9

1.5.1. Basics Of NMR Spectroscopy 9

1.5.2. Signal Overlap In NMR Spectroscopy 11

1.5.3. Sequential Assignment Of Protein NMR Spectra 12

1.5.4. Using NMR Spectroscopy To Determine Protein Structure 15

1.6. Rationale And Objectives 16

Chapter 2: The Structure Of Apelin In Solution (Based On My Contribution To Manuscript (1)) **27**

2.1. The Potential For Apelin Structure 27

2.1.1. Using NMR Spectroscopy To Study Small Peptides 28

2.2. Materials And Methods 29

2.2.1. Materials 29

2.2.2. Peptide Synthesis & Purification 30

2.2.3. NMR Spectroscopy 30

2.2.4. Structure Calculation & Ensemble Analysis 31

2.3. Results And Discussion 34

2.3.1. Apelin Adopts Multiple Conformations In Solution 34

2.3.2. Generation Of Apelin-17 Structural Ensembles At 5 °C And 35 °C 35

2.3.3. Some Sections Of Apelin-17 Are Structured 37
2.3.4. The Structured Regions Of Apelin-17 Are Clustered Into Several Conformations 39
2.3.5. The Smaller Apelin Isoforms Have A Similar Structure To Apelin-17 40
2.3.6. Structure-Function Correlation Of Apelin 40
2.3.7. Insights Into Apelin-AR Interactions 42
2.4. Summary 43
Chapter 3: Interactions Of Apelin With Membrane Mimetics (Based On My Contribution To Manuscripts (2, 3)) 65
3.1. Introduction 65
3.1.1. Membrane Catalysis Of Peptide-Receptor Binding 65
3.1.2. Common Membrane Mimetics Used In Biophysical Studies 67
3.1.3. The Study Of Peptide-Membrane Interactions With NMR Spectroscopy 68
3.2. Materials And Methods 71
3.2.1. Materials 71
3.2.2. CD Spectropolarimetry 71
3.2.3. PRE Measurements 72
3.2.4. Diffusion Ordered Spectroscopy Measurements 73
3.2.5. Sequential Assignment Of Apelin-17 In SDS And DPC Micelles 74
3.2.6. Structural Calculations Of Apelin-17 Bound To SDS Micelles 75
3.3. Results And Discussion 75
3.3.1. CD Spectropolarimetry Of Apelin-12 And Apelin-17 75
3.3.2. Diffusion Coefficient Determination 76
3.3.3. NMR Assignment And Secondary Chemical Shifts 78
3.3.4. PRE Of Apelin-17 With SDS Micelles 79
3.3.5. Structure Of Apelin-17 With SDS Micelles 80
3.3.6. Implications For Membrane Catalysis Of Apelin-AR Binding 81
3.4. Summary 83
Chapter 4: Structure And Function Of The First TM Segment And N-Terminal Tail Of AR 99
4.1. Introduction 99
4.1.1. Current Structural Data Of AR 99

4.1.2. The Divide And Conquer Approach To Membrane Protein Structure 100
4.2. Materials And Methods 101
4.2.1. Materials 101
4.2.2. Cloning And Expression Of AR55 102
4.2.3. Purification Of AR55 102
4.2.4. NMR Spectroscopy Of AR55 103
4.2.5. Assignment And Structure Calculation Of AR55 104
4.2.6. Mutagenesis Of AR 105
4.2.7. Extracellular Regulated Kinase (ERK) Phosphorylation 105
4.2.8. Immunofluorescence Microscopy 106
4.3. Results And Discussion 107
4.3.1. Purification Of AR55 107
4.3.2. Structure Of AR55 In DPC Micelles 107
4.3.3. G42, G45 And N46 Are Required For Proper Folding And Function Of AR 109
4.3.4. G42, G45 And N46 Are Likely Essential For Interhelix Interactions 110
4.3.5. A Potential Role For T43 And T44? 111
4.3.6. E20 And D23 May Form A Binding Face For Apelin 112
4.3.7. Comparison Of The AR55 Structure With An AR Model 112
4.4. Summary 113
Chapter 5: The Effect Of Membrane Mimetics On AR55 Structure 132
5.1. Introduction 132
5.1.1. The Choice Of Membrane Mimetic Can Affect Protein Structure 132
5.1.2. The Use Of Organic Solvents As A Membrane Mimetic 133
5.1.3. NMR Relaxation 134
5.1.4. Model-Free Analysis Of Relaxation Data 136
5.1.5. Reduced Spectral Density Mapping Of Relaxation Data 138
5.2. Materials And Methods 138
5.2.1. Materials 138
5.2.2. CD Spectropolarimetry 139
5.2.3. NMR Spectroscopy Of AR55 139

5.2.4. Structure Calculation Of AR55 140
5.2.5. PRE Measurements 140
5.2.6. NMR Relaxation Data Collection And Analysis 141
5.3. Results And Discussion 142
5.3.1. CD Spectropolarimetry Indicates That AR55 Has A Similar Structure In All Micelle Conditions 142
5.3.2. NMR Assignment Of AR55 143
5.3.3. The Topology Of The AR55-Micelle Complex Is The Same For All Three Micelle Types 143
5.3.4. Structure Calculation Of AR55 144
5.3.5. The Micelle-Spanning Region Of AR55 Has A Similar Structure In Different Membrane Mimetics 146
5.3.6. The Structure Of The N-Terminal Tail Of AR55 Is Influenced By The Membrane Mimetic 147
5.3.7. The Dynamics Of AR55 Are Influenced By The Membrane Mimetic Choice 147
5.4. Summary 150
Chapter 6: Analysis Of Kinks In Membrane Proteins (Based On My Contribution To Manuscript (4)) 177
6.1. Introduction 177
6.1.1. Previous Analyses Of Membrane Protein Helix Kinks 177
6.2. Materials And Methods 179
6.2.1. Dataset Preparation 179
6.2.2. Monte Carlo Method For Helical Axis Definition 180
6.2.3. Bend Prediction From Primary Structure 182
6.3. Results And Discussion 183
6.3.1. Dataset Characteristics 183
6.3.2. The MC-HELAN Server 183
6.3.3. Convergence Of MC-HELAN 183
6.3.4. Frequencies Of Kinks 184
6.3.5. Comparison Of MC-HELAN To HELANAL And DSSP 185
6.3.6. Architecture Of Helical Kinks 188
6.3.7. Amino Acid Prevalence At Helical Kinks 189
6.3.8. Prediction Of Kinks From Sequence Information 191

6.4. Summary 193
Chapter 7: Conclusion 209
7.1. A Model Of Apelin-AR Interactions 209
7.2. Future Work 211
7.3. Impact 212
Bibliography 215

List Of Tables

1.1 The bioactive forms of apelin 18

2.1 Detailed NMR experiment acquisition parameters of apelin-17 44

2.2 Structural details of the apelin-17 ensembles 46

2.3 Summary of missing NMR assignments for apelin-17 47

2.4 Statistics of the clusters observed in the apelin-17 ensembles....48

3.1 Overview of peptide ligands that bind to micelles....84

3.2 Detailed NMR experiment acquisition parameters of apelin with micelles 85

3.3 Diffusion measurements of apelin-17 with micelles....86

3.4 Summary of missing NMR assignments for apelin-17 with micelles 87

3.5 Summary of the final retained NMR ensemble of apelin-17 with SDS....88

4.1 Thermocycling conditions used for DNA amplification....114

4.2 Detailed NMR experiment acquisition parameters of AR55 with DPC 115

4.3 Primer sequences and methods used for AR mutagenesis 116

4.4 Overview of the resonances assigned for AR55 with DPC 117

4.5 Summary of the final retained NMR ensemble of AR55 with DPC....118

5.1 Detailed NMR experiment acquisition parameters of AR55 with different membrane mimetics 151

5.2 Overview of the resonances assigned for AR55 solubilized in different membrane mimetics 155

5.3 Summary of the final retained NMR ensemble of AR55 solubilized in SDS, DPC or LPPG micelles as well as 50% HFIP....156

5.4 RMSD values of various superpositions of AR55 157

5.5 The backbone dynamics parameters of AR55 solubilized in SDS DPC and LPPG micelles 158

6.1 The relative frequencies of amino acids near a bend 194

6.2 TM helices classified by MC-HELAN, HELANAL and DSSP 195

6.3 The ability of classifiers to predict bends from sequence 196

List Of Figures

1.1 The GPCR signaling cycle 19

1.2 Resolving the problem of spectral overlap with multidimensional NMR 20

1.3 The expected chemical shifts of C^{α} and C^{β} nuclei 21

1.4 Comparison of TOCSY and NOESY modes of magnetization transfer 22

1.5 Comparison of TOCSY and NOESY spectra 23

1.6 Comparison of HNCO and HNcaCO experiments 24

1.7 Assigning spin systems using the backbone walk method 25

1.8 Comparison of an X-ray crystal structure to a NMR ensemble 26

2.1 HPLC analysis of apelin-13 49

2.2 Mass spectrometry analysis of apelin-13 50

2.3 Conformational sampling of apelin-17 51

2.4 Per-residue NOE contacts in apelin-17 52

2.5 Graphical summary of NOE contacts in apelin-17 53

2.6 Secondary chemical shifts of apelin-17 at 35 °C and 5 °C 54

2.7 Average deviations of the ϕ and ψ dihedral angles of apelin-17 55

2.8 RMSD values of different superposed regions of apelin-17 56

2.9 The proportion of apelin-17 in a PP_{II} or β-turn conformation 57

2.10 The structure of apelin-17 and 35 °C and 5 °C 58

2.11 A summary of the structural properties of apelin 59

2.12 Clustering analysis of apelin-17 60

2.13 Visualization of the major clusters of apelin-17 structures 61

2.14 RMSD values of the apelin-17 clusters from P14-F17 62

2.15 Secondary chemical shifts of the apelin isoforms at 5 °C.....63

2.16 Secondary chemical shifts of the apelin isoforms at 35 °C.....64

3.1 Superposed structures of peptides that bind to micelles 89

3.2 The structures of membrane mimetics 90

3.3 The structures of 5-doxylstearic acid and 16-doxylstearic acid.....91

3.4 CD spectra of apelin-12 and apelin-17 with micelles 92

3.5 Secondary chemical shifts of apelin-17 with micelles.....93

3.6 PRE analysis of apelin-17 94

3.7 NOE contacts on a per-residue basis of apelin-17 with SDS micelles 95

3.8 RMSD, ϕ and ψ analysis of apelin-17 bound to SDS micelles.....96

3.9 The β-turns present in apelin-17 when bound to SDS micelles.....97

3.10 The structure of apelin-17 bound to SDS micelles 98

4.1 Inclusion body purification of AR55 as shown by SDS PAGE.....119

4.2 HPLC analysis of AR55.....120

4.3 Mass spectrometry analysis of AR55.....121

4.4 ^{1}H-^{15}N HSQC spectrum of AR55 solubilized in DPC micelles.....122

4.5 Per-residue NOE contacts of AR55 solubilized in DPC micelles 123

4.6 Graphical summary of NOE contacts for AR55 with DPC micelles.....124

4.7 Convergence of ϕ and ψ dihedral angles in AR55 with DPC micelles 125

4.8 The structure of AR55 solubilized in DPC micelles.....126

4.9 The structure of the N-terminal tail of AR55 with DPC micelles 127

4.10 ERK activation by various AR mutants upon apelin stimulation 128

4.11 Cellular distribution of various AR mutants in HEK 293A cells 129

4.12 The structures of solved GPCRs 130

4.13 Titration of AR55 with apelin 131

5.1 CD spectra of AR55 solubilized in SDS, DPC and LPPG micelles 162

5.2 ^{1}H-^{15}N spectra of AR55 solubilized in various membrane mimetics 163

5.3 PRE analysis of AR55 in SDS, DPC and LPPG micelles 164

5.4 Possible topologies of AR55 in micelles 165

5.5 Per-residue NOE contacts for AR55 solubilized in membrane mimetics.....166

5.6 Graphical summary of NOE contacts for AR55 in membrane mimetics 167

5.7 Convergence of ϕ and ψ dihedral angles in AR55 with SDS micelles.....168

5.8 Convergence of ϕ and ψ dihedral angles in AR55 with DPC micelles 169

5.9 Convergence of ϕ and ψ dihedral angles in AR55 with LPPG micelles 170

5.10 Convergence of ϕ and ψ dihedral angles in AR55 in 50% HFIP.....171

5.11 The structure of AR55 in various membrane mimetics 172

5.12 The structure of AR55 in LPPG micelles and 50% HFIP.....173

5.13 R_1, R_2 and NOE measurements of AR55 in SDS, DPC and LPPG micelles.....174

5.14 Values of J(0), J(ω_N) and J($0.87\omega_H$) of AR55 in SDS, DPC and LPPG micelles 175

5.15 The S^2 parameters for AR55 solubilized in SDS, DPC and LPPG micelles 176

6.1 Composition of the culled helical membrane protein dataset 197

6.2 MC-HELAN analysis of a membrane protein 198

6.3 The distribution of helix angles to the membrane normal 199

6.4 Examples of TM helices classified as kinked 200

6.5 Comparison of kink detection by MC-HELAN, HELANAL and DSSP.....201

6.6 Kink detection overlap by MC-HELAN, HELANAL and DSSP.....202

6.7 The depth of bends and disruptions in the membrane 203

6.8 The distribution of kink angles for bends and disruptions.....204

6.9 Structures with a kink angle > 90º 205

6.10 The relative frequencies of amino acids near bends 206

6.11 The proportion of TM bends found near various amino acids.....207

6.12 The proportion of TM disruptions found near various amino acids 208

7.1 A model of apelin binding to AR.....214

Abstract

Apelin, the endogenous ligand to the apelin receptor, is a small peptide involved with cardiovascular regulation. Using nuclear magnetic resonance (NMR) spectroscopy, I demonstrate that at low temperature, residues R6-L9 and G13-F17 of apelin are more structured than the rest of the peptide. I also study the interactions of apelin with sodium dodecylsulphate (SDS), dodecylphosphocholine (DPC) and 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)] (LPPG) micelles. Apelin binds to SDS micelles through residues R6-L9, with structure being induced in this region as well as the C-terminus of the peptide. The binding to micelles along with the corresponding change in structure make it likely that apelin binds to the apelin receptor following the membrane catalysis hypothesis. NMR spectroscopy was used to determine the structure of the N-terminal tail and first transmembrane segment of the apelin receptor (AR55) in DPC micelles. AR55 has two disrupted helices from D14-K25 and from A29-K57. The second helix is the membrane spanning region of AR55 and has a significant kink located at N46. Mutagenesis of the apelin receptor and functional assays indicate that G42, G45 and N46 are essential for the proper trafficking and function of AR. In the N-terminal tail, the functionally critical residues E20 and D23 form an anionic face that could take part in initial binding of apelin to AR. The structure of AR55 was also determined in SDS micelles, LPPG micelles and a 1:1 water: 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) solution. Overall, the micelle spanning region of AR55 has a consistent structure with a kink near N46. The N-terminal tail of AR55 is more variable, having similar structures in the micelle conditions but being largely helical in 50% HFIP. NMR relaxation experiments indicate that the N-terminal tail of AR55 undergoes much more motion in LPPG micelles compared to SDS and DPC micelles. Finally, I created a program named MC-HELAN that characterizes the kinks that occur in protein helices. I used MC-HELAN to analyze all non-redundant membrane protein structures as of March 2010. Membrane protein helix kinks are remarkably common and diverse. Initial attempts to predict membrane protein kinks using only the protein sequence were unsuccessful.

List Of Abbreviations And Symbols Used

5-DSA: 5-doxylstearic acid

16-DSA: 16-doxylstearic acid

γ: gyromagnetic ratio

δ: Chemical shift or a pulsed field gradient duration

$\Delta\delta$: Secondary chemical shift

Δ: Diffusion time

λ: wavelength

ϕ: The phi dihedral angle or the volume fraction of lipids in a sample

ψ: The psi dihedral angle

μ_0: Permeability of free space

σ: Chemical shielding tensor

ω_0: Larmor frequency

AAFC: Agriculture and Agri-Food Canada

APJ: Apelin Receptor

AR: Apelin Receptor

AR55: A peptide consisting of the first 55 residues of AR fused to a polyhistidine tail

ARMRC: Atlantic Region Magnetic Resonance Centre

AT_1: Angiotensin II receptor type 1

B_0: External magnetic field

BSA: Bovine serum albumin

CD: Circular dichroism

COSY: Correlation spectroscopy

CSI: Chemical shift index

D: Translational diffusion coefficient

D_b: D of a micelle

D_f: D of a peptide free in solution

D_{obs}: D for a peptide in the presence of a micelle

DIC: Differential interference contrast

DMEM: Dulbecco's modified Eagle's medium

DOSY: Diffusion ordered spectroscopy

DPC: dodecylphosphocholine

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

DTT: Dithiothreitol

ERK: Extracellular regulated kinase

f_b: Fraction bound

FBS: Fetal bovine Serum

g: Pulsed field gradient power of a DOSY experiment

G-protein: Guanine-nucleotide binding regulatory protein

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

IPTG: Isopropyl β-D-1-thiogalactopyranoside

$J(\omega)$: The spectral density function

PCR: Polymerase chain reaction

PP_{II}: Polyproline II

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

M_{eq}: Equilibrium magnetization

NANUC: National High Field Nuclear Magnetic Resonance Centre

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

PDB_TM: Protein Data Bank of Membrane Proteins

PRE: Paramagnetic relaxation enhancement

pyr: Pyroglutamate

QANUC: Québec/Eastern Canada High Field NMR Facility

r: Distance between two nuclei

R_1: Longitudinal relaxation rate

R_2: Transverse relaxation rate

R_{ex}: Conformational sampling rate

RMSD: Root mean square deviation

ROESY: Rotational frame Overhauser effect spectroscopy

S^2: Generalized order parameter

S_f^2: Fast internal motion order parameter

S_s^2: Slow internal motion order parameter

τ_e: Effective internal correlation time

τ_f: Fast motion internal correlation time

τ_m: Overall correlation time

τ_s: Slow motion internal correlation time

SDS: Sodium dodecylsulphate

SVM: Support vector machine

TFA: Trifluoroacetic acid

TM: Transmembrane

TOCSY: Total correlation spectroscopy

XML: Extensible markup language

Acknowledgements

First of all, I would like to thank my supervisor Dr. Jan Rainey for providing me with a wonderful graduate school experience. My decision to enter graduate school was slightly sporadic and it is wonderful that I was able to begin working in Jan's lab. I strongly believe that Jan has always had my best personal and career interests in mind. Jan has gone out of his way to send me to meetings that has allowed me to grow substantially as a scientist. Finally, Jan has always been flexible and has allowed me significant control in the direction that my work has gone. This has allowed me to experience many different scientific techniques and develop into a well rounded scientist. For all of these things I thank you very much.

A significant part of my work was completed in the laboratories of Dr. Denis Dupré and Dr. Graham Dellaire. Both of you took the time out of your busy schedule to train me in new techniques and make me feel welcome in your labs. Through my time spent with you I have grown an appreciation for the different philosophies that exist in either teaching or managing a lab.

My time in graduate school would have been empty if it was not for all of the people whom I have worked with. In Jan's lab there are almost too many people to mention, but I would like to thank Aaron Banks, Nigel Chapman, Ryan Holstead, Pascaline Ngweniform, Aditya Pandey, Tyler Reddy, Lesley Seto, Kyungsoo Shin, Muzaddid Sarker, Matthew Speckert, Bruce Stewart, Marie-Laurence Tremblay, Michal Wieczorek and Lingling Xu. You all have made graduate school an enjoyable experience for me whether it was through technical advice (Bruce), being very entertaining (Nigel and Adi),

starting fires when I am not around (anonymous) or being there for me to grill during group meeting (everybody). You all have played your part in making graduate school a wonderful experience for me. Go H-squad!

I have received help from numerous other people. I cannot include everyone, but I have received extensive technical advice from Dr. Mike Lumsden at the Nuclear Magnetic Resonance Research Resource, Drs. Ian Burton and Ray Syvitski at the National Research Council Institute for Marine Biosciences, Dr. Chris Kirby at Agriculture and Agri-Food Canada, Dr. Tara Sprules at the Québec/Eastern Canada High Field NMR Facility and Dr. Ryan McKay at the National High Field Nuclear Magnetic Resonance Centre. Drs. Stephen Bearne, Roy Duncan and Younes Anini were all very helpful during my committee meetings. I have also received great advice from Dr. Melanie Dobson and Joyce Chew. Drs. David Waisman, Carmichael Wallace and Devanand Pinto have allowed me access to crucial equipment for my experiments. I have received CGS-D funding through NSERC, making graduate school possible for me.

I thank my parents for being supportive of my decision to enter graduate school. You have always been supportive of whatever I do and that is great. Last but not least I thank my wife, Robyn for her support. She is always there for me, even if I am running away to conferences without her. Over the past two months as I wrote this thesis she has been there to keep me sane.

All of you (and countless others) have shaped my time at Dalhousie University into a great experience. Thank-you!

Chapter 1:

Introduction

Although there are several different subjects discussed in this thesis, the overall theme is the study of membrane proteins and their ligands. In particular, most of my studies are focused on the apelin receptor (AR) and its ligand apelin. In this chapter I introduce the basic physiology of AR as well as fundamental concepts of nuclear magnetic resonance (NMR) spectroscopy, which will help the reader understand the methods employed in my experimental work. After this introduction, I discuss the structure of apelin in solution (chapter 2) and in a membrane bound state (chapter 3), which are based on previous publications (1, 2, 3). I then focus on the structure of a fragment of AR (chapter 4) as well as the effect that the membrane environment has on its structure (chapter 5). Finally, I close with chapter 6, which is based on a previous publication (4) and discusses the characteristics and causes of helix kinks that occur in membrane proteins.

1.1. G-Protein Coupled Receptors

1.1.1. Physiological Importance

Guanine-nucleotide binding regulatory protein (G-protein) coupled receptors (GPCRs) are a large and diverse family of 7 transmembrane (TM) pass membrane embedded proteins that are responsible for converting the extracellular signal of a ligand into an intracellular signal. The importance of GPCRs is highlighted by their abundance, with >750 different receptors identified in humans (5, 6, 7) and with about 500 estimated to be important for odour and taste functions while the rest detect endogenous ligands (6).

GPCRs are divided into 5 families (7, 8), which are named based on their hallmark receptor (glutamate-, rhodopsin-, adhesion-, frizzled-, and secretin-like). The rhodopsin-like and secretin-like families are also called class A and B GPCRs, respectively. Being such a large and diverse family, it is not surprising that GPCRs bind a wide variety of ligands including proteins (9), peptides and small molecules. Beyond their natural physiological import, GPCRs are targeted by ~25% of current drugs (10) and have great potential for pharmaceutical development.

1.1.2. GPCR Signaling

The canonical resting state GPCR signalling complex consists of the receptor bound to a trimeric G protein (G_{α}, G_{β}, G_{γ}, Figure 1.1). The G_{β} and G_{γ} subunits form a $G_{\beta\gamma}$ heterodimer that associates with the GPCR via the G_{α} subunit. There are multiple isoforms of the G_{α}, G_{β} and G_{γ} proteins, each of which having their own signaling specificity. In the resting state signaling complex, the G_{α} subunit is bound to guanosine diphosphate with the intracellular loops of the GPCR forming a guanine nucleotide exchange factor domain that contacts the G_{α} subunit. After an agonist binds to a GPCR, the conformational changes in the activated receptor cause guanosine diphosphate to be exchanged for guanosine triphosphate. This results in dissociation of G_{α} from the $G_{\beta\gamma}$ heterodimer (11). After stimulation, the G-proteins are free to dissociate from the GPCR, but remain membrane associated due to covalent lipid modifications to the G-proteins. Combined, the G_{α} subunit and $G_{\beta\gamma}$ heterodimer directly interact with and regulate a variety of downstream effectors such as adenylyl cyclase ($G_{\alpha i}$ and $G_{\alpha S}$), phospholipase C isoforms ($G_{\alpha q/11}$), RhoGEFs ($G_{\alpha 12/13}$) , and ion channels ($G_{\beta\gamma}$) (12, 13, 14, 15, 16, 17). G-proteins

themselves have slow GTPase activity, so eventually the inactive form of the G-protein is regenerated, allowing the resting state GPCR signaling complex to form again.

There are also mechanisms by which a GPCR can signal independently of G-proteins. Activated GPCRs are targets for GPCR kinases, which phosphorylate the GPCR and trigger the association of other proteins with the activated GPCR (18). The interaction of β-arrestin with an activated GPCR is the best characterized interaction. After β-arrestin associates with a GPCR, internalization of the GPCR occurs, causing desensitization to the ligand. β-arrestin also acts as a scaffold for other effectors, eventually resulting in activation of the ERK signaling pathways (19). Other proteins that cause G-protein independent signaling of GPCRs include the Jak family of kinases, Homer 1a and Na^+-H^+ exchanger regulatory factor (18).

Adding to the complexity of GPCR signaling, there is emerging evidence that the dimerization state of a GPCR effects the signaling which occurs (20). The first convincing evidence for GPCR dimerization was obtained for the γ-amino-n-butric acid receptor subunit 1 and 2 where a heterodimer of both subunits is required for effective signaling to occur (21). Also, both T_1R_2 and T_1R_3 receptors are required to heterodimerize to provide taste function (22). The influence of dimerization with other GPCRs can also be more subtle. For example, non-activated AR suppresses the function of AT_1 while activated AR enhances AT_1 function (23).

Perhaps the most interesting phenomenon associated with GPCRs is biased signaling. As mentioned above, GPCRs are able to signal not only through several types of G-proteins but also through G-protein independent pathways such as β-arrestin. This causes the observation that an inverse agonist for one pathway can act as an agonist for another

(22, 24). This greatly complicates studies of GPCRs because the results of a study can depend strongly on the specific functional readout chosen.

1.2. Physiology Of AR And Apelin

AR is a rhodopsin-like GPCR that was identified in 1993 due to high sequence similarity to the angiotensin II receptor type 1 (AT_1, (25)). AR was not activated by angiotensin-II and remained an orphan GPCR until 1999 when the endogenous peptide ligand was discovered and named apelin (26). AR is a typical class A GPCR with 7 TM spanning helices and has two putative extracellular disulphide bonds between C19 and C281 as well as C102 and C181 (27). Finally AR is likely palmitoylated at C325 and C326 while N15 and N175 are putative glycosylation sites (27).

Apelin is expressed as a 77 amino acid long preproprotein called preproapelin. After translation, the N-terminal signal peptide is cleaved in the endoplasmic reticulum (28) leaving a proprotein that is 55 residues long. Afterwards, through an unknown mechanism, the proapelin is cleaved to produce a number of bioactive peptides which all retain the C-terminus of the protein and range from 36-13 residues in length (Table 1.1, (26, 29)). Apelin-13 is the isoform primarily found in plasma (30). Although the generation of apelin isoforms remains a mystery, angiotensin-converting enzyme-related carboxypeptidase has been found to cleave the C-terminal phenylalanine from apelin isoforms, which provides a potential method for apelin regulation (31). While there is currently no known functional role for the N-terminal fragments of apelin, the 12 C-terminal residues of apelin constitute the core region essential for apelin function (32, 33, 34).

Since the discovery of AR, it has been shown to be widely expressed in humans, with high levels in the central nervous system, cardiovascular system, adipocytes and many other tissues (10). In some tissues both AR and apelin are expressed, suggesting an autocrine or paracrine mode of action, however, expression of apelin in the pituitary gland indicates there to likely be an exocrine function to apelin as well (28). AR signaling occurs mainly through coupling to the $G_{\alpha i}$ (35) and G_q pathways (36). From this point onwards, the activated G-proteins cause a signaling cascade which depends largely on the cell and tissue type, but includes effectors such as phospholipase C (36), ERK (35), Akt (37) and protein kinase C (38). AR also interacts with β-arrestin, which causes internalization of the receptor after activation. Interestingly, the strength of the β-arrestin association with AR depends on the isoform of apelin which causes activation (39), suggesting that the various apelin isoforms have functionally distinct roles.

AR signaling has key roles in the cardiovascular system and is amongst the most potent of identified inotropic agents, increasing heart muscle contractility with a half-maximal effective concentration of ~33 pmol/L in rat heart (40). Blood plasma apelin levels are correlated to heart failure progression (41). Adipocytes produce apelin and AR (42), and there is a correlation between plasma levels of apelin and obesity in humans (42, 43) and mice (42). Human immunodeficiency virus type 1 also uses AR as a co-receptor for CD4-mediated viral fusion to host cells (44). Recent studies have shown that apelin induces neoangiogenesis during tumour formation (45, 46), and AR knockout mice have diminished atherosclerotic lesions despite lacking apolipoprotein-E, suggesting a role for AR signalling in atherosclerotic plaque deposition (47).

In addition to its own functions, studies are beginning to tease out the interactions that AR has with other membrane proteins. Non-activated AR suppresses the function of AT_1 while activated AR enhances AT_1 function (23). Also, the function of both AR and the κ opioid receptor are enhanced by the presence of the other receptor (38). As a whole, the many functions of AR, either acting by itself or through other receptors, have made the apelin-AR system popular as a proposed therapeutic target for obesity and cardiovascular disease (28, 48, 49).

1.3. Structural Studies Of GPCRs

The determination of three-dimensional structures of GPCRs by NMR spectroscopy or X-ray diffraction is very challenging given the poor applicability of many standard protein overexpression protocols to membrane proteins, issues with low solubility, difficulties in purification and the poor crystallizability of GPCRs (50).

Although GPCRs are difficult to characterize structurally, there have been recent successes. Over the past few years, the use of X-ray crystallography has provided high resolution structures for 8 distinct GPCRs, providing unprecedented insight into GPCR function (51, 52, 53, 54, 55, 56, 57, 58). However, it is important to note that generally the flexible N-terminal tail of the structure is not resolved. Also, the third intracellular loop is often replaced by T4 lysozyme in order to facilitate crystallization (59).

NMR spectroscopy has also been used successfully to study GPCRs. There have been many studies where fragments of GPCRs are studied following a 'divide and conquer' approach (reviewed in (60)). In the case of rhodopsin, the NMR based structure of many individual fragments of rhodopsin have been modeled into a full length structure of the

receptor (61) that has a backbone atom root mean square deviation (RMSD) of less than 2.5 Å to the published X-ray structure (60, 62). More recently, NMR spectroscopists have attempted to determine the structure of full length GPCRs. NMR has been successfully used to study the structure of sensory rhodopsin II (63). Although not a GPCR, this study demonstrates that NMR spectroscopy is capable of determining the high resolution structure of 7 TM pass receptors. There have also been many studies probing specific questions about binding or dynamics of full length GPCRs using NMR spectroscopy but fall short of actually determining a high resolution structure ((64, 65) reviewed in (60)).

Through structural studies, GPCRs have been shown to have several defining structural features, the most fundamental being that they all have 7 TM helices with an extracellular N-terminus and an intracellular C-terminus. Most of the solved GPCR structures are in the presence of an antagonist or inverse agonist. In these studies it is clear that the faces of the TM helices create the binding site for a ligand. Although this binding site is accessible to solvent, entrance to the binding site is often regulated by the extracellular loops of the receptor as well (66). However, instead of a set ligand binding pocket, the specific residues and helices involved in ligand binding vary between GPCRs. For example, in the CXCR4 structure (57) the ligand binds to TM segments I, II, III and VII, which is different than observed in the β-2 adrenergic (66) and M-2 muscarinic (58) receptors (TM segments III-VII). Since the ligand binding site occurs at the interface of many different TM segments, it is clear that even slight changes in TM topology or helix orientation can have a large influence on the binding site structure. It has been hypothesized that the most structurally converged helices form the fundamental binding

site, while the more structurally variable helices (TM segments I, III and V) provide binding site plasticity (66).

The extracellular loops of GPCRs are highly variable, in particular extracellular loop II can adopt either an α-helix or β-sheet conformation (57, 66). Although the extracellular loops do not contribute to the ligand binding site, they shape the entrance to it with the aid of the N-terminal tail (66). With both the extracellular loops and the TM region of GPCRs being important for ligand binding, a two-step model of GPCR-ligand binding has been proposed for both class A chemokine GPCRs (67, 68) and class B GPCRs (69). There is substantial evidence that following this two-step model, GPCR ligands initially bind to the extracellular domain of the receptor and then penetrate deeper into the TM domain in order to trigger activation (67, 68, 69, 70, 71). Perhaps the strongest evidence for this scenario comes from Kofuku et al. who determine that the chemokine stromal cell-derived factor-1 binds to its receptor CXCR4 in two stages with the central region of the peptide binding first. The authors determine that the N-terminal region of the peptide then binds and activates CXCR4 in an independent event (68).

The principles of G-protein activation by GPCRs have been more elusive to determine. This is largely because it is clear that a spectrum of active GPCR conformations exist (72). The best model of GPCR activation comes from the β2-adrenergic receptor where structures of both nanobody (73) and G-protein (74) bound active states have been solved. Through these studies it is clear that small structural changes that occur in the ligand binding site during activation (2.1 Å movement of TM V) are transduced into large structural changes at the cytoplasmic side of the receptor (14 Å movement of TM

VI). This displacement is facilitated through the packing interactions of residues from TM segments III, V, VI and VII.

1.4. Circular Dichroism (CD) Spectropolarimetry

CD spectropolarimetry is an excellent tool for the rapid determination of protein secondary structure content. In CD spectropolarimetry, both left and right handed circularly polarized light are passed through a sample. Asymmetric molecules will preferentially absorb either left or right handed polarized light at specific wavelengths, causing the light to be described as elliptically polarized (75). In proteins, if the peptide bonds of the backbone are arranged into arrays, their optical transitions are split due to 'exciton' interactions (76). This results in a typical CD spectrum for various protein structural elements. For example, an α-helix has a characteristic positive band at 193 nm with negative bands at 208 and 222 nm (77). By collecting a CD spectrum and then comparing the spectral features to those that are expected for various secondary structures, the secondary structure composition of a protein can quickly be assessed either visually or by using computer algorithms to calculate secondary structure content (78).

1.5. Using NMR Spectroscopy To Study Proteins

1.5.1. Basics Of NMR Spectroscopy

NMR spectroscopy has been instrumental in providing detailed structural information on molecules for decades and is based on the principles of nuclear magnetism (79). NMR spectroscopy relies on the interaction between nuclei that possess nuclear spin angular momentum and an external magnetic field. Nuclear spin angular momentum is a

fundamental property of a nucleus, and requires a non-zero value of nuclear spin. Considering the typical nuclei in proteins, ^{12}C and ^{16}O have a nuclear spin of 0, while ^{1}H, ^{13}C and ^{15}N have a spin of 1/2 and ^{2}H and ^{14}N have a spin of 1 (79). Since NMR relies on a nucleus having spin angular momentum, it is not surprising that all nuclei with spin = 0 cannot be observed using NMR spectroscopy. On the other hand, nuclei with spin > 1/2 have complex interactions with the magnetic field, leading to complex spectra. For this reason, in NMR spectroscopy the nuclei of the sample are often replaced with 'NMR friendly' spin 1/2 nuclei. For example, in proteins the high natural abundance ^{12}C (spin 0) and ^{14}N (spin 1) nuclei are replaced with ^{13}C and ^{15}N.

If a nucleus has spin angular momentum, it also possesses a magnetic moment, meaning it will act as a tiny bar magnet (80). In an external magnetic field (B_0), the nuclear magnetic moment will tend to align itself with the magnetic field, causing the sample to become magnetized in the direction of B_0 until an equilibrium magnetization (M_{eq}) is reached. Radiofrequency irradiation can then be used to apply a torque and change the direction of M_{eq}. Whenever M_{eq} is not parallel with B_0 it will rotate about B_0. This precession of M_{eq} about B_0 occurs at a characteristic frequency called the Larmor frequency:

$$\omega_0 = -\gamma B_0 \qquad 1.1$$

ω_0 is the Larmor frequency in radians/second and γ is the gyromagnetic ratio, which is a constant for a given nucleus (80). As is clear from equation 1.1, the Larmor frequency is influenced by the magnitude of B_0. It is critical to realize that although B_0 is generated by the static magnet of the NMR spectrometer, the magnetic field that a particular nucleus

experiences also depends on molecular structure. Electrons are able to partially shield the nucleus from B_0 and a simple modification of equation 1.1 can take this into account:

$$\omega_0 = -\gamma B_0(1-\sigma) \qquad 1.2$$

Where σ represents a chemical shielding tensor for a particular nucleus (81). The result of equation 1.2 is that most nuclei in a sample have a unique Larmor frequency due to the unique chemical shielding that they experience. This allows NMR spectroscopy to act as a sensitive probe of molecular environment for all of the chemically unique nuclei in a sample.

At the most basic level, in an NMR experiment, radiofrequency pulses are used to change the direction of M_{eq} so that it is not parallel to B_0, after which the Larmor frequency of the nuclei are measured and plotted as B_0-independent chemical shift values (δ).

1.5.2. Signal Overlap In NMR Spectroscopy

As mentioned above, NMR spectroscopy measures the Larmor frequency of the nuclei in a sample. In polypeptides larger than just a few residues, there is a problem as the sheer number of NMR active nuclei causes the peaks in a typical one-dimensional NMR spectrum to be overlapped and unusable for structural analysis (Figure 1.2 (A)). The practical solution for this is to correlate the NMR measurement between multiple nuclei, which increases the dimensionality of the experiment. For example, in the ^{1}H-^{15}N heteronuclear single quantum coherence spectroscopy (HSQC) experiment the nuclear magnetization is transferred from ^{1}H to ^{15}N and then back again for signal detection (80, 82). This means that only ^{1}H nuclei that are able to transfer magnetization to a ^{15}N

nucleus are detected in an HSQC experiment. In addition, each measured signal has both the ^{1}H and ^{15}N frequencies associated with it. By plotting both frequencies along the x and y axes, a crowded one-dimensional spectrum becomes a better resolved two-dimensional spectrum (Figure 1.2 (B)). This concept can also be extended to higher dimensionality. For example, an HNCO (83) experiment relies on the transfer of magnetization between the amide proton, ^{15}N amide and the ^{13}C carbonyl carbon in a protein. Since magnetization is transferred between these three nuclei, the resulting signal has ^{1}H, ^{13}C and ^{15}N frequencies associated with it. Plotting these data in three dimensions results in a further decrease in signal overlap and is very amenable to analysis (Figure 1.2 (C)).

1.5.3. Sequential Assignment Of Protein NMR Spectra

In this section and subsequent sections, I will be using the Greek alphabet nomenclature as used in Cavanaugh to describe the nuclei of amino acids (82). In this nomenclature the backbone nitrogen, carbonyl carbon and alpha carbon are referred to as N, CO and C^{α} respectively. The side chain heavy nuclei are annotated a similar way with the Greek letter incrementing through the Greek alphabet further out in the side chain. For example, the carbon connected to C^{α} is C^{β}. Protons are designated based on the carbon they are attached to (*e.g.* H^{α} is covalently bound to C^{α}). Finally H^{N} corresponds to the backbone amide proton.

Although each NMR active nucleus in a sample is capable of producing an NMR signal, identifying which specific nucleus is responsible for each NMR signal is a difficult task that requires extensive analysis to determine. There are several techniques

used to assign protein NMR spectra, but all of them rely on the fundamental principle that different nuclei have different expected chemical shifts. Each amino acid has expected chemical shifts based both on the size of the amino acid and its unique chemical composition (*e.g.* the C^{α} and C^{β} shifts, Figure 1.3, (84)). Although the example shown is for C^{α} and C^{β} chemical shifts, any combination of chemical shifts can be used to help identify a spin system.

A common experiment which allows NMR resonance assignment is correlation spectroscopy (COSY). In ^{1}H-^{1}H COSY, magnetization is transferred between protons which are connected to each other through chemical bonds (no more than 3). For example, an H^{N} is able to transfer its magnetization to H^{α}, and H^{α} transfers its magnetization to both H^{N} and H^{β}. In this way, all of the proton resonances of an amino acid can be assigned. A similar experiment to COSY is total correlation spectroscopy (TOCSY (85)). In homonuclear ^{1}H TOCSY experiments, the magnetization from a proton is transferred to all other protons that are connected to it through chemical bonds. So long as every heteroatom (generally ^{13}C or ^{15}N) between two protons has at least one proton attached to it, magnetization may be transferred (Figure 1.4 (A)). For both COSY and TOCSY experiments, in the case of an amino acid all of the protons of an amino acid interact with each other and form a "spin system" (Figure 1.5). In the case of proteins, each residue conveniently forms its own spin system. This is because the backbone carbonyl carbon of each peptide bond has no attached protons and prevents COSY and TOCSY magnetization transfer between residues. After each spin system is identified, the amino acid type of each spin system can be determined based on the chemical shifts and number of resonances in the spin system. After this step, nuclear Overhauser effect

(NOE) spectroscopy (NOESY) experiments are acquired. In NOESY there is direct transfer of magnetization from one nucleus to another through space (Figure 1.4 (B), (80, 82)). Residues that neighbour each other in the sequence are also close in space and will have cross peaks to each other in NOESY spectra. In particular, through-space connections are generally seen between the H^N nuclei with the H^{α} and/or H^{β} nuclei of the preceding residue (82). Through the use of TOCSY and NOESY experiments, sequential assignment can be completed for many proteins.

Another approach to sequential resonance assignment is called the main chain directed approach (82), commonly known as the 'backbone walk'. This approach relies on the collection of complementary three-dimensional experiments such as the HNCO (83) and HNcaCO (86) experiments (Figure 1.6). Similar to TOCSY, these experiments rely on the direct transfer of magnetization to neighbouring nuclei through chemical bonds. In the HNCO experiment, a correlation is produced between the amide nitrogen and amide proton of a residue (i) to the carbonyl carbon of the residue before it (i-1) giving a signal in three dimensions with resonance frequencies corresponding to the ^{1}H, ^{13}C and ^{15}N nuclei. In the HNcaCO experiment, the same amide proton and amide nitrogen correlate to CO of both residue i-1 and i. In this way, by looking at both HNCO and HNcaCO spectra, the ordering of the spin systems becomes apparent (Figure 1.7). There are other paired experiments that correlate to the chemical shifts of the C^{α} (HNCA (83), HNcoCA (87)) or the C^{α} and C^{β} (HNCACB (88), HNcoCACB (88)) in the place of the CO chemical shift. Note that for these experiment names I use capital letters to denote nuclei that have their chemical shift detected, while lowercase letters denote nuclei used solely for magnetization transfer. Side chain assignments of the individual spin systems can be

accomplished either through these three dimensional experiments or through the COSY and TOCSY types of experiments as discussed earlier. Compared to using only TOCSY/NOESY spectra for assignment, the backbone walk has the distinct advantage of directly telling the spectroscopist the order in which spin systems are connected to each other.

1.5.4. Using NMR Spectroscopy To Determine Protein Structure

As mentioned previously, the NOE allows for the direct transfer of magnetization from one nucleus to another through space (Figure 1.4 (B), (80, 82)). The amount of magnetization that is transferred between nuclei is proportional to the distance between them. Specifically, in a NOESY spectrum the peak volume is related to the distance (r) between the nuclei by a factor of r^{-6}. Each peak in a NOESY spectrum can be used to define a specific distance restraint between two nuclei, where 5-6 Å is generally considered to be the maximum distance over which the NOE may act (80). After assigning an entire NOESY spectrum, there may be thousands of distance restraints implied between the nuclei in the sample.

In order to deal with these data, computer programs such as XPLOR-NIH (89, 90), CYANA (91) and CNS (92) are used to generate molecular structures in agreement with these restraints. Although many software packages have been developed, most of them are based on a simulated annealing protocol (93, 94). The steps of structure calculation based on simulated annealing are:

i) Start with a basic structure of the protein of interest, initially this can be an extended polypeptide with the correct sequence.

ii) Make small random changes to the basic structure to create a new protein structure.

iii) Check this new structure against all of the known distance restraints.

iv) If the structure fits the distance restraints better than the previous structure there is a chance that it will be kept as the new basic structure. The probability that a structure will be retained often varies throughout the simulated annealing protocol.

v) Repeat this process thousands of times.

An important concept to note is that the generation of NMR structures relies on random changes to a starting structure. This means that no two NMR structures are alike, even if identical distance restraints are used. For this reason, NMR structures are represented as ensembles of structures (Figure 1.8), all of which ideally satisfy the experimental restraints. This contrasts strongly with structures generated using X-ray crystallography, where only one structure is often presented even though the electron density can sometimes accommodate alternate configurations.

1.6. Rationale And Objectives

As indicated above, apelin-AR signaling is important for several medically relevant processes. When I began this work, there were no AR agonists or antagonists other than those derived from the apelin peptides. Drug development can be guided by the structure of the target and there is very little structural data about both apelin and AR. With the

strong structural biology background in the Rainey laboratory, the apelin-AR system is a great candidate for study.

The objective of my work was to use biophysical techniques to characterize apelin and AR. When this work began there was very little structural information about GPCRs and their ligands, with only the structure of bovine rhodopsin being known (54, 61). By studying apelin and AR I set out to provide new insights into how ligands may interact with GPCRs. Furthermore, by determining the structure of apelin and AR I will provide the structural data required for intelligent drug design.

Table 1.1: The sequences of the predominant bioactive forms of human apelin. The functionally essential apelin-12 C-terminal region is bolded in each peptide sequence.

Peptide	Sequence
Apelin-36	H_2N-LVQPRGSRNGPGPWQGGRRKFRRQ**RPRLSHKGPMPF**-COOH
Apelin-17	H_2N-KFRRQ**RPRLSHKGPMPF**-COOH
Pyr-apelin-13[a]	HN-pyr-**RPRLSHKGPMPF**-COOH
Apelin-13	H_2N-Q**RPRLSHKGPMPF**-COOH
Apelin-12	H_2N-**RPRLSHKGPMPF**-COOH

[a] pyr represents the amino acid pyroglutamate.

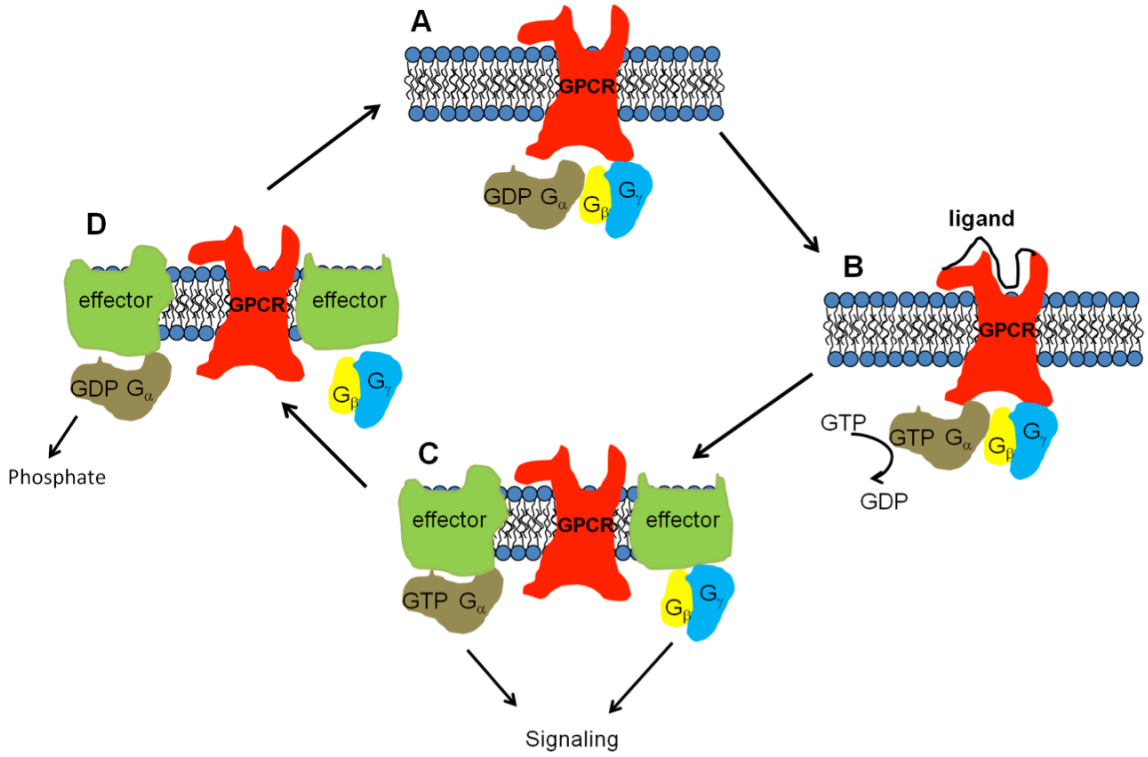

Figure 1.1: The GPCR signaling cycle. When in the resting state, the G_α as well as $G_{\beta\gamma}$ heterodimer are associated with a GPCR (A). When a ligand binds to and activates the GPCR, a conformational change occurs and GDP is exchanged for GTP (B). The G_α and $G_{\beta\gamma}$ proteins are then free to dissociate from the GPCR and interact with various effector proteins (C). Eventually the GTP is hydrolysed to GDP and signaling stops (D) allowing the resting state complex to reform.

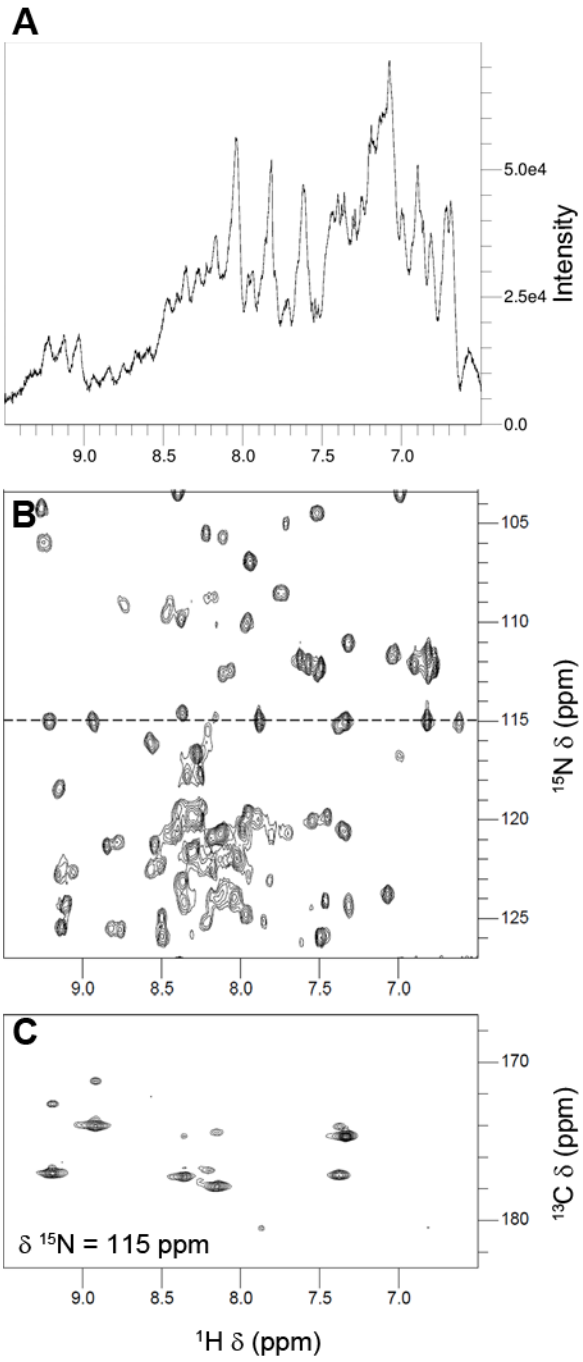

Figure 1.2: Resolving the problem of spectral overlap using multidimensional NMR spectroscopy. In (A), a one dimensional ^{1}H spectrum is shown, while in (B) a two dimensional ^{1}H-^{15}N HSQC spectrum is displayed. In panel (C) a slice from a three dimensional HNCO spectrum is shown. The chemical shift of the ^{15}N dimension in (C) corresponds to the dashed line in (B).

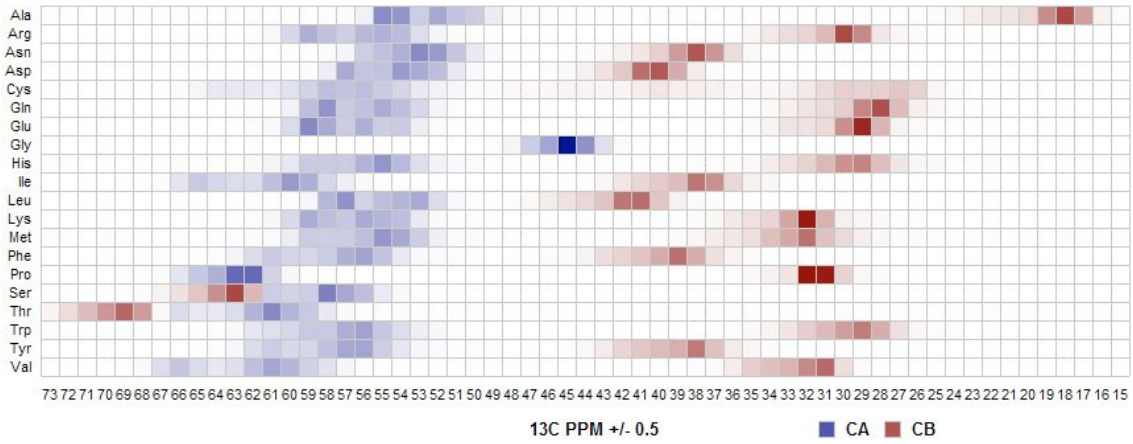

Figure 1.3: The expected chemical shifts of C^{α} and C^{β} nuclei for the 20 naturally occurring amino acids. The darker the blue or red colour, the more likely a particular nucleus will be found at a chemical shift. This table was made with the use of CCPNMR Analysis (95) using the chemical shifts in the Re-referenced Protein Chemical Shift Database (96).

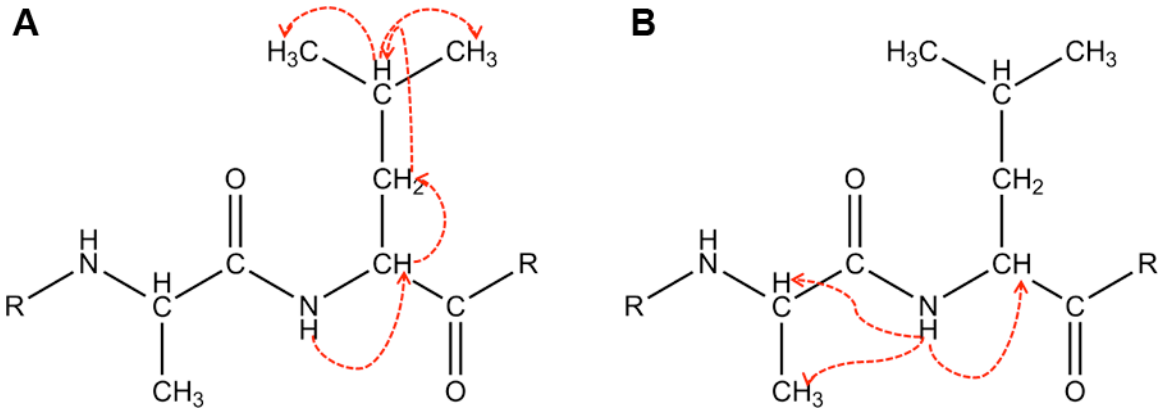

Figure 1.4: A comparison of (A) TOCSY vs. (B) NOESY modes of magnetization transfer. In the TOCSY experiment, magnetization travels through chemical bonds and cannot travel through any nucleus that does not have an attached proton, such as CO. In NOESY, magnetization transfer occurs through space so long as the distance between interacting nuclei is < 6 Å. Red arrows indicate possible routes of magnetization transfer from H^N in both the TOCSY and NOESY experiment.

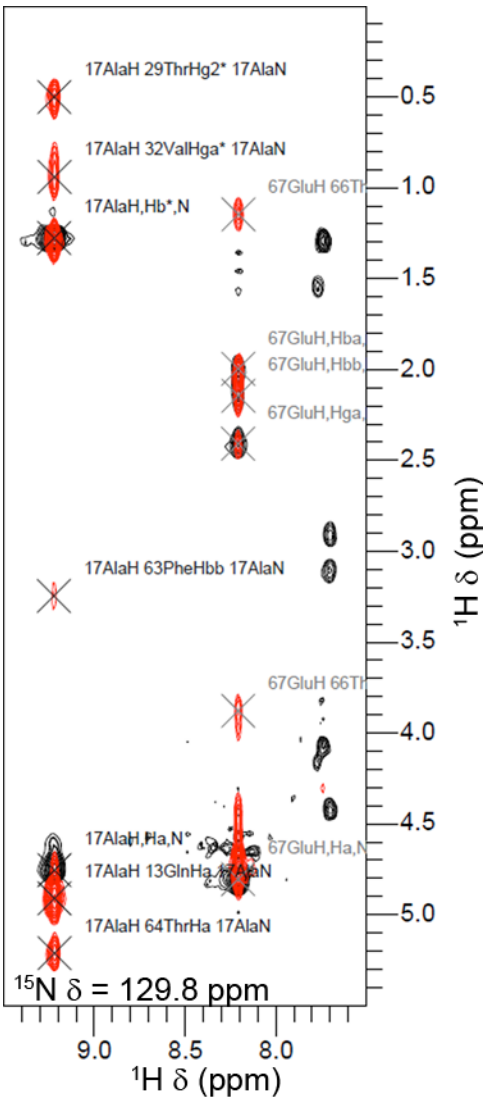

Figure 1.5: Comparison of TOCSY and NOESY spectra. Shown is slice of a TOCSY-^{15}N HSQC spectrum (black) and a NOESY-^{15}N HSQC spectrum (red) at a ^{15}N frequency of 129.8 ppm. Notice that the peaks in a TOCSY spectrum all are the same residue, while for the NOESY spectrum a peak results from all nuclei that are close together in space.

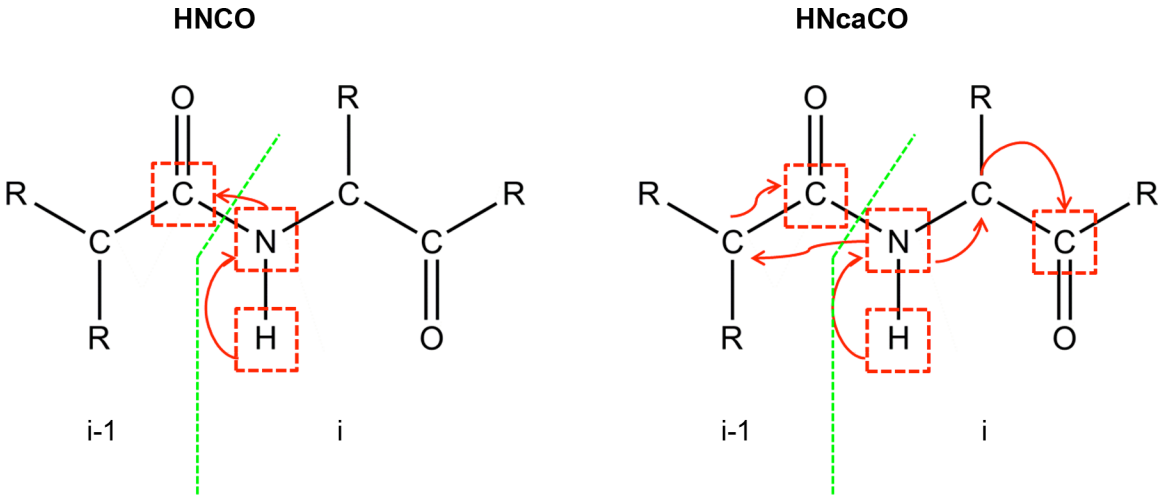

Figure 1.6: A comparison of the HNCO and the HNcaCO experiments. Magnetization transfer is indicated by red arrows and the nuclei surrounded by a red box have their chemical shift observed. Both residue i and i-1 are indicated. In the HNCO experiment, magnetization is transferred from N of residue i to the CO of residue i-1. In the HNcaCO experiment, the magnetization is transferred from the N of residue i to both i-1 and i CO nuclei through the C^{α} nuclei.

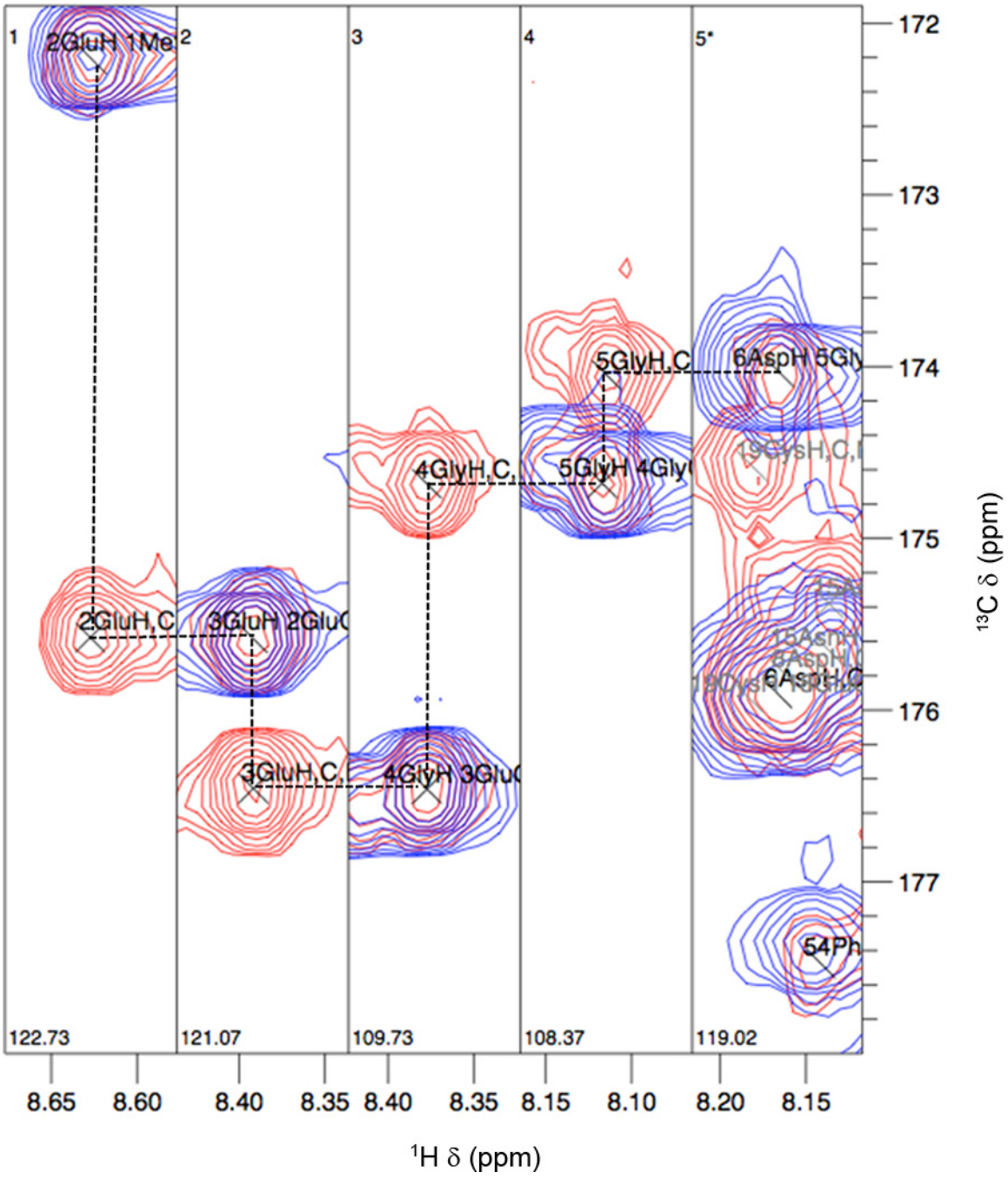

Figure 1.7: Assigning spin systems using the backbone walk method. The HNCO (blue) and HNcaCO (red) spectra are shown. In HNCO only one peak from residue i-1 is visible, while in the HNcaCO two peaks corresponding to residue i-1 and i are present. ^{1}H and ^{13}C chemical shifts are indicated on the x and y axes respectively, while the chemical shift of the associated ^{15}N nucleus is indicated in the lower left corner of each rectangular strip of the figure.

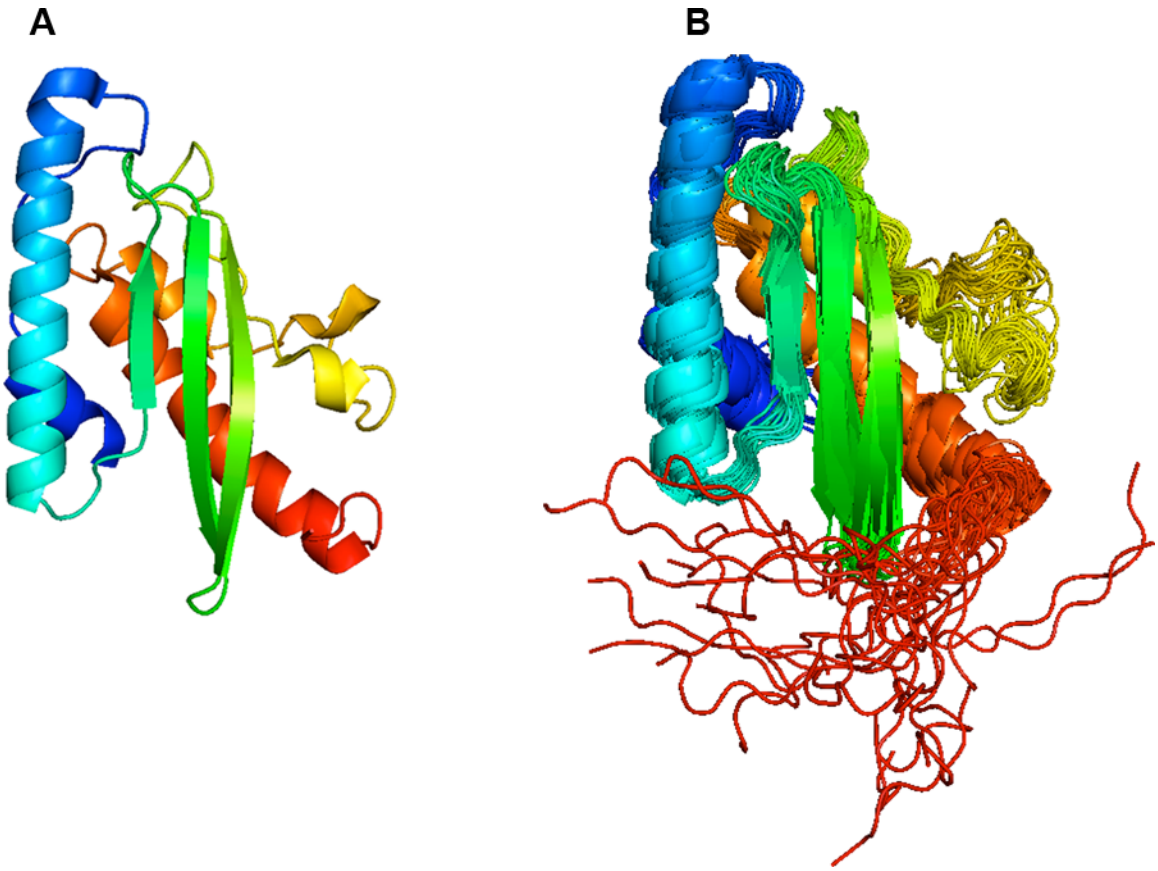

Figure 1.8: Comparison of (A) a single X-ray crystal structure to (B) a NMR ensemble of structures for E2-like ubiquitin-fold modifier conjugating enzyme 1 (97).

Chapter 2:

The Structure Of Apelin In Solution (Based On My Contribution To Manuscript (1))

2.1. The Potential For Apelin Structure

The region of apelin required for binding to and activation of AR, as demonstrated using truncation mutants, is the C-terminal peptide apelin-12 (Table 1.1, (98)), and the C-terminal phenylalanine is critical according to a number of studies (33, 34, 99). Alanine scanning mutagenesis on apelin-13 was published almost simultaneously by two groups ((32, 33), also discussed in (100)). As a whole the mutagenesis studies indicate that most of the 12 C-terminal residues of the apelin isoforms are essential for function. This mutagenesis data and the fact that the apelin peptides (Table 1.1) having at least the 12 C-terminal residues in common are specific ligands for AR (33, 44, 98) suggests that the 12 C-terminal residues of apelin adopt a specific structure to allow binding and activation of AR to occur. Despite this precedence, apelin-13 and apelin-36 adopt a "random conformation" in solution based on studies by CD spectropolarimetry (32). CD spectropolarimetry is an ensemble sum measurement, meaning that structures that only populate a few residues or lower proportion of molecules may not be clearly visible. With this in mind, β-turns and polyproline-II (PP_{II}) are secondary structures potentially consistent with the reported CD spectra (32).

Both β-turns and PP_{II} have been associated with receptor-ligand binding. First, it has been hypothesized that β-turns are ubiquitous and important for GPCR binding peptides (101). Second, PP_{II} is often found in short "unstructured" peptide ligands as indicated by many of these peptides having CD spectra which are PP_{II} in nature (102). PP_{II} is also postulated to play a critical role in intermolecular interactions in part since its extended nature easily allows the side chains to interact with a binding surface (102, 103). The CD study by Fan et al. was conducted at ambient temperature, and it is reasonable to expect that a short peptide at ambient or physiological temperature will be free to undergo a great deal of conformational sampling (104), allowing access to many structural elements.

2.1.1. Using NMR Spectroscopy To Study Small Peptides

NMR spectroscopy is an ideal method for studying the structure and dynamics of small peptides such as apelin (104, 105). The chemical shifts of the nuclei themselves can be used to predict peptide secondary structure since chemical shift has a strong dependence on secondary structure (106). In addition, the use of NOESY experiments provides structural information on a per-residue basis. This allows small regions of converged structure to be accurately characterized in a peptide, even if the rest of it is unstructured. The final structural ensemble generated for peptides or proteins will often contain several different conformations, each of which may be important in some aspect of function (107).

To promote the formation of peptide conformation(s) relevant to receptor binding, a favoured strategy is the use of lowered temperature (0-8 °C) to allow increased sampling of entropically disfavoured states through reduction of the entropic contribution to Gibb's

free energy, *i.e.,* TΔS in ΔG = ΔH - TΔS (105, 108, 109). Also, at low temperatures the formation of hydrogen bonds is favoured and can help pay the entropic penalty of adopting a restrained structure (110). Isolated peptidic ligands at low temperature have been shown to adopt the same structure as the bound-state (111) and several different peptides activating the chemokine GPCR CXCR4 show the same structuring at low temperature (108).

The apelin-AR system is a promising target for therapeutics, and small molecule (44) and peptide (112) based drug design will be facilitated by high-resolution structures for apelin and AR. In this chapter, I present NMR data for a variety of apelin peptides at both physiological and low temperatures allowing determination of the bioactive structure of apelin. These results are placed into a functional context through correlation with previous mutagenesis studies.

2.2. Materials And Methods

2.2.1. Materials

Pyroglutamate modified apelin-13 (pyr-apelin-13), was obtained from AnaSpec (San Jose, CA). Deuterium oxide (D_2O; 99.8 atom % D) and D_2O containing 1% (w/w) sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) were obtained from C/D/N isotopes (Pointe-Claire, QC). All other chemicals were obtained at biotechnology, high performance liquid chromatography (HPLC) or reagent grade, as appropriate, from Sigma-Aldrich Canada Inc. (Oakville, ON).

2.2.2. Peptide Synthesis & Purification

Peptide synthesis of apelin-12, -13 and -17 (Sequence of apelin-17 is KFRRQRPRLS HKGPMPF, Table 1.1) was carried out by Dr. Jan Rainey following standard protocols (1, 113). After lyophilization, each reaction mixture was dissolved in water and examined by analytical reversed-phase HPLC (Beckman System Gold, Fullerton, CA) using a water:acetonitrile (A:B) gradient with B progressing from 2 to 80% on a C_{18} column (5 μm particle, 4.6 mm × 250 mm Spirit Peptide column, AAPPTec) monitored at 210 nm. A semi-preparative C_{18} column (5 μm particle, 10 mm × 250 mm Spirit Peptide column, AAPPTec) was used for purification. Final product purity was assessed with HPLC (Figure 2.1). Product masses and amino acid composition were confirmed by positive mode electrospray ionization mass spectrometry (Figure 2.2, DalGen Proteomics Core Facility, Halifax, NS) and quantitative amino acid analysis (Alberta Peptide Institute, Edmonton, AB and Hospital for Sick Children, Toronto, ON), respectively. For amino acid analysis, exact quantification was achieved using a weighted average of the arginine, proline, leucine and phenylalanine content in the sample.

2.2.3. NMR Spectroscopy

Solutions of apelin-12, apelin-13, pyr-apelin-13 and apelin-17 were prepared with 3-4 mM peptide in a 90% H_2O/ 10% D_2O mixture with 20 mM $Na^+CD_3COO^-$, 1 mM NaN_3, and 1 mM DSS as an internal standard. Sample pH in all cases was adjusted to 5.00 ± 0.05 using DCl and NaOD. NMR experiments were performed at the Atlantic Region Magnetic Resonance Centre (ARMRC, Halifax, NS) on a 500 MHz Bruker AVANCE spectrometer (Milton, ON) and the National High Field Nuclear Magnetic Resonance

Centre (NANUC, Edmonton, AB) on 500, 600 and 800 MHz Varian INOVA spectrometers (Palo Alto, CA). The 800 MHz spectrometer is equipped with a cryogenically cooled probe. One-dimensional ^{1}H, natural abundance gradient enhanced ^{1}H-^{13}C HSQC, ^{1}H-^{1}H NOESY (400 ms mixing time), ^{1}H-^{1}H TOCSY (60 ms mixing time with DIPSI spin lock) and double-quantum filtered COSY experiments were acquired for apelin-17 at 5 °C and 35 °C. Detailed acquisition parameters are found in Table 2.1. Note that the data for apelin-13 at 35 °C were acquired and analyzed by Dr. Jan Rainey and are not included. Similar sets of experiments were performed for the other peptides except that ^{1}H-^{1}H ROESY (200 ms mixing time) replaced the NOESY experiment, and for apelin-12 and pyr-apelin-13, experiments were completed at 6 °C instead of 5 °C due to instrumental constraints. NMR data were processed using NMRPipe (114) and manually assigned in Sparky 3 (T.D. Goddard and D.G. Kneller, University of California, San Francisco).

2.2.4. Structure Calculation & Ensemble Analysis

Peak volumes for NOESY spectra of apelin-17 were estimated using Sparky's integration algorithm with Gaussian line shapes and no peak motion allowed. For peaks where this algorithm failed, the sum over ellipse method was used (T.D. Goddard and D.G. Kneller, University of California, San Francisco). Peak volumes for both Gaussian and sum over ellipse fits were independently calibrated so that the smallest volume represented a 6 Å distance while the largest volume peaks represented a 1.8 Å distance. Peaks were binned by volume to be strong (1.8-2.8 Å), medium (1.8-4 Å), weak (1.8-5 Å) or very weak (1.8-6 Å). The python scripting interface of XPLOR-NIH version 2.18 (89) was used for simulated annealing structure calculations. Ambiguous assignments

were accommodated with the XPLOR-NIH 'or' statement. The NOE term used a square well potential with a scaling factor of 20 for the high temperature stage that was ramped from 1-30 for the cooling stage. The J coupling potential used a scaling factor of 10 throughout the protocol. Families of structures were generated through simulated annealing using a high temperature stage at 3500 K for 20 ps and a cooling stage going from 3500 K to 100 K over 20 ps. For both the high temperature and cooling stage, 10 fs time steps were used.

Three apelin-17 structural ensembles were refined and calculated: one for the NMR data at 35 °C and two separate conformers based on the 5 °C data. The two structural ensembles for the 5 °C NMR data were produced using two separate sets of NOE contacts assignable to apelin-17 conformers in slow conformational exchange in the G13-F17 region (Figure 2.3). For all cases, NOE restraint refinements were performed through iterative calculation of 100-member structural ensembles. Briefly, ensembles were analyzed using an in-house tcl/tk script that calculated the magnitude and frequency of violating NOEs. Starting from a violation length of >0.5 Å in >50% of ensemble members, the stringency was gradually increased to a final value of 0.05 Å in 15% (35 °C) or 10% (5 °C) of ensemble members until the average energy of each ensemble reached a consistent minimum (iterative structure calculation statistics shown in Table 2.2). A final ensemble of 200 structures was calculated, with the 80 lowest energy structures retained for analysis. In addition, simulated annealing of apelin-17 without inclusion of any experimentally derived restraints was used to generate an ensemble of 1000 random structures. This was used as a control for both meaningful RMSD values of

the peptide backbone atoms and for frequency of β-turns and PP_{II} in an unrestrained peptide.

For each final ensemble of structures: (i) Procheck-NMR (115) was used to tabulate residues in favoured, allowed, generously allowed and disallowed areas of the Ramachandran plot; (ii) an in-house tcl/tk script was used to determine the residues with dihedral angles corresponding to a PP_{II} conformation according to (ϕ, ψ) definitions of $-90° \leq \phi \leq -20°$ and $50° \leq \psi \leq 240°$ derived by Best et al. through statistical analysis of the Protein Data Bank (116); and (iii) Promotif-NMR was used to determine frequency and location of β-turns (117). The LSQKAB software of the CCP4 suite (118) was used to iteratively superpose the clusters over 4-17 residue stretches while an in-house tcl/tk script was used to calculate the RMSD of the backbone atoms compared to a representative structure within the superposed region. The most converged areas, determined by comparison to RMSDs calculated for the ensemble of 1000 unrestrained apelin-17 structures described above, were analyzed for preferred conformation within the ensemble using Clusterpose (119). RMSD values for each cluster based upon the R1, R2 and R4 methods of Clusterpose (119) were parsed into NEXUS format (120) by an in-house python script and visualized as phylograms using Figtree (A. Rambaut, University of Edinburgh). A group of structures were considered to be part of the same cluster if: (i) the R1, R2 and R4 clustering methods all placed these structures into nearly or exactly the same cluster; (ii) the resulting cluster constituted ≥10% of the total ensemble; and (iii) the increase in RMSD caused by superposing one cluster onto another is large compared to the RMSD of structures within a cluster. Pymol (Delano Scientific, San Carlos, CA, USA) was used for molecular visualization. Chemical shifts, restraint files, and the final

80-member ensembles for apelin-17 at 35 °C and for conformers A and B of apelin-17 at 5 °C have been deposited at the BioMagResBank (121) using SMSDep (BMRB Accession Codes 20029, 20030 and 20031).

2.3. Results And Discussion

2.3.1. Apelin Adopts Multiple Conformations In Solution

For apelin-17 as well as the shorter apelin peptides examined, slow conformational exchange was apparent at both 5 °C and 35 °C in the C-terminal region (residues G13-F17 of apelin-17). This gave rise to multiple spin systems for each of these residues, with a clearly predominant conformer and multiple minor conformers (example shown for M15 in Figure 2.3). Based on peak areas in 1D ^{1}H and volumes in 2D ^{1}H-^{1}H TOCSY spectra, ~80% of the apelin-17 adopts one conformation at (Conformer A) and the other ~20% adopts another conformation (Conformer B) at 5 °C. This estimate neglects the other less populated conformations that are estimated at a combined total of <5% for all remaining conformers.

The conformational sampling was isolated to G13-F17, a region containing two proline residues. Proline is unique in that its *cis-trans* peptide bond isomerization is much more favourable than for any other amino acid due to its cyclic nature. In protein structures, 5% of X-proline peptide bonds are in a *cis* conformation compared to 0.3% of all other peptide bonds (122). With this in mind it is clear that proline peptide bond isomerization may be responsible for the observed conformational sampling in apelin-17. Schubert et al. (123) show that an accurate indicator of the proline N-terminal peptide bond isomerization state can be determined using:

$$\Delta^{13}C = \delta(C^{\beta})-\delta(C^{\gamma}) \quad 2.1$$

where $\delta(C^{\beta})$ and $\delta(C^{\gamma})$ are the chemical shifts of the indicated proline nuclei. Using this metric, the N-terminal peptide bonds of P14 and P16 are both in the *trans* form in conformer A ($\Delta^{13}C$ (P14) = 4.97 ppm, $\Delta^{13}C$ (P16) = 4.64 ppm) and in the *cis* form in conformer B ($\Delta^{13}C$ (P14) = 9.53 ppm, $\Delta^{13}C$ (P16) = 9.81 ppm).

2.3.2. Generation Of Apelin-17 Structural Ensembles At 5 °C And 35 °C

Sequential assignment (124) of apelin-17 resonances at 35 °C and 5 °C was carried out using natural abundance gradient-enhanced ^{1}H-^{13}C HSQC and homonuclear TOCSY, NOESY and DQF-COSY experiments. Based on NOE build-up measurements (124), a 400 ms mixing time was deemed optimal for the NOESY experiment on apelin-17. Apelin-12, -13, and pyr-apelin-13 were also sequentially assigned at both temperatures using ^{1}H-^{13}C HSQC, TOCSY and ROESY experiments. Nearly complete ^{1}H and ^{13}C resonance assignment was possible at both 35 °C and 5-6 °C for all forms of apelin (Table 2.3).

The shorter apelin peptides are in the molecular weight range where the NOE is minimal (124), meaning that NOESY experiments are not viable. ROESY spectra of these peptides are highly overlapped. When combined with convolution of the oppositely phased J-coupled vs. dipolar coupled peaks, ROESY assignment was greatly hindered. Hence, full structural analysis and calculation using NOESY assignments and $^{3}J_{HNH^{\alpha}}$ couplings was pursued only for apelin-17. NOE assignments for the major conformer (conformer A) and for the most populated minor conformer (conformer B) were made for

apelin-17 at 5 °C, but not at 35 °C due to a lack of observable NOE contacts arising from the lower-intensity resonances. Each conformer shows distinct sets of sequential and medium ($|i - j| \leq 4$) range NOE contacts over residues G13-F17 (26 for conformer A and 10 for conformer B; Figure 2.4), with K12 showing NOE contacts to G13 of both conformer A and B.

For apelin-17 at 35 °C and conformers A and B at 5 °C, Table 2.2 displays a summary of the NOE and DQF-COSY based $^{3}J_{HNH\alpha}$ coupling restraints used as well as average energies and violations of the final 80-member ensembles (80 lowest energy members from 200 calculated structures) of apelin at 5 °C and 35 °C. Notably, almost double the number of NOE contacts was observed at 5 °C vs. 35 °C (Table 2.2) with NOE contacts spread along the length of the peptide (Figure 2.4). Graphical summaries of NOE restraints (Figure 2.5) show connectivity over the length of the peptide at both 5 °C and 35 °C without the hallmark features of α-helix or β-sheet structure (124).

Chemical shifts for apelin-17 at both 35 °C and 5 °C (Figure 2.6) show no continuous deviation from random coil values (84). This is not surprising given the strong random coil of CD spectra presented in the literature (1, 32). Although $^{3}J_{HNH\alpha}$ coupling values were measured and used as restraints (with minimal violations in final structural ensembles, Table 2.2) at both 35 °C and 5 °C, most residues still displayed large average deviations for their dihedral angles (Figure 2.7). Very low dispersion was observed for the dihedrals of the proline residues, as would be anticipated based on the restriction in conformation imparted by the cyclized side-chain of proline (102). Several type I, II, VI, VIII and many type IV β-turns were detected by Promotif-NMR (Table 2.2, (117)). A

significant proportion of residues found in “disallowed” regions of the Ramachandran plot (Table 2.2) is consistent with regions having high type IV β-turn prevalence (125).

2.3.3. Some Sections Of Apelin-17 Are Structured

The apelin-17 ensembles did not demonstrate consistent structuring over the length of the peptide. Rather, a great deal of conformational sampling is apparent. Building on analysis methods detailed previously (104, 126), superposition of each structural ensemble was carried out over 3-6 residue regions starting on each amino acid of apelin-17. Backbone RMSD values for each superposed region were compared to those of a large (1000 member) apelin-17 ensemble calculated without any NOE or ^{3}J-coupling restraints to determine those segments with RMSDs that are significantly better than the unrestrained ensemble (*i.e.*, RMSD of the unrestrained ensemble higher than the RMSD + average deviation of NMR ensemble). This analysis demonstrates clear regions of structural convergence in each apelin-17 ensemble (Figure 2.8, K1-R4 at 35 °C, R6-L9 in all ensembles and G13-F17 or P14-F17 at 5 °C). Note that in some cases, a 4 residue superposition neighbouring well-converged regions (F2-Q5 neighbouring K1-R4 at 35 °C and Q5-R8 neighbouring R6-L9 for conformer A at 5 °C) have lower RMSD than the 1000 member ensemble. However, extension of these two superpositions to 5 residues showed poor structural convergence. I therefore treat only the lower of a pair of neighbouring RMSDs as a truly converged segment. The neighbouring superposition with decreased RMSD is driven to this value through its proximity to the better converged structure. Note that although conformer B at 5 °C does not display a well converged structure over residues K1-R4, unlike conformer A at 5 °C, the error in RMSD for conformers A and B overlap. This implies that the ensemble of structures for

conformer B is just marginally less converged than in conformer A in this region and is likely due to slight differences in the iterative NOE refinement of these two ensembles or the choice of reference structure for RMSD calculation.

Increased apelin-17 structuring upon temperature decrease, as indicated by more NOE contacts and lower RMSD values, was somewhat unexpected in light of NMR chemical shift comparisons. In particular, NMR chemical shifts are minimally different between 35 °C and 5 °C (Figure 2.6), indicative of minimal secondary structuring change. Note that a well-defined structuring of the C-terminus at 5 °C would also not normally be anticipated, since polypeptide termini are typically less structured than the remainder of the polypeptide (124). The two proline residues help structure this region through reduced conformational freedom due to their cyclic nature (102). Further evidence for the convergence of F17 comes from CD spectropolarimetry studies completed in the Rainey lab where at low temperature positive bands appear at ~194 nm which are attributed to side chain transitions of F17. These bands are much more apparent at 5 °C where F17 is presumably structurally constrained (1).

There is an increased proportion of PP_{II} conformation in the structural ensembles at 5 °C vs. 35 °C (Table 2.2 and Figure 2.9 (B)). PP_{II} occurrence in all three apelin-17 ensembles is highest at R4, R6-L9, H11 and P14-M15, while S10, which is not converged, shows the least evidence of PP_{II} (Figures 2.9 and 2.10). Imposing PP_{II} dihedral restraints (using $-110° \leq \phi \leq -40°$ and $130° \leq \psi \leq 180°$, (127)) on a residue-by-residue basis did not reduce mean ensemble energy, suggesting that the PP_{II} content of these regions is persistent but not fixed. Note that the PP_{II} restraints placed upon residues during simulated annealing were stricter than the bioinformatically observed PP_{II} region

of Best et al. (116) used to detect the presence of PP_{II} in the ensemble of structures (Figure 2.9 (B)). A slight increase in PP_{II} content of apelin-17 at 5 °C has also been observed through CD spectropolarimetry in the Rainey lab. Although this band is convoluted by side chain transitions from F17, in a F17A apelin analogue there is still a slight PP_{II} appearance to the CD spectra at 5 °C (1).

β-turns (mostly Type IV) were frequently observed in the final ensembles (Table 2.2 and Figure 2.9 (A)). In comparison to β-turn occurrence in the unrestrained ensemble of 1000 apelin-17 structures, the prevalence of β-turns is highest over the C-terminally converged region and elevated over R6-R8 (shown schematically in Figure 2.11 and detailed in Figure 2.9 (B)). Therefore, both PP_{II} and β-turns are prevalent in the most structured regions of apelin-17. Since PP_{II} has been proposed as an important motif for intermolecular interactions (103) and β-turns have been proposed to be a common motif recognized by GPCRs (101), the two structurally converged regions of apelin-17 at 5 °C are likely candidates for specific binding to AR.

2.3.4. The Structured Regions Of Apelin-17 Are Clustered Into Several Conformations

Examination of the structurally converged regions (K1-R4, R6-L9 for apelin-17 at 35 °C; and R6-L9 and G13-F17 for 5 °C data) demonstrated regions of preferred backbone conformation in each converged region within each ensemble. Therefore, cluster analysis (119) was performed on these regions (Figure 2.12). In order to consider a clustering result significant for further analysis, the cutoff was imposed that the final clusters must each contain >10% of the ensemble members. According to this cutoff, the G13-F17

region in both conformers A and B of apelin-17 at 5 °C was clustered (Figures 2.10 (B,C) and 2.13). No other structurally converged region contained significant clustering. Details of the observed clustering and cluster properties are provided in Table 2.4, with side chain RMSD values of the heavy side chain atoms of these clusters in Figure 2.14.

2.3.5. The Smaller Apelin Isoforms Have A Similar Structure To Apelin-17

The striking similarity of the H^{α}, C^{α} and C^{β} chemical shifts of the shorter apelin peptides to those of apelin-17 (Figures 2.15 and 2.16) suggests that the structure observed in apelin-17 (Figure 2.10) is retained in the smaller apelins. This is not surprising given that all forms of apelin bind to and activate AR, but is the first experimental evidence that the functionally critical apelin-12 region has a defined structure retained between apelin isoforms.

2.3.6. Structure-Function Correlation Of Apelin

The residues in apelin (discussed here with apelin-17 numbering) that are critical for binding and activation of AR are highlighted in Figure 2.11. These structural results provide a new context for the previous mutagenesis studies of apelin. Residues in the structurally defined segments at 5 °C (Figure 2.10 (B,C)) correspond very well to the functionally critical regions identified in the literature (Figure 2.11, (32, 33, 34, 98, 99)). The NMR data supports the idea that R6-L9 are functionally important because of the specific backbone structure they adopt at both 5 °C and 35 °C. Fan et al. found S10 and K12, found here to be in the flexible linker between structured regions (Figure 2.10), to be functionally important but not critical (32). These residues may facilitate interaction

between apelin and AR, without being essential for binding or activation. A specific structuring in this region, not observed here, may also take place once the receptor-ligand complex is formed. Finally, the importance of G13, P14, M15 and F17 is echoed in the well-converged structure over these residues at 5 °C (Figures 2.8 and 2.10 and clustering details in Table 2.4 and Figures 2.13 and 2.14). The evident proline *cis-trans* isomerization in this region, introduces the new issue of which conformer of apelin is required for binding to or activation of AR.

P16 is consistently structured in apelin-17 at 5 °C, but was found to be nonessential by Fan et al. (32), the only study where a P16A substitution has been made. It is unclear how perturbing to structure mutating P16 to alanine would be. It is notable, however, that contrary to other studies (33, 34, 99), Fan et al. also did not find the C-terminal phenylalanine to be essential (32). This may represent a difference in the mechanism of AR response to apelin in the human AR HEK293 transfected cells produced and tested by Fan et al. (32) compared to those of Medhurst et al. (33) or the rat AR transfected CHO cells used by De Mota et al. (99). In a recent study using CHO cells by Iturrioz et al., F17 was found to be essential for internalization of AR but not signalling (128). This highlights how mutagenesis data can vary depending on the system and functional readout chosen, likely due to biased signaling of AR in the different contexts. Since the C-terminal phenylalanine was not susceptible to substitution in the study of Fan et al. but was found to be essential in other physiological assays, a P16A mutant tested differently may indeed perturb apelin-AR binding or activation.

2.3.7. Insights Into Apelin-AR Interactions

A hypothetical mode of interaction between apelin and AR is suggested by the differential structuring of apelin-17 at 35 °C and 5 °C. At 35 °C, apelin may be undergoing conformational selection, similar to that observed for ubiquitin (107), facilitating specific binding to AR. AR binding may be initiated in the R6-L9 region of apelin, which is well structured at both 35 °C and 5 °C. All four residues in this region (RPRL) are highly sensitive to alanine substitution. Also the N-terminal region of AR identified as critical using both truncation mutants and alanine scanning mutagenesis by Zhou et al. (100) is quite acidic. Therefore, the cationic R6 and R8 residues, which generally fall on the same face of apelin-17 in the structural ensembles, provide a likely binding target for the N-terminal region of AR. The L9 side chain, which is solvent exposed in every ensemble member, may act to strengthen the binding of apelin and AR through a hydrophobic interaction. Furthermore, the reduced entropy of the R6-L9 region in the free peptide arising from its increased structuring relative to the remainder of apelin, even at 35 °C, is likely to provide a decreased entropic penalty for binding of apelin to AR vs. a binding interaction occurring elsewhere in the peptide.

Although these results cannot confirm or deny this mode of binding, the work of Sykes and coworkers (108, 111) demonstrates that it is a reasonable hypothesis that the structuring in the R6-L9 region and at the C-terminus of apelin-17 at 5 °C is representative of the apelin-AR bound state. Initial binding of the R6-L9 region of apelin to AR would bring the C-terminal region of apelin into proximity with AR, facilitating its binding, which fits well with a two-step model of GPCR-ligand binding. Determination of the active conformation of apelin in this C-terminal region, including the *cis-trans*

isomer forms of P14 and P16, will be critical for understanding apelin function and for developing specific targets for therapeutic design. The potential for a two-step interaction between apelin and AR provides a key structural starting point for detailed structure-function dissection of the apelin-AR system.

2.4. Summary

All five bioactive forms of apelin examined exhibit almost identical chemical shifts at physiological (35 °C) and low temperatures (5-6 °C), indicating similarity in structure between these five forms of apelin. The apelin-17 structures should, therefore, be representative of the other, shorter apelin isoforms. The R6-L9 region of apelin-17 exhibits nascent structuring at both 35 °C and 5 °C. Residues G13-F17 show structuring at 5 °C, with clustered backbone conformations. The apelin peptides all show strong evidence of *cis-trans* isomerization of the proline residues in the PMPF motif at the C-terminus, with predominant conformers having either both *trans* (~80%) or both *cis* (~20%) peptide bonds. Both the R6-L9 and G13-F17 structured regions correlate very well to functionally critical regions from previous alanine substitution studies and are likely important for binding to AR.

Table 2.1: Detailed experimental parameters used for NMR data acquisition of the indicated apelin peptide and temperature.

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	^{1}H frequency (MHz)	Facility
Apelin-12 6 °C							
^{1}H-^{1}H TOCSY[a]	1	16	^{1}H:4096 ^{1}H:256	^{1}H:11990 ^{1}H:8000	^{1}H:4.95 ^{1}H:4.95	800	NANUC
^{1}H-^{1}H ROESY[a]	1.5	32	^{1}H:2048 ^{1}H:512	^{1}H:11990 ^{1}H:11990	^{1}H:4.95 ^{1}H:4.95	800	NANUC
^{1}H-^{13}C HSQC	1	96	^{1}H:1534 ^{13}C:128	^{1}H:11990 ^{13}C:28160	^{1}H:4.95 ^{13}C:35.26	800	NANUC
Apelin-12 35 °C							
^{1}H-^{1}H TOCSY[b]	2	16	^{1}H:6656 ^{1}H:272	^{1}H:5482 ^{1}H:5482	^{1}H:4.68 ^{1}H:4.68	500	ARMRC
^{1}H-^{1}H ROESY[a]	1.5	32	^{1}H:2048 ^{1}H:512	^{1}H:11990 ^{1}H:11990	^{1}H:4.68 ^{1}H:4.68	800	NANUC
^{1}H-^{13}C HSQC	1.5	64	^{1}H:1024 ^{13}C:164	^{1}H:5482 ^{13}C:16340	^{1}H:4.68 ^{13}C:62.63	500	ARMRC
Apelin-13 5 °C							
^{1}H-^{1}H TOCSY[c]	2	80	^{1}H:44032 ^{1}H:126	^{1}H:5482 ^{1}H:5482	^{1}H:4.96 ^{1}H:4.96	500	ARMRC
^{1}H-^{1}H NOESY[c]	2	128	^{1}H:4096 ^{1}H:144	^{1}H:5482 ^{1}H:5482	^{1}H:4.96 ^{1}H:4.96	500	ARMRC
^{1}H-^{13}C HSQC	1.5	48	^{1}H:1024 ^{13}C:192	^{1}H:5482 ^{13}C:16340	^{1}H:4.96 ^{13}C:62.65	500	ARMRC
Pyr-Apelin-13 6 °C							
^{1}H-^{1}H TOCSY[a]	1	16	^{1}H:4096 ^{1}H:512	^{1}H:11990 ^{1}H:8000	^{1}H:4.95 ^{1}H:4.95	800	NANUC
^{1}H-^{1}H ROESY[a]	1.5	32	^{1}H:2048 ^{1}H:512	^{1}H:11990 ^{1}H:11990	^{1}H:4.95 ^{1}H:4.95	800	NANUC
^{1}H-^{13}C HSQC	1	96	^{1}H:1534 ^{13}C:256	^{1}H:11990 ^{13}C:28160	^{1}H:4.95 ^{13}C:35.26	800	NANUC
Pyr-Apelin-13 35 °C							
^{1}H-^{1}H TOCSY[c]	2.5	40	^{1}H:2876 ^{1}H:320	^{1}H:7201 ^{1}H:5993	^{1}H:4.68 ^{1}H:4.68	600	NANUC
^{1}H-^{1}H ROESY[a]	1.5	32	^{1}H:2048 ^{1}H:512	^{1}H:11990 ^{1}H:11990	^{1}H:4.68 ^{1}H:4.68	800	NANUC
^{1}H-^{13}C HSQC	1.2	64	^{1}H:878 ^{13}C:392	^{1}H:7201 ^{13}C:21102	^{1}H:4.68 ^{13}C:34.99	600	NANUC

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	^{1}H frequency (MHz)	Facility
Apelin-17 5 °C							
^{1}H-^{1}H TOCSY[a]	2	32	^{1}H:4096 ^{1}H:512	^{1}H:11990 ^{1}H:7998	^{1}H:4.96 ^{1}H:4.96	800	NANUC
^{1}H-^{1}H COSY[a]	2	16	^{1}H:4096 ^{1}H:1024	^{1}H:11990 ^{1}H:8000	^{1}H:4.96 ^{1}H:4.96	800	NANUC
^{1}H-^{1}H NOESY[c]	3.5	32	^{1}H:3520 ^{1}H:384	^{1}H:8805 ^{1}H:7998	^{1}H:4.96 ^{1}H:4.96	800	NANUC
^{1}H-^{13}C HSQC	1	224	^{1}H:2048 ^{13}C:192	^{1}H:11990 ^{13}C:28160	^{1}H:4.96 ^{13}C:35.18	800	NANUC
Apelin-17 35 °C							
^{1}H-^{1}H TOCSY[b]	2	72	^{1}H:6656 ^{1}H:384	^{1}H:5482 ^{1}H:5480	^{1}H:4.68 ^{1}H:4.68	500	ARMRC
^{1}H-^{1}H COSY[b]	3	64	^{1}H:22016 ^{1}H:196	^{1}H:5482 ^{1}H:5001	^{1}H:4.68 ^{1}H:4.68	500	ARMRC
^{1}H-^{1}H NOESY[c]	3.5	48	^{1}H:3362 ^{1}H:392	^{1}H:8403 ^{1}H:7998	^{1}H:4.68 ^{1}H:4.68	800	NANUC
^{1}H-^{13}C HSQC	1.5	64	^{1}H:1024 ^{13}C:192	^{1}H:5482 ^{13}C:16340	^{1}H:4.68 ^{13}C:125.7	500	ARMRC

[a] Water suppression was achieved using WATERGATE 3-9-19
[b] Water suppression was achieved using excitation sculpting
[c] Water suppression was achieved using presaturation

Table 2.2: Summary of iterative structure calculation protocols, restraints and structural statistics of the lowest energy 80 structures (out of 200) for each of the three apelin-17 ensembles produced.

	35 °C	**Conformer A**	**Conformer B**
Structure calculation[a]			
Rounds of NOE refinement prior to ^{3}J-coupling incorporation	3	6	7
Total rounds of NOE refinement	12	37	32
Unique NOE restraints			
Total	285	562	529
Intra-residue	146	224	226
Sequential	108	222	194
Medium Range ($\lvert i-j \rvert \leq 4$).	22	62	54
Long Range ($\lvert i-j \rvert > 4$).	0	0	0
Ambiguous	9	54	55
^{3}J-coupling restraints			
Number of restraints	10	12	12
Ramachandran plot statistics			
Core	226	171	150
Allowed	457	431	395
Generously allowed	86	166	196
Disallowed	111	112	139
In PPII conformation (116)	280/1500	345/1500	323/1500
Number of type I β-turns	9	0	1
Number of type II β-turns	2	9	2
Number of type IV β-turns	123	96	193
Number of type VIb β-turns	0	0	3
Number of type VIa β-turns	0	0	5
Number of type VIII β-turns	10	0	15
XPLOR-NIH energies (kcal/mol)[b]			
Total	45.14 ±5.39	53.45±5.34	61.74±6.40
NOE	0.60±0.49	0.67±0.59	0.61±0.62
J Coupling	2.55±1.30	1.83±0.88	2.71±1.22
Violations			
NOE Violations > 0.5 Å	0	0	0
NOE Violations of 0.3-0.5 Å	0	3	0
NOE Violations of 0.2-0.3 Å	4	3	7
^{3}J-Coupling Violations	11	4	6

[a] Iterative NOE contact refinement was carried out as described previously (129) through calculation of the given total number of 100-member ensembles where the initial rounds were performed without ^{3}J-couplings incorporated.

[b] Ranges are given by average deviations for XPLOR-NIH energies.

Table 2.3: Summary of ^{1}H and ^{13}C NMR assignments that were either not observable or that were ambiguously assigned for each apelin isoform (sequences in Table 1.1).

Dataset	**Missing Resonance Assignments**
Apelin-12 6 °C	R1:H^{N},H^{η}
Apelin-12 35 °C	R1:H^{N},H^{η} R3:H^{η} K7:H^{ζ}
Apelin-13 5 °C	Q1:H^{N} R2:H^{ε}, H^{η}* R4:H^{ε}, H^{η}* K8:H^{ζ}
Apelin-13 35 °C	Q1:H^{N} R2:C^{α}, H^{η} R4:H^{η} H7:H^{α},C^{α} K8:H^{ζ} M11:C^{α}
Pyr-Apelin-13 6 °C	Q1:H^{N} R4:H^{η}
Pyr-Apelin-13 35 °C	Q1:H^{N} R2:H^{η} R4:H^{η} H7:H^{α}, C^{α} K8:H^{ζ}
Apelin-17 5 °C conformer A	K1:H^{N},H^{ζ} F2:H^{ε}, C^{δ}, C^{ε} R3:H^{η}* R4:H^{η}* R6:H^{η}* R8:H^{η}* F17:H^{ε}, C^{ε}
Apelin-17 5 °C conformer B	K1:H^{N},H^{ζ} F2:H^{ε}, C^{δ}, C^{ε} R3:H^{η}* R4:H^{η}* R6:H^{η}* R8:H^{η}* P14:C^{δ} F17:H^{ε}, C^{ε}, H^{ζ}, C^{ζ}
Apelin-17 35 °C	K1:H^{N}, H^{ζ} R3:H^{η} R4:H^{η} R6:H^{η} R8:H^{η} K12:H^{ζ}

*Ambiguous assignment (chemical shift assigned, but not uniquely) made for this resonance

Table 2.4: Statistics of the clusters observed in the C-terminal region of the indicated apelin-17 ensemble of structures.

	Apelin-17 5 °C Conformer A		**Apelin-17 5 °C Conformer B**	
Cluster size and consistency				
Residues cluster is defined over	G13-F17		G13-F17	
Number of structures in R1 cluster	33	47	17	61
Number of consistent structures	32	47	17	60
Number of inconsistent structures[a]	1	0	0	1
Average RMSD of cluster[b]				
Before clustering using R1	1.65		1.64	
Calculated using R1	1.12	1.18	0.92	1.27
Calculated using R2	0.76	0.84	0.63	0.85
Calculated using R4	1.43	1.45	1.12	1.43
Ramachandran plot statistics				
Core	38	81	23	123
Allowed	198	226	68	314
Generously allowed	85	100	81	137
Disallowed	42	110	15	117

[a] Structure is inconsistent if it is not included into the same cluster by all three clustering methods of Clusterpose (R1,R2 and R4).

[b] RMSD values as calculated only over those residues which the cluster is defined to cover using the R1, R2 and R3 methods of Clusterpose.

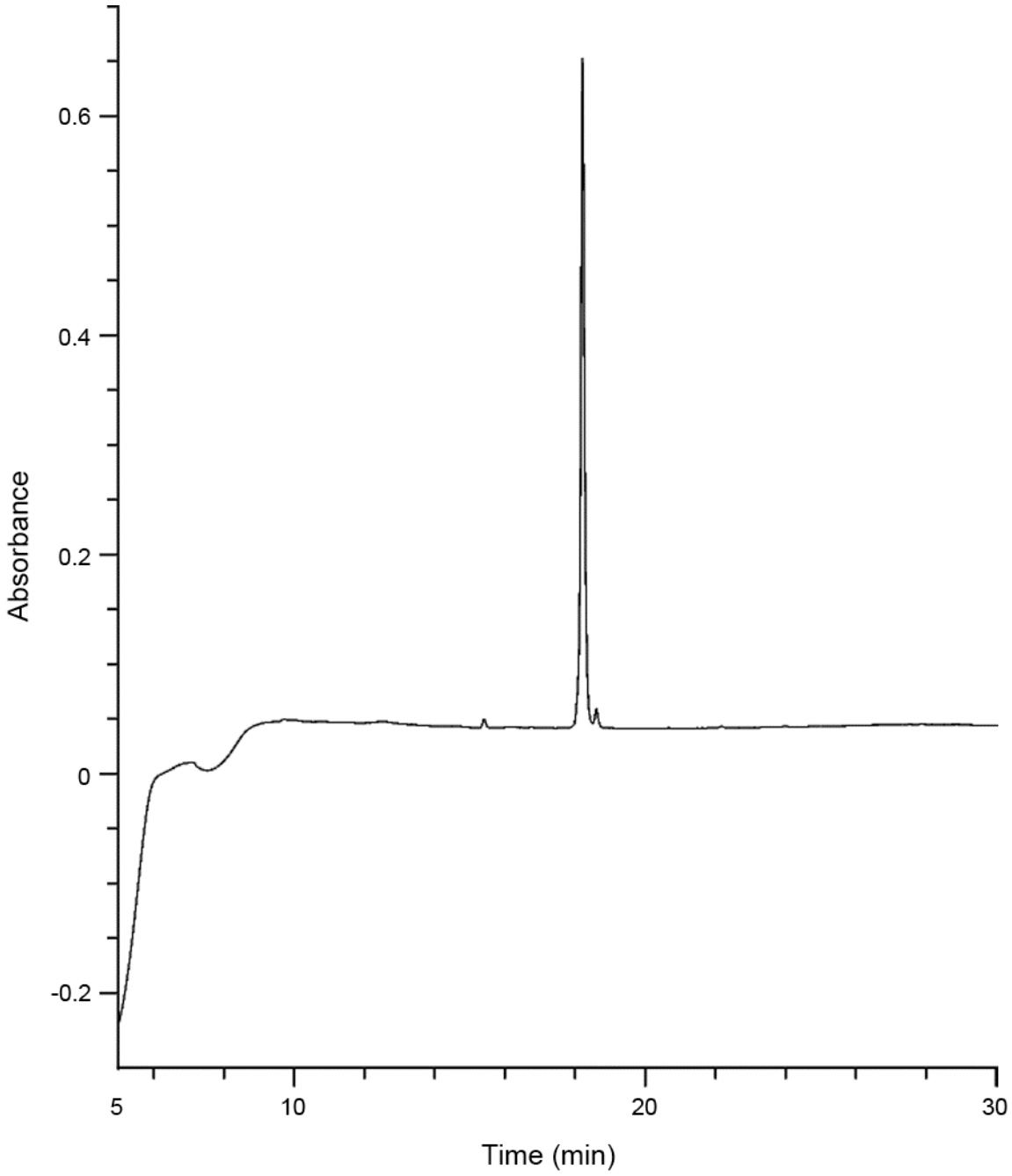

Figure 2.1: Reversed-phase HPLC analysis of purified apelin-13. HPLC was carried out using a C_{18} column (5µm particle, 4.6 mm × 250 mm Spirit Peptide column, AAPPTec) with a water:acetonitrile solvent mixture progressing from 2% to 60% B over 29 min. The major peak corresponds to apelin-13 and elutes at ~18 min.

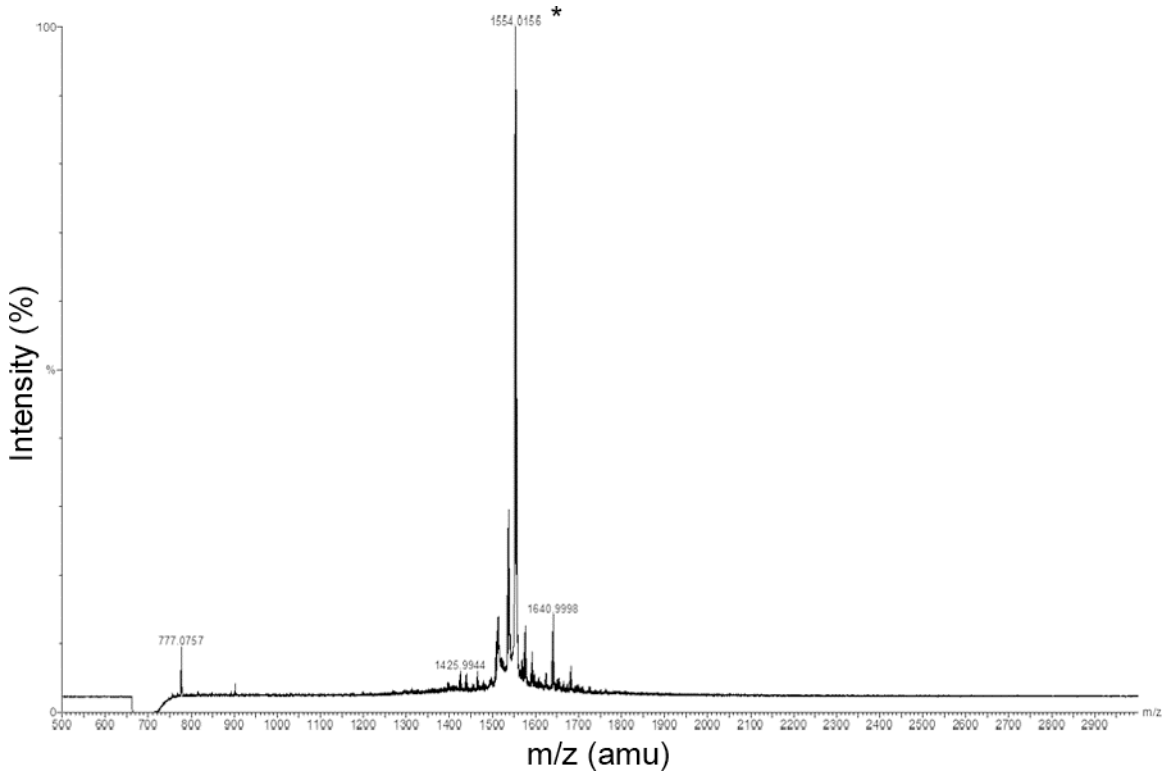

Figure 2.2: Positive mode electrospray ionization mass spectrometry analysis of the sample shown in Figure 2.1. The peak denoted by * corresponds to apelin-13 (m/z = 1554 amu, expected mass is 1555 amu).

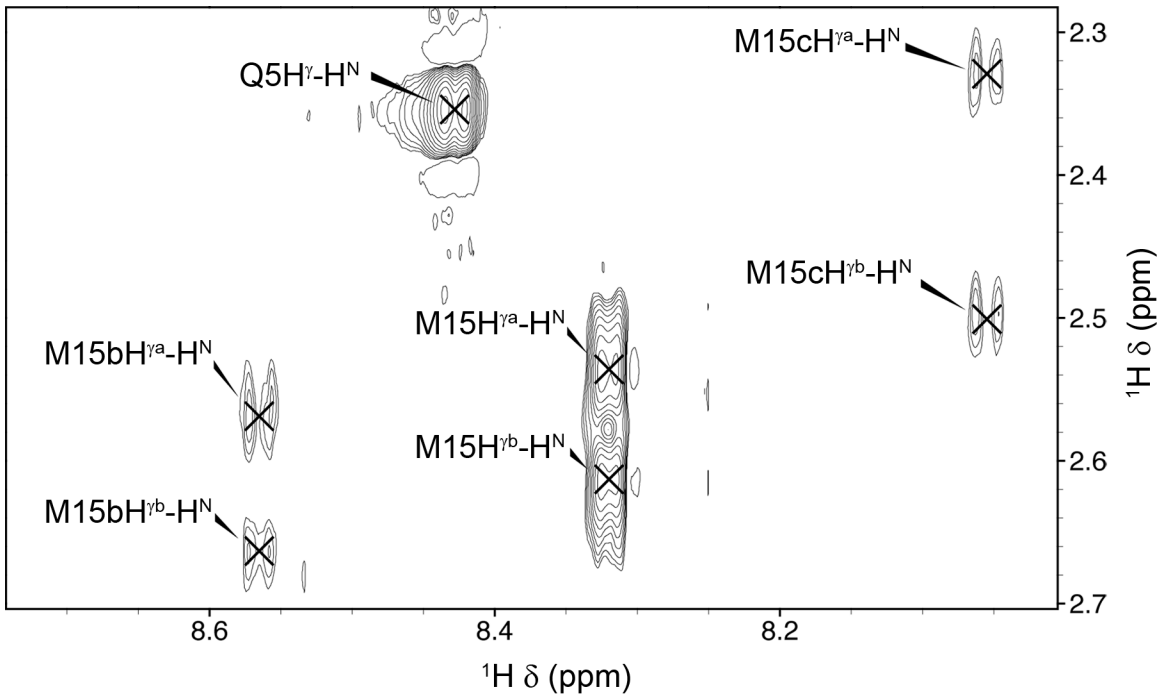

Figure 2.3: The H^{γ}-H^N connectivity region of a ^1H–^1H TOCSY spectrum (60 ms mixing time, 800 MHz field strength) acquired for apelin-17 at 35 °C. Multiple spin systems of M15 are denoted by M15b and M15c.

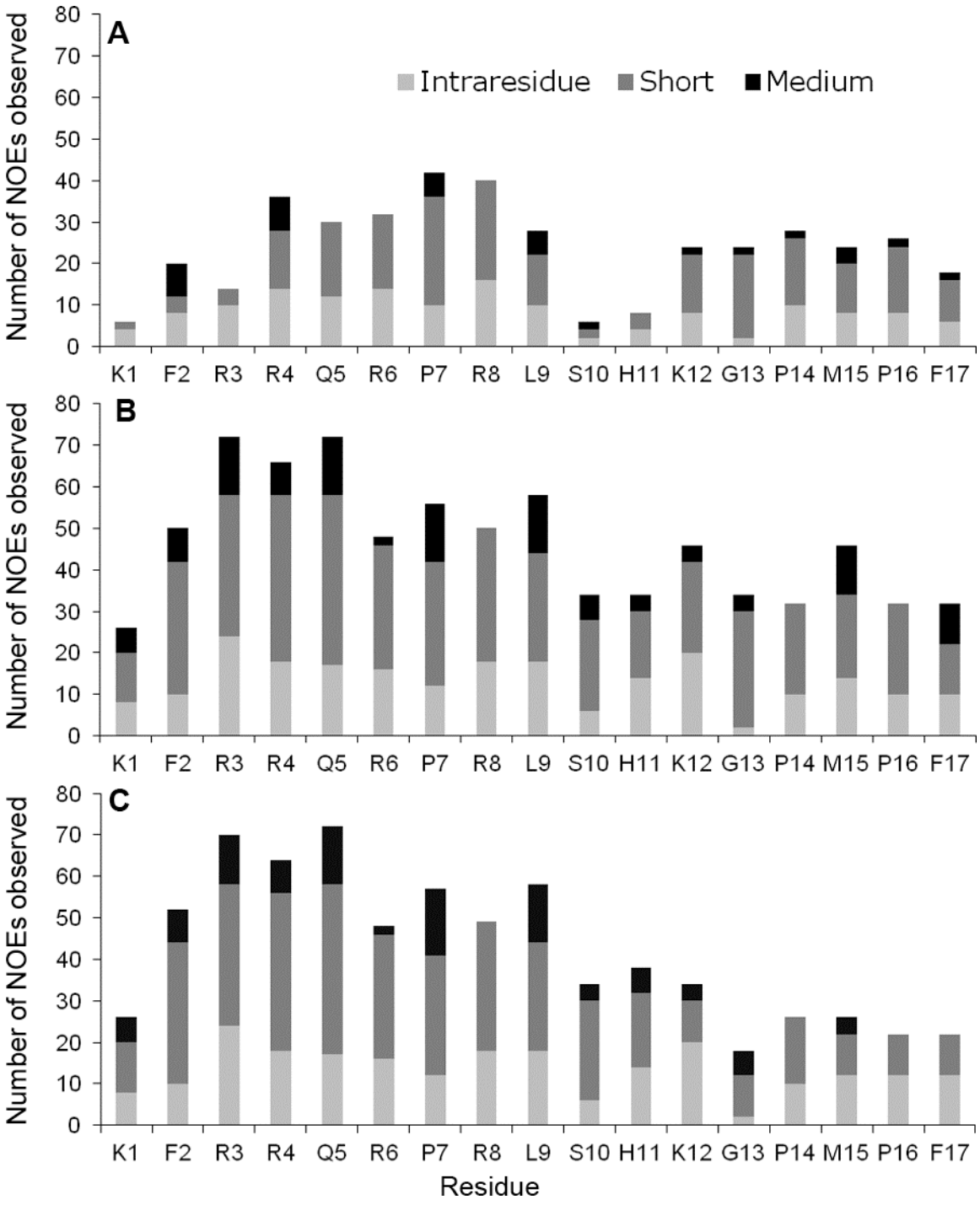

Figure 2.4: A breakdown of NOE contacts per amino acid residue used to calculate the final ensembles of structures, with short (sequential) and medium ($|i-j| \leq 4$) NOE interactions shown for (A) apelin at 35 °C and for conformers A and B at 5 °C (shown in (B) and (C), respectively). Note that these plots do not include the ambiguous NOE contacts shown in Table 2.2.

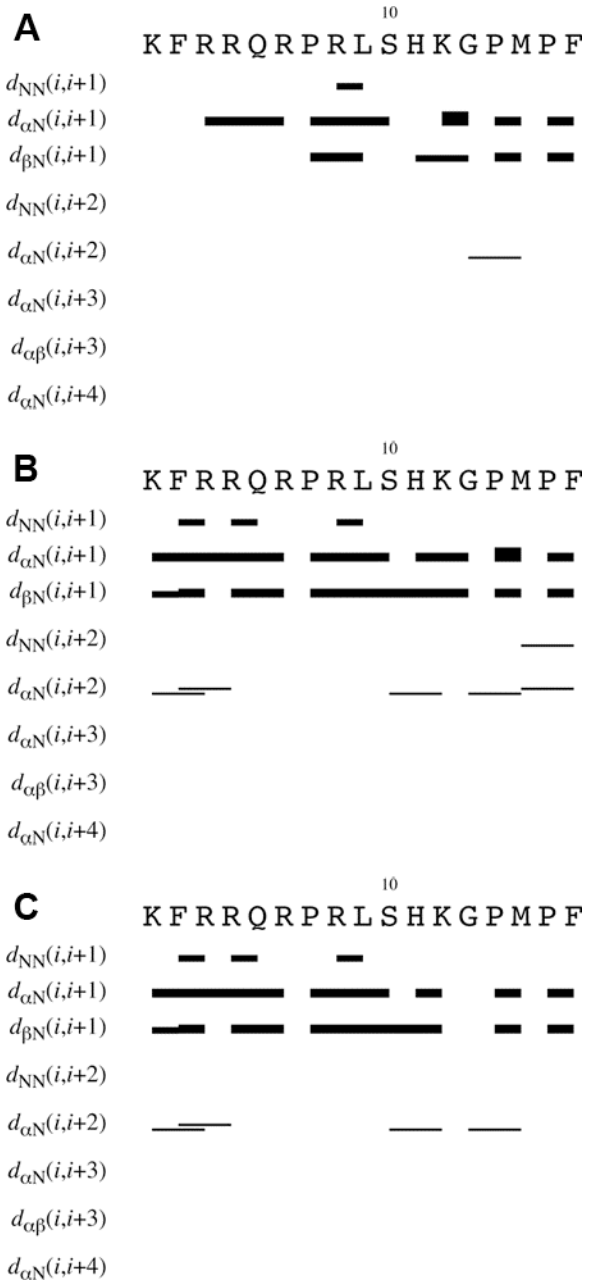

Figure 2.5: Graphical summaries of NOE contacts observed between the pairs of protons identified in subscripts and in brackets, as described in (124), for (A) apelin-17 at 35 °C and for conformers A and B of apelin-17 at 5 °C ((B) and (C) respectively). Note that these plots do not include ambiguous restraints and only represent a small subset of the total unique NOE contacts assigned and used for structure calculation (Table 2.2). Plots were produced using CYANA (L.A. Systems Inc.).

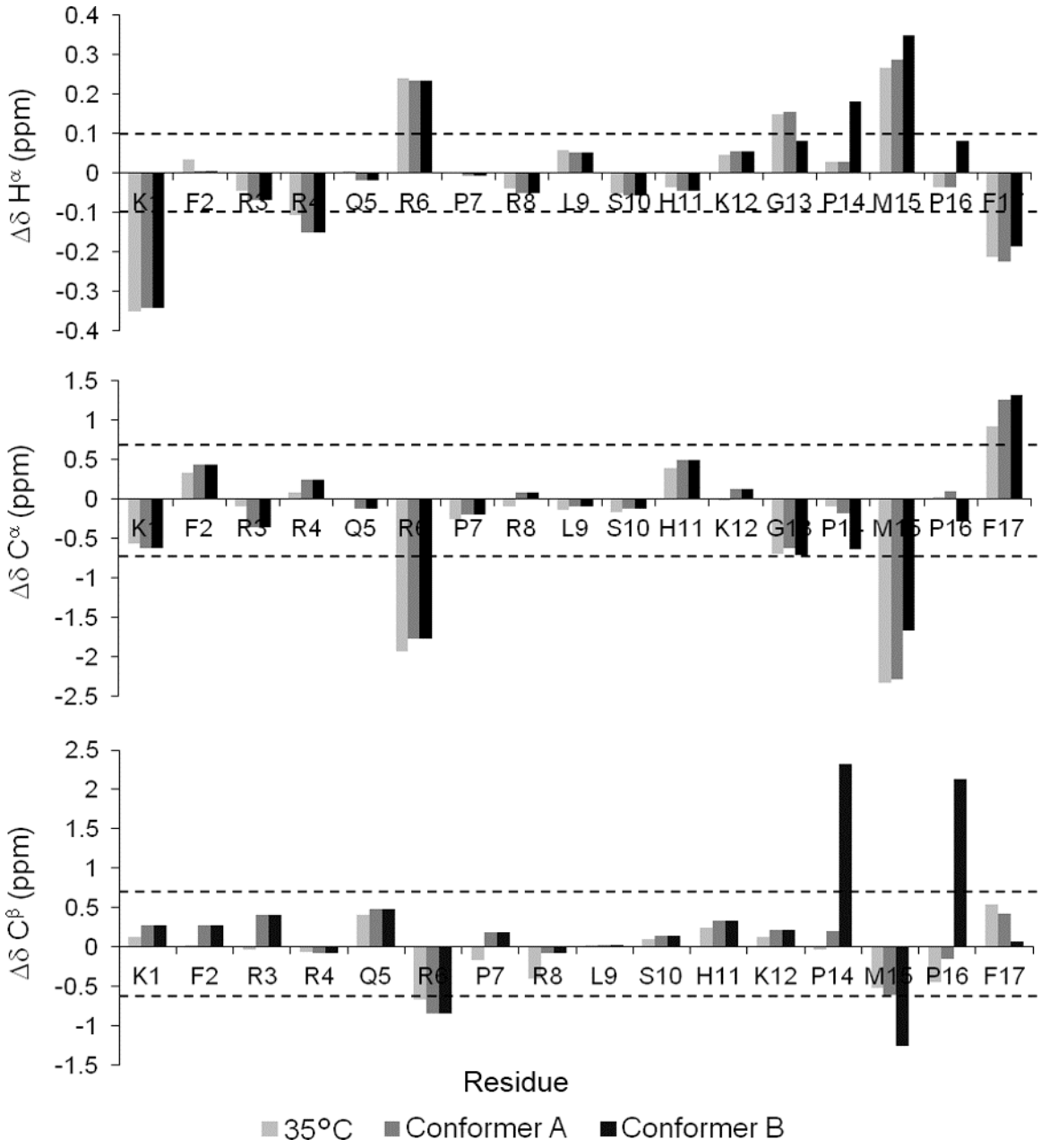

Figure 2.6: Secondary chemical shifts (Δδ=δ(observed)-δ(random-coil) using random-coil values of Wishart et al. (84) for apelin-17 at 35 °C and at 5 °C in conformers A and B for H^{α}, C^{α} and C^{β}, as indicated. The horizontal lines show the values of Δδ identified as significant for secondary structuring by the ^{1}H and ^{13}C chemical shift indices (refs (106) and (130) respectively). The C^{α} and C^{β} resonances for proline have identified ranges of ±4 ppm (not shown), unlike the other residues.

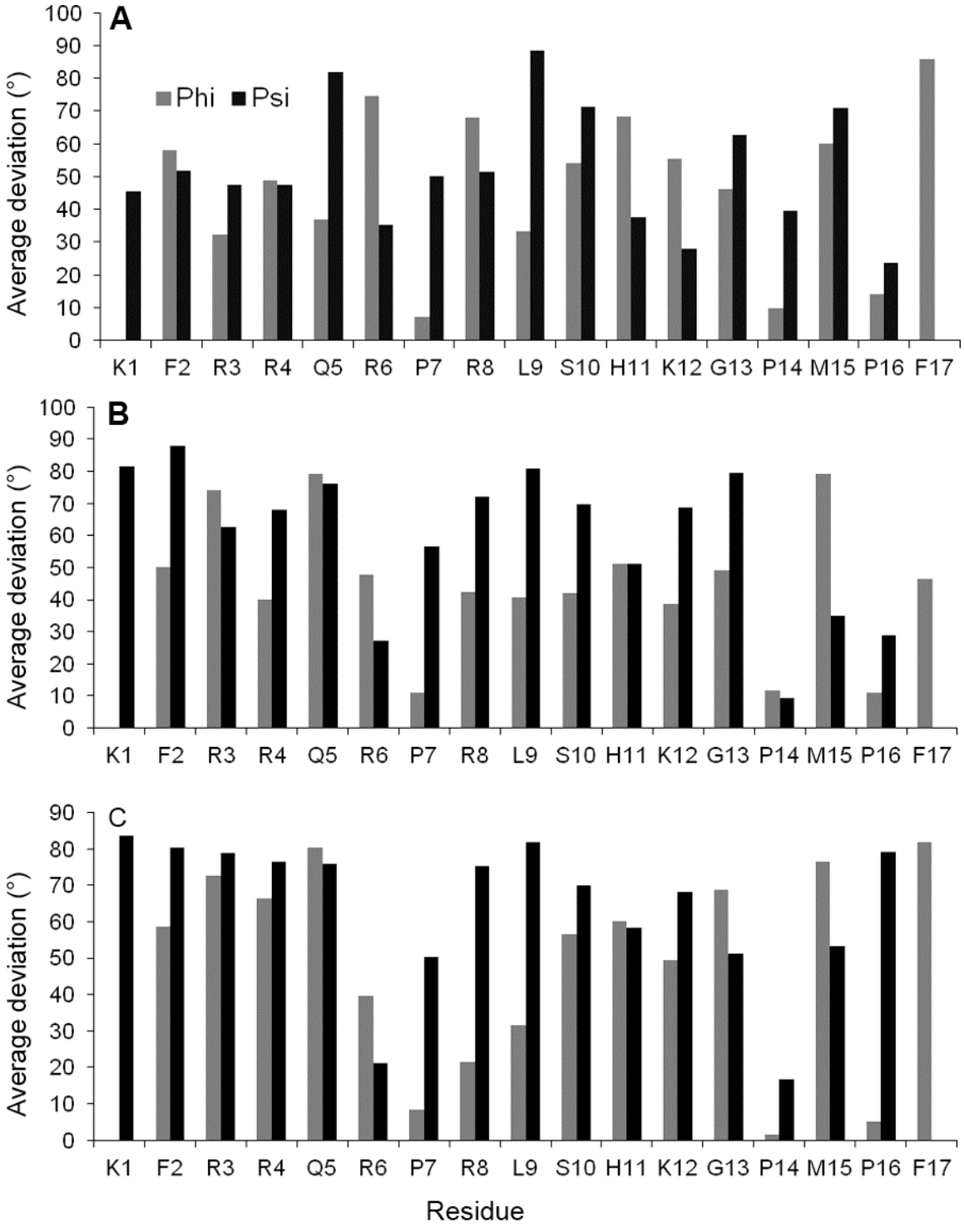

Figure 2.7: Average deviations of the dihedral angles ϕ and ψ on a per-residue basis for the final ensembles of 80 NMR structures calculated for (A) apelin-17 at 35 °C and for conformers A and B at of apelin-17 at 5 °C ((B) and (C) respectively).

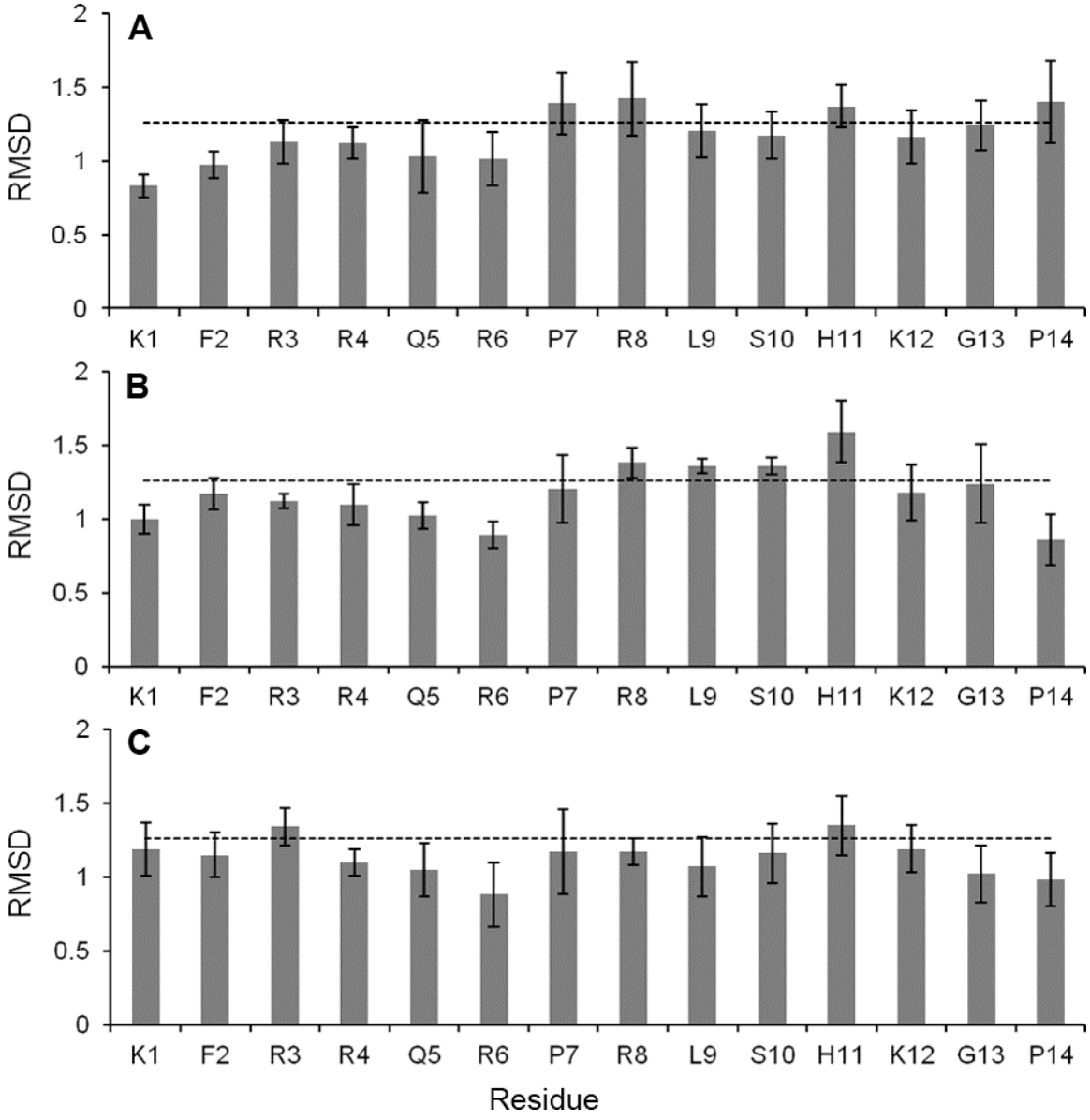

Figure 2.8: RMSD values calculated iteratively for superpositions over 4 residue segments for indicated 80-member NMR-based structural ensembles of (A) apelin-17 at 35 °C, (B) of the primary apelin-17 conformer (conformer A; *trans* peptide bonds at P14 and P16) at 5 °C and (C) of the second most populated apelin-17 conformer (conformer B; *cis* peptide bonds at P14 and P16) at 5 °C. Bars correspond to a superposition starting at the indicated residue with the average deviation for that superposition indicated by the error bar. The average RMSD value of an unconstrained ensemble of 1000 calculated apelin-17 structures superposed over the 4 residue segments is represented by the dashed line.

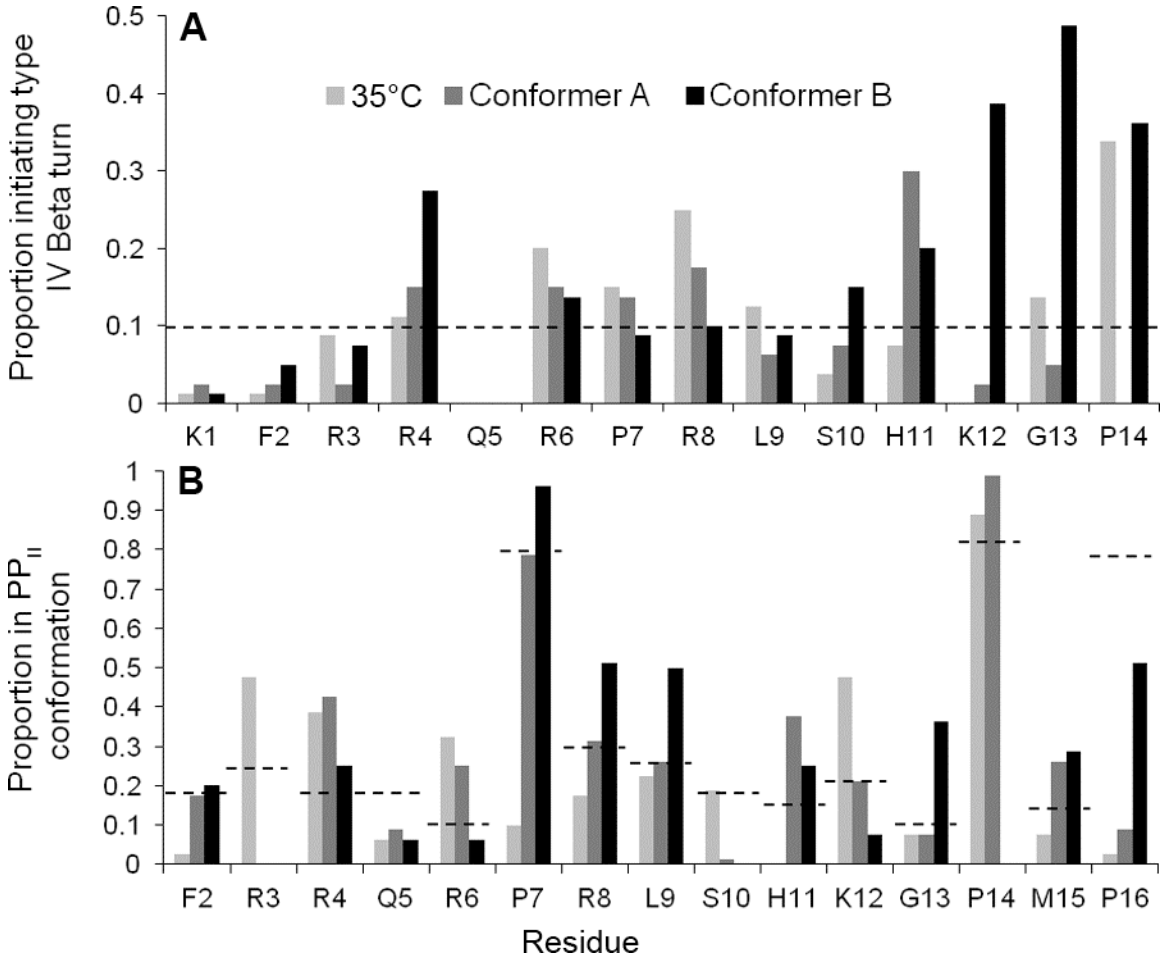

Figure 2.9: (A) Proportion of ensemble members out of the 80-member ensembles with type IV β-turns initiated at the given residue as determined by Promotif-NMR (117). Data is shown for all 3 ensembles of apelin-17 structures (35 °C and conformers A & B at 5 °C). The horizontal line represents the average number of type-IV β-turns found in an ensemble of 1000 structures for an unconstrained XPLOR-NIH (89) simulated annealing run performed under otherwise identical conditions. (B) Proportion of ensemble members found in PP_{II} conformation at given residue. Horizontal lines at each residue show the proportion of that residue adopting a PP_{II} conformation in the 1000-member unrestrained ensemble.

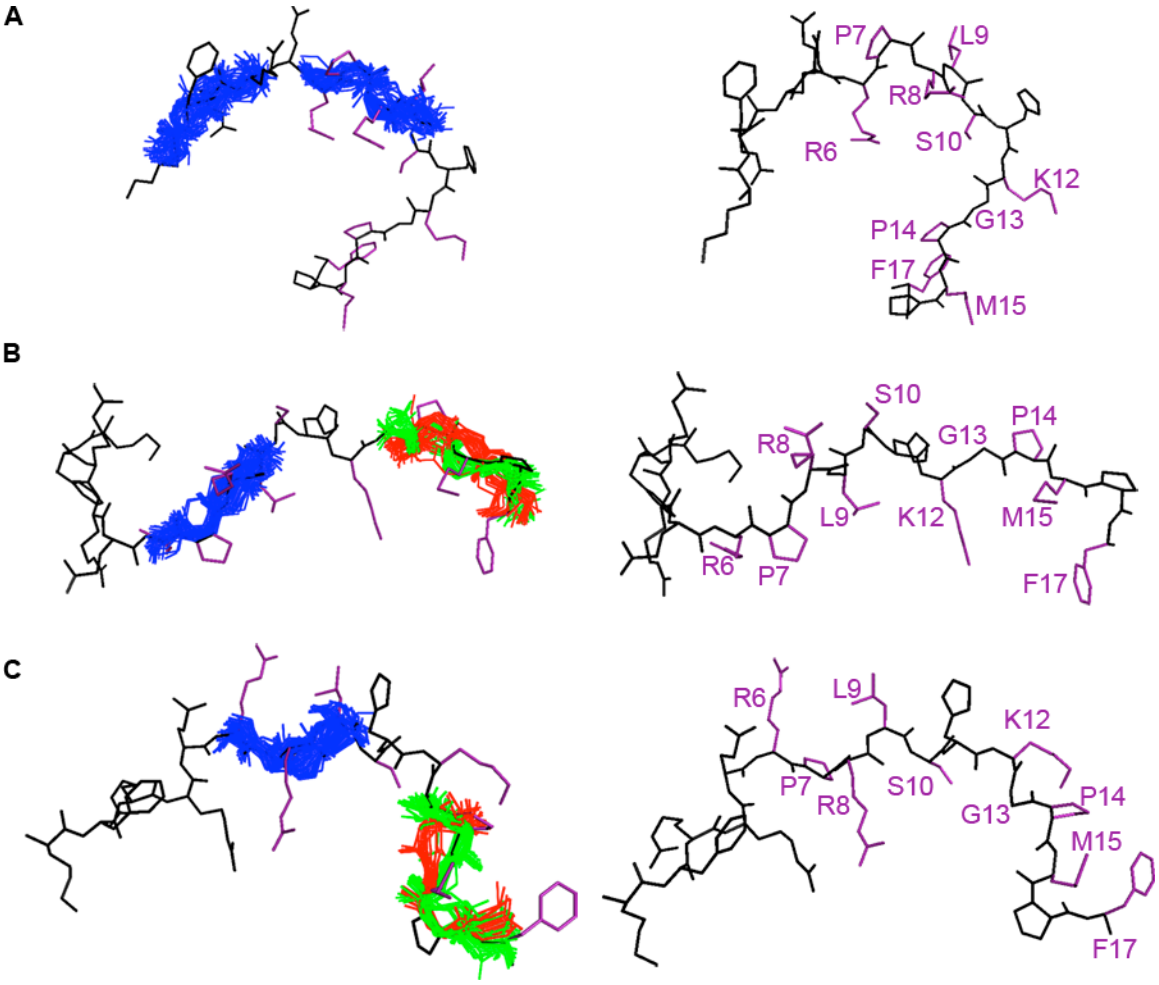

Figure 2.10: Representative structures for apelin-17 structural ensembles (A) at 35 °C, (B) of the primary conformer (conformer A; *trans* peptide bonds at P14 and P16) at 5 °C and (C) of the second most populated conformer (conformer B; *cis* peptide bonds at P14 and P16) at 5 °C. On the left hand side converged regions are superposed, while on the right hand side the lowest energy ensemble member is shown with functionally important residues indicated. The backbone atoms of the lowest energy structure are shown in black. Residues R6-L9 are coloured blue and in (B) and (C) and over residues G13-F17 the two clusters are coloured green and red. The side chains of functionally critical residues (32, 33, 34, 98, 99) identified by alanine substitution studies are labeled and coloured purple.

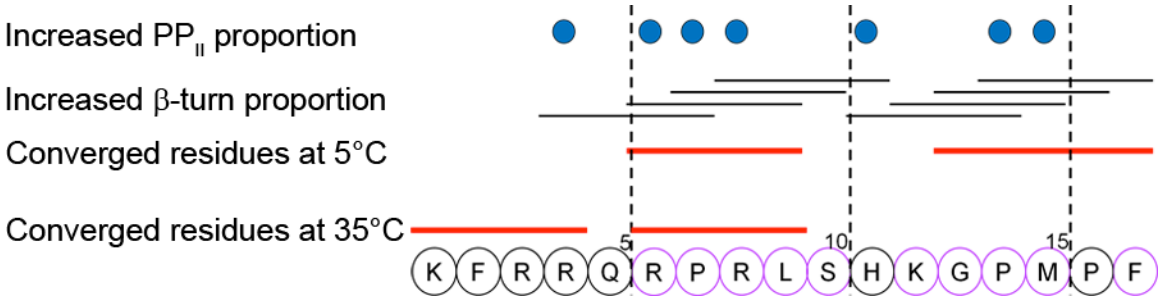

Figure 2.11: A summary of structural properties of apelin relative to the apelin-17 sequence. Structurally converged regions are indicated by a red bar. Increased PP_{II} and β-turn abundance based on parallel examination of all three ensembles (detailed in Figure 2.9) are indicated, with all four residues involved in a given β-turn indicated by a bar covering those residues. Functionally important residues are indicated with purple circles.

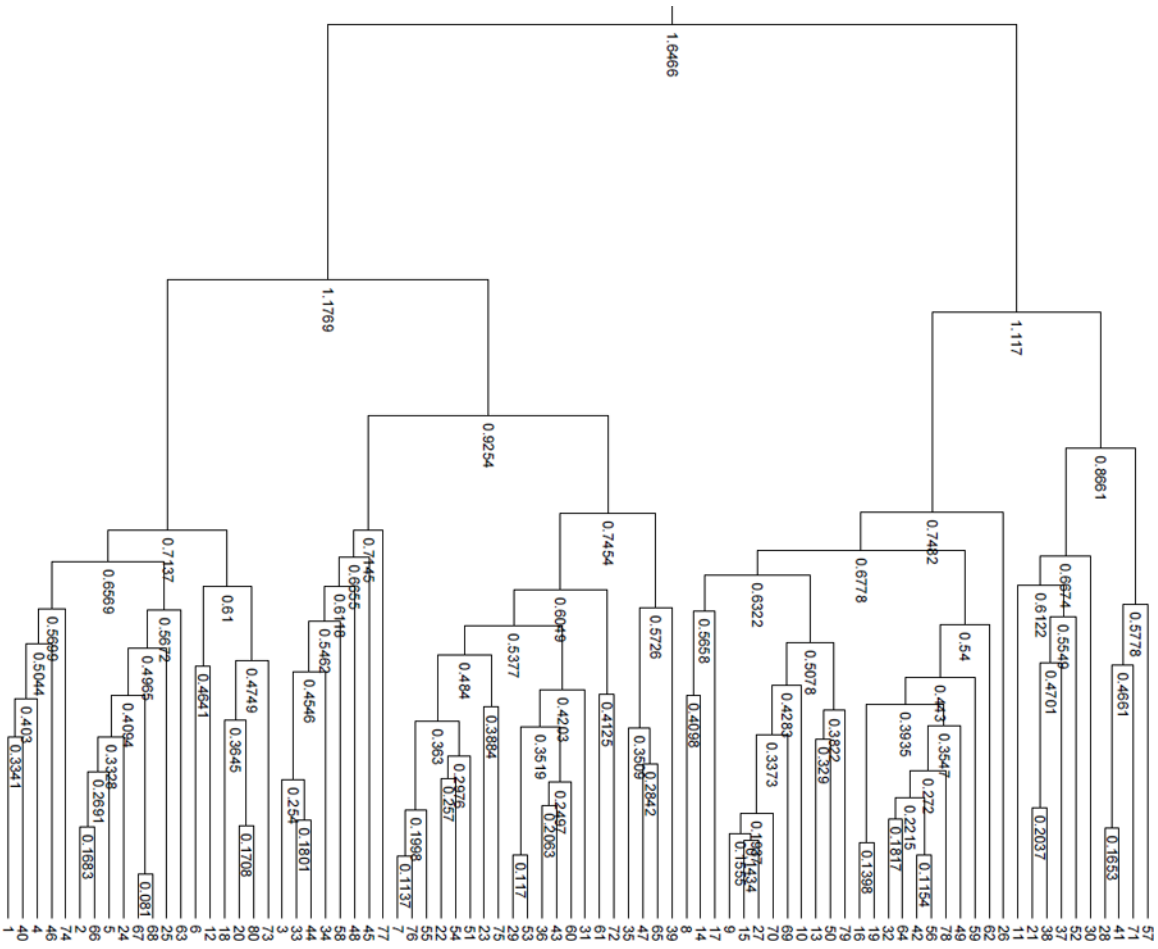

Figure 2.12: Clustering of the apelin-17 conformer A ensemble. Each ensemble member is represented by a number at the tips of the tree (bottom). The number at each node represents the RMSD of all ensemble members that it contains. The R1 method of Clusterpose (119) was used for RMSD calculation as well as clustering the structures. Figtree (A. Rambaut, University of Edinburgh) was used for visualization of the data generated by Clusterpose.

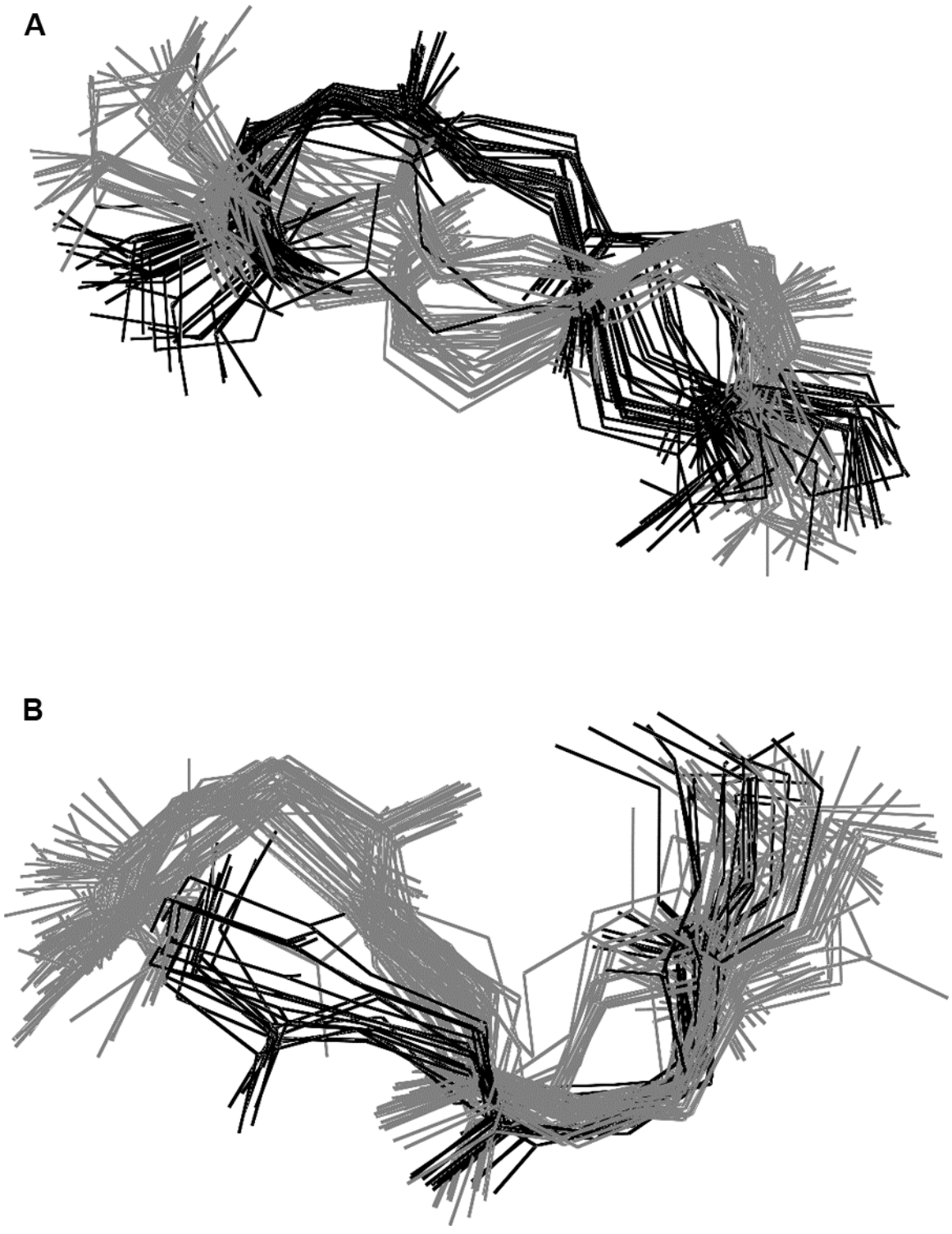

Figure 2.13: Visualization of the two major clusters found over residues G13-F17 in (A) conformer A and (B) conformer B of apelin-17 at 5 °C. Cluster statistics are provided in Table 2.4.

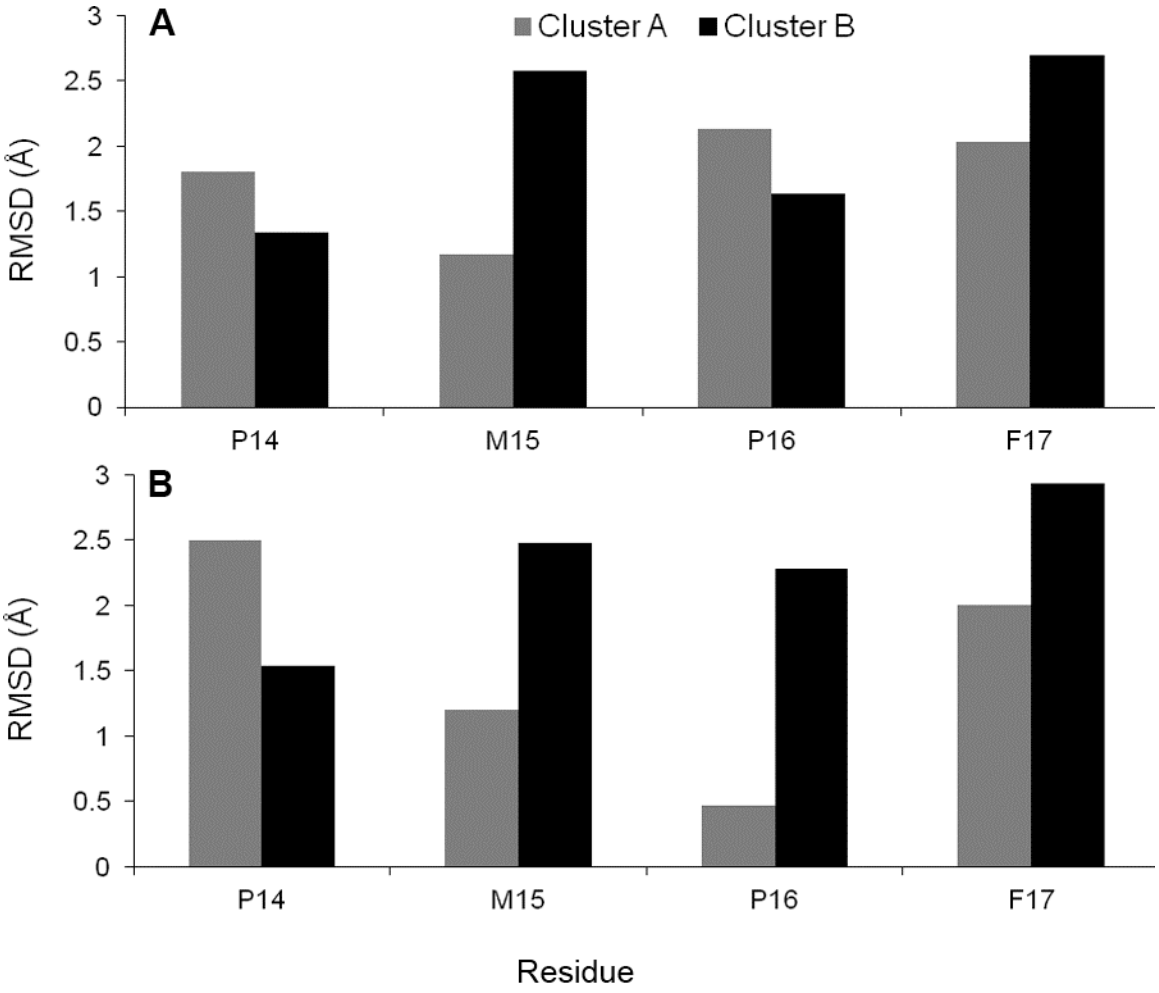

Figure 2.14: Side chain heavy atom RMSD values for G13-F17, the regions of apelin-17 conformers A and B (shown in (A) and (B) respectively) where two major clusters of backbone conformation were observed (Table 2.4).

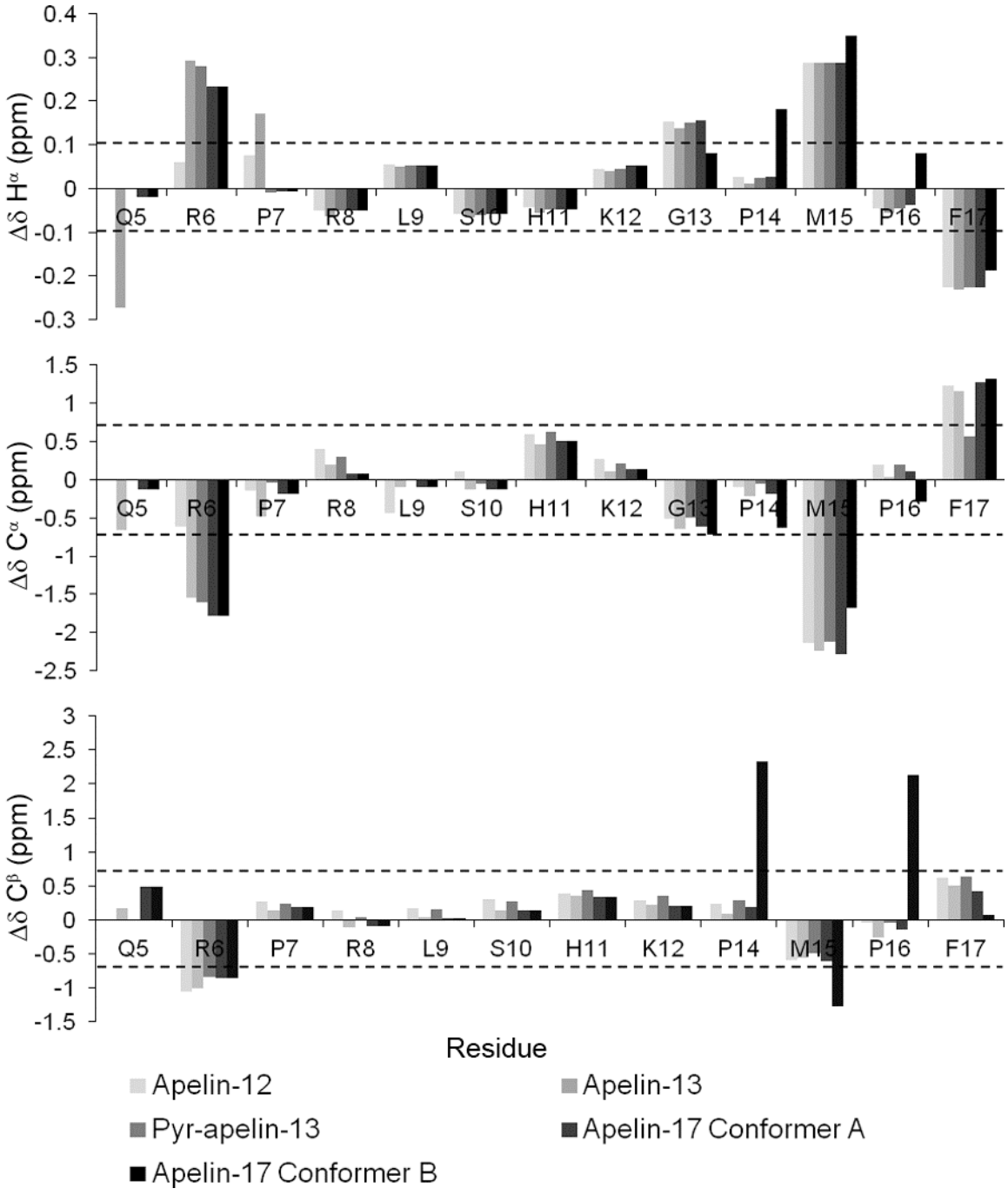

Figure 2.15: Secondary chemical shifts using random-coil values of Wishart et al. (84) at 5 °C of apelin-12**,** apelin-13**,** pyr-apelin-13 and apelin-17 conformers A and B for H^{α}, C^{α} and C^{β}, as indicated. The horizontal lines show the values of Δδ identified as significant for secondary structuring by the ^{1}H and ^{13}C chemical shift indices (refs (106) and (130), respectively). The C^{α} and C^{β} resonances for proline have identified ranges of ±4 ppm (not shown), unlike the other residues.

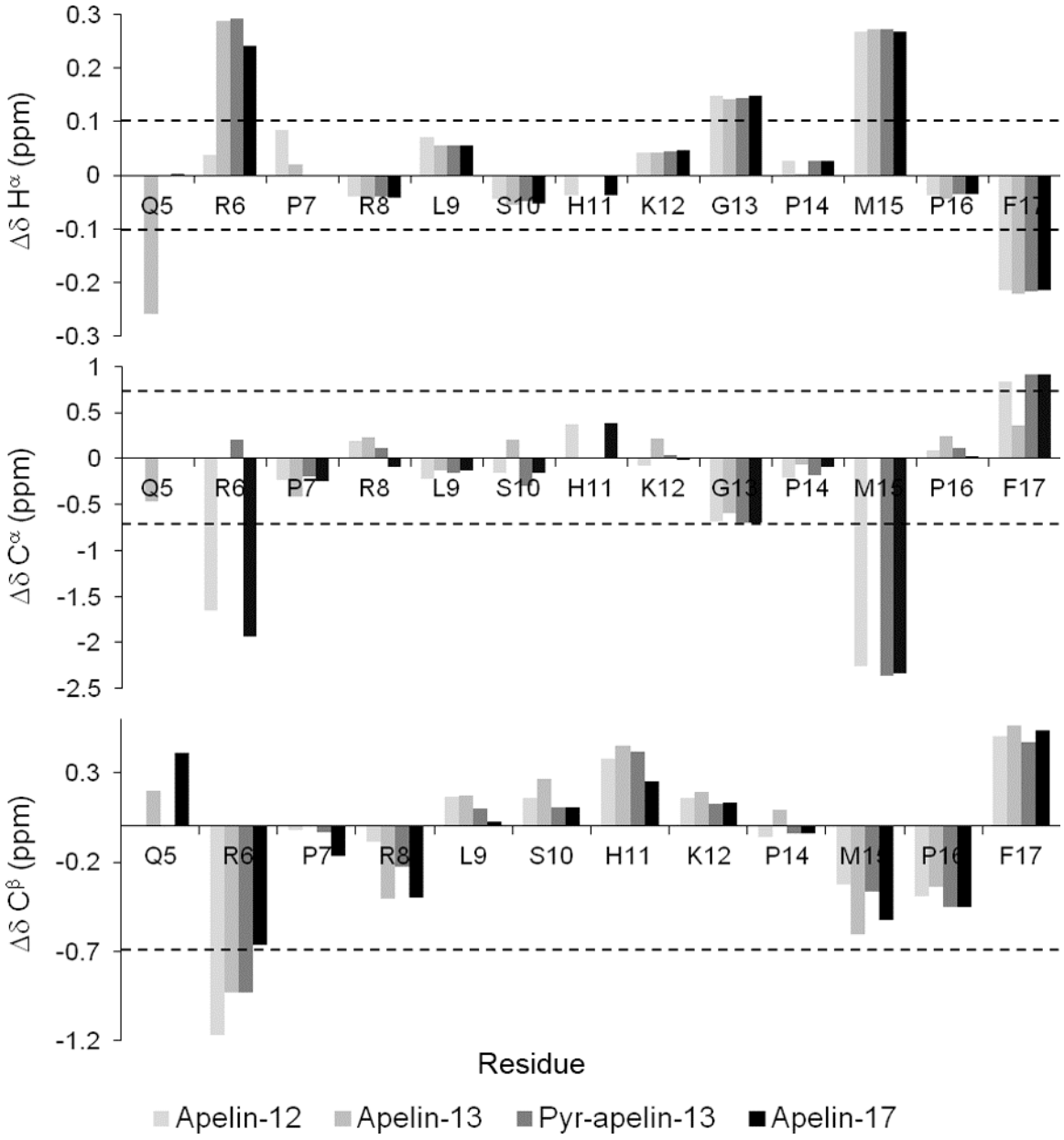

Figure 2.16: Secondary chemical shifts using random-coil values of Wishart et al. (84) of apelin-12, apelin-13, pyr-apelin-13 and apelin-17 at 35 °C for H^{α}, C^{α} and C^{β}, as indicated. The horizontal lines show the values of Δδ identified as significant for secondary structuring by the ^{1}H and ^{13}C chemical shift indices (refs (106) and (130), respectively). The C^{α} and C^{β} resonances for proline have identified ranges of ±4 ppm (not shown), unlike the other residues.

Chapter 3:

Interactions Of Apelin With Membrane Mimetics (Based On My Contribution To Manuscripts (2, 3))

3.1. Introduction

In the previous chapter I characterized the structure of apelin-17 in solution. However, apelin and AR do not interact with each other in isolation. Instead, AR exists in a complex environment that includes, among other things, the lipid molecules of the plasma membrane. Indeed, apelin is much more likely to interact with the many other components of the cell membrane before it comes into contact with AR. Realizing this, Sargent and Schwyzer (131) suggested that the cell membrane is an essential component of ligand-receptor interactions. In this chapter, I discuss the membrane catalysis theory introduced by Sargent and Schwyzer and characterize the interactions that apelin has with detergent micelles.

3.1.1. Membrane Catalysis Of Peptide-Receptor Binding

Sargent and Schwyzer (131) formulated the membrane catalysis hypothesis by extending the ideas of Adam and Delbrück that membranes could cause accelerated receptor tracking due to reduced diffusion dimensionality (132). The membrane is proposed to act as a receptor-ligand catalyst by i) increasing the local concentration of the

ligand thereby improving probability of diffusional collision between the receptor and ligand; and/or ii) inducing conformational change in the peptide to increase its affinity for the receptor. In this way, the energy of a peptide binding to its receptor is broken down into several steps including the surface accumulation of ligands, ligand-membrane interactions, potential conformational changes of the ligand and finally binding of the ligand to the receptor. The membrane catalysis model is attractive as it helps to explain the large association constants (pM affinity) and high association/dissociation rates found with membrane receptors and their ligands compared to soluble enzymes (131). This model also suggests that differences of ligand affinity to the same receptor in different tissues could simply be due to the composition of the plasma membrane in the vicinity of the receptor.

There are many examples of peptide hormones that bind to membrane mimetics (133, 134, 135, 136, 137) and in some cases they have been shown to adopt a specific structure when bound to the membrane (138, 139, 140). A striking example is calcitonin, a peptide hormone involved in calcium homeostasis (138). In the presence of sodium dodecylsulphate (SDS) micelles, calcitonin assumes an α-helical structure while being unstructured in solution. Furthermore, deletion of the critical F16 residue disrupts both the induction of the α-helical structure and the function of calcitonin. Perhaps the least intuitive part of the membrane catalysis hypothesis is that ligands can undergo specific conformational changes that enhance binding to the receptor while bound to the membrane. Although difficult to prove, a lot of work has been done to characterize the membrane bound structures of peptides. Table 3.1 provides a list of peptide hormones

that have been shown to bind to membrane mimetics with the high resolution structures of the membrane-bound peptides shown in Figure 3.1.

3.1.2. Common Membrane Mimetics Used In Biophysical Studies

Biological membranes are a very complex mixture. There are many lipids in cell membranes that vary not only by species (141) but also by cell type. Adding to this complexity, there are numerous proteins and small molecules embedded into or interacting with the plasma membrane. Both the lipids in the membrane and the proteins associated with it can also be glycosylated. Protein folding is greatly affected by the location of the protein within the membrane. In the center of the membrane, the folding environment is dominated by the hydrophobic tails of the lipid molecules while near the surface of the membrane, the environment changes to include the headgroups of the lipids and water (reviewed in (142, 143)).

Despite this inherent complexity, membrane mimetic models have emerged as tools for *in vitro* classification of membrane associated proteins (141). Several types of common membrane mimetics used are organic solvents, micelles, bicelles, supported bilayers and liposomes (reviewed in (141)). Despite the fact that these membrane mimetics contain at best only a few types of lipids, there are almost countless studies demonstrating the relevance of these membrane mimetics to biological systems (3, 51, 52, 55, 56, 57, 61, 126, 134, 139, 140, 144, 145, 146, 147, 148, 149, 150, 151).

For structural studies by either NMR or CD spectroscopy, a membrane mimetic that allows fast tumbling of the protein of interest is preferred since fast motion of the

mimetic allows for sharp lines in NMR spectra (82). For this reason, in my studies that required membrane mimetics, I worked primarily with micelles (Figure 3.2).

Micelles are composed of one or two different types of amphipathic molecules. At concentrations above the critical micelle concentration, these detergents assemble into spherical but labile structures with the hydrophobic tail of the detergent located in the center of the micelle. The orientation of the detergent headgroups at the surface of the micelle with the hydrophobic tail located in the center provides a rough approximation of biological membranes (141). Various types of detergents are commonly used to construct micelles for the purpose of protein characterization such as SDS (152), dodecylphosphocholine (DPC, (153)) and 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)] (LPPG, Figure 3.2, (154)).

3.1.3. The Study Of Peptide-Membrane Interactions With NMR Spectroscopy

NMR spectroscopy is a great tool for analyzing interactions between a peptide and micelle. In addition to simply determining the structures of peptides both free in solution and bound to the micelle, experiments such as diffusion ordered spectroscopy (DOSY, (155)) and paramagnetic relaxation enhancement (PRE) provide invaluable information.

In DOSY, pulsed field gradients are used to quantify the translational diffusion coefficient of molecules or supramolecular assemblies (155). Briefly, a pulsed field gradient (of power g and duration δ) is applied to the sample, which causes each of the nuclear spins to acquire a phase shift that is dependent on the position of the spin in the NMR tube. After a set amount of diffusion time (Δ) a second pulsed field gradient is

applied that reverses the phase shift of the first pulse. If all of the spins in the NMR sample have their phases refocused perfectly, the result is a strong NMR signal. Of course, since diffusion is occurring, the magnetization is not exactly refocused because the individual nuclear spins are in different places before and after Δ. It is intuitive that as more diffusion occurs during Δ, the less the NMR signal from the bulk sample will be refocused. In DOSY the loss of signal intensity can be modelled as:

$$I=I_0\exp[-D\gamma^2g^2\delta^2(\Delta-\delta/3)] \quad 3.1$$

where D is the translational diffusion coefficient and I and I_0 are the peak intensities when the experiment is performed with and without a gradient, respectively. In DOSY, multiple NMR experiments are completed with different values of g, and then peak intensities are fit to equation 3.1 in order to calculate D. It may seem counterintuitive to vary the gradient strength instead of Δ, but it is important to avoid the effects of relaxation by not varying the length of the experiment.

Since there are specific signals from both the micelle and the peptide, D of both can be measured and binding between the peptide and micelle can be quantified (156, 157). In my work I use a fast two-site exchange model (156) to quantify apelin binding to micelles:

$$D_{obs} = f_bD_b + (1-f_b)D_f \quad 3.2$$

where D_f is the translational diffusion coefficient of free apelin, D_{obs} is the observed diffusion coefficient of apelin-17 in the presence of micelles and D_b is the diffusion coefficient of the micelles. These values can be used to provide an estimate of the fraction of apelin-17 bound to micelles (f_b).

In paramagnetic relaxation enhancement (PRE), a compound that contains an unpaired electron is added to the sample. Free electrons themselves generate relatively large randomly fluctuating local magnetic fields (80). These fluctuating local magnetic fields cause any transverse magnetization along the xy axis to lose coherence quickly, resulting in a loss of NMR signal. In order to study protein-micelle interactions various reagents can be added to the sample to understand the topology of the micelle-protein complex. For example, Mn^{2+} partitions into the aqueous phase of the sample causing signal to disappear from solvent exposed parts of the peptide (158). Similarly, 5-doxylstearic acid (5-DSA) and 16-doxylstearic acid (16-DSA, Figure 3.3) partition into the headgroup and interior of micelles, respectively (159). With the large variety of paramagnetic reagents available, a detailed topology of the protein-micelle complex can be determined through PRE (158, 160).

Given the work that supports the membrane catalysis model, there is strong incentive to investigate the potential interactions of apelin with membrane mimetics. Herein, I use CD spectropolarimetry and DOSY to probe interactions of apelin-12 and apelin-17 with SDS, DPC and LPPG micelles. I present an NMR spectroscopy-based structure of apelin-17 bound to SDS micelles and use PRE to give an overview of the topology of the apelin-17/micelle complex. Finally, I discuss the significance of these findings in relation to the membrane catalysis model.

3.2. Materials And Methods

3.2.1. Materials

LPPG was purchased from Avanti Polar lipids (Alabaster, AL). Deuterium oxide (D_2O; 99.8 atom% D), SDS-d_{25}, DPC, DPC-d_{38} and D_2O containing 1% (w/w) DSS were obtained from C/D/N isotopes (Pointe-Claire, QC). The apelin-12 and apelin-17 used for this study are the same peptides as in chapter 2. All other chemicals were obtained at biotechnology, high performance liquid chromatography or reagent grade, as appropriate, from Sigma-Aldrich Canada Inc. (Oakville, ON).

3.2.2. CD Spectropolarimetry

Far-ultraviolet CD spectra of apelin-12 and apelin-17 were recorded at 35 °C using a Jasco J-810 spectropolarimeter (Easton, MD). Solutions of apelin peptides (apelin-12: 55.2±0.7 μM, apelin-17 64.5±5.2 μM) were prepared from a single stock solution. Exact quantification of each peptide was achieved through amino acid analysis (Hospital for Sick Children), using a weighted average of the arginine, proline, leucine and phenylalanine content in the sample. For each apelin peptide, samples were prepared with no lipid added, 16 mM SDS, 19 mM DPC or 38 mM LPPG in 20 mM sodium phosphate adjusted to pH 7.00 ± 0.05. CD spectra were acquired from 260 nm downwards (1 nm steps), with reliable ellipticity values observed at >190 nm based on spectropolarimeter photomultiplier tube voltages, in 0.5 cm path-length quartz cuvettes (Hellma, Müllheim, Germany). All measurements were collected in triplicate. Machine data were converted to mean residue ellipticity, averaged over all trials, blank subtracted and subjected to

sliding-window averaging over 3 nm stretches so that the mean residue ellipticity reported at a given wavelength (λ) is:

$$[\theta] = \frac{1}{4}[\theta]_{\lambda-1} + \frac{1}{2}[\theta]_{\lambda} + \frac{1}{4}[\theta]_{\lambda+1} \quad 3.3$$

Finally, each spectrum was baseline adjusted by subtracting the average ellipticity value over 250-260 nm.

3.2.3. PRE Measurements

Three identical NMR samples were prepared with 1 mM apelin-17 in a 90% H_2O/10% D_2O mixture with 20 mM $Na^+CD_3COO^-$, 1 mM NaN_3, 90 mM SDS and 1 mM DSS. Each sample was titrated with $MnCl_2$, 5-DSA or 16-DSA up to a maximum of 8 mM paramagnetic agent at 35 °C. Titrations were halted when ~50% reduction in intensities of ~50% peaks were observed using ^1H-^1H TOCSY relative to a reference spectrum. For each residue, intensities of one or two non-overlapped cross-peaks were measured (depending on availability of non-overlapped peaks). Titration of 16-DSA was at 800 MHz using a Varian INOVA spectrometer (Palo Alto, CA) at NANUC (Edmonton, AB) while both Mn^{2+} and 5-DSA were titrated at 700 MHz on a Bruker Avance III spectrometer (Milton, ON) at the National Research Council Institute for Marine Biosciences (NRC-IMB, Halifax, NS). Both spectrometers are equipped with cryogenically-cooled triple-resonance probes. NMR data were processed using NMRPipe (114) and peak intensities measured using a Gaussian lineshape in Sparky 3 (T.D. Goddard and D.G. Kneller, University of California, San Francisco).

3.2.4. Diffusion Ordered Spectroscopy Measurements

Four samples of apelin-17 were prepared for DOSY experiments with 1 mM apelin-17, 20 mM $Na^+CD_3COO^-$, 1 mM NaN_3, and 1 mM DSS in 90% H_2O/10% D_2O, with three samples containing one of 81 mM SDS-d_{25}, 94 mM DPC-d_{25} or 187 mM LPPG. Identical samples were also prepared without apelin. Detergent concentrations were chosen to give ~1.25:1 micelle:peptide ratios. Sample pH was adjusted to 5.00 ± 0.05 in each case using DCl and NaOD. DOSY spectra were acquired at 35 °C using a stimulated echo sequence with bipolar gradients (161) and a longitudinal eddy current decay of 5 ms at 500 MHz on a Bruker AVANCE II spectrometer with TXI probe at NMR-3 (Halifax, NS). Susceptibility matched tubes (Shegemi, Allison Park, PA) were used to minimize diffusion artefacts due to gradient non-linearity or convection currents. Signal attenuation at the maximum gradient strength was adjusted to be 95%, with 32 increments collected in the diffusion dimension with the gradient strength ranging from 1.7 G/cm to 32.0 G/cm. DOSY experiments were performed in triplicate.

Diffusion coefficients for apelin-17 in each condition alongside those of each micelle in absence of apelin were obtained using the relaxation module of Topspin 2.1 (Bruker) to fit the relaxation data to equation 3.1. Due to what appeared to be a consistent artefact, the first (lowest gradient strength) increment of each experiment was discarded, and 31 diffusion time points were used for fitting. For apelin-17 and each micelle, D was calculated using three separate ^{1}H-NMR peaks for each spectral dataset (*i.e.* 9 measurements per condition) and averaged to give a final value of D with associated standard error.

I made the assumption that the diffusion coefficient of a micelle, D_b, does not differ significantly with the presence or absence of apelin-17 due to the much larger size of the micelle. The viscosity of each solution was not measured. Instead, the expected effect of the presence of micelles on the value of D_f was estimated via the following expression for small molecules obstructed by spherical particles (162):

$$D_f = D^0_f/(1+\phi/2) \qquad 3.4$$

where D^0_f is the measured diffusion coefficient of apelin-17 in water and ϕ is the volume fraction of lipids in the sample, which can be estimated by using the weight fraction (162). Using equations 3.2 and 3.4 as well as the values of D obtained using Topspin, the fraction of apelin-17 bound to each micelle was estimated.

3.2.5. Sequential Assignment Of Apelin-17 In SDS And DPC Micelles

Two 600 μL NMR samples were made with 20 mM $Na^+CD_3COO^-$, 1 mM NaN_3, 1 mM DSS, 0.5 mM apelin-17 and either 81 mM SDS-d_{25} or 94 mM DPC-d_{38}. ^1H-^1H TOCSY (80 ms mixing time), ^1H-^1H NOESY (350 ms mixing time), ^1H-^{13}C HSQC and ^1H-^{15}N HSQC experiments were acquired at 700 MHz (NRC-IMB Avance III) with a TCI cryoprobe at 35 °C. Detailed data acquisition parameters for all NMR experiments are provided in Table 3.2. NMR data were processed using NMRPipe (114) and manually assigned in Sparky 3 (T.D. Goddard and D.G. Kneller, University of California, San Francisco). Secondary chemical shifts, $\Delta\delta$ = δ(observed) - δ(random coil) (106, 130, 163), of the H^α, C^α and C^β nuclei were calculated using the random coil chemical shifts of Wishart and coworkers (84).

3.2.6. Structural Calculations Of Apelin-17 Bound To SDS Micelles

XPLOR-NIH version 2.18 (89) was used for final structure determination of apelin-17 bound to SDS micelles using the methods of NOE intensity calculation and refinement outlined in chapter 2. An ensemble of 200 structures was generated, with the 80 lowest energy structures retained as the final ensemble. The LSQKAB software of the CCP4 suite (118) was used to iteratively superpose the ensemble over 2-17 residue stretches. The final ensemble was analyzed for structural quality, the presence of converged regions and the presence of β-turns as outlined in chapter 2. All chemical shift data for apelin-17 with SDS and DPC micelles as well as the final structural ensemble for apelin-17 with SDS micelles have been deposited to the Biological Magnetic Resonance Bank (121) using SMSDep (accession numbers 20082 and 16275 for apelin-17 with SDS and DPC micelles, respectively).

3.3. Results And Discussion

3.3.1. CD Spectropolarimetry Of Apelin-12 And Apelin-17

The far-ultraviolet CD spectra from 260-190 nm of both apelin-12 and apelin-17 in buffer and in the presence of SDS, DPC or LPPG micelles are shown in Figure 3.4. For any given condition, the CD spectra for both apelin-12 and apelin-17 have similar features. As described previously (1), the CD spectra for both apelin-12 and apelin-17 in buffer at 35 ºC are random coil in nature (75), convoluted with positive bands at 195 nm and 218 nm that may be attributable to the ^{1}B and $^{1}L_{a}$ transitions (1, 164, 165) of the C-

terminal phenylalanine. For both apelin-12 and apelin-17, DPC micelles do not strongly perturb the CD spectrum relative to buffer (Figure 3.4). In contrast, dramatic spectral changes occur with both apelin isoforms in the presence of either SDS or LPPG micelles. The CD spectra of apelin with these anionic micelles are convoluted by an α-helix-like band pattern (difference spectra in Figure 3.4), which may be attributed either to formation of α-helical secondary structure (75) or to the formation of β-turns (166, 167).

Although the main trends observed in the CD spectra of apelin-12 and apelin-17 are identical, there are a few minor differences between isoforms. In the presence of DPC micelles, a slightly larger deviation from buffer conditions is observed for apelin-12 than for apelin-17. Also, the CD spectra of apelin-17 in the presence of LPPG and SDS are nearly identical, while for apelin-12 the CD spectrum in LPPG shows a larger deviation from buffer conditions in comparison to SDS (Figure 3.4). Despite these slight differences, the underlying trends in the CD spectra of apelin-12 and apelin-17 are the same, with DPC micelles causing little to no change in the spectra relative to buffer versus SDS and LPPG, which produce increasing amounts of perturbation of the CD spectrum relative to buffer. This trend was also observed for apelin-13, pyroglutamate-apelin-13 and apelin-36 (Bebbington and Rainey, unpublished), suggesting that all of the apelin peptides undergo similar structural changes in the presence of anionic micelles.

3.3.2. Diffusion Coefficient Determination

Translational diffusion coefficients (D) were measured for SDS, DPC and LPPG micelles, for apelin-17 in buffer and for apelin-17 in the presence of each micelle (Table 3.3). Reported literature diffusion values (DOSY determined) are 1.2×10^{-10} m^2/s for

DPC micelles (151), 9.5 × 10^{-11} m^2/s for SDS micelles (168) and 6.62 × 10^{-11} m^2/s for LPPG micelles (169). Once adjusted for temperature, viscosity and the presence of D_2O using literature viscosity values (170) to match the experimental conditions, these measurements become 1.35 × 10^{-10}, 1.47 × 10^{-10} and 8.49 × 10^{-11} m^2/s, respectively. Although the reported D values are slightly higher, the agreement with previous D values is reasonable. Discrepancies may be due to effects of the different temperature and buffer conditions in the study upon the aggregation number of micelles. In particular, the aggregation number of micelles decreases with increasing temperature (171) and this study was at 35 °C while the reported values of D for SDS and LPPG were at 25 °C and that for DPC at 30 °C. Also, the observed D for a micelle is often an overestimate due to contributions from detergent monomers. Under buffer conditions, however, the critical micelle concentrations of detergents are lowered and, for each micelle employed herein, the contribution of monomeric detergent to D was minor since only ~ 5%, ~1% and ~0.03% of SDS, DPC and LPPG are estimated to be present as monomers (based upon critical micelle concentration values of ~4 mM (172), 1.1 mM (173) and 0.06 mM (173) for SDS, DPC and LPPG respectively).

Upon determination of D, the f_b of apelin-17 with each micelle was calculated using equations 3.2 and 3.4, indicating relatively weak binding of apelin-17 to DPC micelles and stronger binding with both SDS and LPPG micelles (Table 3.3), in good agreement with the CD spectropolarimetry results. This interaction is somewhat predictable due to the strongly cationic nature of apelin-17, which contains 8 basic residues to facilitate its interaction with anionic headgroups. However, without experimental demonstration such

as this, it is not predictable that a peptide as hydrophilic as apelin-17 will interact so strongly with detergent micelles in solution.

3.3.3. NMR Assignment And Secondary Chemical Shifts

Nearly complete assignment of ^{1}H, ^{13}C and ^{15}N was obtained for apelin-17 (Table 3.4) in the presence of both SDS and DPC micelles using standard methods (124) allowing comparison of Δδ (84) for H^{α}, C^{α} and C^{β} nuclei in the presence of SDS and DPC micelles to those in buffer reported in chapter 1 (Figure 3.5). For both H^{α} and C^{β}, there is no clear relationship between the value of the secondary chemical shift and the condition used. However, these Δδ values are significantly different in the presence of SDS micelles than either the buffer or DPC conditions over K1-K12, while Δδ in the C-terminal of apelin-17 show little perturbation in the presence of micelles. The large changes in Δδ values of C^{α} nuclei in the presence of SDS suggest that apelin-17 is binding to the SDS micelles rather than forming a different, canonical 2° structure since H^{α} and C^{β} would also be expected to have a consistent Δδ perturbation with 2° structure formation. Determination of the effect of LPPG micelles on Δδ values was not possible as deuterated LPPG is not routinely commercially available, making spectral assignment infeasible for apelin without isotopic enrichment.

In all cases, apelin-17 showed a single, predominant set of chemical shifts at 35 °C with D values between those of free apelin and those of each detergent micelle indicative of fast exchange on the NMR timescale between the free and bound states (*i.e.* sub-millisecond exchange (174)). A second conformation of apelin-17 was also evident in the NMR data, as previously observed in buffer (chapter 2), with a minor conformation in the

P14-F17 region populated at <10% relative to the major one. Only the major conformer was sequentially assigned as there were few NOE contacts in the minor conformer. CD spectropolarimetry, DOSY and chemical shift analysis clearly demonstrate that the apelin peptides bind weakly and transiently to zwitterionic DPC micelles but bind favourably to anionic SDS and LPPG micelles.

3.3.4. PRE Of Apelin-17 With SDS Micelles

To determine the region of apelin-17 that is interacting with anionic micelles, PRE NMR spectroscopy experiments were used. The ^{1}H-^{1}H TOCSY peak attenuation of apelin-17 in SDS micelles with various concentrations of 5-DSA, 16-DSA or Mn^{2+} is summarized in Figure 3.6. These reagents would be expected to attenuate the NMR signals for nuclei in the tailgroup core of the micelle (16-DSA), just below the anionic headgroup (5-DSA) or readily accessible to solution (Mn^{2+}). For both 5-DSA and 16-DSA, there is no clear trend of peak attenuation over apelin-17, even at relatively high concentrations of paramagnetic agent, suggesting that apelin-17 does not associate in the hydrophobic core of the micelles or just below the headgroups of SDS (Figure 3.6 (B)). In contrast Mn^{2+} attenuates the signal of apelin-17. At 0.25 mM Mn^{2+}, many residues of apelin-17 show large attenuation of peak intensity, notably at the C-terminal region (Figure 3.6 (A)). After titration with 1 mM Mn^{2+}, only K1 and R6-L9 have observable signal, indicating that these residues of apelin-17 interact most strongly with SDS micelles since they are partially shielded from the effects of Mn^{2+}. However, even these residues are largely attenuated by Mn^{2+}, indicating that although they interact with the micelle, they are still partially solvent exposed and very near the micelle surface, as supported by the lack of peak attenuation induced by 5-DSA or 16-DSA.

3.3.5. Structure Of Apelin-17 With SDS Micelles

In total, 522 unique NOE contacts were observed. This is an almost 2-fold increase in observed NOE contacts in comparison to apelin-17 in buffer at 35 ºC (285 restraints), indicating an increase in structuring of apelin-17 bound to SDS micelles vs. free in solution. Of these contacts, the majority that are medium range occur C-terminal to R6 (Figure 3.7), implying that this region of apelin-17 is most structured in the presence of SDS micelles. Eight rounds of structure calculation were performed with these NOE restraints, with 100 structures generated in rounds 1-7 and 200 in round 8. Of the 200 structures in the final round, the 80 structures with the lowest total energy were retained as the final ensemble, with minimal NOE violations and excellent Ramachandran plot statistics (Table 3.5). Examination of both the average deviation and order parameters as defined by Hyberts et al. (175) of the ϕ and ψ dihedral angles for each residue show a stretch of seven residues from R6-K12 with both low deviation and high order parameter (Figure 3.8 (A,B)), which is indicative of structural convergence (104).

Independent of this, the ensemble of apelin-17 structures was superposed in 7 residue stretches and the RMSD of backbone atoms was calculated (Figure 3.8 (C)). At superposition lengths of 8 residues or greater, no RMSDs below 0.88 ± 0.2 Å were observed. For a 7 residue superposition, the backbone atom RMSD is 0.72 ± 0.17 over R6-K12, suggesting that this region is better converged and mirrors the observation of highly defined ϕ and ψ dihedral angles. RMSDs over 7 residues indicate that both the N- and C-terminal regions of apelin-17 are relatively unstructured in the presence of SDS micelles. Using Promotif-NMR (125), 45 type I and 374 type IV β-turns were detected in the final ensemble (Table 3.5) with 98% of type I β-turns being initiated between R6 and

L9. Out of 80 structures, 30 have a type I β-turn initiation at P7. Type IV β-turns appear to be distributed throughout apelin-17, although there is a very well defined type IV β-turn initiated at S10 in 77/80 of the ensemble members (Figure 3.9). PRE experiments (Figure 3.6) indicate that the structural convergence over R6-K12 is the result of apelin binding to the micelle. In the superposed structural ensemble, the cationic side-chains of R6 and R8 also tend to lie on the same face of the peptide, providing a cationic surface that may interact with the anionic micelle headgroups (Figure 3.10 (B)). The NMR ensemble generated for apelin-17 interacting with SDS is also representative of the apelin-17 structure when bound to LPPG micelles, since CD spectropolarimetry suggests the structures are the same.

Upon analysis of superpositions for shorter regions of apelin-17, an extremely well converged region of apelin-17 was observed from M15-F17 (Figure 3.10 (C)), with a heavy atom RMSD of 0.37 ± 0.19 Å and well converged ϕ and ψ dihedrals (Figure 3.8 (A,B,D)). This region is just as converged as some of the residues between R6-K12 in terms of both atom positions and dihedral angles but shows no canonical secondary structuring by Promotif-NMR analysis. Notably, this structural convergence was not observed in apelin-17 in solution at 35 ºC, indicating that this structure is induced by apelin-micelle binding even though it is not in contact with the micelle.

3.3.6. Implications For Membrane Catalysis Of Apelin-AR Binding

This study implies that apelin is likely to fit with the membrane catalysis model put forward by Sargent and Schwyzer (131). Since apelin interacts almost equally well with

two quite different anionic detergent headgroups, binding to biological membranes is likely. Furthermore, the structure observed with apelin bound to SDS micelles should represent the bound-state with a biological membrane. Although a micelle surface is curved, in this system the curvature is negligible. The radius of a SDS micelle is ~20 Å (176) while the binding region of apelin-17 (R6-L9) spans ~6.7 Å, or ~5%, of the micelle circumference.

Upon membrane interaction, the membrane catalysis model hypothesizes that a peptide will adopt a conformation that will accelerate binding to its receptor. The abundant type I and type IV β-turns that are induced by the anionic headgroups at the SDS and LPPG micelle surface may increase apelin's affinity to the AR receptor, in line with the suggestion of Tyndall et al. that β-turns are a ubiquitous motif recognized by GPCR's (101). However, the high degree of structuring for M15-F17, although not β-turn in character, is a very striking micelle-induced structural element that may allow specific recognition of apelin by AR since this portion of the peptide is not directly membrane-associated and therefore free to bind to the receptor.

Taken together, the structural changes that occur to apelin when bound to anionic micelles could be essential for initiating interactions with AR. Since both regions of apelin-17 observed to undergo significant conformational restriction reside within the 12 residue functional core retained in all bioactive apelin isoforms, this binding hypothesis is immediately applicable to the other apelin peptides.

3.4. Summary

The various active isoforms of apelin interact similarly with anionic SDS and LPPG micelles but not with the zwitterionic DPC micelle. Key to this interaction is the formation of a converged structure from R6-K12 with significant type I and type IV β-turn initiation between R6-L9 and at S10 respectively, while the residues N-terminal to this region are largely disordered. Part of this converged structure, R6-L9, interacts the most strongly with anionic micelles. Binding of apelin-17 to SDS also induces structuring of M15-F17, even though this region is not in direct contact with the micelle surface. These results indicate that it is possible that AR activation by apelin follows the membrane catalysis model, since apelin binds to membranes and undergoes a conformational change.

Table 3.1: An overview of studies on peptide ligands that bind to micelles where both a high-resolution structure is available and the residues involved in membrane binding were determined.

Peptide	PDB entry	Micelle type	Residues structured	Residues bound to micelle
Amylin nucleation site (177)	1KUW	SDS	F23-I26: β–turn	N21, N22, F23, I26, L27
Bovine pancreatic polypeptide (178)	1LJV	DPC	M17-L31: α-helix	L3, E4, E6, G9, A12, A18, A21, L24, R25, R33, P34, R35
Glucagon (179, 180)[a]	1KX8	DPC	T5-Y10, L14-R17: extended Y10-L14: helix-like turn R17-T29: irregular α-helix	Hydrophobic residues
Glucagon antagonist (181)	1NAU	DPC	T5-L12: Irregular helix A17-T27: α-helix	Y8, Y11, L12, F20, V21, W23, L24, M25
Neuromedin B(182)	1C98	SDS	W4-M10: α-helix	W4, H8 and F9
Neuropeptide K (183)	2B19	DPC	A2-K8: β-turns or 3-10 helix Q9-G18 and H27-V33: α-helix H19-R26: non-canonical β–turn	F32, V33, G34, L35 and M36 (hydrophobic residues)
Neurotensin (184)	1OYV	DPC	L2-N5: extended P10-L13: close to β-turn	P10, Y11, I12 and L13
Pituitary adenylate cyclase activating polypeptide (185)	1D2P	DPC	I5-L27: α-helix	I5, F6, Y10, Y13, M17, Y22, L23, V26 and L27
Porcine Neuropeptide Y (186, 187)	1F8P	DPC	Y21-T32: α-helix	L17, Y20, Y21, L24, R25, Y27, I28, N29, I31, T32 and Y36
Porcine peptide YY (187)	1RUU	DPC	L17-V31: α-helix	L17, Y20, Y21, L24, Y27, L28, L30, V31, T32, R33, Y36
Proadrenomedullin N-terminal peptide (188)[b]	2FLY	SDS	R2-A17: α-helix	R2, L3, V5, F9, K12, W13 W16 and R20
Rat islet amyloid polypeptide (189)	2KJ7	DPC	A5-L23: α-helix	N3, T6, A8, T9, L12, A13, L16

[a] Glucagon (PDB entry 1KX8) did not superpose well over the identified converged regions.
[b] Structure published in PDB (entry 2FLY) in trifluoroethanol, with direct comparison to SDS micelle bound state in paper.

Table 3.2: Detailed experimental parameters used for data acquisition of apelin-17 in either SDS or DPC micelles.

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	^{1}H frequency (MHz)	Facility
Apelin-17 SDS							
^{1}H-^{1}H TOCSY[a]	16	2	^{1}H:8192 ^{1}H:256	^{1}H:8418 ^{1}H:7228	^{1}H:4.68 ^{1}H:4.68	700	NRC-IMB
^{1}H-^{1}H NOESY[a]	2	32	^{1}H:8192 ^{1}H:256	^{1}H:8418 ^{1}H:7228	^{1}H:4.68 ^{1}H:4.68	700	NRC-IMB
^{1}H-^{13}C HSQC	1.5	96	^{1}H:4096 ^{13}C:128	^{1}H:11261 ^{13}C:14085	^{1}H:4.68 ^{13}C:37.6	700	NRC-IMB
^{1}H-^{15}N HSQC	1.5	512	^{1}H:2048 ^{15}N:32	^{1}H:8418 ^{15}N:2258	^{1}H:4.68 ^{15}N:118.0	700	NRC-IMB
Mn^{2+} and 5' DSA titration	2	8	^{1}H:8192 ^{1}H:192	^{1}H:8418 ^{1}H:7228	^{1}H:4.68 ^{1}H:4.68	700	NRC-IMB
^{1}H-^{1}H TOCSY[a] 16' DSA titration	1.5	16	^{1}H:4096 ^{1}H:256	^{1}H:11990 ^{1}H:10000	^{1}H:4.68 ^{1}H:4.68	800	NANUC
^{1}H-^{1}H TOCSY[b] DOSY ^{1}H 1D[c]	3	8	^{1}H: 8192	^{1}H: 5482	^{1}H: 4.68	500	NMR-3
Apelin-17 DPC							
^{1}H-^{1}H TOCSY[a]	16	2	^{1}H:8192 ^{1}H:256	^{1}H:8418 ^{1}H:7228	^{1}H:4.68 ^{1}H:4.68	700	NRC-IMB
^{1}H-^{1}H NOESY[a]	2	16	^{1}H:8192 ^{1}H:256	^{1}H:8418 ^{1}H:7228	^{1}H:4.68 ^{1}H:4.68	700	NRC-IMB
^{1}H-^{13}C HSQC	1.5	96	^{1}H:4096 ^{13}C:128	^{1}H:11261 ^{13}C:14085	^{1}H:4.68 ^{13}C:37.6	700	NRC-IMB
^{1}H-^{15}N HSQC	1.5	544	^{1}H:2048 ^{15}N:32	^{1}H:8418 ^{15}N:2258	^{1}H:4.68 ^{15}N:118.0	700	NRC-IMB
DOSY ^{1}H 1D[c]	3	8	^{1}H: 8192	^{1}H: 5482	^{1}H: 4.68	500	NMR-3

[a] Water suppression was achieved using excitation sculpting
[b] Water suppression was achieved using WATERGATE 3-9-19
[c] Water suppression was achieved using presaturation

Table 3.3: Fraction of apelin-17 bound (f_b) to indicated micelle types as determined through DOSY. Topspin (Bruker) was used to determine diffusion coefficients (D_{obs}) of three peaks of the component of interest for three replicate experiments to give an average D_{obs} with associated average deviation. Assuming two-state fast exchange for apelin-17, f_b was estimated using equations 3.2 and 3.4. The experimental settings for diffusion time (Δ) and gradient duration (δ) for each experiment are indicated. Pulsed field gradients were varied from 1.7 G/cm to 32.0 G/cm.

Component	**Δ(ms)**	**δ(ms)**	**D_{obs} (m^2/s)**	**f_b**
SDS micelles	60	6.4	$1.7 \times 10^{-10} \pm 1.6 \times 10^{-12}$	N/A
DPC micelles	60	6.4	$1.8 \times 10^{-10} \pm 1.6 \times 10^{-12}$	N/A
LPPG micelles	85	7.0	$9.6 \times 10^{-11} \pm 2.8 \times 10^{-13}$	N/A
Apelin-17	55	4.4	$3.6 \times 10^{-10} \pm 3.4 \times 10^{-12}$	N/A
Apelin-17 with SDS micelles	65	5.2	$2.0 \times 10^{-10} \pm 1.2 \times 10^{-11}$	82 ± 5%
Apelin-17 with DPC micelles	60	5.2	$2.8 \times 10^{-10} \pm 3.4 \times 10^{-12}$	46 ± 5%
Apelin-17 with LPPG micelles	70	6.6	$1.0 \times 10^{-10} \pm 1.2 \times 10^{-12}$	98 ± 5%

Table 3.4: Summary of unassigned and ambiguously assigned ^{1}H, ^{13}C and ^{15}N chemical shifts for apelin-17 with SDS or DPC micelles.

Condition	**Missing and ambiguous resonance assignments**
Apelin-17 with SDS micelles*	K1:H^{N}, H^{ε}, N F2:C^{α} R3:$H^{\eta\dagger}$ R4:$H^{\eta\dagger}$, $N^{\dagger}$ Q5:$N^{\dagger}$ R6:C^{α}, $N^{\dagger}$ R8:$H^{\eta\dagger}$, $N^{\dagger}$ L9:$N^{\dagger}$ H11:C^{α}, C^{ε} K12:$N^{\dagger}$ P14:C^{α} M15:C^{α},$N^{\dagger}$ F17:C^{α}
Apelin-17 with DPC micelles*	K1:H^{N}, H^{ε}, N F2:$N^{\dagger}$ R3:H^{η}, $N^{\dagger}$ R4:H^{η} Q5:$N^{\dagger}$ R6:H^{η}, $N^{\dagger}$ R8:H^{η} L9:$N^{\dagger}$ H11:H^{α}, C^{α}, $N^{\dagger}$ K12:H^{ε} M15:C^{α}

*The only side chain ^{15}N shift determined was Q5-N^{ε}.

†Ambiguous assignment (chemical shift assigned, but not uniquely for this resonance)

Table 3.5: Summary of the restraints employed for the final ensemble of 80 retained structures from 200 calculated structures, XPLOR-NIH energies and violation occurrences. Structural statistics for apelin-17 in buffer at 35 ºC are shown for comparison.

	Apelin-17 with SDS	**Apelin-17 in buffer**
Total rounds of refinement	8	12
Unique NOE restraints		
Total	522	285
Intra-residue	248	146
Sequential	188	108
Medium Range ($\|i-j\| \leq 4$).	46	22
Long Range ($\|i-j\| > 4$).	0	0
Ambiguous	40	9
Ramachandran plot statistics		
Core	33%	26%
Allowed	51%	52%
Generously allowed	10%	10%
Disallowed	6%	13%
Number of type I β-turns	45	9
Number of type IV β-turns	274	123
Number of type VIII β-turns	6	10
XPLOR-NIH energies (kcal/mol)[a]		
Total	26.9 ± 2.1	45.1 ± 5.4
NOE	4.6 ± 1.0	0.6 ± 0.5
Violations		
NOE Violations > 0.5 Å	0	0
NOE Violations of 0.3-0.5 Å	1	0
NOE Violations of 0.2-0.3 Å	4	4

[a] Ranges are given by average deviations for XPLOR-NIH energies.

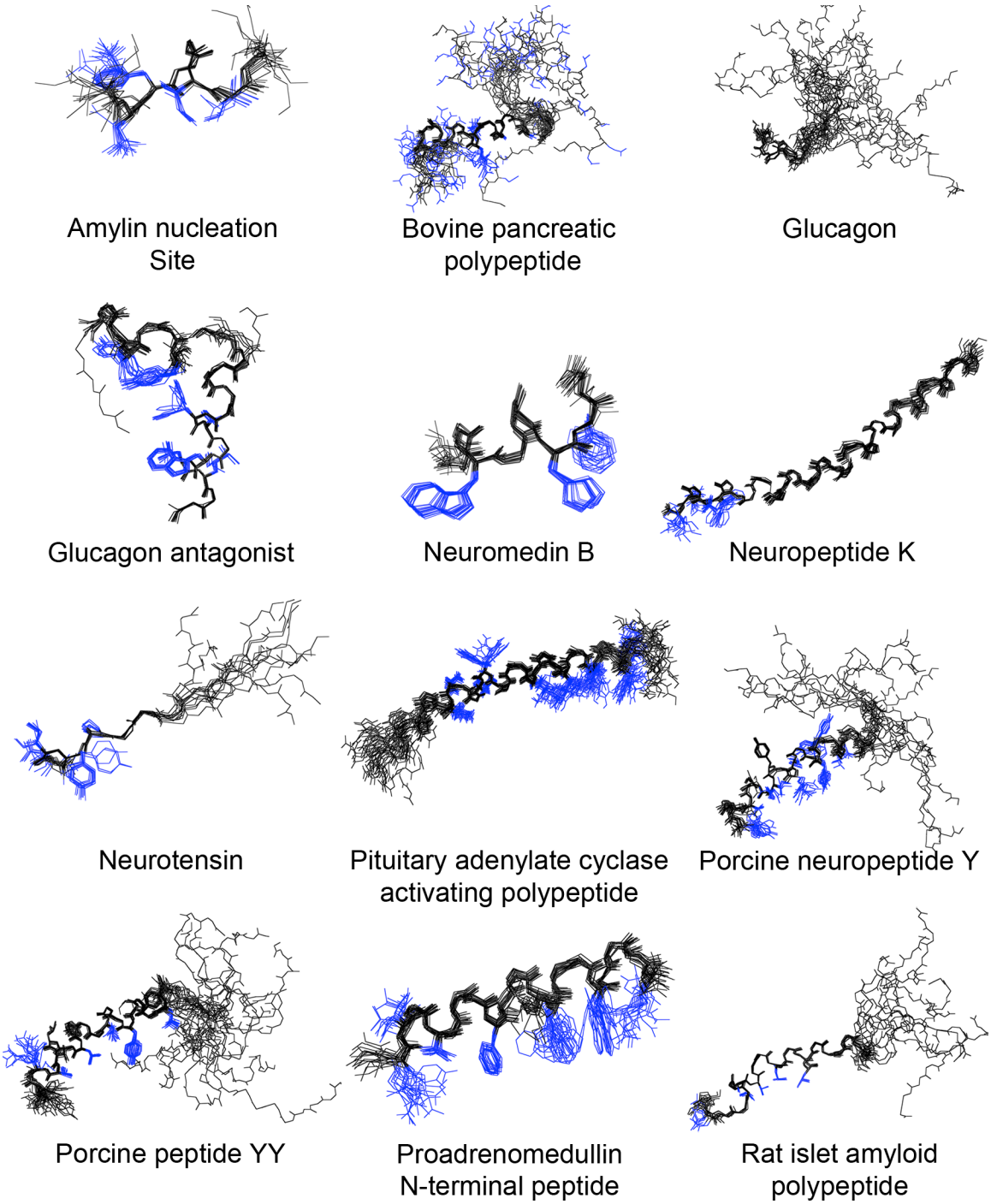

Figure 3.1: Superposed structures of the peptides listed in Table 3.1. Where possible, all members of an ensemble have been superposed over the converged regions using LSQKAB (118). Backbone atoms are shown in black and the side chains of residues that are implicated in membrane-binding are in blue, whenever the NMR ensemble was well superposed.

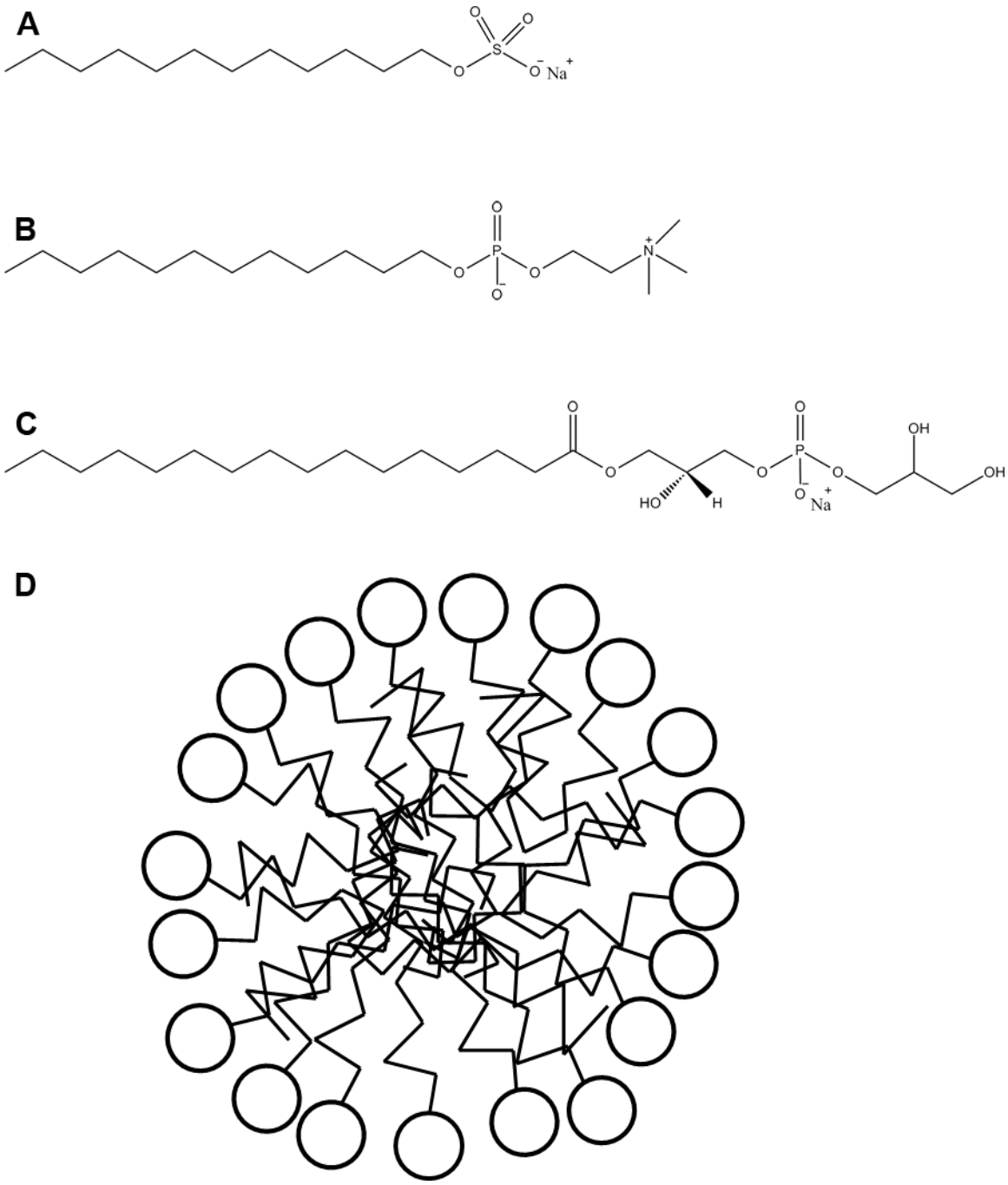

Figure 3.2: The structure of membrane mimetics used in my studies. Structures are shown for (A) SDS, (B) DPC and (C) LPPG micelles. A cross-section of a spherical micelle is shown in (D) with the lipid headgroups shown as circles and hydrophobic tails shown as lines.

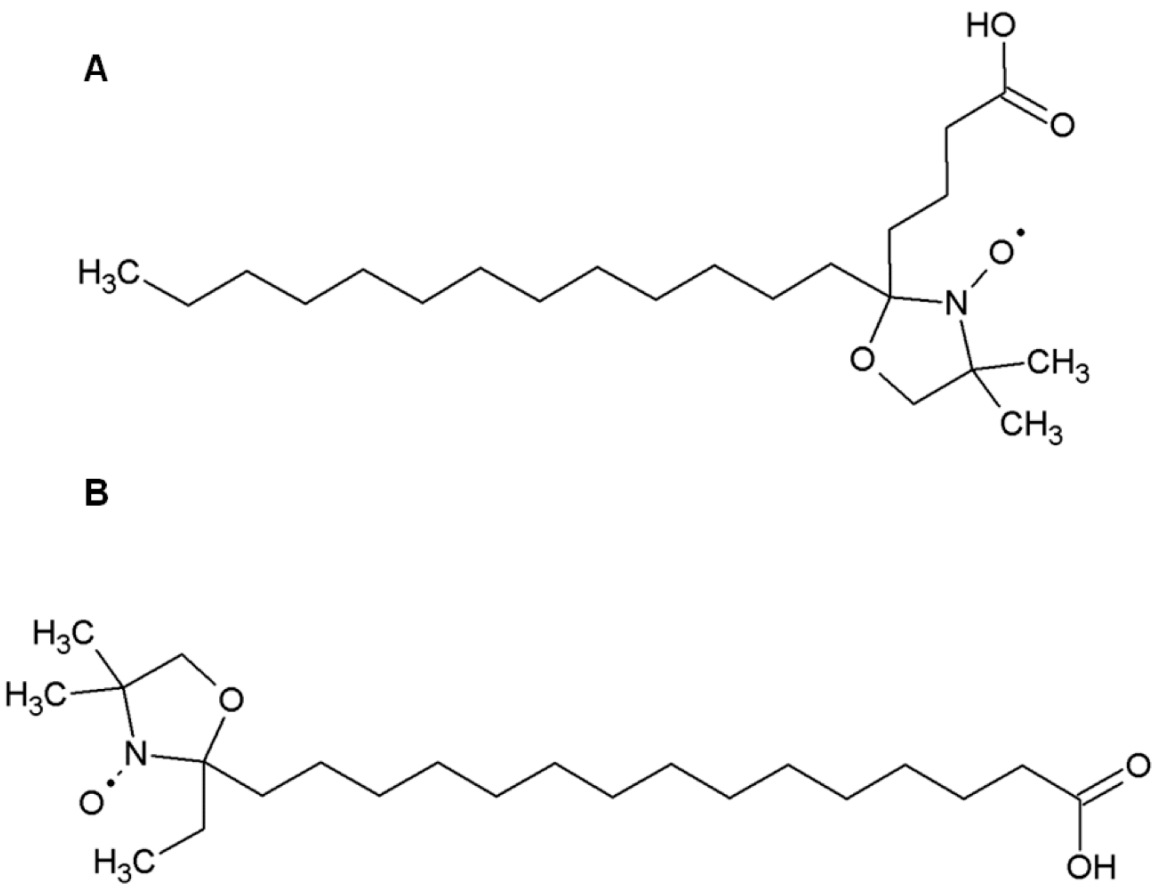

Figure 3.3: The structures of 5-doxylstearic acid (A) and 16-doxylstearic acid (B).

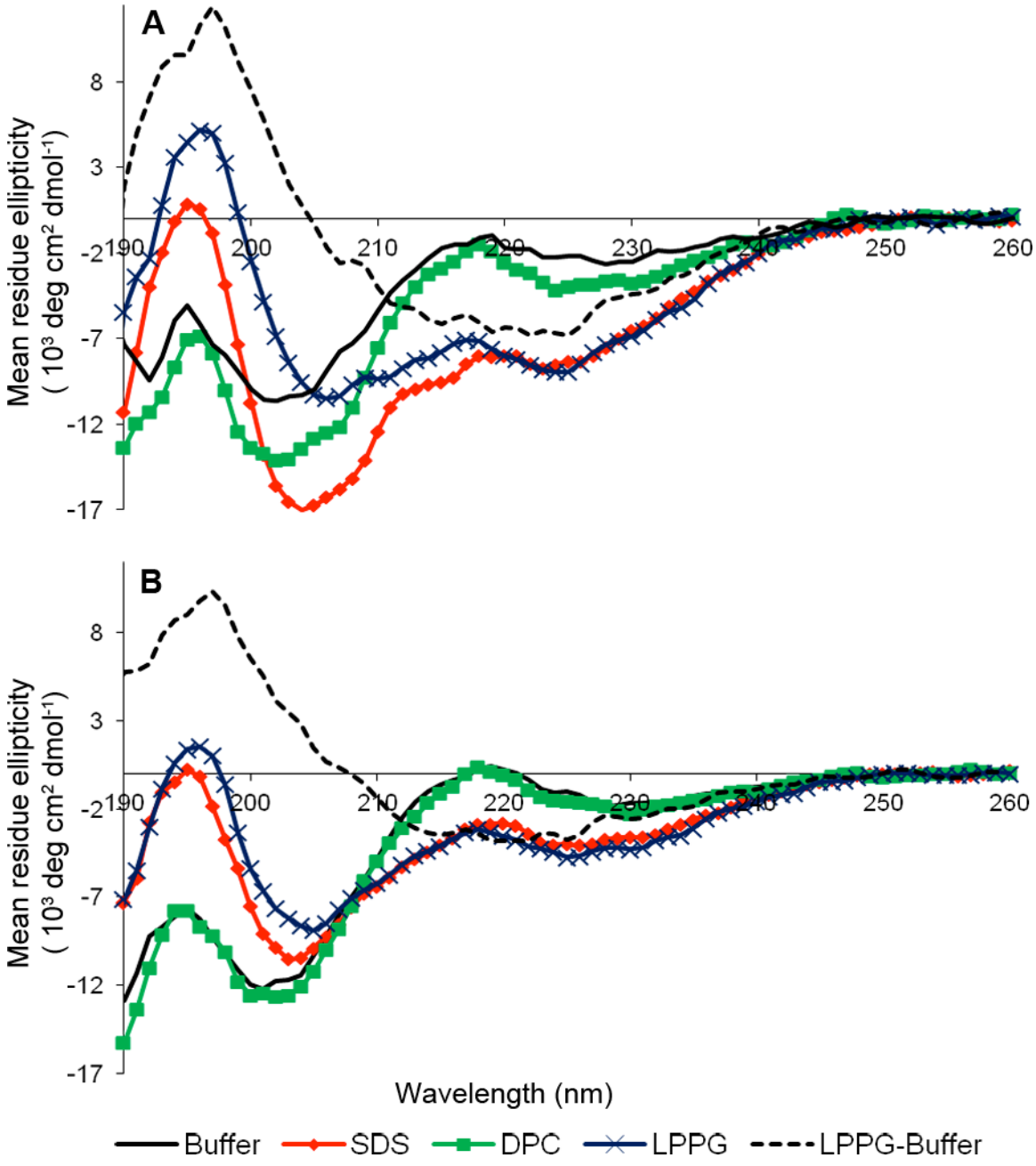

Figure 3.4: Far-ultraviolet CD spectra of (A) apelin-12 and (B) apelin-17 in 20 mM phosphate, SDS micelles, DPC micelles and LPPG micelles alongside a difference spectrum between LPPG and buffer.

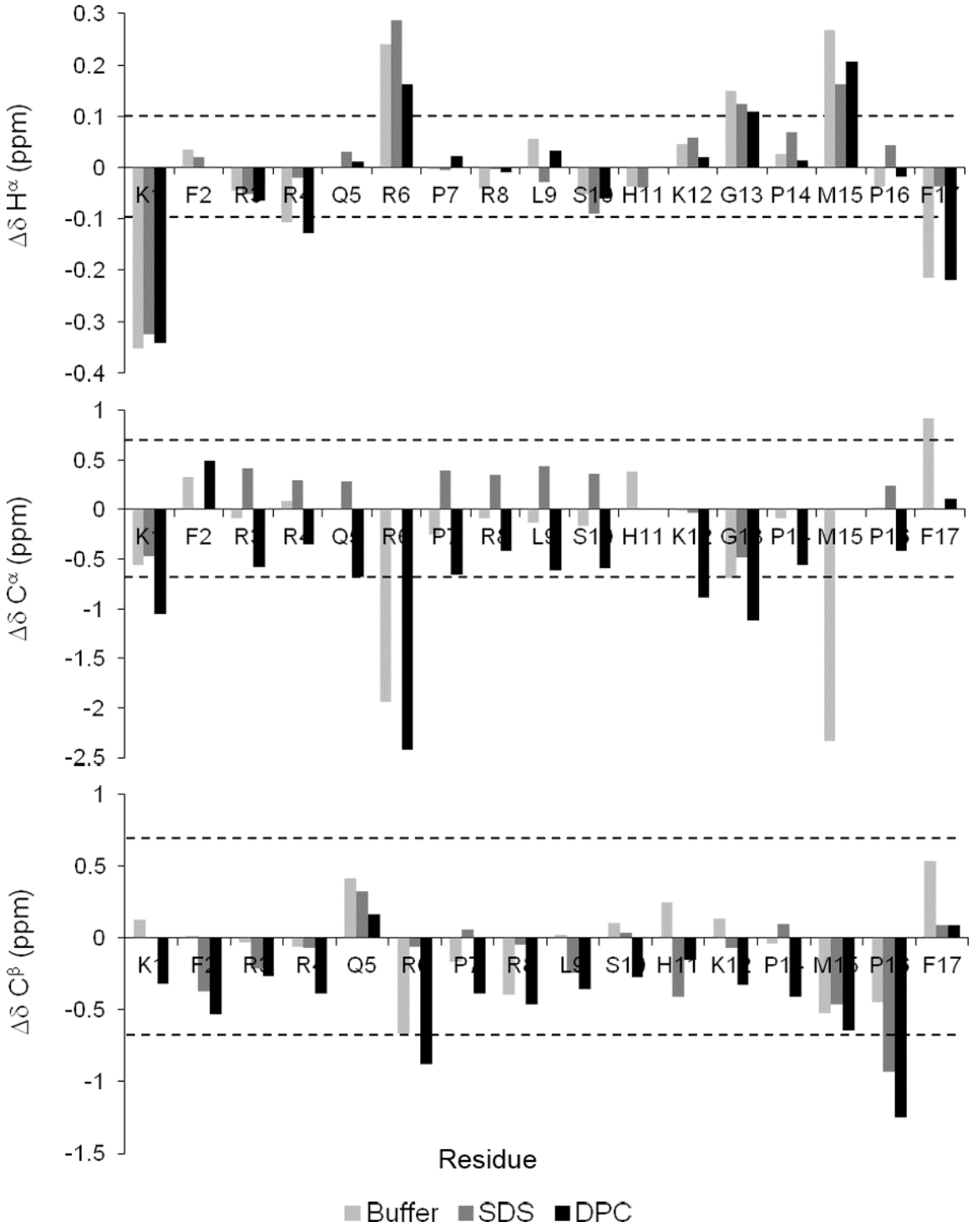

Figure 3.5: Apelin-17 secondary chemical shifts for H^{α}, C^{α} and C^{β} at 35 °C in buffer, SDS micelles and DPC micelles. Horizontal lines show the Δδ cut-off significant for secondary structuring (106) Note that C^{α} and C^{β} for proline have identified significance ranges of ±4 ppm (not shown).

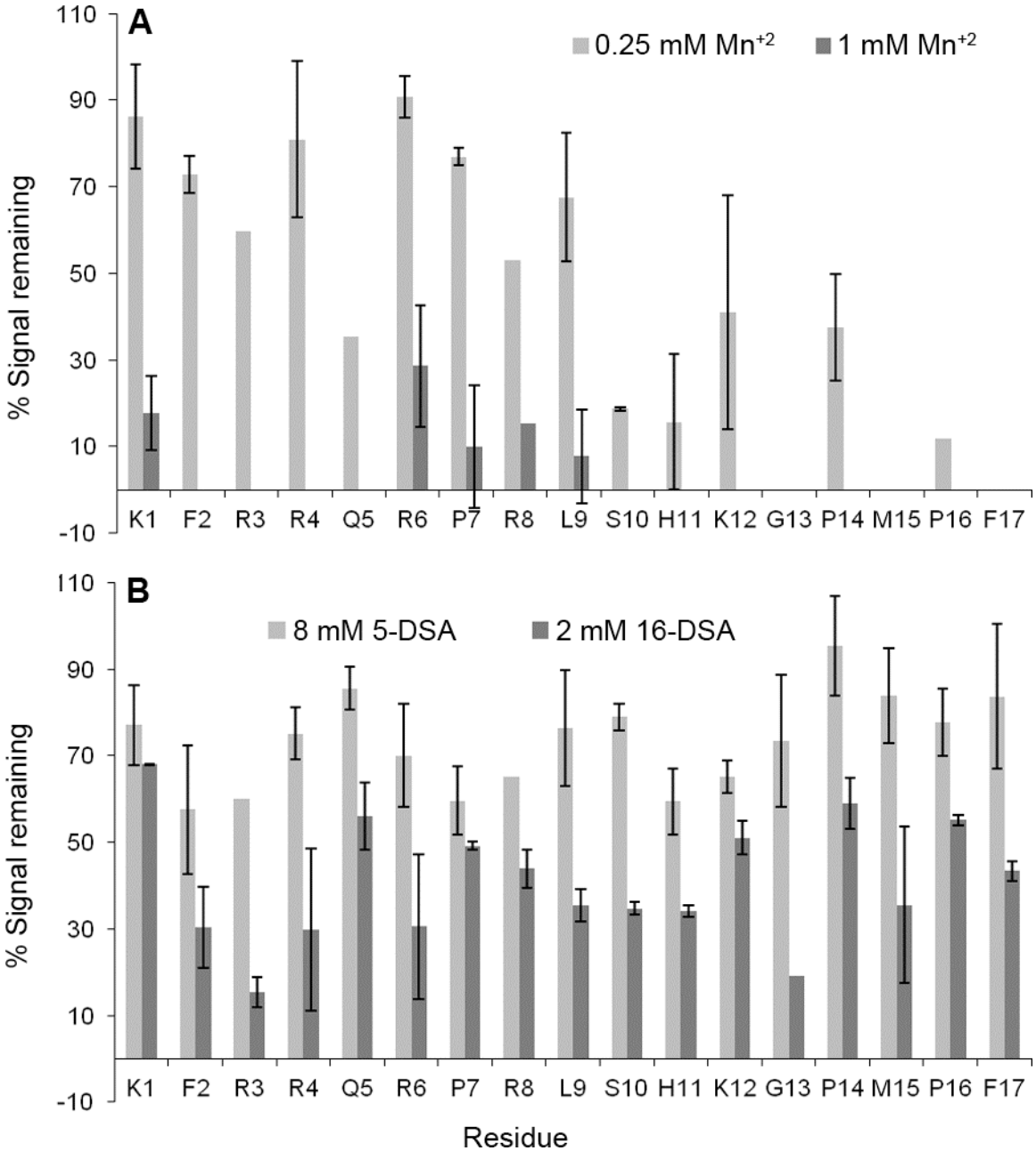

Figure 3.6: Retained peak intensity (error bar is average deviation) of apelin-17 in SDS micelles with indicated amounts of (A) Mn^{2+} or (B) 5-DSA, 16-DSA calculated from intensities of ^{1}H-^{1}H TOCSY peaks relative to a reference spectrum.

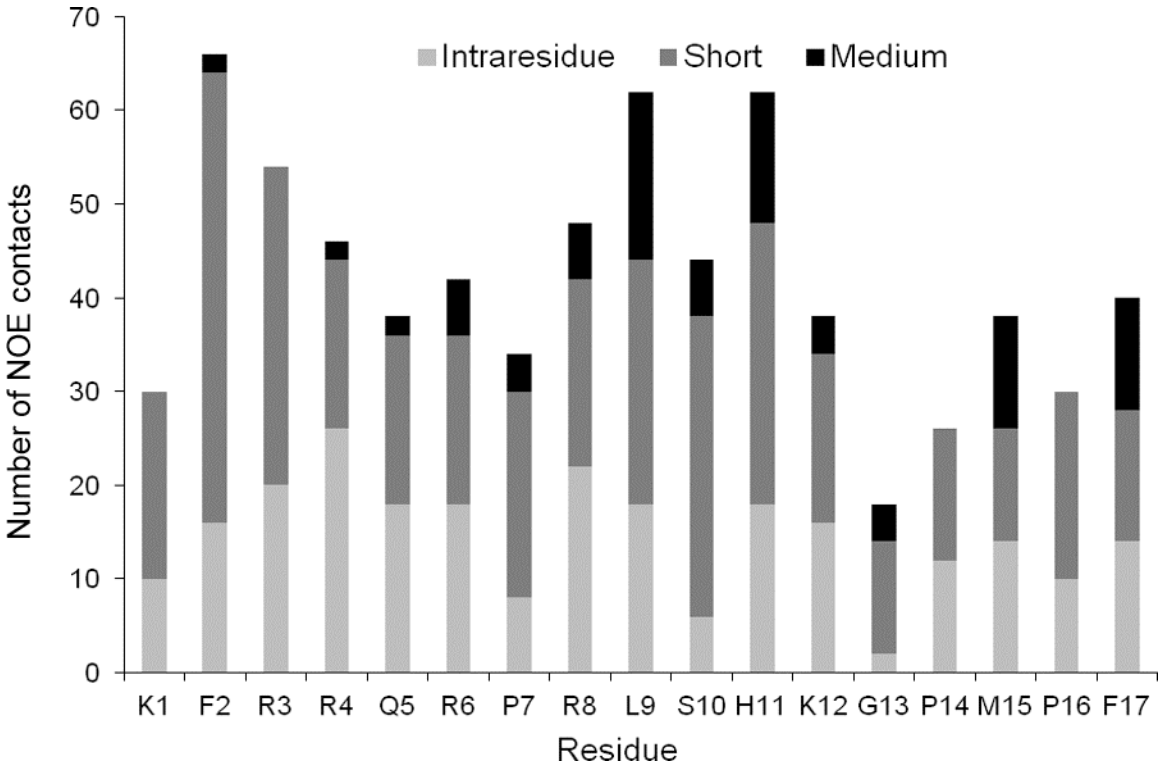

Figure 3.7: NOE contacts on a per-residue basis used in the calculation of the final structural ensemble of apelin-17 with SDS. Intraresidue, short (sequential) and medium ($|i-j| \leq 4$) contacts are shown.

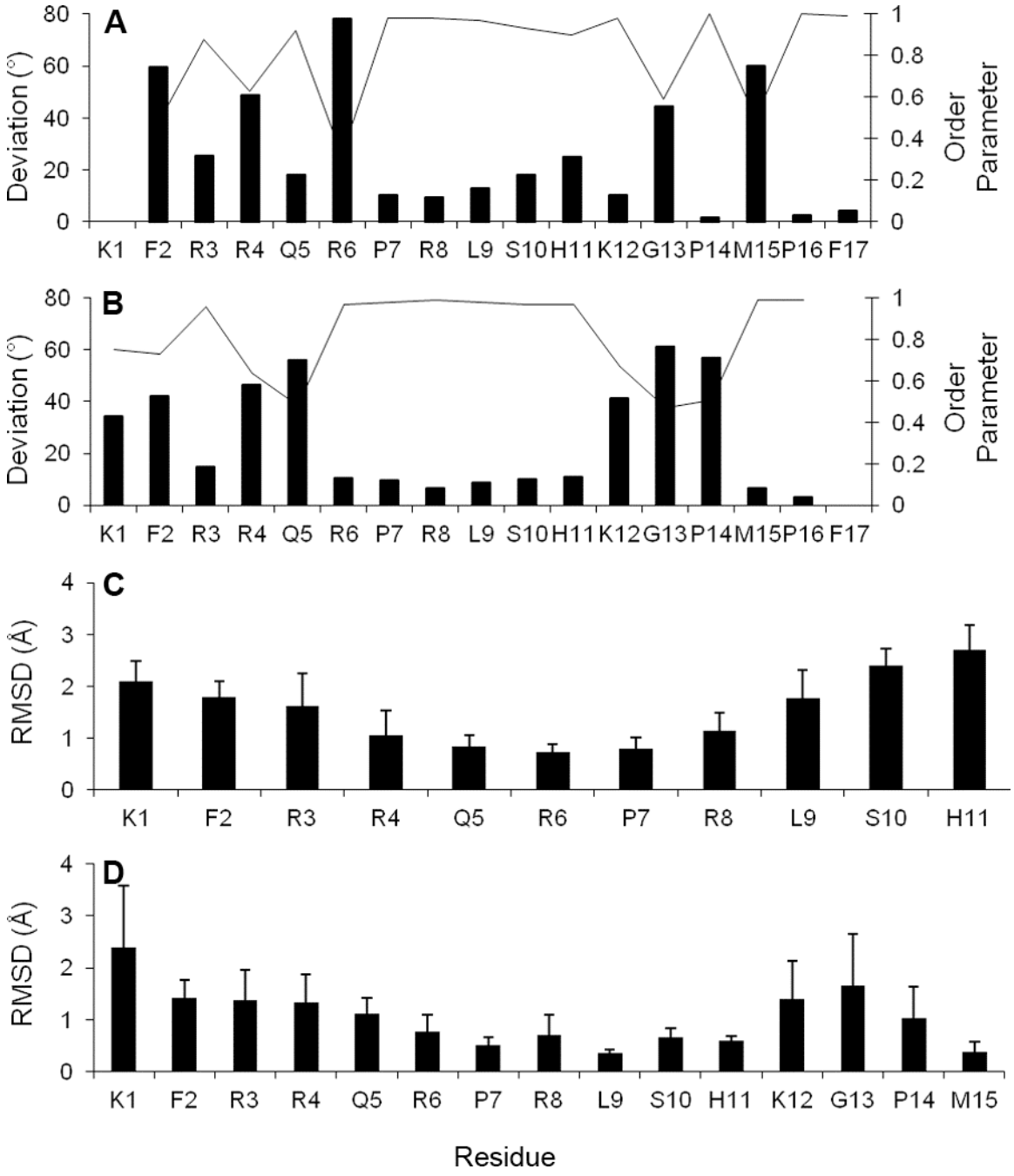

Figure 3.8: Analysis of the 80 member NMR-based structural ensemble of apelin-17 bound to SDS micelles: (A) φ and (B) ψ angle average deviation (bars) and order parameters (lines, (175)). RMSD calculated iteratively for (C) backbone atom superpositions over 7 residue segments and (D) all heavy atoms over 3 residue superpositions with each bar (error bar is average deviation) plotted at the first residue in a superposition.

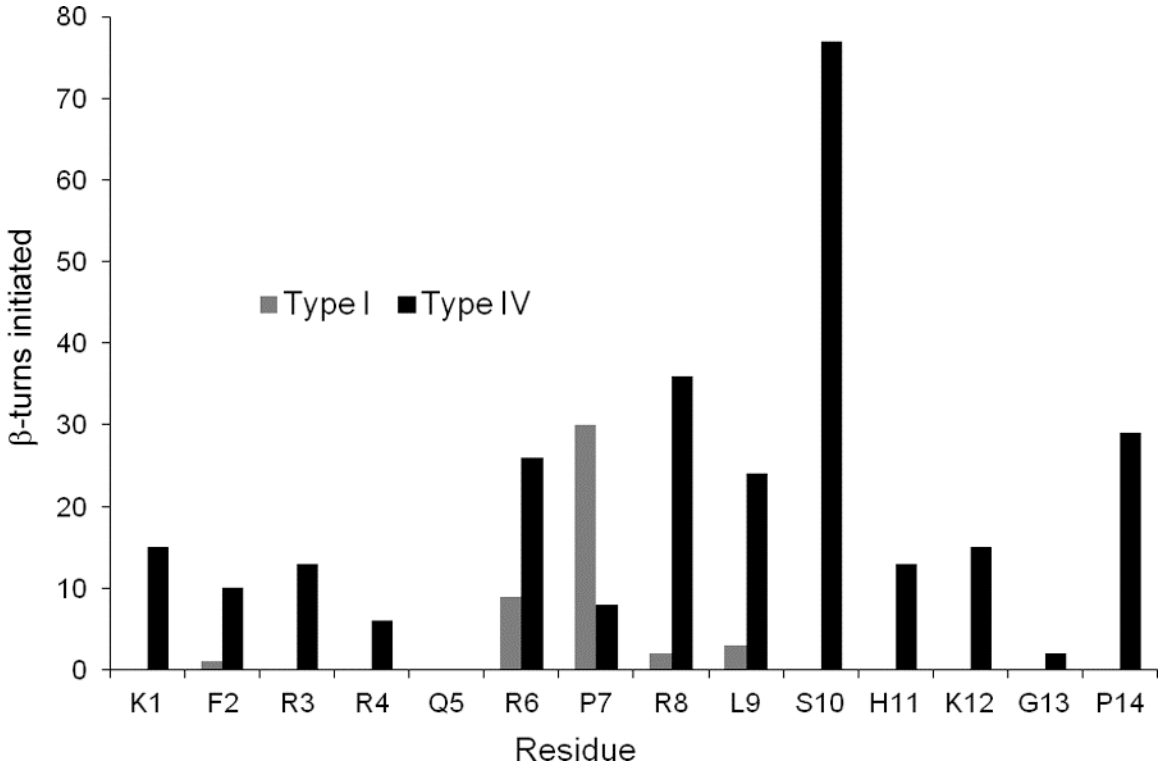

Figure 3.9: The presence of β-turns in the final 80 member ensemble as determined by Promotif-NMR. The number of times a type I or type IV β-turn is initiated at each residue in apelin-17 is shown.

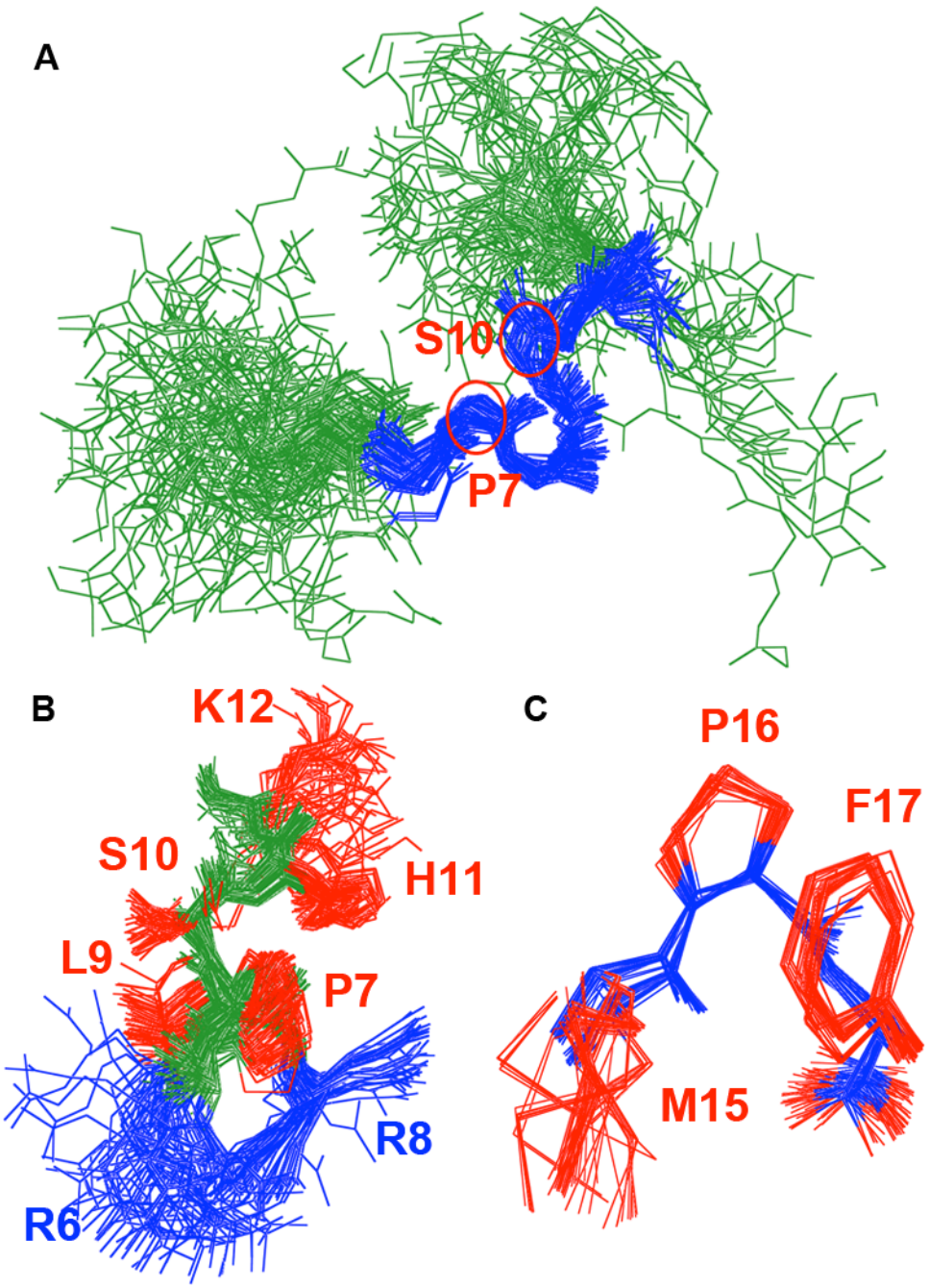

Figure 3.10: Structure of apelin-17 bound to an SDS micelle: A) superposition of the final ensemble of structures from R6-K12 (R6-K12 coloured blue, remainder green) with P7 and S10, initiation points of type I and type IV β-turns, respectively, indicated. B) Zoom of superposition in (A) (backbone atoms green) showing cationic side chains of R6 and R8 (blue; all other side chains red) falling on the same face of apelin-17. C) Superposition of all heavy atoms of M15-F17 for all ensemble members, with the peptide backbone shown in blue and side chains coloured red.

Chapter 4:

Structure And Function Of The First TM Segment And N-Terminal Tail Of AR

4.1. Introduction

In the previous two chapters I have described the structure of apelin-17 in solution and how it changes when binding to anionic membrane mimetics. Although this information is helpful for understanding how apelin-AR interactions take place, it does not provide information on AR itself. In this chapter, I continue my characterization of the apelin-AR system by determining the structure of a fragment of AR and by completing functional assays on full length AR.

4.1.1. Current Structural Data Of AR

Currently there is no published experimental structural data for AR. There have been two mutagenesis studies on AR that probe the binding to and activation of AR by apelin. However, the coverage of the receptor is too sparse to make any conclusions about the ligand binding site or mechanism of activation. Zhou et al. (100) showed that deletion of the first 10 amino acids has no effect on function while deletion of the first 20 residues completely disrupts binding between apelin and AR. Furthermore, mutations of E20 and D23 to alanine prevent apelin from binding despite normal receptor localization. Iturrioz et al. (128) examined W152, F255 and W259, demonstrating that mutation of W152 to alanine or phenylalanine in TM IV affected folding. Mutation of F255 to alanine or

tryptophan and W259 to alanine or phenylalanine in TM VI indicates the importance of these residues for internalization of AR but not apelin binding.

4.1.2. The Divide And Conquer Approach To Membrane Protein Structure

Full-length membrane proteins are generally challenging to characterize structurally (50). The "divide and conquer" approach provides a tractable avenue to collect structural information about a membrane protein of interest. In this approach, individual segments of membrane proteins are studied in a membrane mimetic environment and are assumed, following from the multi-step cycle of membrane protein folding (150, 190, 191, 192), to have structural integrity with the same segment in the context of a full-length protein (193). The clear pitfall of the divide and conquer technique is that tertiary and quaternary information of the receptor in question is lost. However, the divide and conquer approach has been extensively applied to many GPCRs (60, 61, 144, 145, 146, 147, 194, 195, 196).

The problem of lost tertiary information has been addressed in a variety of ways by using other techniques to determine tertiary structure, such as using electron paramagnetic resonance spin labels to generate long range distance restraints in bacteriorhodopsin (197). In another approach, NMR studies can also focus on a fragment containing multiple TM segments, which then provides some tertiary information about the full length receptor (154).

In this chapter I provide the high resolution structure of the N-terminal tail and first TM segment of AR fused to a polyhistidine tag (AR55) solubilized in DPC micelles. The structural data were used to guide mutagenesis and functional assays on full length AR. I

then discuss the significance of the current and previous mutagenesis data in the context of the AR55 structure.

4.2. Materials And Methods

4.2.1. Materials

pEXP5-CT vector, reagents for polymerase chain reaction (PCR) amplification of DNA and fetal bovine serum (FBS) were supplied by Invitrogen (Burlington, ON). Bacterial codon optimized AR55 as well as DNA primers for PCR amplification and mutagenesis were purchased from BioBasic Inc. (Markham, ON). HPLC grade acetonitrile, trifluoroacetic acid (TFA), dithiothreitol (DTT), ampicillin, bovine serum albumin (BSA), isopropyl β-D-1-thiogalactopyranoside (IPTG) and reagents for making LB media were purchased from Fisher Scientific (Ottawa, ON). DPC-d_{38}, DTT-d_6, deuterium oxide (D_2O; 99.8 atom % D) and D_2O containing 1% (w/w) DSS were obtained from C/D/N Isotopes (Pointe-Claire, QC). Uniformly ^{13}C labelled glucose and ^{15}N labelled ammonium chloride were obtained from Cambridge Isotope Laboratories (St. Leonard, QC). Non-cross reactive anti-rabbit Dylight 549 conjugated and anti-mouse Dylight 649 conjugated antibodies were purchased from Rockland (Gilbertsville, PA). The apelin-17 used in this study is the same as was prepared in chapter 2. Dulbecco's modified Eagle's medium (DMEM) and all other chemicals were obtained from Sigma-Aldrich (Oakville, ON).

4.2.2. Cloning And Expression Of AR55

PCR was used to amplify the purchased deoxynucleotide sequence encoding AR55 and clone it into pEXP5-CT vector according to the conditions in Table 4.1. The forward and reverse primers used had the sequences 5'-ATGGAAGAAGGCGGCGAT-3' and 5'-TTTGCGAAACACGGTCCA-3' respectively. The PCR product was then directly ligated into pEXP5-CT vector. Sequencing (BioBasic Inc.) was used to check for correct vector insertion of the PCR product to code for the following amino acids: MEEGGDFDNY YGADNQSECE YTDWKSSGAL IPAIYMLVFL LGTTGNGLVL WTVFRKKGHH HHHH, corresponding to the first 55 residues of AR as well as a C-terminal His_6-tag. High yield expression of uniformly $^{13}C/^{15}N$-labelled AR55 was achieved with BL-21(DE3) *E. coli* grown in LB medium to an OD_{600} of 0.6 and transferred to a one half equivalent volume of minimal medium (100 mM NaH_2PO_4 , 40 mM K_2HPO_4 , 4 mM $MgSO_4$, 1.8 μM $FeSO_4$, 2g/L uniformly ^{13}C labelled glucose, 1g/L of $^{15}NH_4Cl$ and 100 mg/L ampicillin, titrated with NaOH to pH 7.3), as previously described (198). Expression was induced by adding IPTG to 0.5 mM followed by shaking (37 °C, 4 h) and centrifugation of the cells (3000x*g*, 15 min). Expression and initial purification of AR55 was monitored by SDS-PAGE.

4.2.3. Purification Of AR55

Cell pellets were resuspended (50 mM Tris-HCl buffer, pH 8.0, 25% w/v sucrose, 1 mM EDTA, 0.1% w/v NaAzide, 5 mM MgCl, 0.03 mg/mL Dnase I and 1% Triton-X 100), sonicated and centrifuged (15,000x*g*, 30 min). The pellet was washed 4x (50 mM Tris-HCl buffer, pH 8.0, 0.5% Triton-X100, 100 mM NaCl, 1 mM EDTA). The resulting

inclusion body-containing pellet (199) was resuspended in acetonitrile:H_2O (1:1 v:v) containing 0.1% TFA, filtered (0.2 μm filter, Millipore, Billerica, MA) and reduced with DTT at ~100 mM prior to reversed-phase HPLC (Cosmosil C_{18} column, 5 μm particle, 20 × 250 mm) with a linear gradient from 45% to 65% acetonitrile in H_2O, 0.1% TFA over 15 min using a Beckman System Gold (Mississauga, ON)). Sample purity was assessed by analytical HPLC (AAPPTec Sprit Peptide C_{18} column, 5 μm particle, 4.6 mm × 250 mm) with a linear gradient from 2% to 80% acetonitrile in H_2O, 0.1% TFA over 39 min. Peaks were collected and immediately lyophilized, with AR55 peak identity confirmed by matrix assisted laser desorption ionization mass spectrometry using a Micromass M@LDI instrument (Waters, Mississauga, ON) at the NRC-IMB.

4.2.4. NMR Spectroscopy Of AR55

An NMR sample was prepared with 1 mM $^{13}C/^{15}N$ AR55 in 90% H_2O/10% D_2O containing 150 mM DPC-d_{38}, 20 mM $Na^+CD_3COO^-$, 1 mM DSS, 1 mM NaN_3 and 10 mM DTT-d_6 at pH 4.00±0.05, without accounting for deuterium isotope effects. ^1H-^{15}N HSQC, HNCO, HNcaCO, HNCA, HNcoCA, HNCACB, HNcoCACB (500 MHz) as well as ^{15}N-edited NOESY (800 MHz, 125 ms mixing time) experiments were acquired at the Québec / Eastern Canada High Field NMR facility (Montreal, QC) using Varian INOVA instruments. The HcCH TOCSY and a ^{13}C-edited NOESY (125 ms mixing time) experiments were collected using a 700 MHz Bruker Avance III spectrometer at NRC-IMB. Table 4.2 provides detailed acquisition parameters for the experiments. A second NMR sample was prepared with 0.5 mM ^{15}N AR55 and 77 mM DPC in the same buffer as the previous sample. ^1H-^{15}N HSQC spectra were acquired before and after the addition

of 2 mM apelin-17 on a Bruker 700 MHz Avance III spectrometer (NRC-IMB) in order to test for binding between AR55 and apelin-17.

4.2.5. Assignment And Structure Calculation Of AR55

The data were processed using NMRPipe (114) and assigned manually with CCPNMR Analysis version 2.1.5 (95). Assignments were made following the main chain directed approach (82). NOESY peak volumes were binned into distance restraint classes of 1.8-2.8 Å, 1.8-4 Å, 1.8-5 Å and 1.8-6 Å using CCPNMR Analysis. XPLOR-NIH version 2.19 (89, 90) was used for generation of 100 member structural ensembles. After each round of structure calculation, a Tcl/Tk script was used to identify restraint violations that were then either reassigned or lengthened. Eleven rounds of iterative structure calculation led to an ensemble of structures that converged, as indicated by reported XPLOR-NIH total energies. The 40 lowest energy structures were retained as the final ensemble. The final ensemble was then analyzed i) using Procheck-NMR (115) to determine the proportion of residues in favoured, allowed, generously allowed and disallowed regions of the Ramachandran plot; and ii) with an in house tcl/Tk script to calculate both the deviation and order parameter (as defined by Hyberts et al. (175)) of the ϕ and ψ dihedral angles. The LSQKAB software of the CCP4 suite (118) was used to superpose the AR55 ensemble over structurally converged regions and an in-house tcl/tk script was used to calculate the RMSD of the backbone atoms within the superposed region. The final structural ensemble of AR55 solubilized in DPC micelles has been deposited in the PDB ((200), accession #: 2LOU) using SMSDep. The chemical shift, spectral peak list and restraint data have been deposited to the BMRB ((121), accession # 18225).

4.2.6. Mutagenesis Of AR

3xHA-tagged AR was purchased from the Missouri University of Science and Technology cDNA resource center and was mutated using Phusion (New England Biolabs, Pickering, ON) or QuikChange (Stratagene, Wilmington, DE) mutagenesis kits according to the conditions in Table 4.1. Both alanine and isoleucine single point mutants were generated at G42, T43, T44, G45, N46 and G47. In addition N46D and N46Q mutants were constructed. All constructs were sequenced (Genewiz Inc, Cambridge, MA) for exactitude. The primer sequences and site directed mutagenesis kit used for each mutation are detailed in Table 4.3.

4.2.7. Extracellular Regulated Kinase (ERK) Phosphorylation

HEK 293A cells were subcultured into 6 well plates and grown in DMEM containing 10% FBS. At ~50% confluence, cells were transfected by adding 100 μL of a solution containing 2 μg of DNA, 20 μg of polyethylenimine and 100 μL DMEM. The media was aspirated and replaced with DMEM 8 hours post-transfection. Stimulation occurred 24 hours post-transfection with 1 μM apelin-17 for 5 min. After stimulation, the cells were harvested with 1 mL of cold phosphate buffered saline (PBS). Cell pellets were resuspended in 80 μL of radioimmune precipitation assay buffer (50 mM Tris/HCl, pH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, 0.5% sodium deoxycholate, 1% Nonidet P-40, 0.1% SDS, Complete Protease inhibitors (Roche, Laval, QC), DNase I, 2 mM sodium orthovanadate and 1 mM sodium fluoride) and 20 μL of 5x SDS sample loading buffer and then heated to 65 °C for 10 min. The lysate was resolved by SDS-PAGE at 200 V for

45 min, with transfer onto nitrocellulose membrane at 115 V for 75 min. The membranes were blocked in blocking buffer (Tris buffered saline with 2% w/v BSA) for 1 hour and then incubated in blocking buffer with primary antibodies for ERK (polyclonal rabbit, 1:1000) and phospho-ERK (monoclonal mouse, 1:500) at 4 °C overnight. Secondary antibodies (1:2000 donkey anti-rabbit and 1:1000 donkey anti-mouse) were applied for 1 hour in blocking buffer. After final washing and drying the membrane was imaged on a VersaDoc imaging system (Biorad, Mississauga, ON).

4.2.8. Immunofluorescence Microscopy

HEK 293A cells were grown on coverslips in DMEM containing 10% FBS until about 50% confluence. Transfection occurred as earlier except only 12 µg of PEI was used and media was refreshed with DMEM with 10% FBS. Cells were washed in warm PBS and then fixed in 2% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA) for 10 min 24 hours post-transfection. Coverslips were blocked with PBS containing 5% normal donkey serum for 20 min then incubated with primary antibody (1:1000 mouse anti HA, Covance, Montreal, QC) for 1 hour. After washing, the coverslips were incubated with secondary antibody (1:500 Cy5 conjugated Donkey anti mouse, Jackson Immunoresearch, West Grove, PA) for 30 min. Fluorescence mounting medium (Dako, Burlington, ON) was used for mounting the coverslips. Fluorescence images were acquired on a custom built Zeiss Axio Observer Z1 inverted microscope (Intelligent Imaging Innovations (3i), Boulder, CO), equipped with a solid state Spectra light source (Lumencor, Beaverton, OR). The cells were observed using a 40X objective (1.3 N.A) and images were recorded using an HQ2 charge-coupled device camera (Photometrics, Tucson, AZ, USA) and Slidebook 5.0 software (3i, Boulder CO). For each AR mutant, at

least three representative fields of view were collected. Further image processing was done using Photoshop CS5 (Adobe).

4.3. Results And Discussion

4.3.1. Purification Of AR55

Expression of $^{13}C/^{15}N$ uniformly labelled AR55 was successfully induced in BL-21(DE3) *E. coli* with IPTG. An initial inclusion body purification provided AR55 as the main component of the inclusion body pellet (Figure 4.1). AR55 was further enriched to > 95% purity through HPLC (Figure 4.2), with a final yield of ~5 mg/L. Mass spectrometry analysis indicated the mass of purified $^{13}C/^{15}N$ AR55 to be 7658.8 Da (Figure 4.3). The expected masses of unlabelled and 100% $^{13}C/^{15}N$ labelled AR55 are 7296 Da and 7708 Da, respectively, indicating ~88% efficiency of ^{13}C and ^{15}N incorporation.

4.3.2. Structure Of AR55 In DPC Micelles

The ^{1}H-^{15}N HSQC spectrum of AR55 solubilized in DPC micelles has extensive resonance overlap (Figure 4.4), however, spectral resolution using 3D NMR experiments was still sufficient for resonance assignment. Nearly complete assignment of ^{1}H, ^{13}C and ^{15}N for AR55 was accomplished using the main chain directed approach, TOCSY and NOESY spectra (Table 4.4). Through analysis of the ^{13}C- and ^{15}N- edited NOESY spectra, 1608 unique NOE contacts were observed for AR55 (Table 4.5). The NOE restraints were not evenly distributed throughout the peptide with residues L30-K57 having the most medium range ($|i-j| \leq 4$) restraints (Figure 4.5), suggesting that the TM

region of AR55 is more structured than the N-terminal tail. Residues G42-G47 also have a reduced number of NOE restraints, indicating that this region may be less structured than the area around it. This notion is supported by the pattern of NOE contacts, chemical shift index (CSI, (106)) and DANGLE (201) predictions (Figure 4.6). Both CSI and DANGLE predict protein secondary structure based solely on the chemical shifts of its nuclei. Canonical α-helical NOE contacts (αN (i, i+3), αN (i,i+4) and αβ (i, i+3), (124)) are observed from N15-W24, A29-L41 and T44-R55 while analysis of the chemical shift deviation from random coil (106) predicts extensive helical structure from I31-K57. Finally, the DANGLE algorithm (201) predicts the TM region of AR55 to be composed of two helices with a disruption in helical character from G42-N46 and the presence of a third helix in the N-terminal tail of AR from N15-K25 (Figure 4.6).

Using the experimental distance restraints and iterative refinement, 11 sets of 100 member structural ensembles were generated. The 40 lowest energy structures of the last round were taken as the final 40 member structural ensemble of AR55, which agrees with the experimental data and has good Ramachandran plot statistics (Table 4.5). Analysis of dihedral angle variability in the structural ensemble (Figure 4.7) indicates that the D14-C19, E20-K25, A29-N46 and G47-K57 regions of AR55 are converged, displaying high ϕ and φ dihedral angle order parameters over these ranges. Each of these regions is superposed on a representative ensemble member in Figure 4.8, with backbone atom RMSD values of 1.30, 0.97, 0.68 and 0.65 Å respectively. The remainder of AR55, including parts of the N-terminal tail and the His_6-tag, is seen to be largely disordered.

TM I in AR55 has a helix-kink-helix structure, with the kink occurring in the relatively polar region (GTTGNG) around N46. The angle of the kink is highly variable, with

measurement by the MC-HELAN algorithm (chapter 6, (4)) providing an ensemble average kink angle of 69° ± 20°. The N-terminal tail of AR55 has a converged helix from residues D14-K25, with a flexible point at C19-E20. In this region the E20 and D23 side chains are on the same side of a helix and produce an anionic surface (Figure 4.9).

4.3.3. G42, G45 And N46 Are Required For Proper Folding And Function Of AR

After analysis of the AR55 structure, it was clear that the polar residues surrounding the kink may be important for function. With this in mind, residues G42-G47 were selected for functional characterization by mutagenesis. Following previous demonstration of AR stimulation leading to ERK phosphorylation in transfected HEK 293A cells (35), this phosphorylation process was monitored for transiently transfected 3xHA-AR mutants in HEK 293A cells upon stimulation with 1 μM apelin-17 for 5 min. Residues G42, G45 and N46 showed significant sensitivity to mutation while T43, T44 and G47 did not (Figure 4.10). The residue at each position was substituted for both alanine and isoleucine, with isoleucine generally having a more pronounced effect. N46 was particularly sensitive to mutation; the N46A, N46D and N46I mutants all had severely impaired function while the N46Q mutant functioned well.

Through immunofluorescence microscopy, the general localization of both wild-type 3xHA-AR and each of the AR mutants was established. For the wild-type receptor, proper membrane trafficking and localization of 3xHA-AR was apparent (Figure 4.11). There were 5 mutants (G42I, G45I, N46A, N46D, N46I) where the surface localization of the receptors appeared to be impaired relative to the wild-type receptor. GPCR

maturation begins after translation when it is post-translationally modified in the endoplasmic reticulum and Golgi apparatus (202). Therefore, the abnormal surface distribution seen may already begin when AR is being processed intracellularly. Many GPCRs are present in homo- and heterodimers and it is likely that for these mutants these protein-protein protein interactions are dysfunctional. The data that I have collected is insufficient to determine if the inactive AR mutants are inactive due to improper folding of the receptor, or due to the receptor undergoing completely non-native protein-protein interactions. However, given the importance of N46 for interhelical hydrogen bonding within AR there is likely to be improper folding of AR.

4.3.4. G42, G45 And N46 Are Likely Essential For Interhelix Interactions

As is clear from both ERK phosphorylation (Figure 4.10) and AR localization (Figure 4.11), mutations at G42, G45 or N46 severely affect the folding of AR. These residues are in or proximal to the kink in TM I that is observed in the AR55 NMR structure (Figure 4.8). N46 is completely conserved in class A GPCRs (203). In all of the currently solved crystal structures of GPCRs, the side chain of this asparagine is H-bonded to the backbone carbonyl oxygen of a serine (B_1AR, B_2AR, A2A, dopamine, histamine and M2 muscarinic receptors) (51, 52, 53, 55, 56, 58), cysteine (CXCR4) (57) or alanine (bovine rhodopsin) (54) in TM VII, highlighting its importance for proper folding of AR. The N46D mutant is inactive, while stimulation of the N46Q mutant causes the same response as wild-type AR. This indicates that the amide moiety of the side chain is essential, but that there is ability to accommodate variation in size. Given this, the reason an N to Q

substitution at residue 46 is not more prevalent in nature is an interesting question for future consideration.

Although the structure of TM I is quite variable in GPCRs, both bovine rhodopsin and the CXCR4 chemokine receptor have a kink near the N46 position in TM I (Figure 4.12). There is also little mutational data in GPCRs at this conserved asparagine. Since the conserved asparagine is H-bonded to TM VII, which changes conformation during GPCR activation (73), the proximal kink in TM I may simply allow flexibility for the conformational changes required for activation to occur.

Residues G42 and G45 are highly conserved in class A GPCRs (40-60% and 60-80% respectively, (203)) but do not show specific conserved inter-helix interactions as seen with N46. Rather, these are generally on the same face of the TM I helix, implying that the small glycine side chains may be generally employed in TM I to allow both proper packing the proximal TM VII segment (204).

4.3.5. A Potential Role For T43 And T44?

Sequence analysis has provided a map of evolutionary conservation in GPCRs. Although their position in the GPCR sequence is somewhat conserved (40-60%), T43 and T44 appear to be rather unique, being present in only ~3% of class A GPCRs. Even in the closely related AT_1 receptor, the amino acids at the corresponding positions are hydrophobic residues. However, in this study neither T43 nor T44 appeared to be very sensitive to mutation. In related GPCR structures, the corresponding residues face the lipid bilayer. If T43 and T44 face the lipid bilayer, this would render them accessible for intermolecular interactions and/or lead to association of water molecules that could in turn disrupt lipid tailgroup packing and allow for increased flexibility of the TM I kink.

4.3.6. E20 And D23 May Form A Binding Face For Apelin

The 12-residue region of the N-terminal tail that showed structural convergence contains the residues E20 and D23 previously shown to be essential for apelin binding to AR (100). These residues are structured such that both anionic side-chains are located on the same face of the tail. Taken in context with the two-step model of GPCR-ligand binding, it is possible that the first stage of binding between apelin and AR would be promoted by an electrostatic interaction between the cationic face from R6-L9 of apelin that I discuss in chapters 2 and 3 and the anionic surface created by E20 and D23. TM I of GPCRs have been shown to play a role with ligand binding, where in the dopamine D1 and D5 receptors the N-terminal tail and TM I of the receptor influences ligand specificity and efficacy (205).

In pursuit of quantifying the interaction of AR55 with apelin-17, the two molecules were titrated together. However, no significant perturbations were seen in ^{1}H-^{15}N HSQC spectra of AR55 (Figure 4.13). This suggests that E20 and D23, although essential for apelin to bind to AR, are not sufficient for significant binding to occur.

4.3.7. Comparison Of The AR55 Structure With An AR Model

An atomistic molecular dynamics simulation of AR was generated by Aaron Banks in the Rainey lab. With no structural data present for full length AR, this model helps to put the structure of AR55 into context with the rest of AR. Overall, the observations seen for AR55 hold true in the generated model. The kink at N46 is still present in the AR model and the side chain of N46 forms a hydrogen bond to TM VII which would be expected

based on other GPCR crystal structures. Also T43 and T44 face the lipid bilayer and have several associated water molecules. Although this does not indicate that T43 and T44 are important for protein-protein interactions, this model does clearly indicate that the hydroxyl groups of the threonine residues are not stabilized by AR itself. Finally, the anionic side chains of E20 and D23 continue to create an anionic face in the extracellular region of the AR model. In addition to providing confirmation to the structural features observed in AR55, the agreement between the AR model and AR55 structure suggests that the divide and conquer technique is applicable to AR.

4.4. Summary

Overall, this study illustrates a strong correlation between functionally essential residues and structural features in the N-terminus and first TM segment of AR. AR55 solubilized in DPC micelles consists of two disrupted helices, spanning residues D14-K25 and A29-K57. Residues E20 and D23 form an anionic face, giving mechanistic insight into previous mutagenesis studies identifying these residues as essential for binding of apelin to AR. A kink in the otherwise helical TM I segment is located at the conserved N46 that could provide flexibility during activation, with a pair of proximal glycine residues allowing both flexibility and close inter-helical packing.

Table 4.1: The reaction composition and thermocycling conditions used for PCR amplification of DNA.

	AR55 cloning	**Phusion site directed mutagenesis of AR**	**QuikChange site directed mutagenesis of AR**
Reaction composition			
Volume (μL)	50	20	50
Buffer	1x PCR Buffer no Mg^{2+}	1x Phusion HF reaction buffer	1x *Pfu* Ultra HF reaction buffer
Template DNA (ng)	20	1	100
[Primers] (μM)	0.5	0.5	0.2
Final [Mg^{2+}] (mM)	2	1.5	2
[DNTPs] (μM)	200	200	250
Polymerase type and amount	Platinum *Taq*, 2 U	Phusion, 0.8 U	*Pfu* Ultra polymerase II, 2.5 U
Cycling conditions			
Initial denaturation	94 °C, 5 min	98 °C, 30 sec	95 °C, 1 min
Number of cycles	39	25	15
Melting step	94 °C, 1 min	98 °C, 10 sec	95 °C, 30 sec
Annealing step	50 °C, 2 min	N/A (2 step PCR)	55 °C, 30 sec
Extension step	72 °C, 1 min	72 °C, 3 min	72 °C, 2 min
Final Extension	72 °C, 15 min	72 °C, 10 min	72 °C, 10 min

Table 4.2: Detailed experimental parameters used for data acquisition of AR55 solubilized in DPC 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	1.2	4	^{1}H:1024 ^{13}C:256	^{1}H:8000 ^{13}C:10053	^{1}H:4.66 ^{13}C:34.9	500	QANUC
^{1}H-^{15}N HSQC	1.2	4	^{1}H:1024 ^{15}N:128	^{1}H:8000 ^{15}N:1418	^{1}H:4.66 ^{15}N:117	500	QANUC
HNCO	1.2	4	^{1}H:1024 ^{13}C:64 ^{15}N:32	^{1}H:8000 ^{13}C:1257 ^{15}N:1418	^{1}H:4.66 ^{13}C:175 ^{15}N:117	500	QANUC
HNCA	1.2	8	^{1}H: 1024^{13}C:64 ^{15}N:60	^{1}H:8000 ^{13}C:3770 ^{15}N:1418	^{1}H:4.66 ^{13}C:56 ^{15}N:117	500	QANUC
HNcoCA	1.2	16	^{1}H:1024 ^{13}C:64 ^{15}N:52	^{1}H:8000 ^{13}C:3770 ^{15}N:1165	^{1}H:4.66 ^{13}C:56 ^{15}N:115.5	500	QANUC
HNCACB	1.2	16	^{1}H:1024 ^{13}C:100 ^{15}N:52	^{1}H:8000 ^{13}C:8796 ^{15}N:1165	^{1}H:4.66 ^{13}C:46 ^{15}N:115.5	500	QANUC
HcCH_TOCSY	1.5	8	^{1}H:4096 ^{1}H:128 ^{13}C:48	^{1}H:9804 ^{1}H:6161 ^{13}C:6689	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:29.6	700	NMR-3
NOESY-^{13}C HSQC	1.5	8	^{1}H:2048 ^{1}H:170 ^{13}C:56	^{1}H:9804 ^{1}H:6161 ^{13}C:6689	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:29.6	700	NMR-3
NOESY-^{15}N HSQC	1.2	8	^{1}H:1024 ^{1}H:220 ^{15}N:60	^{1}H:8000 ^{1}H:6246 ^{15}N:1418	^{1}H:4.66 ^{1}H:4.66 ^{15}N:117	500	QANUC

Table 4.3: Primer sequences and mutagenesis method used to construct the AR mutants.

Mutant	Forward Primer (5'-3')	Reverse Primer (5'-3')	Mutagenesis kit used
G42A	GGTCTTCCTCCTGGCCAC CACGGGAAACGG	CCGTTTCCCGTGGTGGCC AGGAGGAAGACC	QuikChange
G42I	GTTGGTCTTCCTCCTGATC ACCACGGGAAACGG	CCGTTTCCCGTGGTGATC AGGAGGAAGACCAAC	QuikChange
T43A	CTTCCTCCTGGGCGCCAC GGGAAACG	CGTTTCCCGTGGCGCCC AGGAGGAAG	QuikChange
T43I	GTCTTCCTCCTGGGCATC ACGGGAAACGGTCTG	CAGACCGTTTCCCGTGA TGCCCAGGAGGAAGAC	QuikChange
T44A	CTCCTGGGCACCGCCGGA AACGGTCTG	CAGACCGTTTCCGGCGG TGCCCAGGAG	QuikChange
T44I	CCTCCTGGGCACCATCGG AAACGGTCTGGTG	CACCAGACCGTTTCCGA TGGTGCCCAGGAGG	QuikChange
G45A	CTGGGCACCACGGCAAAC GGTCTGGTG	CACCAGACCGTTTGCCG TGGTGCCCAG	QuikChange
G45I	CTCCTGGGCACCACGATC AACGGTCTGGTGCTC	GAGCACCAGACCGTTGA TCGTGGTGCCCAGGAG	QuikChange
N46A	GGCACCACGGGAGCCGG TCTGGTGCTC	CAGGAGGAAGACCAACA TGTAGATG	Phusion
N46D	GACGGTCTGGTGCTCTGG ACC	TCCCGTGGTGCCCAGGA G	Phusion
N46Q	CAGGGTCTGGTGCTCTGG ACC	TCCCGTGGTGCCCAGGA G	Phusion
N46I	GGCACCACGGGAATCGGT CTGGTGCTC	CAGGAGGAAGACCAACA TGTAGATG	Phusion
G47A	ACCACGGGAAACGCCCTG GTGCTCTGG	GCCCAGGAGGAAGACCA ACATGT	Phusion
G47I	ACCACGGGAAACATCCTG GTGCTCTGG	GCCCAGGAGGAAGACCA ACATGT	Phusion

Table 4.4: The number of possible assignments and the numbers of assigned resonances for different isotopes and resonance types of AR55 in DPC micelles. Data for this table were generated using CCPNMR Analysis (95).

Category	**Total possible**	**AR55_DPC assignments**	
		#	**%**
Carbon	315	270	86
Proton	363	358	99
Nitrogen	88	70	80
Amide	126	124	98
Backbone	254	249	98
Backbone non-^{1}H	192	187	97
Side chain ^{1}H	301	296	98
Side chain non-^{1}H	211	153	73

Table 4.5: Summary of iterative structure calculation protocols, restraints and structural statistics of the lowest energy 40 structures (out of 100) for AR55 solubilized in DPC micelles.

Rounds of structure calculation	11
Unique NOE restraints	
Total	1608
Intraresidue	788
Sequential	424
Medium range ($\|i-j\| \leq 4$)	322
Long range ($\|i-j\| > 4$)	0
Ambiguous	74
Residue Ramachandran plot statistics	
Core	39.2%
Allowed	47.3%
Generously allowed	9.3%
Disallowed	4.2%
XPLOR-NIH energies (kcal/mol)[a]	
Total	39.8 ± 1.7
NOE	1.1 ± 0.6
Violations	
NOE violations > 0.5 Å	0
NOE violations of 0.3-0.5 Å	1
NOE violations > 0.2-0.3 Å	0

[a] Ranges are given by average deviations for XPLOR-NIH energies.

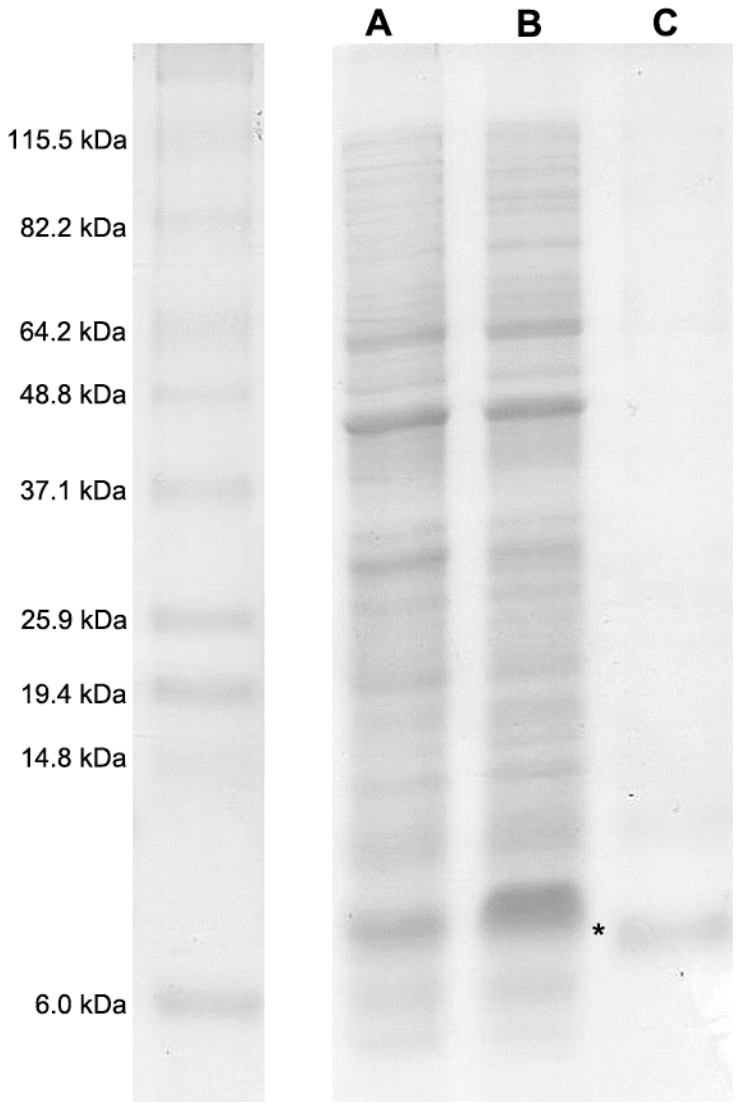

Figure 4.1: A 15% SDS-PAGE gel after electrophoresis and coomassie blue staining demonstrating successful expression and inclusion body purification of AR55. The Benchmark protein ladder (Invitrogen) is shown on the left, with cell pellets from BL-21 (DE3) *E. coli* (A) without induction, (B) 4 hours after induction with 0.5 mM IPTG and (C) the resolubilized pellet after the inclusion body purification. The band corresponding to AR55 is denoted by *.

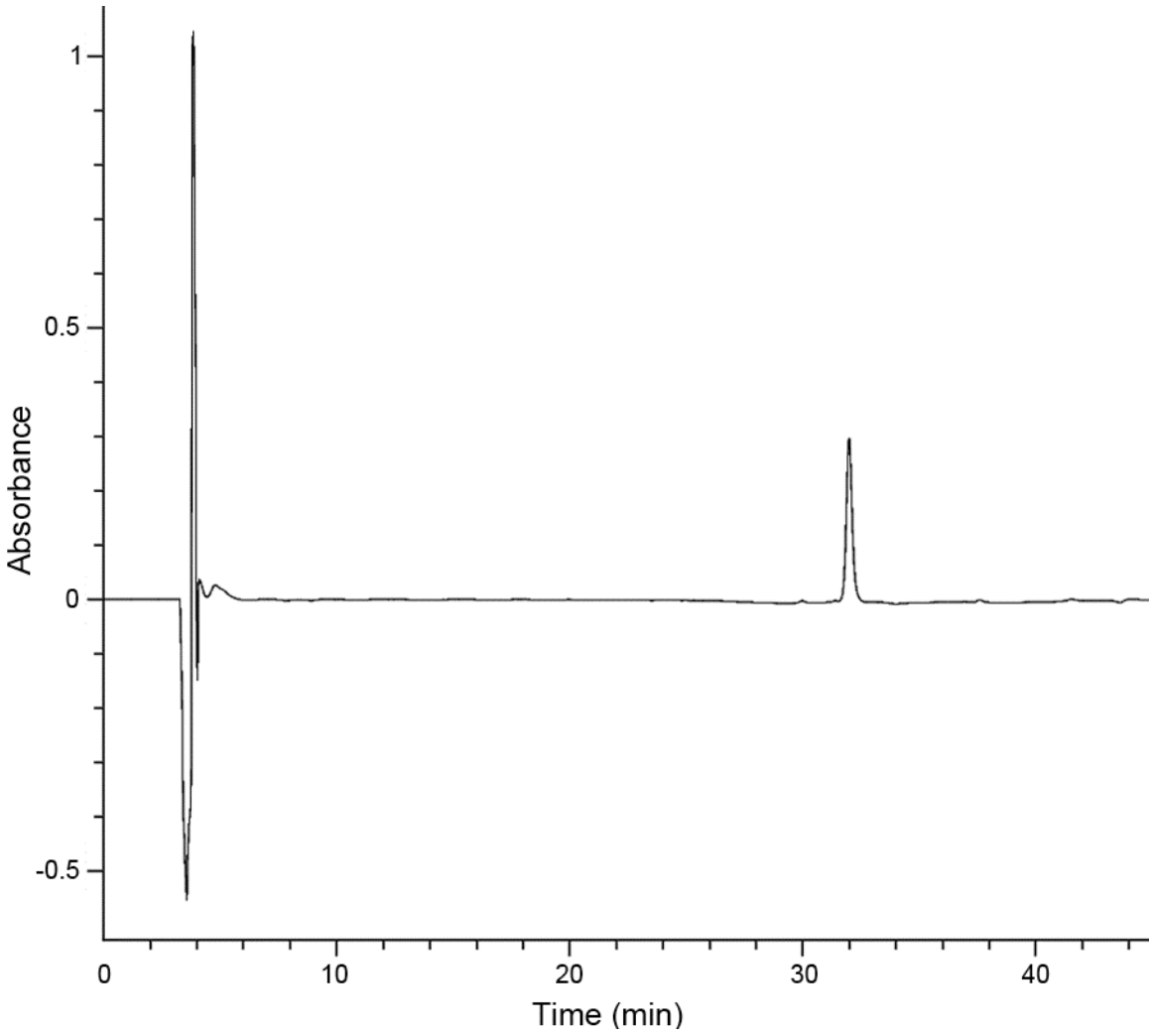

Figure 4.2: Reversed-phase HPLC analysis of purified AR55. HPLC was carried out using a C_{18} column (5µm particle, 4.6 mm × 250 mm Spirit Peptide column, AAPPTec) with a water:acetonitrile solvent mixture progressing from 2% to 80% B over 39 min. Only one major peak is visible that corresponds to AR55 (Figure 4.3) and elutes at ~32 min. The initial peak at 2 min is due to injection of the sample.

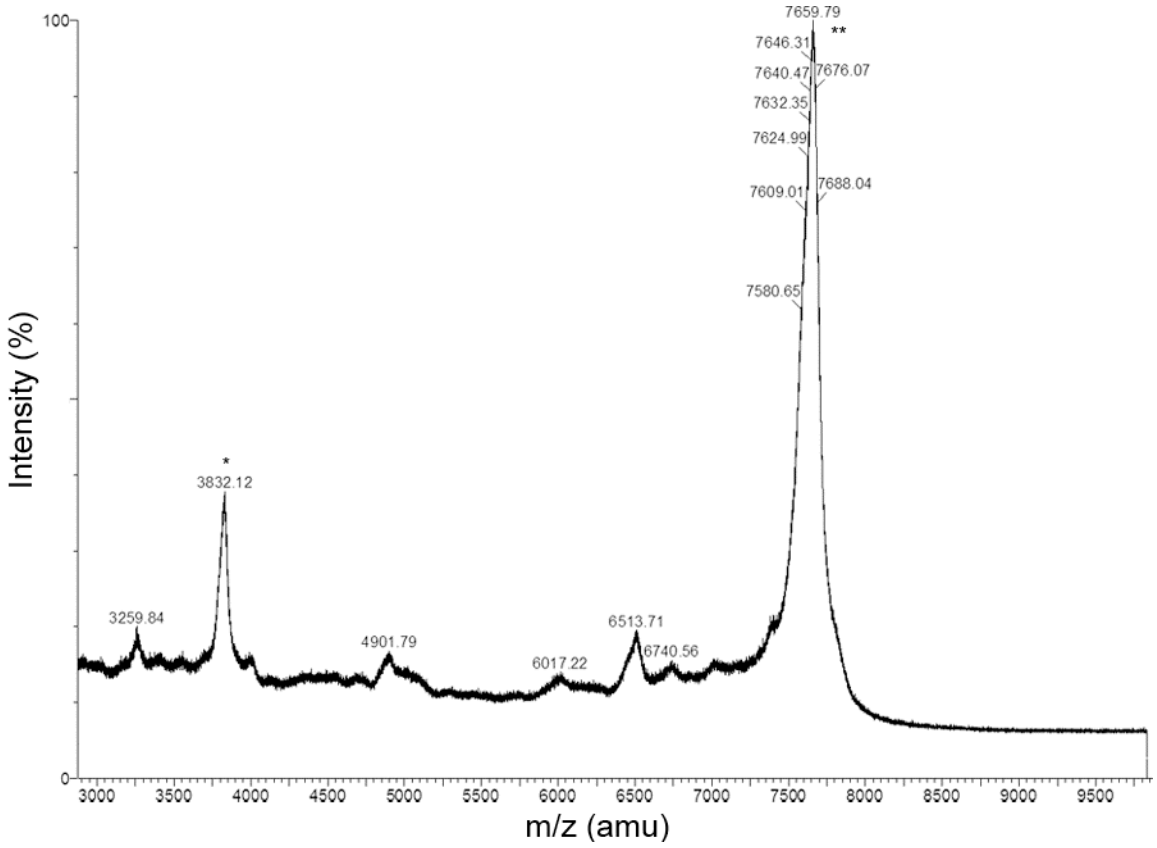

Figure 4.3: Matrix assisted laser desorption ionization mass spectrometry analysis of the sample shown in Figure 4.2. The peak denoted by ** corresponds to $^{13}C/^{15}N$ uniformly labelled AR55 (m/z = 7659.8 amu, expected mass of 100% $^{13}C/^{15}N$ labelled AR55 is 7708 amu). The second major peak denoted by * likely corresponds to the +2 ion (m/z = 3832 amu, ~ half the value of the major peak).

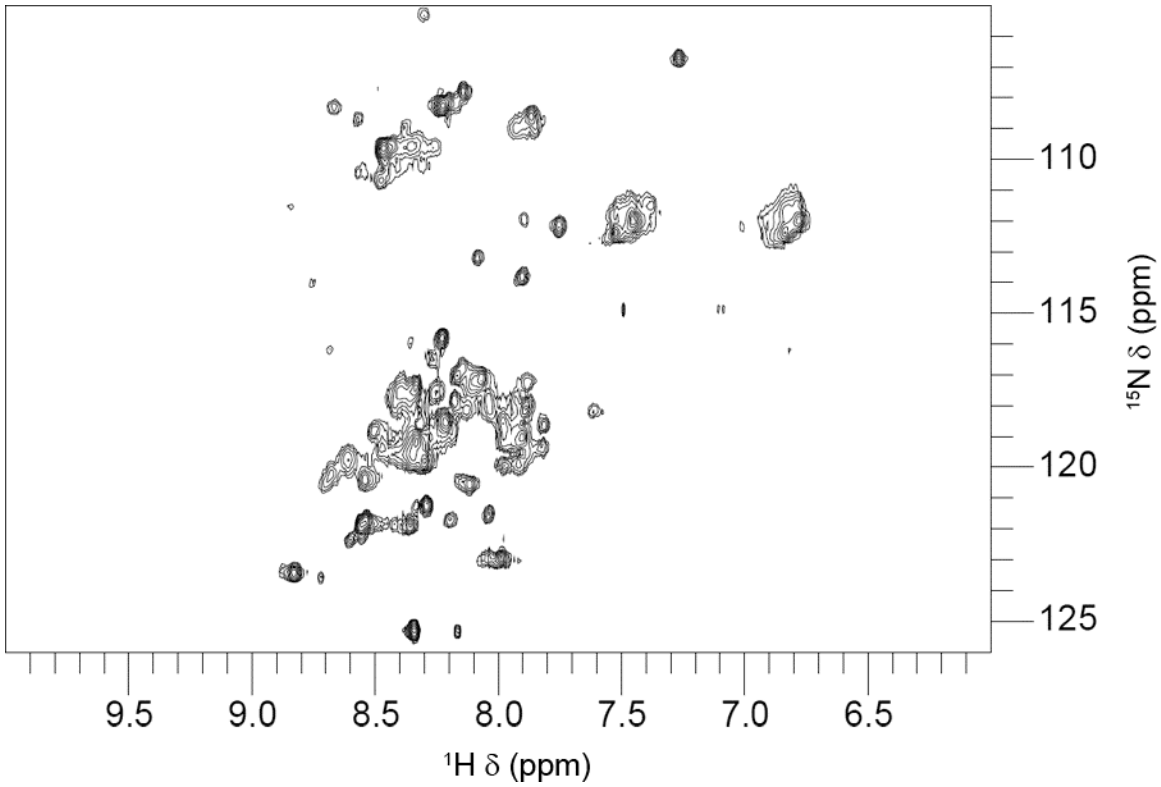

Figure 4.4: ^{1}H-^{15}N HSQC spectra of AR55 solubilized in DPC micelles at 37 °C. The data were acquired at 700 MHz at NRC-IMB.

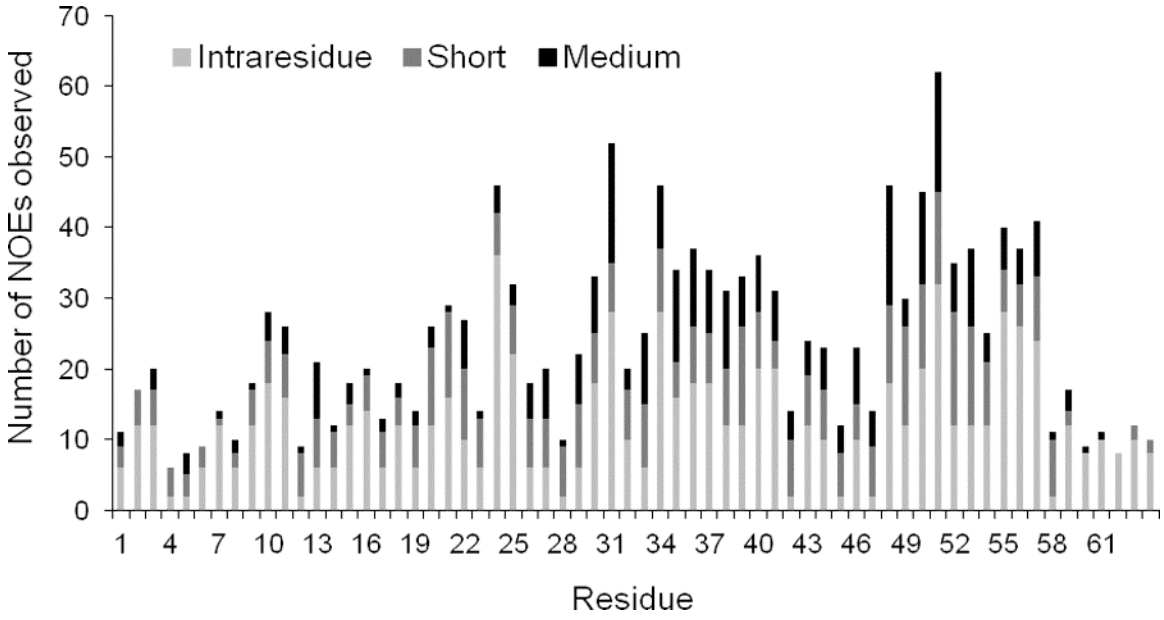

Figure 4.5: A breakdown of per-residue NOE contacts used to calculate the final ensemble of structures, with short (sequential) and medium ($|i-j| \leq 4$) NOE interactions shown for AR55 solubilized in DPC micelles. Note that these plots do not include the ambiguous NOE contacts shown in Table 4.5.

Figure 4.6: Through-space contacts observed for AR55 solubilized in DPC micelles. NOE contacts between indicated resonances are indicated by horizontal lines. The thicker the horizontal line, the closer the two resonances are to each other. Values of CSI (106) as well as secondary structure predictions using the DANGLE algorithm (201) are shown. This figure was generated using CCPNMR Analysis (95).

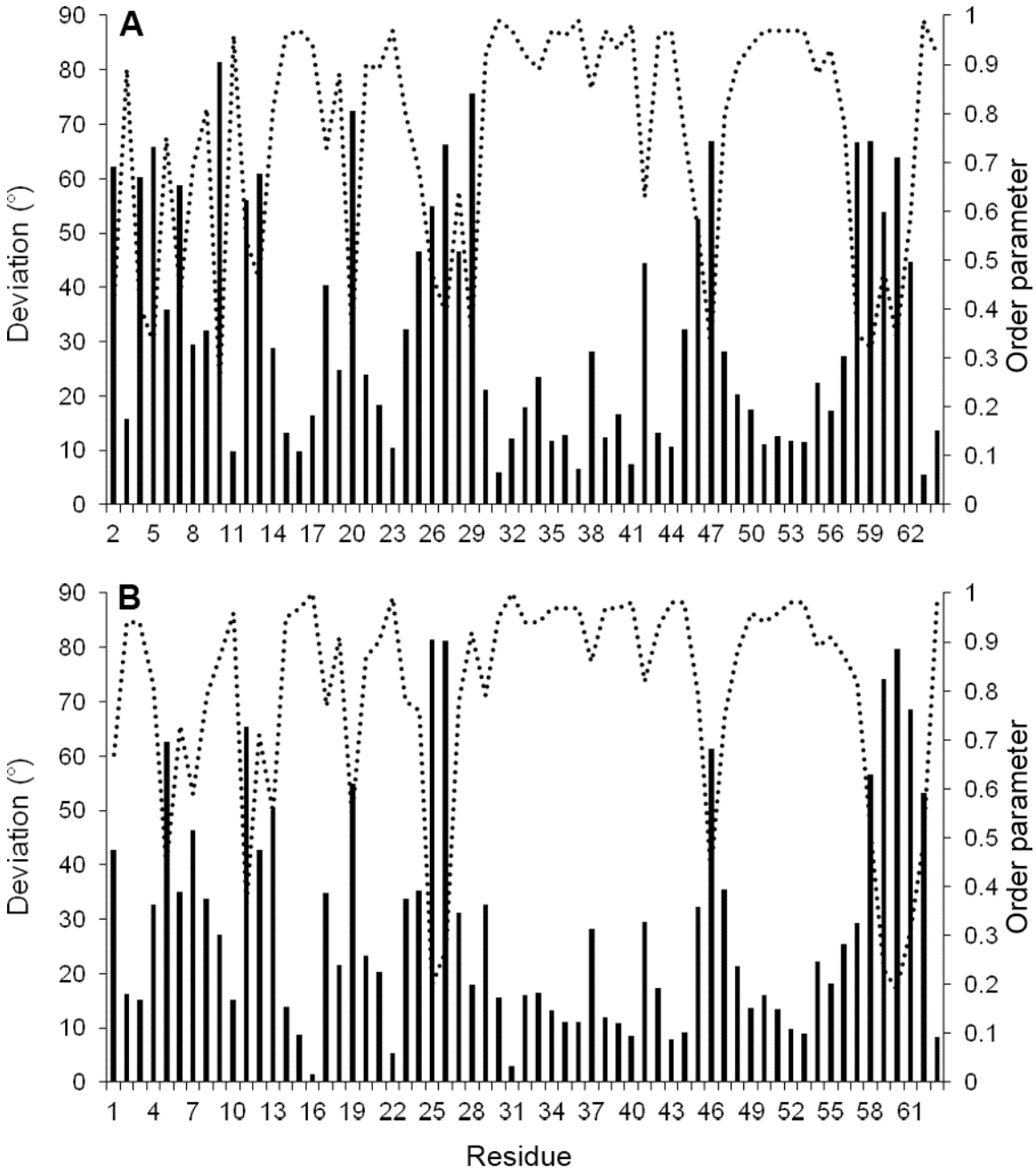

Figure 4.7: An overview of the convergence of the ϕ (A) and φ (B) angles of the members of the AR55 structural ensemble. Both the deviation (bars) and order parameter (dashed line) are indicated.

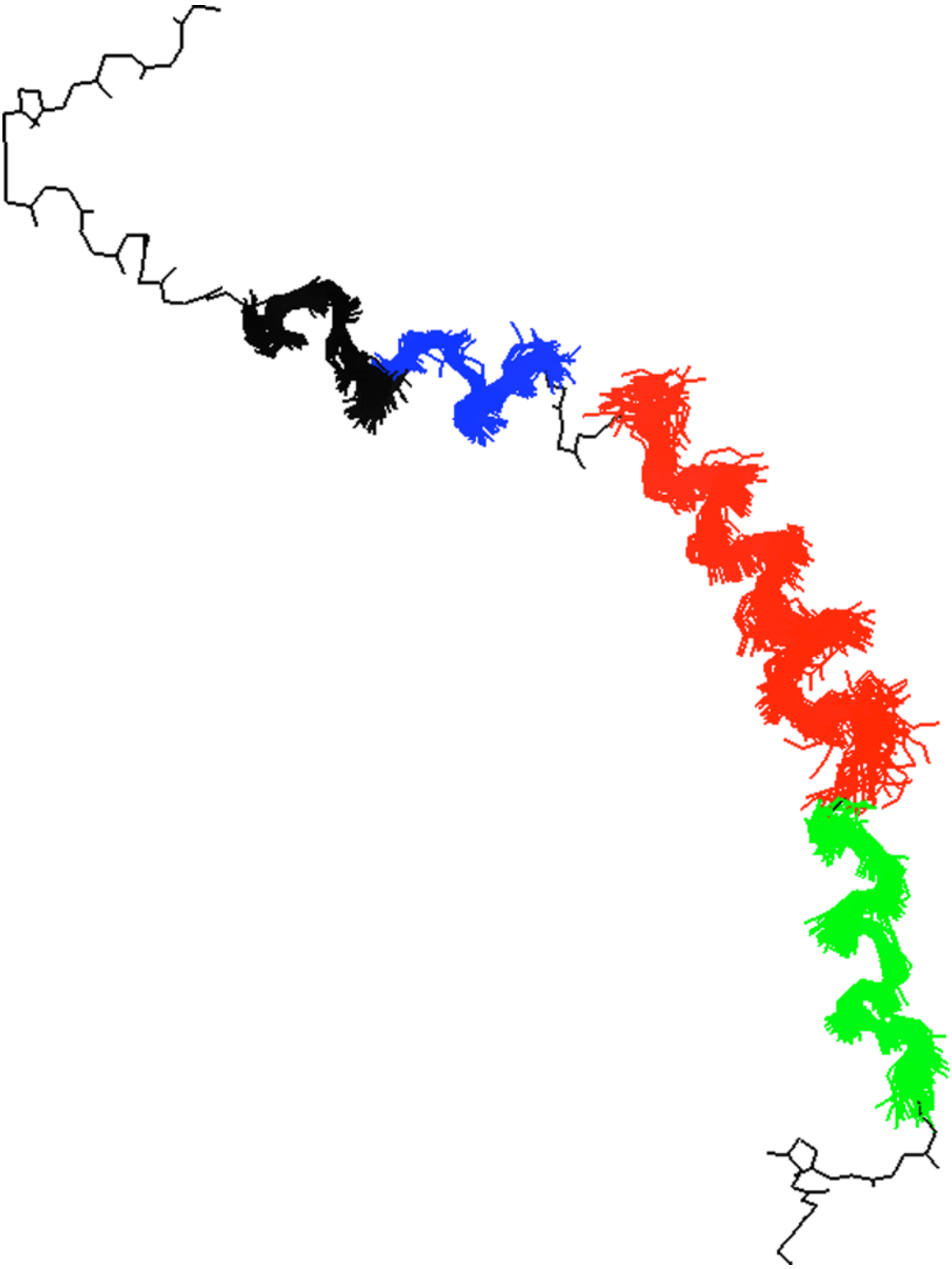

Figure 4.8: A superposition of 40 members from the AR55 structural ensemble created using Pymol. The main chain of one structure is shown in black. Over areas of structural convergence the backbone of all 40 members are superposed (D14-C19 in black, E20-K25 in blue, A29-N46 in red and G47-K57 in green).

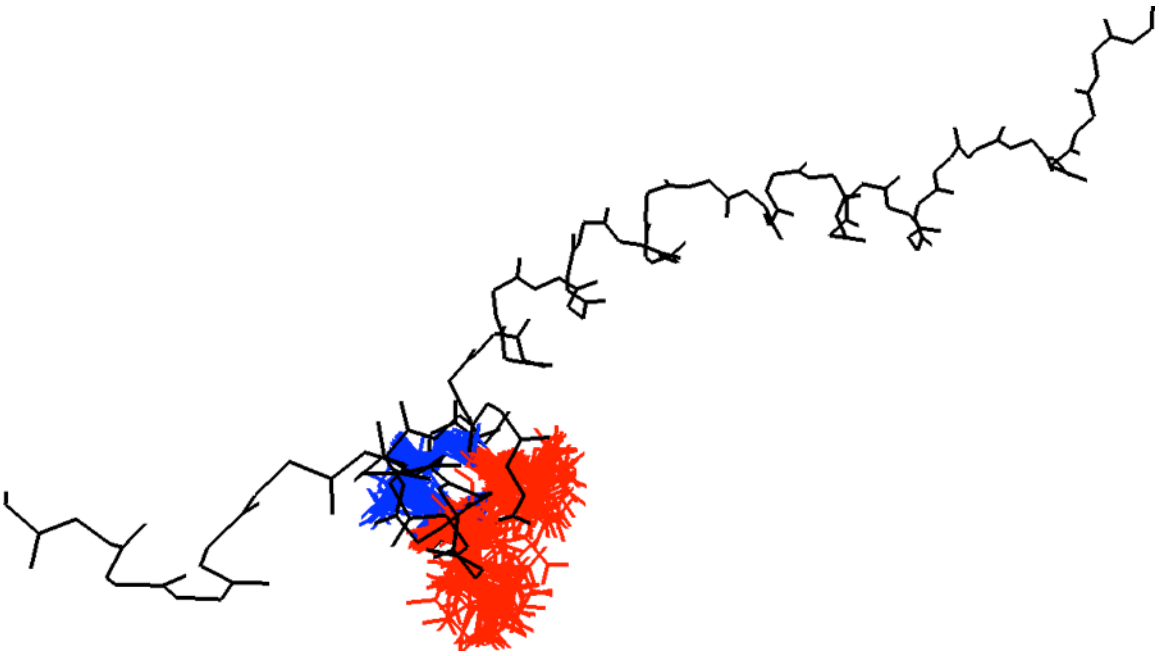

Figure 4.9: A view of AR55 with only E20-K25 superposed. The backbone residues are shown in black, with all 40 ensemble members superposed in blue over the converged region. The side chains of E20 and D23 are shown in red and fall on the same face of the helix.

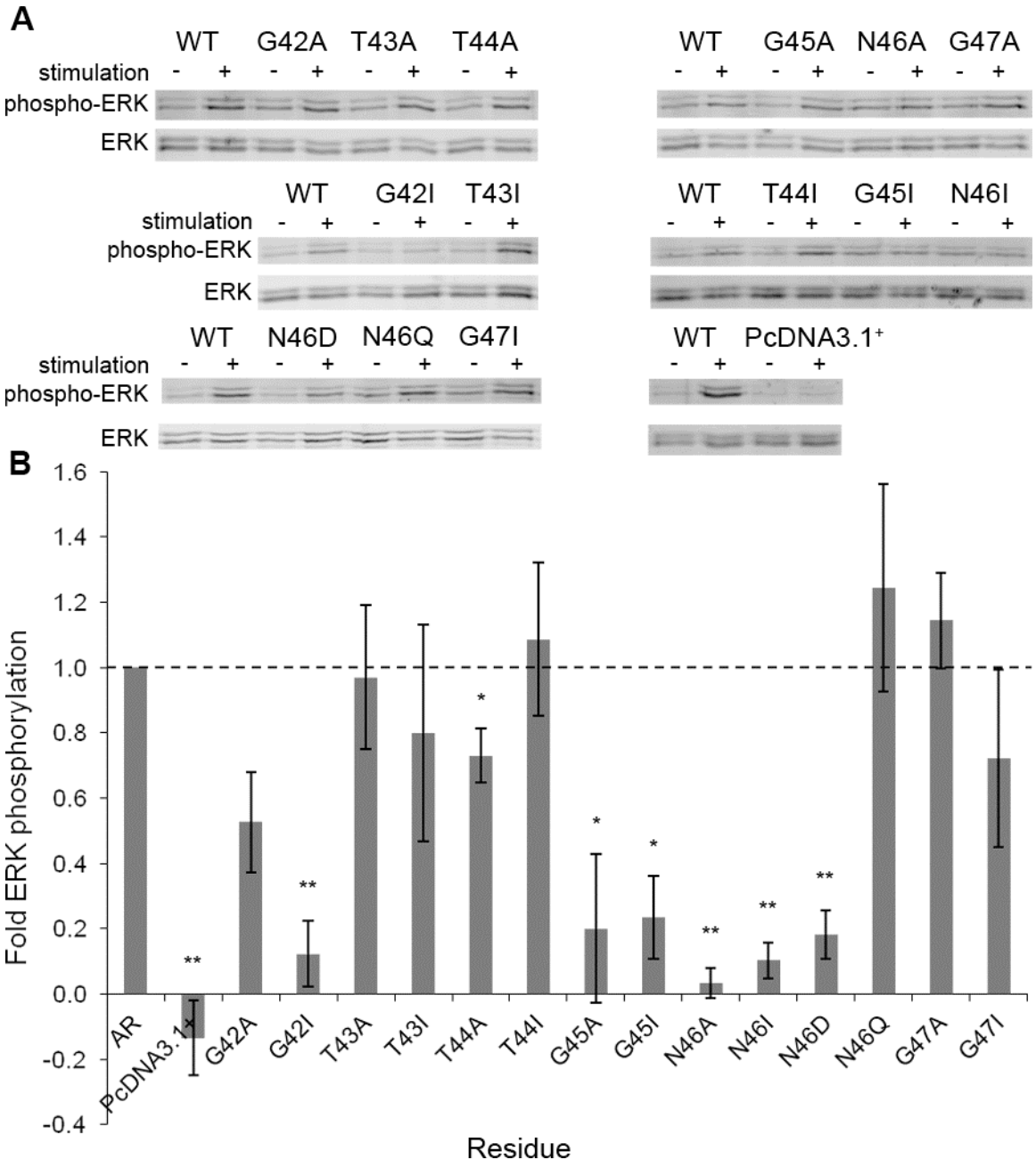

Figure 4.10: Functional analysis of 3xHA-AR mutants. For empty vector and each mutant, the fold ERK phosphorylation compared to wild-type receptor after stimulation by apelin-17 is plotted (B). Representative western blots are shown in (A). Error is shown as standard error and for each mutant the experiment was replicated at least 3 times (*= $p< 0.05$; **=$p<0.01$).

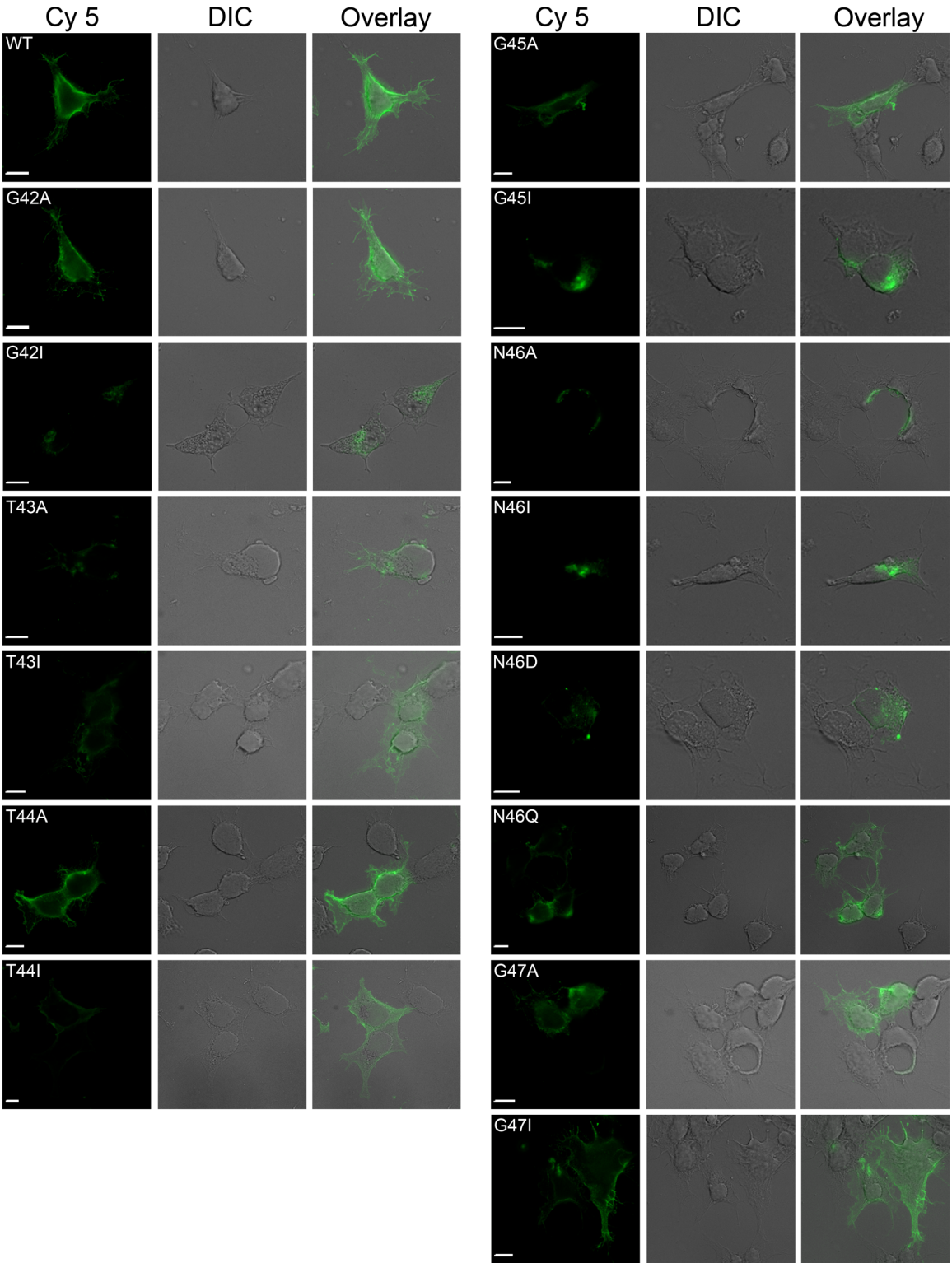

Figure 4.11: Distribution of 3xHA-AR in transiently transfected HEK 293A cells as shown by surface immunofluorescence microscopy. A differential interference contrast (DIC) field of view overlaid by the signal from Cy5 bound to 3xHA-AR is shown. The scale bar in corner of each image corresponds to 10 μm.

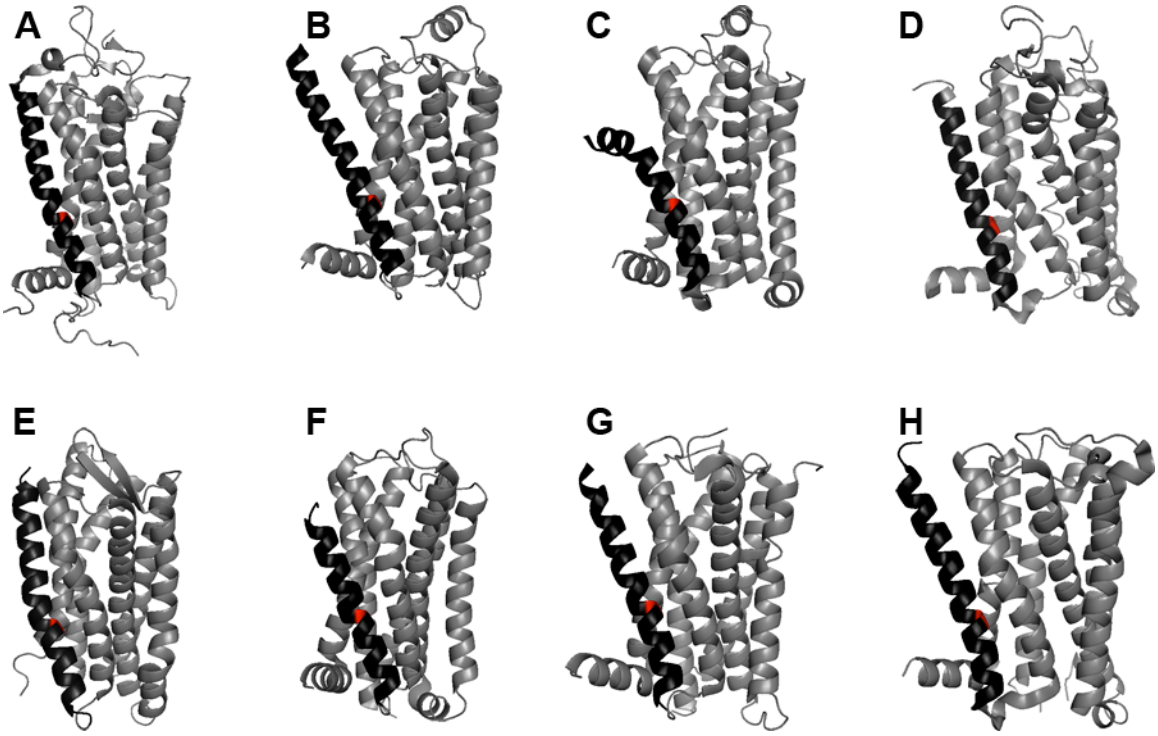

Figure 4.12: The structures of solved GPCRs. In each panel, TM I of the GPCR is black with the conserved asparagine residue shown in red. The GPCRs shown above are: (A): Bovine rhodopsin (1F88, (54)), (B): β2-adrenergic receptor (2RH1, (56)), (C): β1-adrenergic receptor (2VT4, (53)), (D): A2A adenosine receptor (3EML, (52)), (E): CXCR4 chemokine receptor (3OE6, (57)), (F): Dopamine D3 receptor (3PBL, (51)), (G): Histamine H1 receptor (3RZE, (55)) and (H): M2 muscarinic receptor (3UON, (58)).

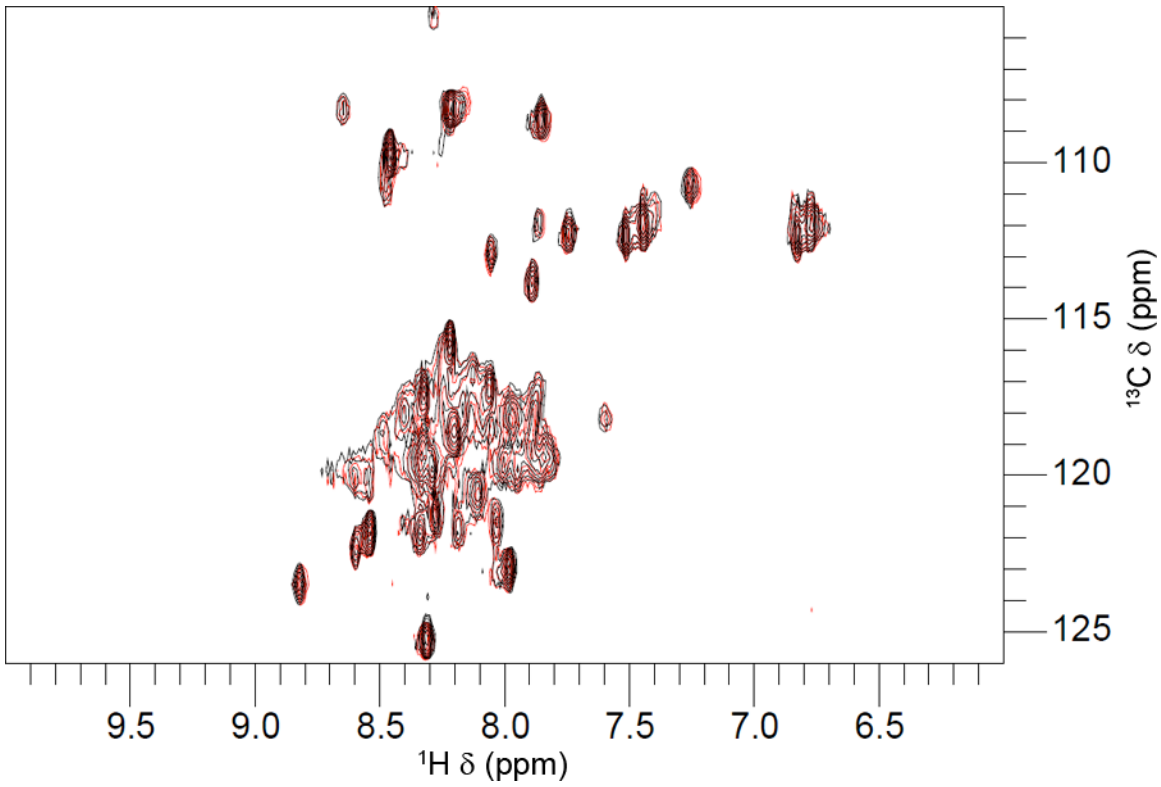

Figure 4.13: ^{1}H-^{15}N HSQC spectra of 0.5 mM AR55 solubilized in DPC micelles before (black) and after (red) addition of 2 mM apelin-17. These spectra were acquired at 700 MHz at NRC-IMB.

Chapter 5:

The Effect Of Membrane Mimetics On AR55 Structure

5.1. Introduction

In the previous chapter I described the kink observed in the TM region of AR55 solubilized in DPC micelles. One question that I have not yet addressed is if this could be an artefact due to the membrane mimetic that I used. This leads to the even bigger question of how much of an influence the choice of membrane mimetic has on the structure of AR55, or on a membrane protein in general. Although organic solvents are generally known to induce helical character in peptides, the effect that various micelle types has on polypeptide structure is largely unknown. In this chapter I attempt to address these concerns by comparing the structure of AR55 solubilized in DPC micelles that I described in chapter 4 with the structures of AR55 solubilized by SDS, LPPG and a 1:1 mixture of water:1,1,1,3,3,3-hexafluoroisopropanol (HFIP).

5.1.1. The Choice Of Membrane Mimetic Can Affect Protein Structure

Despite the ability of a membrane mimetic to approximate a membrane environment, there are definite instances where the conformation of a protein varies with the membrane mimetic. For instance, α-synuclein forms an extended helix in liposomes and bicelles (206) while forming two anti-parallel helices in micelles (207, 208, 209, 210). In another

example, the populations of a peptide having either *cis* or *trans* isomerization of its proline residues are different in SDS and DPC micelles (211). In bilayers, there is a correlation of hydrophobic mismatch between a TM peptide and the bilayer with the tilt angle of the TM peptide (142, 212). Finally, self-association of membrane proteins occurs when helix-helix interactions are more favourable than helix-lipid interactions, which suggest that tertiary structure and interprotein interactions are very dependent on the membrane mimetic chosen (reviewed in (142)).

5.1.2. The Use Of Organic Solvents As A Membrane Mimetic

In addition to the detergent and lipid based membrane mimetics, organic solvent based solvents such as dimethyl sulphoxide, trifluoroethanol/water, HFIP/water and chloroform/methanol/water are often used as a membrane mimetic. Even though this environment is not nearly as structurally complex as other membrane mimetics, TM proteins often adopt a structure similar to the native state or other membrane mimetics (213, 214, 215, 216, 217). Although these solvents are in theory isotropic, there is evidence that they actually phase separate in the vicinity of the protein, thus creating the anisotropic environment required for proper protein folding (218, 219). For NMR studies, membrane mimetic solvents have the advantage of providing sharp resonance lines of the peptide due to faster tumbling of a peptide vs. peptide embedded in a micelle.

A well-known disadvantage of membrane mimetic solvents is the induction of helical character in a polypeptide. The fluorinated solvents in particular are known to be effective at stabilizing secondary structure (220). The reason for the induction of helical structure is that water is displaced from the surface of the protein by the organic molecules (219, 221, 222). This coating of organic molecules both reduces the number of

water molecules available to form hydrogen bonds and lowers the dielectric of the environment. Both of these factors favour the formation of intrapeptide hydrogen bonds and stabilize secondary structure formation.

5.1.3. NMR Relaxation

Although determination of protein structure is clearly valuable, the motions of a protein are what make biology work. NMR spectroscopy is an ideal method to characterize the motions of biomolecules on a wide range of time scales. The R_1, R_2 and heteronuclear NOE relaxation measurements of nuclei provide an ideal method to gain information of ps-ns dynamics of chemical bonds and information about μs-ms chemical exchange. The utility of NMR relaxation measurements to biomolecular NMR has been reviewed extensively (223, 224, 225, 226).

Understanding NMR relaxation is a matter of understanding a system's return to thermal equilibrium (79). In a relaxed thermal equilibrium state, two conditions are met: i) The populations of the two nuclear spin states are given by their Boltzmann distribution and ii) there is no coherence between the nuclear spins. During an NMR experiment, both of these two conditions are broken and the system moves to re-establish them. The spin-lattice relaxation is the re-establishment of condition (i) and is characterized by a rate constant R_1. Spin-spin relaxation refers to the system's loss of coherence over time and is characterized by a rate constant R_2. The heteronuclear NOE is the enhancement of magnetization in a spin by radio frequency irradiation of a covalently linked spin. The R_1, R_2 and heteronuclear NOE relaxation rates are affected by several factors, including molecular motion (82). It follows then that measurement and comparison of R_1, R_2 and

the heteronuclear NOE provide sensitive information of the types of motions that NMR active nuclei experience.

Although relaxation measurements can be made for any NMR active nucleus, in protein NMR the ^{15}N nucleus is commonly used due the low spectral overlap in its associated spectra (*e.g.* ^{15}N HSQC) and the simplicity in the theoretical treatment compared to ^{13}C. The dipole-dipole interactions with the amide proton and the chemical shift anisotropy of the amide bond are well characterized, allowing effective analysis of the heteronuclear NOE. By analyzing ^{15}N relaxation parameters in proteins, at least one probe of molecular motion is available for each non-proline amino acid, providing a high-resolution picture of protein dynamics.

As alluded to above, the experimental values of R_1, R_2 and the heteronuclear NOE for a ^{15}N nucleus can be correlated to the underlying motions of the protein (227, 228):

$$R_1 = \left(d^2/4\right)\left[J\left(\omega_H - \omega_N\right) + 3J\left(\omega_N\right) + 6J\left(\omega_H + \omega_N\right)\right] + c^2 J\left(\omega_N\right) \qquad 5.1$$

$$R_2 = \left(d^2/8\right)\left[4J(0) + J\left(\omega_H - \omega_N\right) + 3J\left(\omega_N\right) + 6J\left(\omega_H\right) + 6J\left(\omega_H + \omega_N\right)\right] + \left(c^2/6\right)\left[3J\left(\omega_N\right) + 4J(0)\right] + R_{ex} \qquad 5.2$$

$$NOE = 1 + \left(d^2/4\right)\left(\gamma_H/\gamma_N\right)\left[6J\left(\omega_H + \omega_N\right) - J\left(\omega_H - \omega_N\right)\right]\left(1/R_1\right) \qquad 5.3$$

where $d = \left|\mu_0 h \gamma_N \gamma_H / \left(8\pi^2\right)\right|\left\langle 1/r_{NH}^3 \right\rangle$, $c = \left(\omega_N/\sqrt{3}\right)\left(\sigma_{\|} - \sigma_{\perp}\right)$, ω_N and ω_H are the Larmor frequencies of the ^{15}N and ^{1}H nuclei, γ_N and γ_H are the gyromagnetic ratios ^{15}N and ^{1}H, μ_0 is the permeability of free space, h is Plank's constant r_{NH} is the length of the amide bond vector and $\sigma_{\|}$ and $\sigma_{\perp}$ are the parallel and perpendicular components of the (assumed to be axially symmetric) chemical shift tensor. Typically a value of -160 ppm is used for $\left(\sigma_{\|} - \sigma_{\perp}\right)$ (229). J(ω) is the spectral density function, and is a Lorentzian function that describes how much motion occurs at a specific frequency. For R_2 measurements,

conformational sampling on a μs-ms time scale can cause an apparent increase in R_2, this has been modelled as a rate R_{ex} (228). As mentioned earlier, the values of R_1, R_2 and the NOE can easily be experimentally determined and thus J(ω) can be fit to the observed relaxation parameters.

5.1.4. Model-Free Analysis Of Relaxation Data

The model-free formalism was introduced by Lipari and Szabo (230, 231) and was further extended by Clore and coworkers (232). In the extended model-free formalism (232), J(ω) is expressed as:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S_f^2)\tau_f'}{1+(\omega\tau_f')^2} + \frac{(S_f^2-S^2)\tau_s'}{1+(\omega\tau_s')^2} \right) \qquad 5.4$$

where $\tau_f' = \tau_f\tau_m/(\tau_f+\tau_m), \tau_s' = \tau_s\tau_m/(\tau_s+\tau_m)$, τ_m is the overall rotational correlation time of the molecule and can be calculated directly from the R_1 and R_2 relaxation rates (233, 234). τ_f and τ_s are the effective correlation times for internal motions on a fast ($\tau_f <$ 100-200 ps) and slow ($\tau_f < \tau_s < \tau_m$) time scales respectively. S^2 is the square of the generalized order parameter while S_f^2 and S_s^2 are the squares of the order parameters for motion on the fast and slow time scales, respectively. Note that $S^2 = S_s^2 S_f^2$. The overall effect of the model-free approach is that amide bond motion can be expressed as a generalized order parameter and correlation times. S^2 can take a value between 0 (a completely unrestrained bond vector) and 1 (a static bond vector). The correlation time is a measurement of how fast this bond vector is changing orientation.

The R_1, R_2 and NOE values are dependent on 6 different parameters (5 in equation 5.4 plus the R_{ex} term in equation 5.2). However, the number of free parameters can be

reduced if equation(s) 5.2 and/or 5.4 are simplified according to 5 different molecular motion models (228). Model 1 assumes that $S_s^2 = 1$ (slow time scale motions are negligible) and $\tau_f \rightarrow 0$ (fast bond motions are < 20 ps) and is given by the following formula:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} \right) \qquad 5.5$$

Model 2 is the original formulation of J(ω) by Lipari and Szabo (230, 231) and only assumes that $S_s^2 = 1$. The simplified model-free equation for model 2 is shown below.

$$J(\omega) = \frac{2}{5}\left(\frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(1 - S^2)\tau_f'}{1 + (\omega \tau_f')^2} \right) \qquad 5.6$$

Models 3 and 4 are derived from models 1 and 2 respectively, with the addition of a non-zero chemical exchange term, R_{ex} in equation 5.2. In model 5 the only assumption is that $\tau_f \rightarrow 0$ and is fit to the equation below.

$$J(\omega) = \frac{2}{5}\left(\frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(S_f^2 - S^2)\tau_s'}{1 + (\omega \tau_s')^2} \right) \qquad 5.7$$

These 5 models each contain a maximum of 3 bond order parameters and a single internal time scale parameter. Since a model can only contain either τ_f or τ_s, either of these variables can be referred to as the effective internal correlation time, τ_e.

5.1.5. Reduced Spectral Density Mapping Of Relaxation Data

In instances of high protein disorder, it is difficult to obtain meaningful values of S^2, in part because it is difficult to assign a single τ_m to a disordered protein. In these instances the reduced spectral density mapping approach to analyzing relaxation data can be used. In reduced spectral density mapping, the equations for R_1, R_2 and NOE are simplified (235):

$$R_1 = (d^2/4)[3J(\omega_N) + 7J(0.870\omega_H)] + c^2 J(\omega_N) \quad 5.8$$

$$R_2 = (d^2/8)[4J(0) + 3J(\omega_N) + 13J(0.870\omega_H)] + (c^2/6)[3J(\omega_N) + 4J(0)] \quad 5.9$$

$$NOE = 1 + (d^2/4)(\gamma_H/\gamma_N)[5J(0.870\omega_H)](1/R_1) \quad 5.10$$

By measuring the values of R_1, R_2 and the NOE the values of J(0), J(ω_N) and J(0.87ω_H) can be determined. This provides a quantitative measurement of the amplitude of motion occurring at these three frequencies.

5.2. Materials And Methods

5.2.1. Materials

LPPG was purchased from Avanti Polar lipids (Alabaster, AL). SDS-d_{25} and HFIP-d_2 was purchased from C/D/N isotopes (Pointe-Claire, QC). Unless otherwise detailed, AR55 and all other reagents were obtained as detailed in chapter 4.

5.2.2. CD Spectropolarimetry

Far-ultraviolet CD spectra of AR55 were recorded at 37 °C using a Jasco J-810 spectropolarimeter (Easton, MD). Aliquots of AR55 (100 μM) were prepared from a single stock solution. Each sample contained either 20 mM SDS, 22 mM DPC or 32 mM LPPG in 20 mM sodium phosphate and 10 mM DTT adjusted to pH 5.00 ± 0.05 for the SDS and LPPG samples. The sample containing DPC was adjusted to pH 4.00 ± 0.05 due to insolubility at pH 5. Additional samples were prepared at pH 7.00 ± 0.05. For these samples the same buffer was used and 10 μM AR55 was solubilized in 20 mM SDS, 10 mM DPC or 10 mM LPPG.

CD spectra were acquired from 260 nm downwards (1 nm steps), with reliable ellipticity values observed at > 180 nm (pH 4 and 5) or > 195 nm (pH 7) based on spectropolarimeter photomultiplier tube voltages, in 0.1 mm (pH 4 or 5) or 1 mm (pH 7) path-length quartz cuvettes (Hellma, Müllheim, Germany). All measurements were collected in triplicate. Ellipticity data were processed the same way as in chapter 3 except that each spectrum was not baseline adjusted by subtracting the average ellipticity value over 250-260 nm.

5.2.3. NMR Spectroscopy Of AR55

All NMR solutions with micelles contained 20 mM $Na^+CD_3COO^-$, 1 mM DSS, 1 mM NaN_3, 10 mM DTT-d_6 and were pH adjusted to 5.00 ± 0.05 (SDS and LPPG) or 4.00 ± 0.05 (DPC) without accounting for deuterium isotope effects. Solutions of AR55 with SDS (0.8 mM AR55, 96 mM SDS, 95% H_2O, 5% D_2O) and AR55 with LPPG (0.56 mM AR55, 72 mM LPPG, 95% H_2O, 5% D_2O) and a NMR sample containing HFIP (0.9 mM

AR55, 1 mM DSS, 5 mM DTT-d_6, 10% D_2O, 50% H_2O, 40% HFIP-d_2) were prepared. ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, HNCO, HNcaCO, HNCA, HNcoCA, HNCACB, ^{15}N-edited TOCSY, HcCH TOCSY, ^{15}N-edited NOESY and ^{13}C-edited NOESY experiments were collected for the NMR samples at 700 MHz (Bruker Avance III, NRC-IMB) and 800 MHz (Varian INOVA, NANUC). In addition, for the HFIP sample a CNH NOESY (236) was collected. All NOESY experiments for AR55 in SDS or LPPG were collected with a 125 ms mixing time, while a 150 ms mixing time was used for AR55 in 50% HFIP. All NMR data collection occurred at 37 °C and detailed acquisition parameters for all experiments are listed in Table 5.1.

5.2.4. Structure Calculation Of AR55

For each sample condition, data were processed, analyzed, and structures were generated as detailed in chapter 4 for AR55 with DPC micelles. For all experimental conditions, the 40 member final structural ensembles of AR55 have been deposited in the PDB (200) using SMSDep (accession #: 2LOT, 2LOV and 2LOW for AR55 solubilized with SDS, LPPG and 50% HFIP respectively). The chemical shift, spectral peak and restraint data were deposited into the BMRB ((121), accession #:18224, 18226 and 18227 for AR55 solubilized with SDS, LPPG and 50% HFIP respectively)

5.2.5. PRE Measurements

NMR samples of uniformly ^{15}N-enriched AR55 were prepared as before except with 10% D_2O and 90% H_2O and either 77 mM DPC or 125 mM LPPG with 0.5 mM AR55 or 110 mM SDS with 1 mM AR55. The samples were titrated with Mn^{2+} up to 1 mM with ^{1}H-^{15}N HSQC spectra collected at each point. Data for AR55 solubilized in SDS micelles

were collected at Agriculture and Agri-Food Canada (AAFC, Charlottetown, PE) on a 600 MHz Bruker Avance III instrument equipped with a 1.7 mm room temperature TXI probe. Data for AR55 solubilized in DPC and LPPG micelles were collected on a 700 MHz Bruker Avance III spectrometer equipped with a 1.7 mm TCI cryoprobe at NRC-IMB. For all datasets PRE was measured as the ratio of peak heights in attenuated (I_{att}) and reference (I_{ref}) spectra. Error (σ_{PRE}) was calculated by evaluating the standard deviation of the PRE as:

$$\sigma_{PRE} = |PRE|((\sigma_{I_{att}}/I_{att})^2 + (\sigma_{I_{ref}}/I_{ref})^2)^{1/2} \quad 5.11$$

Where $\sigma_{I_{att}}$ and $\sigma_{I_{ref}}$ represent the standard deviations of the spectral noise (as measured using NMRDraw (114)) in the I_{att} and I_{ref} spectra.

5.2.6. NMR Relaxation Data Collection And Analysis

NMR samples of $^{13}C/^{15}N$ labelled AR55 (1 mM AR55, 20 mM $Na^+CD_3COO^-$, 1 mM DSS, 5 mM DTT, 1 mM NaN_3, 90% H_2O/10% D_2O) were prepared with either 102 mM DPC or 150 mM LPPG. In addition, the previously prepared $^{13}C/^{15}N$ sample of AR55 in SDS micelles was used. For all samples, ^1H-^{15}N HSQC (237) spectra were collected to measure the R_1 and R_2 relaxation rates as well as the NOE. For R_1 measurements, recovery delays of 50, 100, 250, 600, 1000 and 1500 ms were used. For R_2 measurements delays of 10, 30, 50, 70, 90, 130 and 150 ms were used. For AR55 solubilized in DPC micelles a R_1 measurement at 250 ms was not collected. For AR55 solubilized in SDS micelles the R_1 time points at 50 and 100 ms were removed due to an experimental artefact.

The ^{1}H-^{15}N NOE was measured as the ratio of a saturated spectrum to a reference spectrum and error was propagated by measuring the spectral noise, as shown in equation 5.11. Mathematica (Wolfram Research Inc., Champaign, IL) notebooks (238) were used to fit the R_1 and R_2 relaxation data, calculate τ_m (assuming isotropic diffusion) and calculate the relevant model-free or reduced spectral density mapping parameters. For the model-free analysis each residue of AR55 was fit to one of the 5 motional models (introduced above) using Akaike's information criteria (239). 100 Monte Carlo simulations were performed for both the model-free and reduced spectral density mapping approaches to determine the errors of the fitted parameters.

5.3. Results And Discussion

5.3.1. CD Spectropolarimetry Indicates That AR55 Has A Similar Structure In All Micelle Conditions

At pH 7, the CD spectra of AR55 solubilized in SDS and LPPG micelles are very similar with a clearly defined minimum at 208 nm, a negative band (but not a minimum) at ~222 nm and crossover points at 200 nm and 202 nm for AR55 solubilized in SDS and LPPG respectively. The CD spectrum of AR55 solubilized in DPC micelles was quite different, with only a negative band at 215 nm and crossover point at 202 nm. At pH 7, AR55 solubilized in SDS and LPPG micelles is largely helical, while it may be somewhat aggregated in DPC micelles.

At lower pH, the CD spectra of AR55 in all micelle conditions are very similar with a minimum at 209 nm, a maximum at 192 nm , a negative band at ~222 nm and a crossover point at 200 nm, 202 nm and 201 nm for AR55 solubilized in SDS, DPC and LPPG

micelles respectively (Figure 5.1). The CD spectra of AR55 solubilized in SDS and LPPG micelles are nearly identical and corresponds to AR55 adopting a largely helical character, as seen at pH 7. Although the spectrum of AR55 solubilized in DPC micelles has a similar general shape compared to SDS and LPPG, the intensity of the bands are reduced, with the minimum at 209 nm not being nearly as pronounced. These spectral differences could simply be due to unequal AR55 concentrations between samples or may suggest that AR55 adopts a slightly different structure at low pH in DPC micelles. Overall, the structure of AR55 solubilized in SDS and LPPG micelles is very similar at both pH 5 and 7.

5.3.2. NMR Assignment Of AR55

High quality NMR data were collected for each sample of AR55. The ^{1}H-^{15}N HSQC spectra of AR55 in SDS, DPC and LPPG micelles are all very similar, suggesting that AR55 adopts a similar structure in all micelles (Figure 5.2 (A-C)). AR55 solubilized in 50% HFIP produced higher quality spectra (Figure 5.2 (D)). In addition to the striking reduction in peak linewidth, the resonances were also less overlapped than in the micelle conditions. The same techniques used to assign AR55 resonances in DPC were used to assign AR55 resonances in SDS, LPPG and HFIP. All residues and nearly all resonances were successfully assigned for AR55 in each membrane mimetic (Table 5.2).

5.3.3. The Topology Of The AR55-Micelle Complex Is The Same For All Three Micelle Types

Using the assigned NMR resonances, Mn^{2+} was used to attenuate specific resonances in a ^{15}N-^{1}H HSQC spectrum (Figure 5.3). Although not all resonances were resolvable due

to signal overlap, it is clear that the N-terminal tail up to residue L30 as well as residues T43-V49 of AR55 are attenuated by Mn^{2+}, and hence are exposed to water in SDS, DPC and LPPG micelles. This indicates that AR55 is not in a canonical trans-micelle orientation and instead has two micelle embedded sections with residues T43-V49 being exposed to water.

Three primary scenarios can be brought forwards to explain this non-canonical topology (Figure 5.4). First, AR55 may span the same micelle twice, with residues T43-V49 forming a disruption and tight turn at the surface of the micelle. Another option is that AR55 spans two distinct micelle-like objects, with T43-V49 residing between the two objects and hence being solvent exposed. Finally, AR55 may span one typical micelle, but have water associated with residues T43-V49 in the interior of the micelle. Given the plasticity of micelles, any of these scenarios is possible. From the data presented here it is difficult to tell which of these scenarios is correct. It should be noted, though, that the experimental data is not supportive of the first scenario. In that topology, the two helical segments would likely be in proximity, leading to long-range NOE contacts between the helical segments that are not observed. Furthermore, the third scenario is quite unlikely due to the energy penalty of water being present in the interior of a micelle. Overall, AR55 is most likely spanning two micelle-like objects.

5.3.4. Structure Calculation Of AR55

Through analysis of ^{15}N-edited and ^{13}C-edited NOESY experiments, 1556, 1608, 1654 and 2355 NOE contacts (counted for each nucleus involved in each NOE) were determined for AR55 solubilized in SDS, DPC, LPPG and 50% HFIP respectively (Table 5.3). The distribution of NOE contacts throughout AR55 was similar for all membrane

mimetics used, with the majority of NOE contacts occurring in the micelle-spanning region of AR55 (Figure 5.5). The patterns of distance restraints determined for AR55 suggest similar structuring over the micelle-spanning region in each membrane mimetic. In particular, the micelle-spanning regions (L30-G42 and G47-K56) of AR55 have distance restraints of α-helical nature while the solvent exposed region from T43-N46 has fewer canonical α-helical contacts (αN (i, i+3), αN (i,i+4) and αβ (i, i+3) Figure 5.6, (124)). Chemical shift index (CSI) (106) and secondary structure prediction by DANGLE (201) agree with the lack of helical distance restraints and predict a break in helical character from T43-N46. Although AR55 solubilized in 50% HFIP still has some helical distance restraints over this region, there are fewer than in the neighbouring helices.

Using the observed NOE contacts, a final 40 member ensemble was generated for AR55 in SDS, LPPG and HFIP that agreed with experimental distance restraints and has excellent Ramachandran plot statistics (Table 5.3). Overall, the ensembles of AR55 solubilized in micelles have a similar number of distance restraints as well as similar proportions of residues in the core, allowed, generously allowed and disallowed regions of Ramachandran space (Table 5.3). The structure of AR55 solubilized in 50% HFIP appears to have a higher proportion of residues occupying the most favoured region (Table 5.3). This may be due to the higher number of distance restraints available to converge the structure of AR55 in 50% HFIP.

Several regions of converged structure were present in each of the ensembles generated for AR55. Areas of a particular AR55 structure were selected for superposition based on the regions of the structure that have a continuous and high ϕ and φ dihedral angle order parameter (SDS; Figure 5.7, DPC; Figure 5.8, LPPG; Figure 5.9, HFIP; Figure 5.10).

Regions of the AR55 structures selected for superposition with their respective RMSD are listed in Table 5.4 and the AR55 ensembles are shown in Figure 5.11.

5.3.5. The Micelle-Spanning Region Of AR55 Has A Similar Structure In Different Membrane Mimetics

Analysis of the dihedral variability of the AR55 structures (Figure 5.7-5.10) indicates that the micelle-spanning region of AR55 is the most converged portion of the protein in all membrane mimetics. At some point near N46 there is a reduction in dihedral angle convergence, with this reduction being the most pronounced in the SDS and DPC environments.

Ensembles of AR55 in each membrane mimetic studied are shown in Figures 5.11 and 5.12. In all four ensembles, it is clear that both of the micelle-spanning regions from L30-L41 and L50-G58 are helices that superpose quite well compared to the rest of the peptide. A change in helical direction, or kink, is present in all of the structures of AR55 (T43-N46 in SDS, N46-G47 in DPC, G45-N46 in LPPG and L41-G42 in 50% HFIP). The average kink angle of all ensemble members (63° ± 28° in SDS, 69° ± 20° in DPC, 66° ± 17° in LPPG and 27° ± 10° in 50% HFIP, measured using MC-HELAN (4)) indicates the angle of the kink is much better defined for AR55 solubilized in HFIP than it is in any of the micelles. The ensembles of AR55 solubilized in LPPG micelles or 50% HFIP both show high dihedral angle convergence over residues involved in the kink and accordingly these structures were also superposed over the entire micelle-spanning region (Figure 5.12).

5.3.6. The Structure Of The N-Terminal Tail Of AR55 Is Influenced By The Membrane Mimetic

There are both similarities and differences in the structure of the N-terminal tail of AR55 in the different membrane mimetics. In the micelles, the N-terminal tail is largely unstructured with the exception of a helical region from residues E20-K25 in SDS and DPC micelles and T22-K25 in LPPG micelles. AR55 in SDS and DPC micelles have an additional converged structure from A13-E18 and D14-C19, respectively. Although the N-terminal tail of AR55 is structurally similar in all of the detergent types tested, it is significantly different in 50% HFIP where a helix is induced from A13-S27 (Figure 5.11). This is hardly surprising considering that fluorinated solvents are well known to induce helical character in peptides (220, 240). An interesting point is that even though HFIP is able to cause helical character throughout the N-terminus of AR55, the kink in the micelle-spanning region of AR55 is maintained.

5.3.7. The Dynamics Of AR55 Are Influenced By The Membrane Mimetic Choice

From the ^{1}H-^{15}N HSQC spectra, 41, 49 or 53 well resolved resonances were selected for relaxation analysis of AR55 solubilized in SDS, DPC and LPPG micelles respectively. Using Mathematica notebooks the R_1, R_2 and NOE were successfully fit to the data, although the NOE measurement in SDS micelles has more noise due to lower signal in the associated spectra (Figure 5.13). For each micelle condition, AR55 appears to be composed of two regions that behave quite differently. Specifically the N-terminal tail and C-terminus have a larger R_1, smaller R_2 and smaller NOE than the micelle

spanning region of AR55. This almost certainly is due to the increased flexibility of the N-terminal tail compared to the micelle spanning region.

Overall, the R_1, R_2 and NOE measurements of AR55 are similar when solubilized in either SDS or DPC micelles. The only significant difference is that R_2 is slightly lower throughout AR55 in DPC micelles. When solubilized in LPPG, the relaxation properties of AR55 change considerably. Specifically the N-terminal tail of AR55 has lower values of R_2 and the NOE in LPPG micelles. Also, in the micelle spanning region of AR55, R_1 is decreased and R_2 is increased for the LPPG data (Figure 5.13). The τ_m values calculated by the Mathematica notebooks using R_1/R_2 ratios of AR55 solubilized in SDS, DPC and LPPG are 9.96, 12.26 and 16.97 ns, respectively. These values are in agreement with literature values for similar peptides (223). The much higher τ_m of AR55 solubilized in LPPG can be attributed to the larger radius of an LPPG micelle (~20 Å × ~30 Å, (241)) compared to either SDS (~17 Å (242)) or DPC micelles (~20 Å (243)).

For each analyzed residue of AR55, values of J(0), J(w_N) and J(0.87w_H) were fit for the three micelle conditions (Figure 5.14). In general J(0) > J(ω_N) > J(0.87ω_H) which corresponds to theory (235). In the N-terminal tail of AR55, J(0) is largest when AR55 is solubilized in DPC and smallest when solubilized in LPPG micelles. J(ω_N) is very similar throughout AR55 for all of the micelle conditions except that over the micelle spanning region of LPPG micelles J(ω_N) is reduced. Finally J(0.87ω_H) is similar in the micelle spanning region of AR55 for all membrane mimetics, but is increased in the N-terminal tail for AR55 in LPPG micelles.

Model-free analyses were completed for the R_1, R_2 and NOE data and detailed dynamical parameters are found in Table 5.5. As an overall trend, S^2 is much larger in the

micelle spanning region of AR55 than the N-terminal tail, emphasizing the high degree of flexibility in the N-terminal tail. S^2 is similar in the micelle spanning region in SDS, DPC and LPPG micelles (Figure 5.15). However, the value of S^2 is much lower in the N-terminal tail of AR55 when solubilized in LPPG micelles compared to either SDS or DPC micelles.

Overall, model-free analysis and spectral density mapping provide an overview of AR55 dynamics in SDS, DPC and LPPG micelles. The N-terminal tail is seen to be much more flexible than the micelle spanning region of AR55 as indicated by the low S^2 in this region as well as the increase of high frequency motions ($J(0.87\omega_H)$) and decrease of $J(0)$. This is hardly surprising but does confirm the structural results shown above. Interestingly, neither model-free analysis nor reduced spectral density mapping suggest that the kink in the micelle spanning region is undergoing motion different from the helices surrounding it. Although S^2 is the same in the micelle spanning region of all 3 micelle types, $J(0)$ and $J(\omega_N)$ are comparatively less for AR55 in LPPG. This suggests that in all micelles the micelle spanning region is well structured, but that in LPPG micelles the motions that do occur are of lower frequency. This may be an effect of the LPPG micelle being larger than both SDS and DPC micelles. An unexpected trend is the increased motion on fast time scales of the N-terminal tail of AR55 when solubilized in LPPG micelles. In this case although the AR55 is tumbling slower in solution, the N-terminal tail experiences faster motions than if it were solubilized in a smaller micelle. This is reflected in the structures of AR55 (Figure 5.11) where the N-terminal tail is the least structured in LPPG micelles.

5.4. Summary

The structure of AR55 is surprisingly consistent in SDS, DPC and LPPG micelles as well as 50% HFIP. In all of the micelle conditions, the micelle-spanning region of AR55 consists of a helix-kink-helix motif with the kink and N-terminal tail of AR55 being exposed to water. This kink persists even in the helix inducing solvent HFIP. The N-terminal tail of AR55 is also quite similar in the different micelles with structural convergence from E20-K25 for SDS and DPC and T22-K25 in LPPG. However, AR55 in SDS and DPC shows additional N-terminal structural elements that were not observed when solubilized in LPPG. In contrast the N-terminal tail of AR55 in 50% HFIP is largely helical. Dynamics data demonstrates that the N-terminal tail of AR55 is less ordered and experiences larger amplitude motion on a ps-ns time scale than the micelle spanning region. Also, the N-terminal tail experiences the highest amplitude ps time scale motion when solubilized in LPPG micelles. This trend cannot be predicted based on the characteristics of the micelle systems and demonstrates how a peptide in a membrane mimetic may not behave as expected.

Table 5.1: Detailed experimental parameters used for data acquisition of AR55 solubilized in SDS, DPC and LPPG micelles as well as 50% HFIP.

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	^{1}H frequency (MHz)	Facility
AR55 in SDS							
^{1}H-^{13}C HSQC	1.5	8	^{1}H:1024 ^{13}C:256	^{1}H:8418 ^{13}C:14085	^{1}H:4.66 ^{13}C:37.6	700	NRC-IMB
^{1}H-^{15}N HSQC	1.5	4	^{1}H:1024 ^{15}N:64	^{1}H:11261 ^{15}N:1490	^{1}H:4.66 ^{15}N:115	700	NRC-IMB
HNCO	1.5	8	^{1}H:2048 ^{13}C:32 ^{15}N:32	^{1}H:9804 ^{13}C:1761 ^{15}N:1490	^{1}H:4.66 ^{13}C:176 ^{15}N:115	700	NRC-IMB
HNcaCO	1.5	8	^{1}H:2048 ^{13}C:32 ^{15}N:32	^{1}H:9804 ^{13}C:1761 ^{15}N:1490	^{1}H:4.66 ^{13}C:176 ^{15}N:115	700	NRC-IMB
HNCA	1.5	8	^{1}H: 1024 ^{13}C:32 ^{15}N:32	^{1}H:9804 ^{13}C:2993 ^{15}N:1490	^{1}H:4.66 ^{13}C:59.6 ^{15}N:115	700	NRC-IMB
HNcoCA	1.5	8	^{1}H: 2048 ^{13}C:32 ^{15}N:32	^{1}H:11261 ^{13}C:2993 ^{15}N:1490	^{1}H:4.66 ^{13}C:59.6 ^{15}N:115	700	NRC-IMB
HNCACB	1.5	8	^{1}H:2048 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:9950 ^{15}N:1490	^{1}H:4.66 ^{13}C:43.4 ^{15}N:115	700	NRC-IMB
HNcoCACB	1.5	8	^{1}H:2048 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:9950 ^{15}N:1490	^{1}H:4.66 ^{13}C:43.4 ^{15}N:115	700	NRC-IMB
HcCH_TOCSY	1.5	8	^{1}H:2048 ^{1}H:96 ^{13}C:64	^{1}H:9804 ^{1}H:6020 ^{13}C:10121	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:44.4	700	NRC-IMB
TOCSY-^{15}N HSQC	1.5	8	^{1}H:2048 ^{1}H:96 ^{15}N:32	^{1}H:9804 ^{1}H:6020 ^{15}N:1490	^{1}H:4.66 ^{1}H: 4.66 ^{15}N:115	700	NRC-IMB
NOESY-^{13}C HSQC	1.5	8	^{1}H:1024 ^{1}H:96 ^{13}C:74	^{1}H:9804 ^{1}H:6020 ^{13}C:11099	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:42.1	700	NRC-IMB
NOESY-^{15}N HSQC	1.5	8	^{1}H:2048 ^{1}H:96 ^{15}N:32	^{1}H:9804 ^{1}H:6020 ^{15}N:1490	^{1}H:4.66 ^{1}H: 4.66 ^{15}N:115	700	NRC-IMB
Mn^{2+} titration (^{15}N HSQCs)	1	196	^{1}H:2048 ^{15}N:64	^{1}H:9615 ^{15}N:1277	^{1}H:4.66 ^{15}N:115	600	AAFC

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	^{1}H frequency (MHz)	Facility
AR55 in SDS							
T1 and T2 relaxation (^{15}N HSQCs)	2	16	^{1}H:2048 ^{15}N:256	^{1}H:11990 ^{15}N:1621	^{1}H:4.66 ^{15}N:115	800	NANUC
NOE (^{15}N HSQCs)	5	32	^{1}H:2048 ^{15}N:128	^{1}H:11990 ^{15}N:1621	^{1}H:4.66 ^{15}N:115	800	NANUC
AR55 in DPC							
Mn^{2+} titration (^{15}N HSQCs)	1.5	64	^{1}H:2048 ^{15}N:32	^{1}H:11261 ^{15}N:1845	^{1}H:4.66 ^{15}N:115	700	NRC-IMB
T1 and T2 relaxation (^{15}N HSQCs)	2	16	^{1}H:2048 ^{15}N:256	^{1}H:11990 ^{15}N:1800	^{1}H:4.66 ^{15}N:118	800	NANUC
NOE (^{15}N HSQCs)	5	32	^{1}H:2048 ^{15}N:256	^{1}H:11990 ^{15}N:2500	^{1}H:4.66 ^{15}N:118	800	NANUC
AR55 in LPPG							
^{1}H-^{13}C HSQC	1.5	4	^{1}H:1024 ^{13}C:128	^{1}H:8418 ^{13}C:14085	^{1}H:4.66 ^{13}C:37.6	700	NRC-IMB
^{1}H-^{15}N HSQC	1.5	4	^{1}H:1024 ^{15}N:64	^{1}H:11261 ^{15}N:1490	^{1}H:4.66 ^{15}N:115.5	700	NRC-IMB
HNCO	1	8	^{1}H:1024 ^{13}C:32 ^{15}N:32	^{1}H:9804 ^{13}C:1761 ^{15}N:1490	^{1}H:4.66 ^{13}C:176 ^{15}N:115	700	NRC-IMB
HNcaCO	1	8	^{1}H:1024 ^{13}C:32 ^{15}N:32	^{1}H:9804 ^{13}C:1232 ^{15}N:1490	^{1}H:4.66 ^{13}C:175 ^{15}N:115.5	700	NRC-IMB
HNCA	1	16	^{1}H: 1024 ^{13}C:32 ^{15}N:32	^{1}H:11261 ^{13}C:3170 ^{15}N:1490	^{1}H:4.66 ^{13}C:50.6 ^{15}N:115	700	NRC-IMB
HNcoCA	1	16	^{1}H: 1024 ^{13}C:32 ^{15}N:32	^{1}H:11261 ^{13}C:3170 ^{15}N:1490	^{1}H:4.66 ^{13}C:50.6 ^{15}N:115	700	NRC-IMB

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	^{1}H frequency (MHz)	Facility
AR55 in LPPG							
HNCACB	1	8	^{1}H:1024 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:8977 ^{15}N:1490	^{1}H:4.66 ^{13}C:41.1 ^{15}N:115	700	NRC-IMB
HNcoCACB	1	8	^{1}H:1024 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:8977 ^{15}N:1490	^{1}H:4.66 ^{13}C:41.1 ^{15}N:115	700	NRC-IMB
HcCH_TOCSY	1	8	^{1}H:1024 ^{1}H:144 ^{13}C:64	^{1}H:9804 ^{1}H:5602 ^{13}C:12330	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:42.6	700	NRC-IMB
TOCSY-^{15}N HSQC	0.95	16	^{1}H:1024 ^{1}H:222 ^{15}N:60	^{1}H:6983 ^{1}H:5998 ^{15}N:1520	^{1}H:4.66 ^{1}H: 4.66 ^{15}N:118	500	NANUC
NOESY-^{13}C HSQC	1	8	^{1}H:1024 ^{1}H:128 ^{13}C:96	^{1}H:9804 ^{1}H:5952 ^{13}C:21142	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:50.6	700	NRC-IMB
NOESY-^{15}N HSQC	0.95	16	^{1}H:1024 ^{1}H:224 ^{15}N:60	^{1}H:6983 ^{1}H:5998 ^{15}N:1520	^{1}H:4.66 ^{1}H: 4.66 ^{15}N:118	500	NANUC
Mn^{2+} titration (^{15}N HSQCs)	1.5	64	^{1}H:2048 ^{15}N:32	^{1}H:11261 ^{15}N:1845	^{1}H:4.66 ^{15}N:115	700	NRC-IMB
T1 and T2 relaxation (^{15}N HSQCs)	2	16	^{1}H:2048 ^{15}N:256	^{1}H:11990 ^{15}N:1800	^{1}H:4.66 ^{15}N:118	800	NANUC
NOE (^{15}N HSQCs)	5	32	^{1}H:2048 ^{15}N:256	^{1}H:11990 ^{15}N:2500	^{1}H:4.66 ^{15}N:118	800	NANUC
AR55 in HFIP							
^{1}H-^{13}C HSQC	1	8	^{1}H:2048 ^{13}C:128	^{1}H: 11261 ^{13}C:13558	^{1}H:4.66 ^{13}C:36.1	700	NRC-IMB
^{1}H-^{15}N HSQC	1	8	^{1}H:2048 ^{15}N:64	^{1}H:11261 ^{15}N:1455	^{1}H:4.66 ^{15}N:115.2	700	NRC-IMB
HNCO	1	8	^{1}H:2048 ^{13}C:32 ^{15}N:40	^{1}H:9804 ^{13}C:2133 ^{15}N:1455	^{1}H:4.66 ^{13}C:175.6 ^{15}N:115.2	700	NRC-IMB
HNcaCO	1	8	^{1}H:2048 ^{13}C:32 ^{15}N:40	^{1}H:9804 ^{13}C:2133 ^{15}N:1455	^{1}H:4.66 ^{13}C:175.6 ^{15}N:115.2	700	NRC-IMB

Experiment	Recovery delay (s)	# of scans	# of complex points	Sweep width (Hz)	Center position (ppm)	^{1}H frequency (MHz)	Facility
AR55 in HFIP							
HNCA	1	8	^{1}H: 2048 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:4930 ^{15}N:1455	^{1}H:4.66 ^{13}C:55.6 ^{15}N:115.2	700	NRC-IMB
HNcoCA	1	8	^{1}H: 2048 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:4930 ^{15}N:1455	^{1}H:4.66 ^{13}C:55.6 ^{15}N:115.2	700	NRC-IMB
HNCACB	1	8	^{1}H:1024 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:10213 ^{15}N:1455	^{1}H:4.66 ^{13}C:43.6 ^{15}N:115.2	700	NRC-IMB
HNcoCACB	1	8	^{1}H:1024 ^{13}C:64 ^{15}N:32	^{1}H:9804 ^{13}C:10213 ^{15}N:1455	^{1}H:4.66 ^{13}C:43.6 ^{15}N:115.2	700	NRC-IMB
HcCH_TOCSY	1	8	^{1}H:2048 ^{1}H:128 ^{13}C:64	^{1}H:9804 ^{1}H:5881 ^{13}C:5810	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:56.6	700	NRC-IMB
TOCSY-^{15}N HSQC	1	8	^{1}H:2048 ^{1}H:128 ^{15}N:32	^{1}H:9804 ^{1}H:7422 ^{15}N:1455	^{1}H:4.66 ^{1}H: 4.66 ^{15}N:115.2	700	NRC-IMB
NOESY-^{13}C HSQC	1	8	^{1}H:2048 ^{1}H:128 ^{13}C:64	^{1}H:9804 ^{1}H:7422 ^{13}C:5810	^{1}H:4.66 ^{1}H: 4.66 ^{13}C:56.6	700	NRC-IMB
NOESY-^{15}N HSQC	1	8	^{1}H:2048 ^{1}H:128 ^{15}N:32	^{1}H:9804 ^{1}H:7422 ^{15}N:1455	^{1}H:4.66 ^{1}H: 4.66 ^{15}N:115.2	700	NRC-IMB
CNH-NOESY	1	8	^{1}H:2048 ^{13}C:128 ^{15}N:32	^{1}H:9804 ^{13}C:10213 ^{15}N:1455	^{1}H:4.66 ^{13}C:43.6 ^{15}N:115.2	700	NRC-IMB

Table 5.2: The number of possible assignments and the numbers of assigned resonances for different isotopes and resonance types of AR55 in various membrane mimetics. Data for this table was generated using CCPNMR Analysis (95).

Category	Total possible	AR55_SDS assignments		AR55_DPC assignments		AR55_LPPG assignments		AR55_HFIP assignments	
		#	%	#	%	#	%	#	%
Carbon	315	262	83	270	86	258	82	258	82
Proton	363	348	96	358	99	350	96	350	96
Nitrogen	88	70	80	70	80	70	80	70	80
Amide	126	124	98	124	98	124	98	124	98
Backbone	254	252	99	249	98	246	97	246	97
Backbone non-H	192	190	99	187	97	184	96	184	96
Side chain H	301	286	95	296	98	288	96	288	96
Side chain non-H	211	142	67	153	73	144	68	144	68

Table 5.3: Summary of the restraints employed for the final ensemble of 40 retained structures from 100 calculated structures. XPLOR-NIH energies, violation occurrences and structural statistics for AR55 solubilized in SDS, DPC, LPPG and 50% HFIP are shown.

	SDS	**DPC**	**LPPG**	**50% HFIP**
Rounds of structure calculation	9	11	18	15
Unique NOE restraints				
Total	1556	1608	1654	2355
Intraresidue	740	788	756	776
Sequential	444	424	470	594
Medium range ($\|i-j\| \leq 4$)	320	322	330	728
Long range ($\|i-j\| > 4$)	2	0	0	6
Ambiguous	50	74	98	251
Residue Ramachandran plot statistics				
Core	43.0%	39.2%	30.6%	63.0%
Allowed	45.5%	47.3%	55.6%	31.4%
Generously allowed	10.1%	9.3%	12.1%	4.2%
Disallowed	1.4%	4.2%	1.7%	1.5%
XPLOR-NIH energies (kcal/mol)[a]				
Total	43.5 ± 1.4	39.8 ± 1.7	42.3 ± 2.5	44.6 ± 2.9
NOE	1.3 ± 0.6	1.1 ± 0.6	1.5 ± 0.9	2.0 ± 1.2
Violations				
NOE violations > 0.5 Å	0	0	0	0
NOE violations of 0.3-0.5 Å	0	1	0	0
NOE violations > 0.2-0.3 Å	1	0	1	5

[a] Ranges are given by average deviations for XPLOR-NIH energies.

Table 5.4: RMSD values of the various superpositions of AR55 illustrated in Figures 5.11 and 5.12.

	Superposed Residues	RMSD (Å)
AR55 in SDS micelles	13-18	1.16
	20-25	0.61
	31-43	0.71
	46-58	0.80
AR55 in DPC micelles	14-19	0.70
	20-25	0.66
	29-46	1.33
	47-57	0.99
AR55 in LPPG micelles	22-25	0.30
	27-45	1.16
	46-58	1.20
	27-58	2.98
AR55 in 50% HFIP	13-27	0.71
	30-41	0.33
	42-59	0.56
	30-59	1.10

Table 5.5: The backbone dynamics parameters for AR55 solubilized in SDS, DPC and LPPG micelles.

Residue	Model	S^2	S_f^2	τ_e (ps)	R_{ex}
AR55 in SDS micelles					
2	5	0.24 ± 0.02	0.81 ± 0.03	0.78 ± 0.06	
3	5	0.29 ± 0.01	0.76 ± 0.02	0.66 ± 0.04	
4	5	0.24 ± 0.01	0.77 ± 0.02	0.69 ± 0.02	
5	5	0.24 ± 0.01	0.80 ± 0.01	0.73 ± 0.02	
7	5	0.31 ± 0.05	0.83 ± 0.04	0.72 ± 0.04	
9	5	0.50 ± 0.02	0.82 ± 0.02	0.99 ± 0.11	
10	5	0.45 ± 0.02	0.85 ± 0.03	0.86 ± 0.08	
12	5	0.27 ± 0.01	0.83 ± 0.01	0.85 ± 0.02	
13	5	0.37 ± 0.02	0.81 ± 0.02	0.87 ± 0.07	
14	5	0.37 ± 0.02	0.74 ± 0.02	0.85 ± 0.05	
17	5	0.30 ± 0.02	0.84 ± 0.02	0.82 ± 0.04	
18	5	0.33 ± 0.04	0.73 ± 0.05	0.91 ± 0.94	
20	2	0.49 ± 0.02		270.53 ± 64.89	
23	5	0.60 ± 0.05	0.83 ± 0.04	0.95 ± 1.32	
24	5	0.65 ± 0.07	0.83 ± 0.05	1.05 ± 3.93	
25	5	0.74 ± 0.06	0.90 ± 0.05	0.43 ± 1.36	
26	5	0.66 ± 0.02	0.89 ± 0.02	0.94 ± 0.22	
27	5	0.71 ± 0.02	0.92 ± 0.02	0.68 ± 0.10	
28	5	0.58 ± 0.02	0.91 ± 0.02	0.77 ± 0.07	
29	5	0.76 ± 0.02	0.94 ± 0.02	0.65 ± 0.09	
30	3	0.96 ± 0.09			1.53 ± 1.50
33	3	0.80 ± 0.06			9.44 ± 1.32
34	4	0.85 ± 0.11		56.35 ± 2107.82	2.78 ± 1.82
37	2	0.89 ± 0.03		74.08 ± 42.95	
40	4	0.83 ± 0.17		61.74 ± 1702.94	6.43 ± 5.29
41	4	0.85 ± 0.15		59.11 ± 1101.12	2.53 ± 2.13
42	4	0.86 ± 0.13		67.12 ± 907.30	4.01 ± 2.34
43	4	0.86 ± 0.14		113.24 ± 1379.79	5.68 ± 2.44
45	4	0.86 ± 0.23		383.19 ± 86.52	0.49 ± 3.60
46	2	0.88 ± 0.00		562.35 ± 151.60	
47	4	0.87 ± 0.04		255.48 ± 977.79	2.08 ± 0.75
49	5	0.71 ± 0.03	0.95 ± 0.02	0.69 ± 0.10	
51	2	0.89 ± 0.02		534.98 ± 183.69	
52	4	0.90 ± 0.09		211.12 ± 1307.69	4.74 ± 1.86
56	3	0.98 ± 0.02			1.09 ± 0.33
57	2	0.92 ± 0.03		323.35 ± 3246.60	
58	2	0.82 ± 0.02		678.44 ± 75.02	
59	5	0.73 ± 0.01	0.98 ± 0.01	0.72 ± 0.03	

Residue	Model	S^2	S_f^2	τ_e (ps)	R_{ex}
AR55 in SDS micelles					
60	5	0.64 ± 0.01	0.93 ± 0.01	0.91 ± 0.04	
63	5	0.39 ± 0.01	0.91 ± 0.01	0.98 ± 0.02	
64	5	0.42 ± 0.02	0.84 ± 0.02	0.95 ± 0.03	
AR55 in DPC micelles					
2	2	0.15 ± 0.01		593.00 ± 3.54	
3	5	0.23 ± 0.12	0.85 ± 0.25	0.58 ± 0.07	
4	5	0.22 ± 0.05	0.78 ± 0.04	0.74 ± 0.03	
5	5	0.26 ± 0.00	0.80 ± 0.00	0.81 ± 0.01	
6	5	0.33 ± 0.03	0.86 ± 0.03	0.80 ± 0.02	
9	5	0.48 ± 0.01	0.87 ± 0.01	0.98 ± 0.03	
10	5	0.50 ± 0.02	0.88 ± 0.02	1.14 ± 0.06	
11	5	0.61 ± 0.02	0.95 ± 0.02	1.02 ± 0.08	
12	5	0.46 ± 0.01	0.90 ± 0.02	1.01 ± 0.05	
13	5	0.39 ± 0.01	0.81 ± 0.01	1.03 ± 0.03	
14	5	0.38 ± 0.01	0.84 ± 0.01	0.89 ± 0.02	
17	5	0.39 ± 0.01	0.83 ± 0.01	0.96 ± 0.02	
18	5	0.43 ± 0.01	0.89 ± 0.01	0.89 ± 0.02	
20	5	0.53 ± 0.01	0.89 ± 0.01	0.99 ± 0.04	
21	5	0.70 ± 0.02	0.97 ± 0.02	0.98 ± 0.06	
22	4	0.73 ± 0.10		862.25 ± 63.10	1.37 ± 3.01
23	5	0.71 ± 0.01	0.96 ± 0.01	0.95 ± 0.05	
24	2	0.79 ± 0.02		802.05 ± 1146.18	
25	4	0.82 ± 0.17		896.51 ± 122.22	1.85 ± 2.97
26	5	0.76 ± 0.01	0.97 ± 0.01	0.82 ± 0.06	
27	2	0.78 ± 0.01		808.26 ± 45.85	
28	2	0.77 ± 0.09		828.14 ± 5014.29	
29	2	0.86 ± 0.01		734.53 ± 85.39	
30	4	0.91 ± 0.04		110.65 ± 1698.80	3.06 ± 1.05
31	2	0.93 ± 0.01		62.58 ± 16.67	
33	3	0.97 ± 0.04		13.78 ± 4.17	
34	1	0.99 ± 0.01			
36	2	0.96 ± 0.02		85.88 ± 1711.27	
38	4	0.94 ± 0.04		86.63 ± 3812.16	2.52 ± 1.02
42	2	0.94 ± 0.02		83.17 ± 54.31	
43	4	0.92 ± 0.03		792.94 ± 2281.57	9.34 ± 1.27
45	2	0.86 ± 0.03		781.33 ± 2318.65	
46	4	0.86 ± 0.01		729.46 ± 61.43	1.29 ± 0.34
47	2	0.92 ± 0.02		609.85 ± 2976.07	
48	2	0.91 ± 0.02		498.03 ± 4227.57	

Residue	Model	S^2	S_f^2	τ_e (ps)	R_{ex}
AR55 in DPC micelles					
49	4	0.90 ± 0.02		703.27 ± 117.48	3.24 ± 0.55
50	4	0.91 ± 0.13		680.11 ± 628.64	3.88 ± 3.89
51	4	0.94 ± 0.02		400.27 ± 746.02	1.01 ± 0.51
52	2	0.92 ± 0.02		842.02 ± 191.87	
53	4	0.91 ± 0.09		769.13 ± 578.94	3.23 ± 2.04
54	4	0.90 ± 0.01		617.99 ± 79.46	2.90 ± 0.33
55	5	0.82 ± 0.02	0.96 ± 0.02	0.81 ± 0.08	
57	5	0.75 ± 0.01	0.98 ± 0.01	0.91 ± 0.04	
58	5	0.48 ± 0.01	0.92 ± 0.01	0.96 ± 0.02	
59	5	0.37 ± 0.01	0.87 ± 0.02	1.12 ± 0.03	
60	5	0.36 ± 0.01	0.83 ± 0.02	1.02 ± 0.04	
61	5	0.34 ± 0.02	0.84 ± 0.02	0.92 ± 0.03	
62	5	0.28 ± 0.04	0.85 ± 0.03	0.86 ± 0.02	
64	5	0.15 ± 0.11	0.84 ± 0.40	0.56 ± 0.05	
AR55 in LPPG micelles					
2	2	0.09 ± 0.01		442.60 ± 3.74	
3	2	0.06 ± 0.00		439.96 ± 1.05	
4	5	0.04 ± 0.14	0.64 ± 0.32	0.61 ± 0.14	
5	2	0.11 ± 0.00		601.88 ± 1.09	
6	2	0.07 ± 0.00		734.40 ± 1.53	
7	2	0.07 ± 0.00		566.69 ± 1.22	
8	2	0.15 ± 0.00		726.99 ± 1.89	
10	5	0.12 ± 0.13	0.76 ± 0.23	0.83 ± 0.09	
11	2	0.12 ± 0.01		799.81 ± 3.59	
12	2	0.11 ± 0.01		716.17 ± 2.08	
13	5	0.00 ± 0.00	0.68 ± 0.05	0.89 ± 0.03	
14	2	0.11 ± 0.01		638.19 ± 1.50	
15	2	0.13 ± 0.00		789.18 ± 2.16	
16	2	0.16 ± 0.00		638.37 ± 1.54	
17	5	0.10 ± 0.02	0.74 ± 0.03	0.72 ± 0.01	
18	2	0.13 ± 0.00		657.41 ± 1.93	
19	5	0.19 ± 0.02	0.80 ± 0.01	0.74 ± 0.01	
20	2	0.21 ± 0.00		806.92 ± 4.11	
21	5	0.30 ± 0.00	0.81 ± 0.01	0.83 ± 0.01	
23	5	0.38 ± 0.00	0.83 ± 0.01	0.94 ± 0.01	
24	5	0.53 ± 0.01	0.84 ± 0.01	0.95 ± 0.03	
25	5	0.67 ± 0.01	0.98 ± 0.01	0.91 ± 0.04	
26	5	0.51 ± 0.00	0.89 ± 0.00	0.88 ± 0.01	
27	5	0.63 ± 0.01	0.91 ± 0.01	0.93 ± 0.02	

Residue	Model	S^2	S_f^2	τ_e (ps)	R_{ex}
AR55 in LPPG micelles					
28	5	0.60 ± 0.01	0.91 ± 0.01	0.91 ± 0.02	
29	5	0.69 ± 0.01	0.91 ± 0.02	0.96 ± 0.04	
30	4	0.88 ± 0.01		648.44 ± 42.51	2.37 ± 0.74
33	4	0.97 ± 0.02		143.76 ± 4542.14	2.88 ± 0.58
34	5	0.85 ± 0.03	0.93 ± 0.03	0.69 ± 0.11	
38	2	0.82 ± 0.01		715.14 ± 31.00	
40	2	0.95 ± 0.01		482.66 ± 5186.52	
41	4	0.96 ± 0.01		387.27 ± 4816.82	0.43 ± 1.67
42	4	0.94 ± 0.01		457.35 ± 1258.73	3.03 ± 1.16
43	4	0.86 ± 0.01		793.12 ± 55.51	4.09 ± 0.76
45	4	0.82 ± 0.01		831.58 ± 42.85	3.70 ± 0.79
46	4	0.84 ± 0.01		768.65 ± 34.99	1.83 ± 0.75
47	4	0.84 ± 0.01		754.69 ± 71.65	9.93 ± 1.83
48	4	0.88 ± 0.01		605.75 ± 37.19	4.82 ± 2.44
49	5	0.55 ± 0.09	0.70 ± 0.09	1.05 ± 0.16	
50	4	0.93 ± 0.01		632.87 ± 1876.63	5.10 ± 0.77
52	4	0.92 ± 0.01		730.34 ± 123.75	6.07 ± 0.90
53	5	0.65 ± 0.19	0.81 ± 0.14	0.96 ± 0.15	
54	5	0.12 ± 0.11	0.73 ± 0.04	16.97 ± 3.61	
55	2	0.92 ± 0.02		834.85 ± 1617.61	
56	4	0.90 ± 0.01		666.19 ± 2424.83	6.57 ± 2.95
57	4	0.85 ± 0.02		690.98 ± 54.48	5.12 ± 1.30
58	5	0.60 ± 0.01	0.91 ± 0.01	0.95 ± 0.02	
59	5	0.47 ± 0.01	0.87 ± 0.01	1.04 ± 0.02	
60	5	0.42 ± 0.01	0.89 ± 0.02	1.07 ± 0.02	
61	5	0.37 ± 0.02	0.92 ± 0.03	1.12 ± 0.02	
62	5	0.32 ± 0.02	0.93 ± 0.03	1.17 ± 0.02	
63	2	0.26 ± 0.01		920.94 ± 9.04	
64	2	0.17 ± 0.01		761.71 ± 2.45	

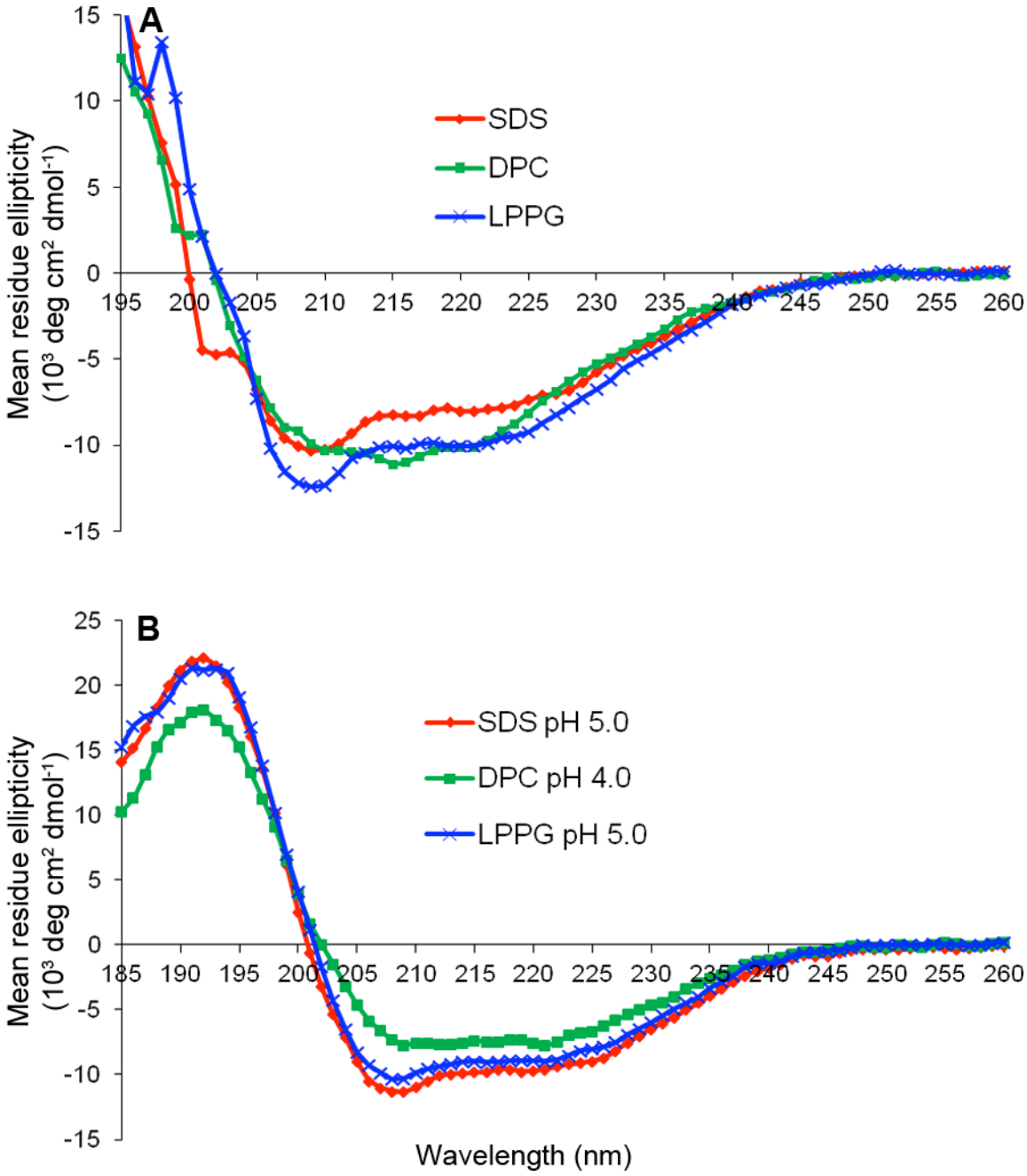

Figure 5.1: CD spectra of AR55 solubilized in SDS, DPC, and LPPG micelles obtained at 37 °C at (A) pH 7 or (B) pH 4-5.

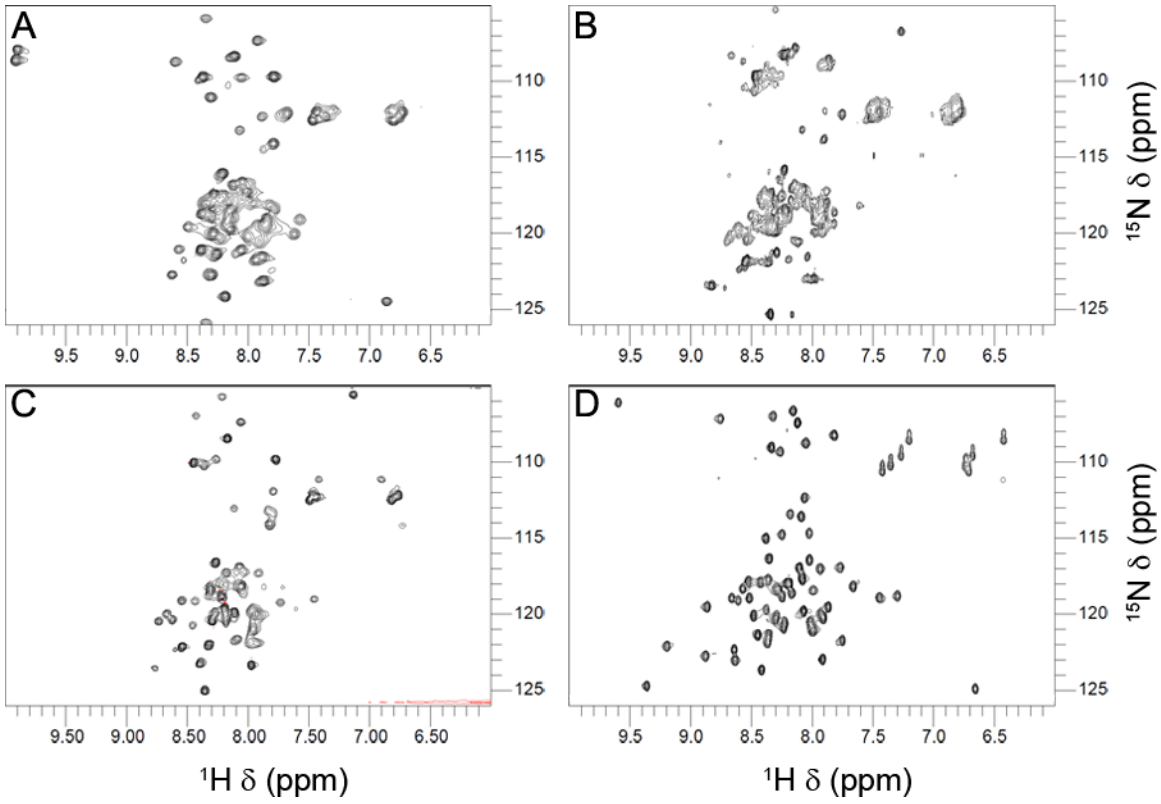

Figure 5.2: ^{1}H-^{15}N HSQC spectra of AR55 solubilized in (A) SDS micelles, (B) DPC micelles, (C) LPPG micelles and (D) 50% HFIP. All spectra were acquired at 700 MHz with 128 indirect points at NRC-IMB.

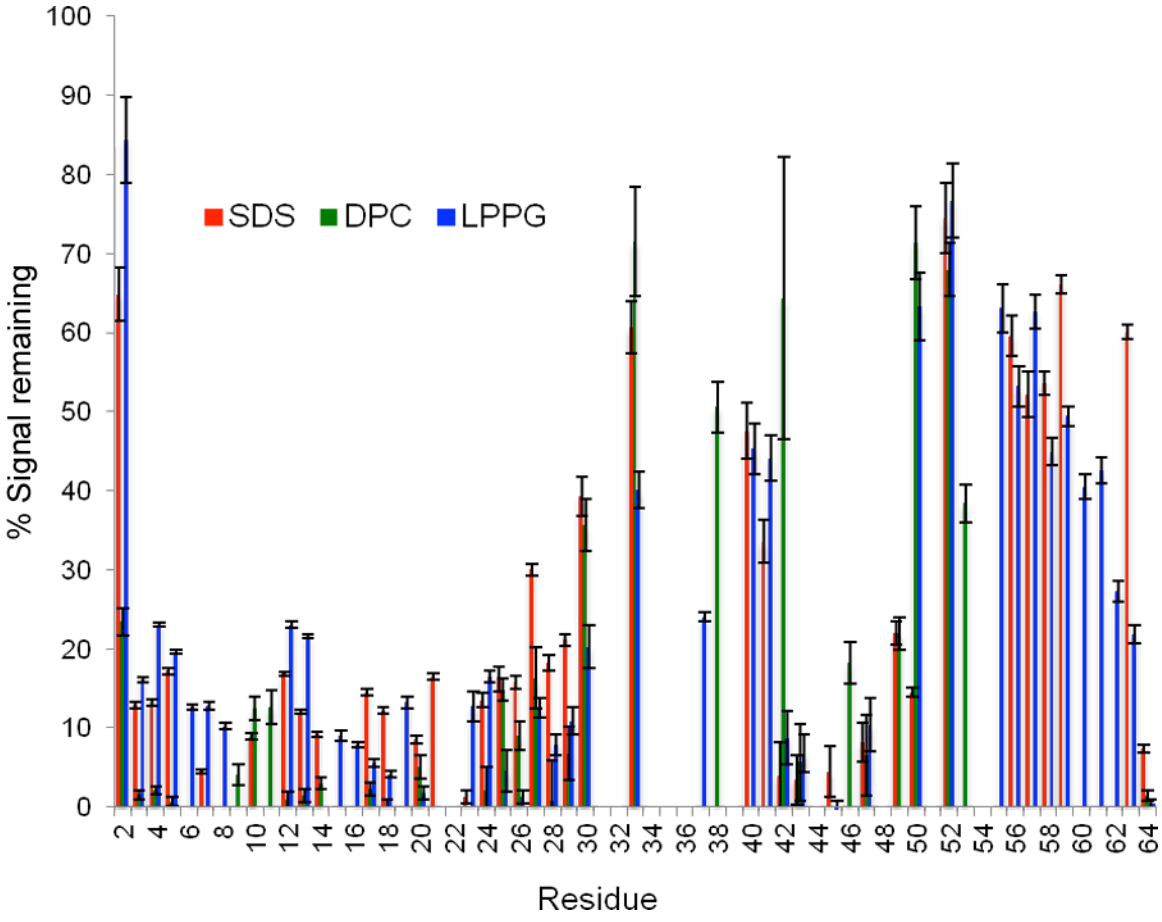

Figure 5.3: The percentage signal intensity remaining for selected signals after addition of Mn^{2+} to the indicated NMR sample. Data is shown for AR55 solubilized in SDS with 150 μM Mn^{2+}, DPC with 1 mM Mn^{2+} and LPPG with 1 mM Mn^{2+}.

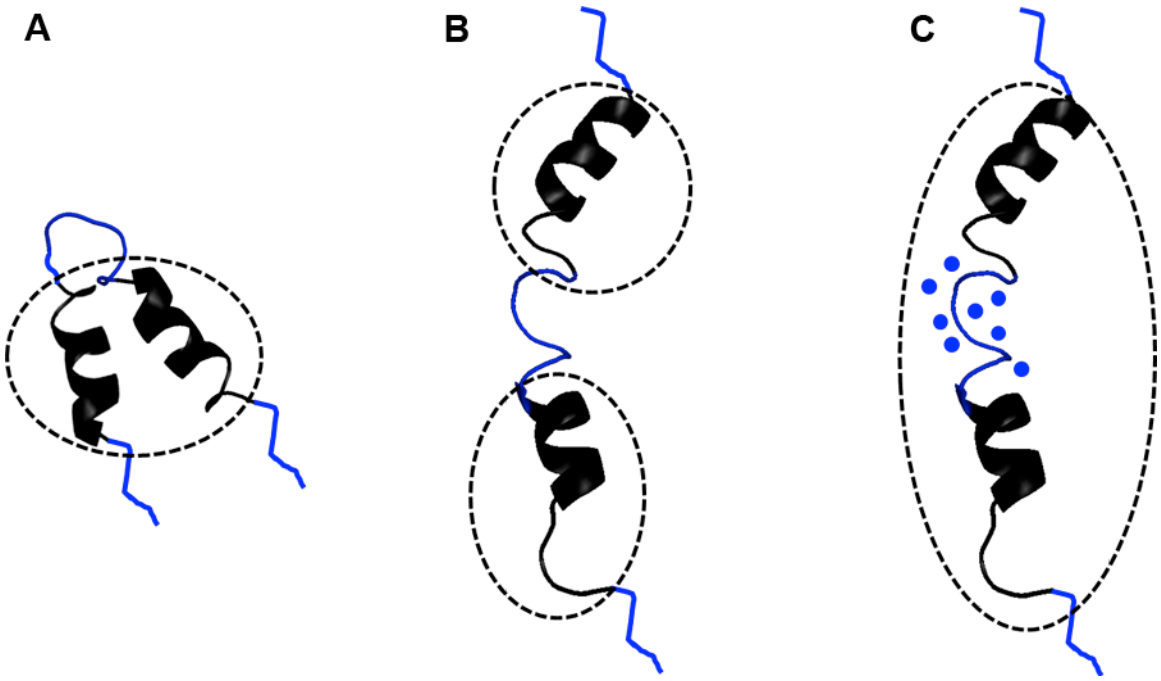

Figure 5.4: AR55 may have 3 different topologies in micelles. In this schematic water protected and water exposed regions of AR55 are shown in black and blue respectively. The approximate outline of a micelle is denoted by a dashed black line and water molecules are shown as small blue circles. In (A) AR55 is in one micelle with a hairpin turn. In (B) AR55 is shown spanning two different micelle like objects and in (C) AR55 is shown spanning one micelle with associated water in the solvent exposed region.

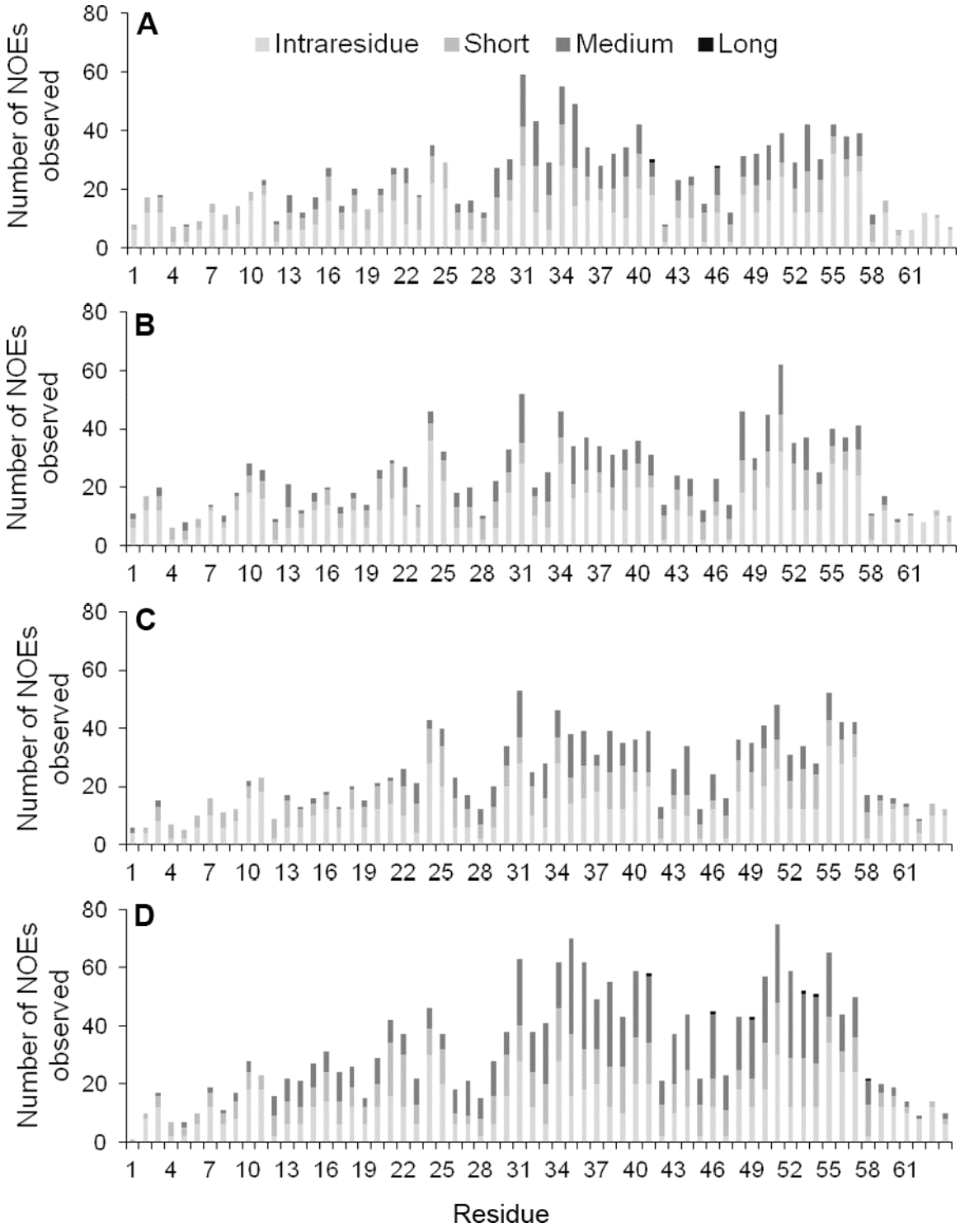

Figure 5.5: A breakdown of per-residue NOE contacts used to calculate the final ensembles of AR55 structures. Intraresidue, short (sequential), medium ($|i - j| \leq 4$) and long range NOE interactions shown for AR55 in (A) SDS, (B) DPC, (C) LPPG and (D) 50% HFIP. Note that these plots do not include the ambiguous NOE contacts listed in Table 5.3.

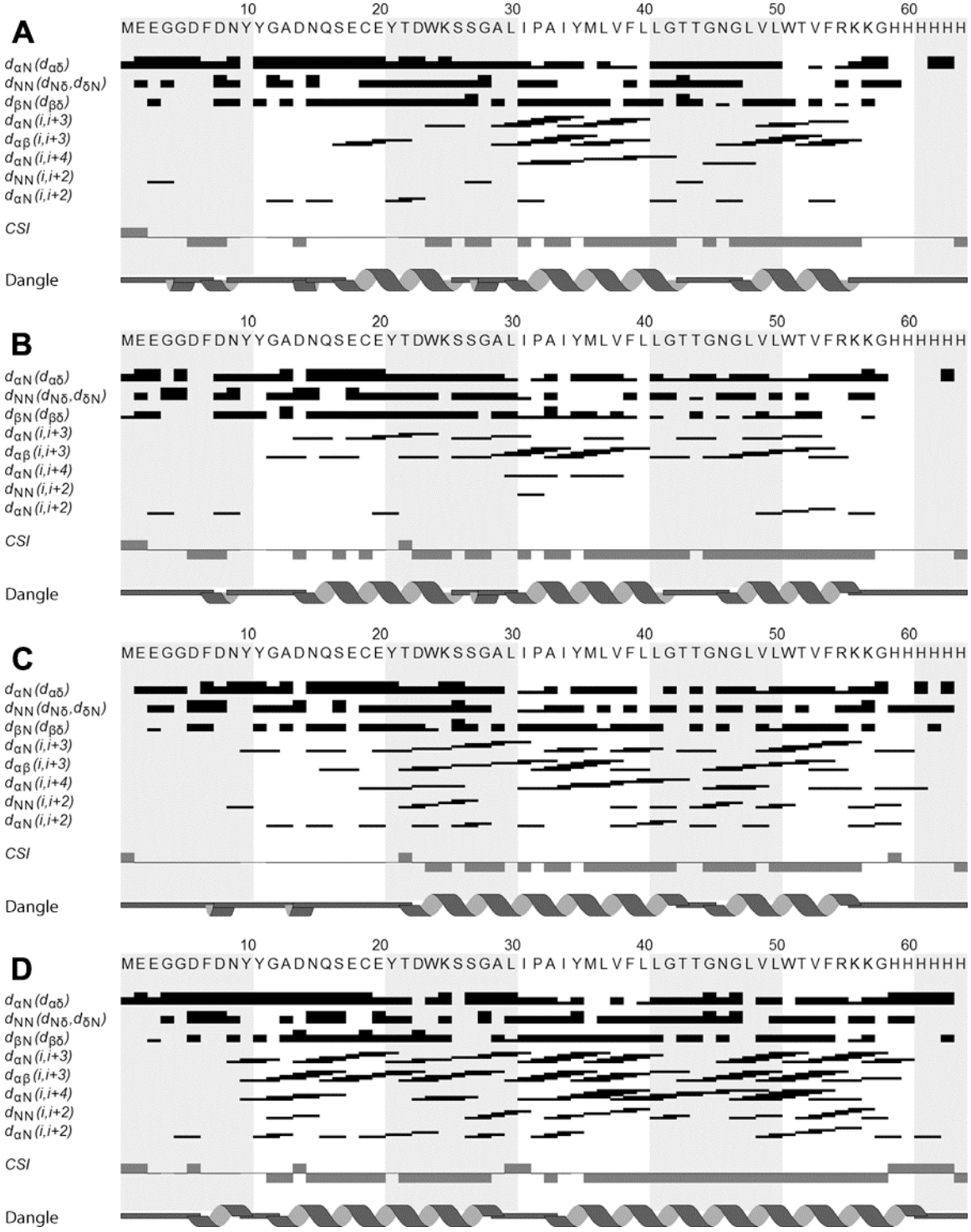

Figure 5.6: Through-space contacts observed for AR55 solubilized in (A) SDS micelles, (B) DPC micelles, (C) LPPG micelles and (D) 50% HFIP. Values of CSI (106) as well as secondary structure predictions using the DANGLE algorithm (201) are shown. This figure was generated using CCPNMR Analysis (95).

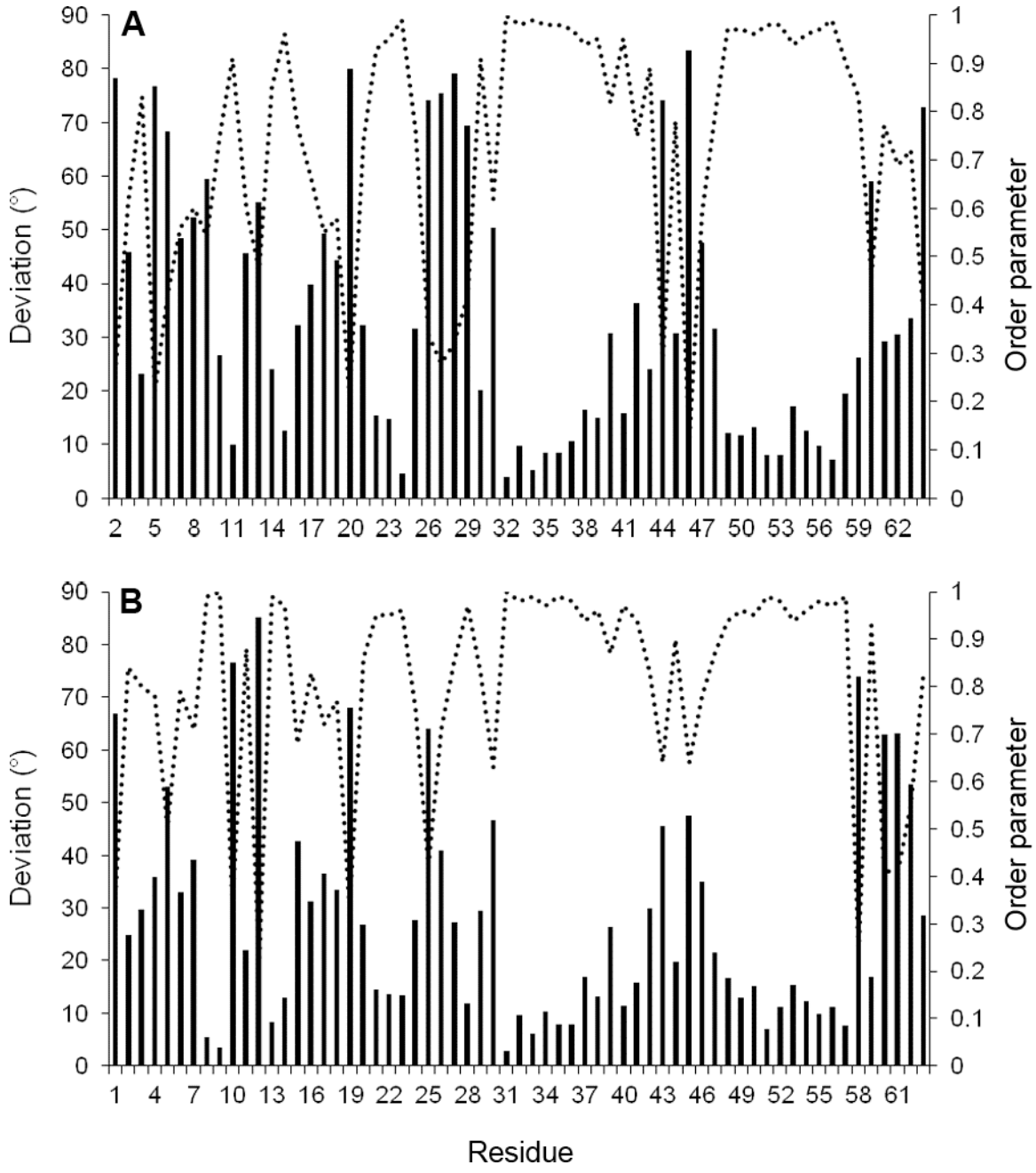

Figure 5.7: An overview of the convergence of ϕ (A) and φ (B) dihedral angles of the 40 members of the AR55 structural ensemble in SDS micelles. Both the deviation (bars) and order parameter (line) are indicated.

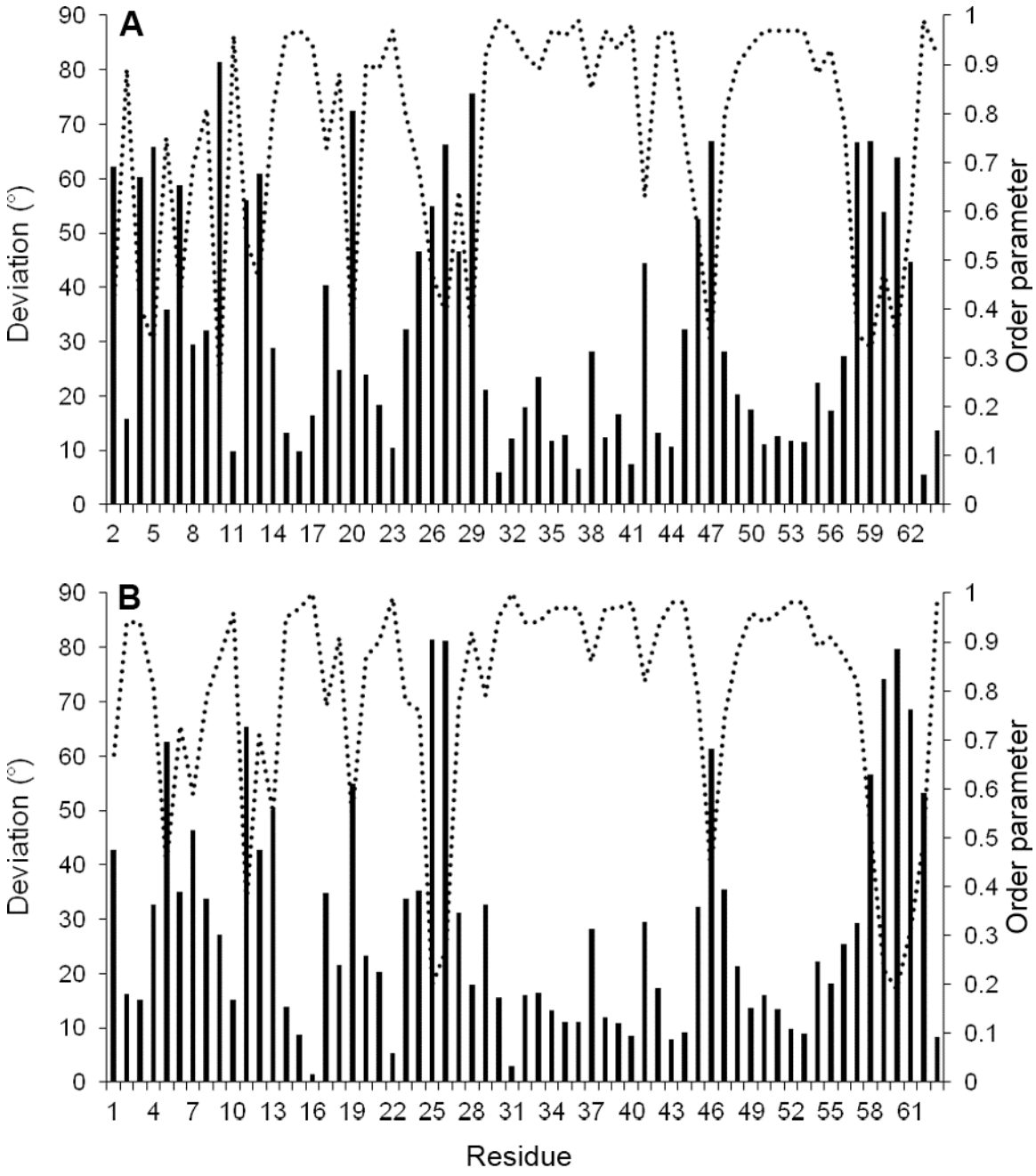

Figure 5.8: An overview of the convergence of ϕ (A) and φ (B) dihedral angles of the 40 members of the AR55 structural ensemble in DPC micelles. Both the deviation (bars) and order parameter (line) are indicated.

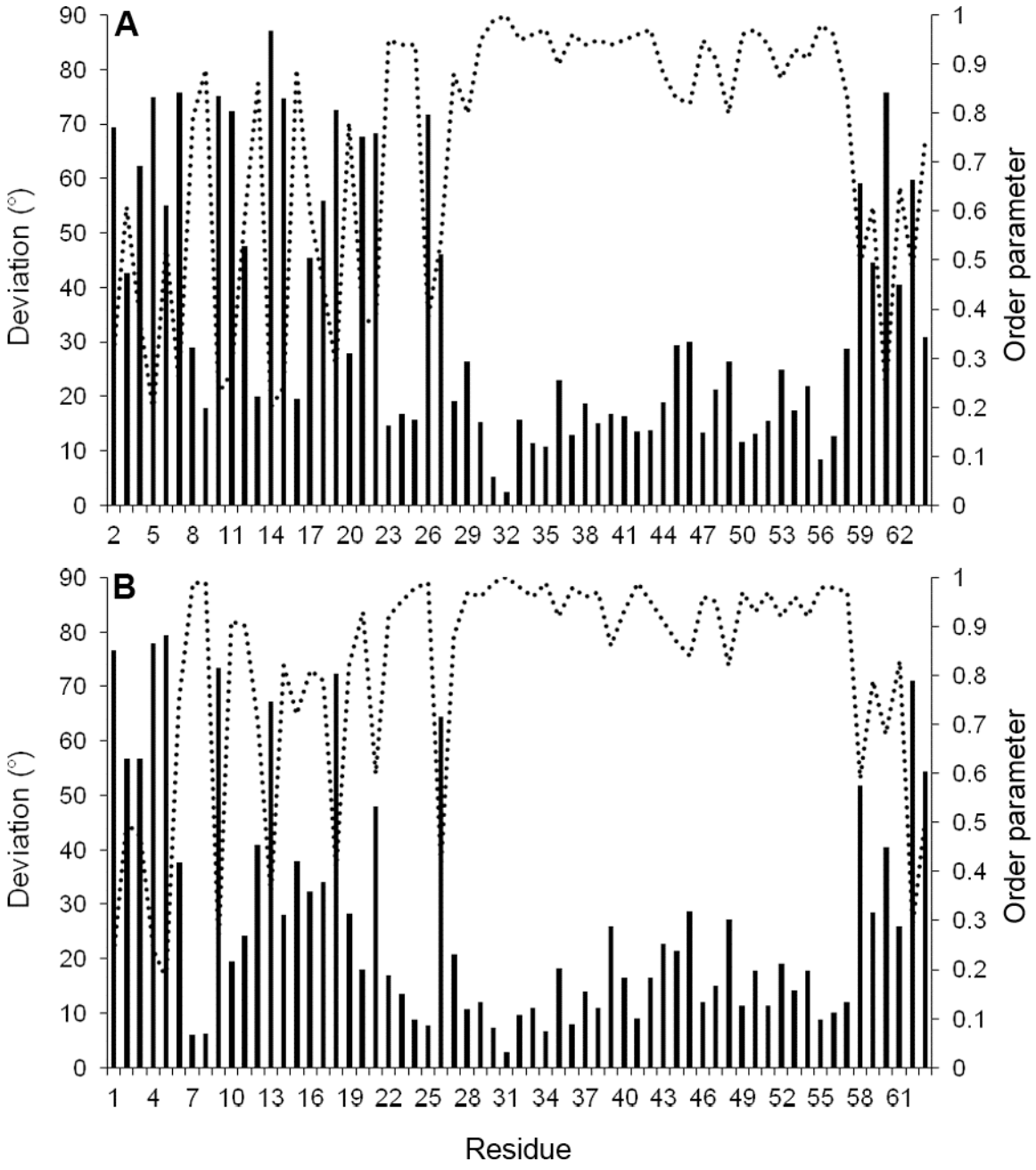

Figure 5.9: An overview of the convergence of ϕ (A) and φ (B) dihedral angles of the 40 members of the AR55 structural ensemble in LPPG micelles. Both the deviation (bars) and order parameter (line) are indicated.

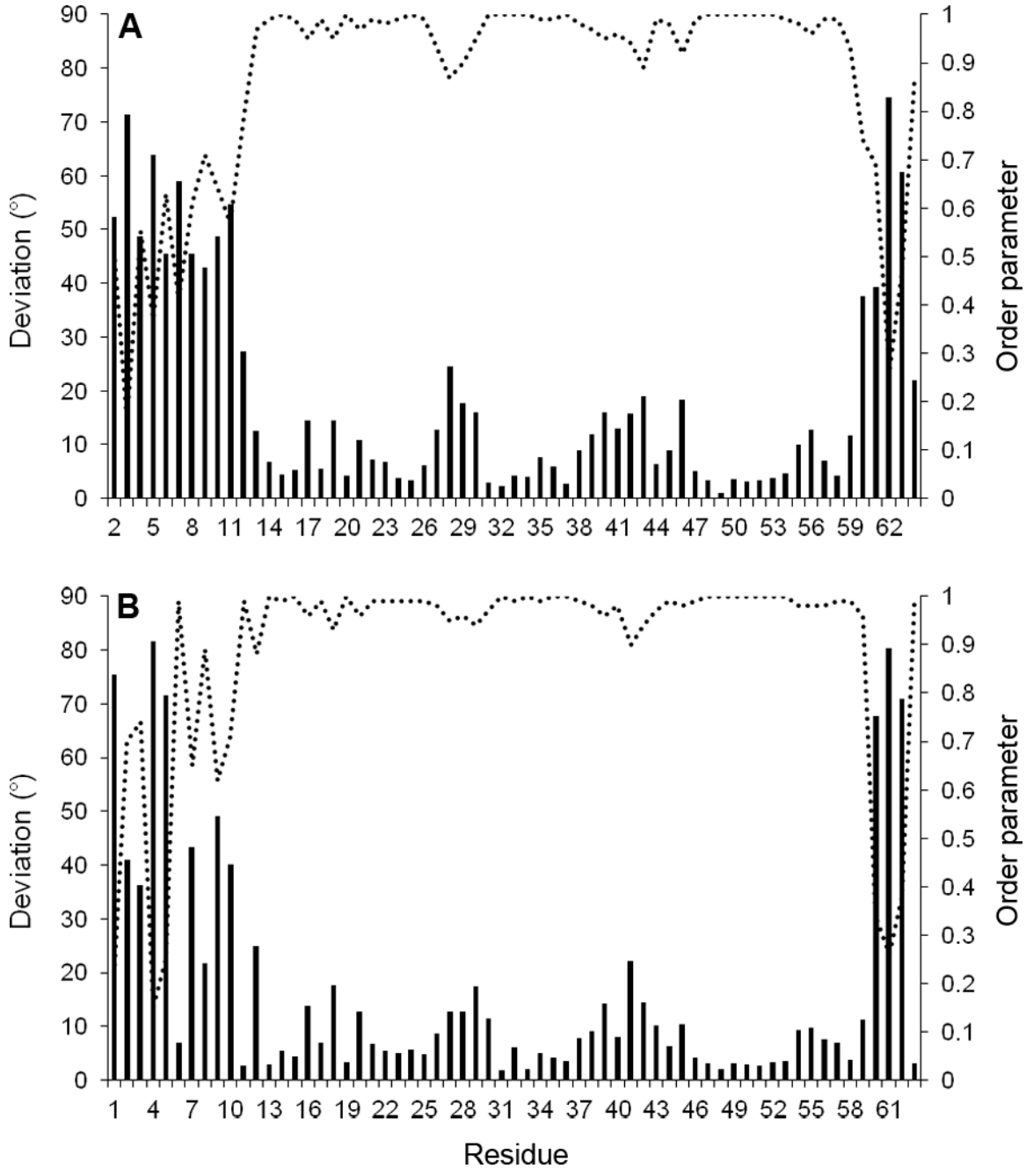

Figure 5.10: An overview of the convergence of ϕ (A) and φ (B) dihedral angles of the 40 members of the AR55 structural ensemble in 50% HFIP micelles. Both the deviation (bars) and order parameter (line) are indicated.

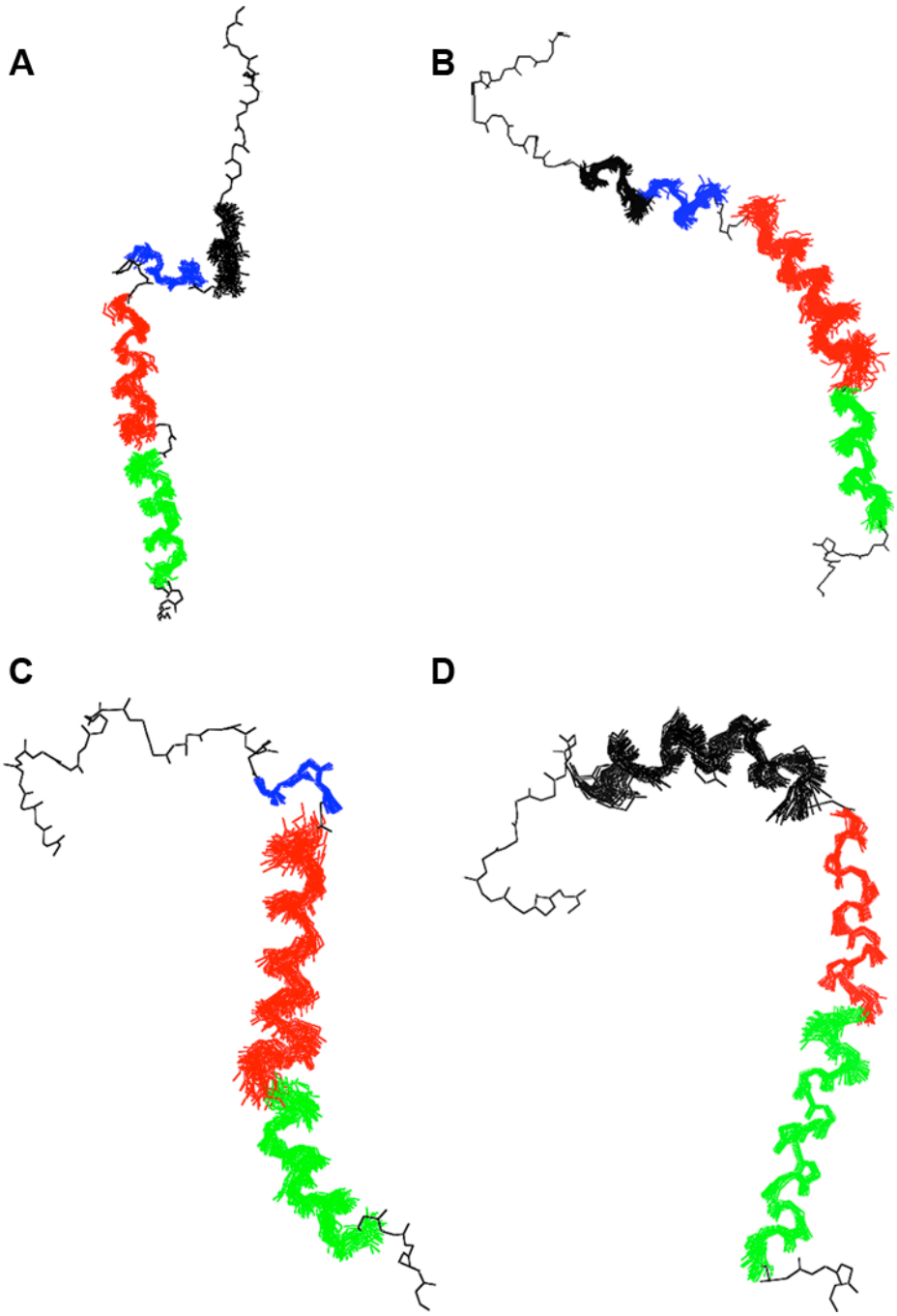

Figure 5.11: An overview of the structures of AR55 in (A) SDS, (B) DPC and (C) LPPG micelles as well as in (D) 50% HFIP. All backbone atoms for one representative structure are shown and for regions that superpose well, all ensemble members are shown. In (A) residues 13-18 are black, 20-25 are blue, 31-43 are red and 46-58 are green. In (B) residues 14-19 are black, 20-25 are blue, 29-46 are red and 47-57 are green. In (C) residues 22-25 are blue, 27-45 are red and 46-58 are green. In (D) residues 13-27 are black, 30-41 are red and 42-59 are green. Images were produced using Pymol (Delano Scientific, San Carlos, CA, USA).

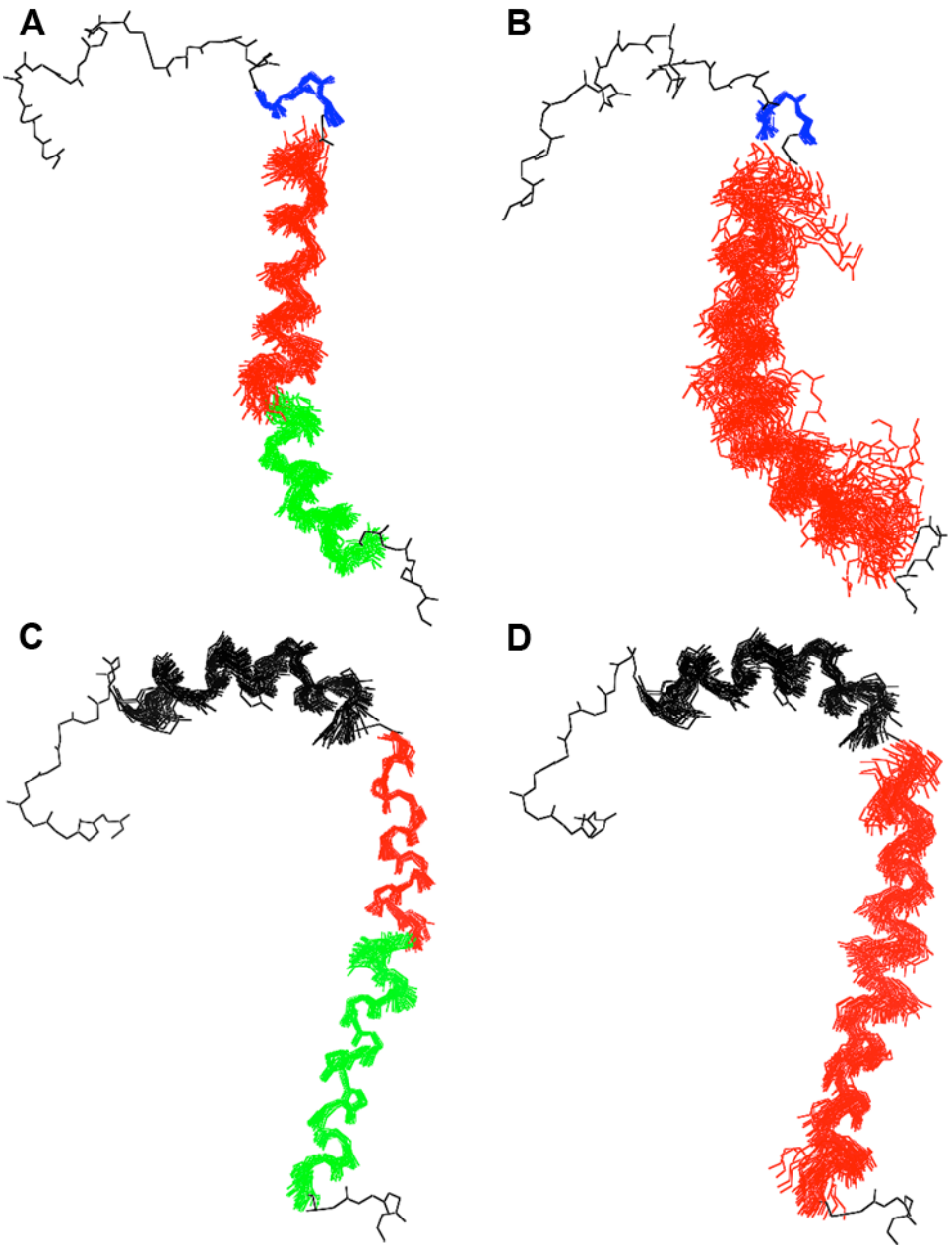

Figure 5.12: The kink of AR55 in LPPG micelles and HFIP is not very flexible. Shown are structures of AR55 in LPPG micelles (A,B) and 50% HFIP (C,D). In (A) and (C) the superposed regions are the same as those shown in Figure 5.11. In (B) residues 27-58 are superposed and in (D) residues 30-59 are superposed. Pymol (Delano Scientific, San Carlos, CA, USA) was used for visualization.

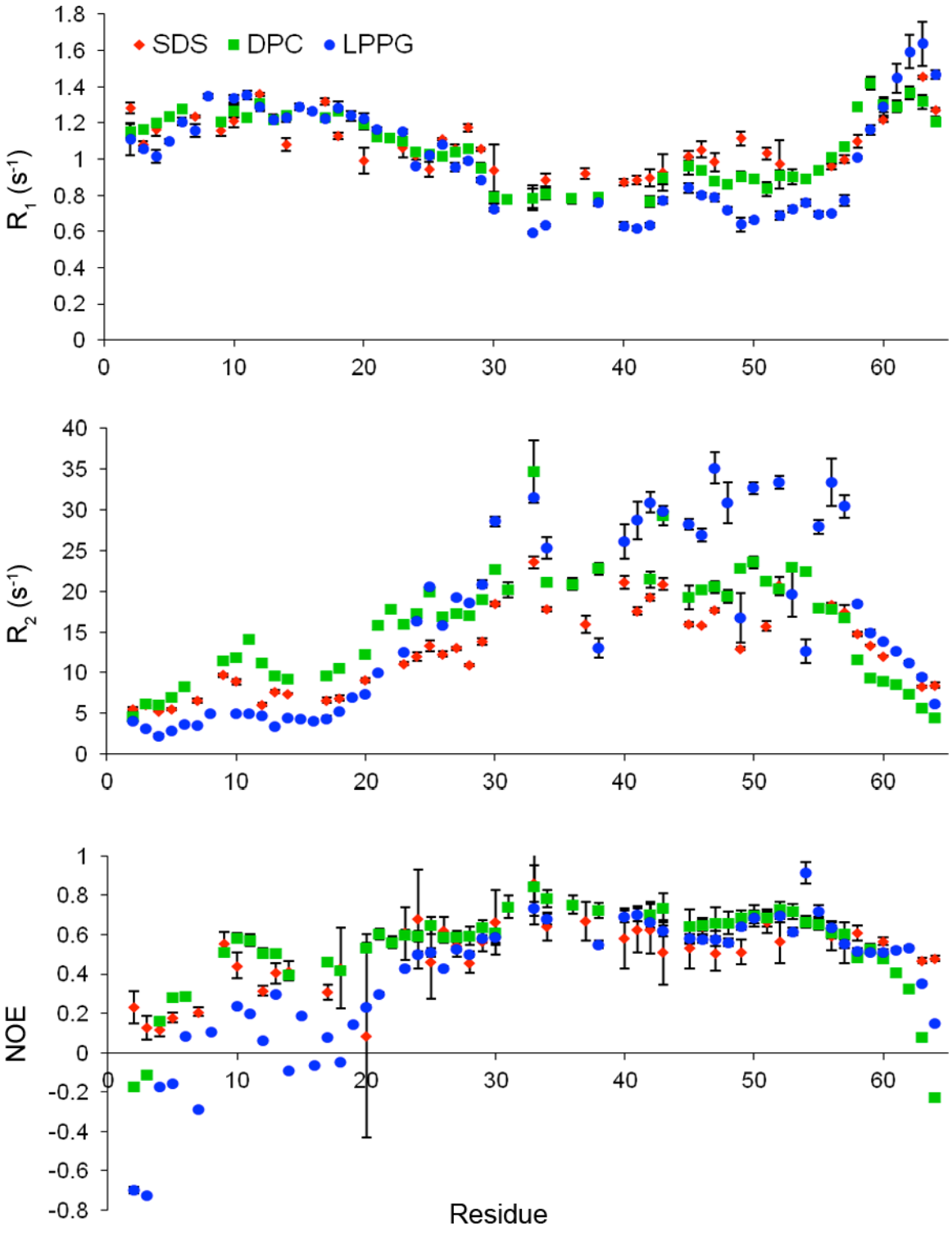

Figure 5.13: The per-residue R_1, R_2 and NOE measurements of AR55 in SDS, DPC and LPPG micelles.

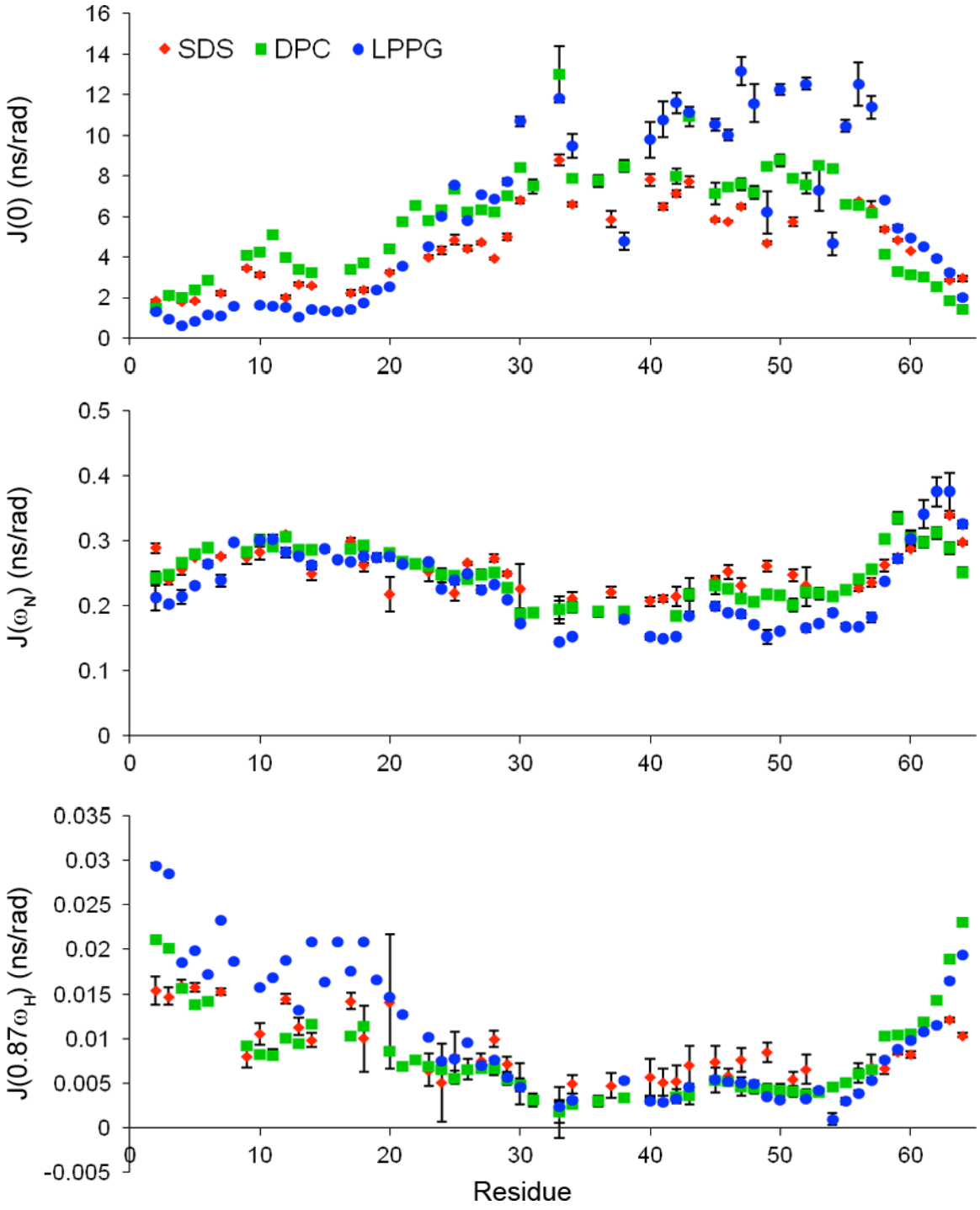

Figure 5.14: The per-residue calculated values of J(0), J(ω_N) and J(0.87ω_H) of AR55 solubilized in SDS, DPC and LPPG micelles.

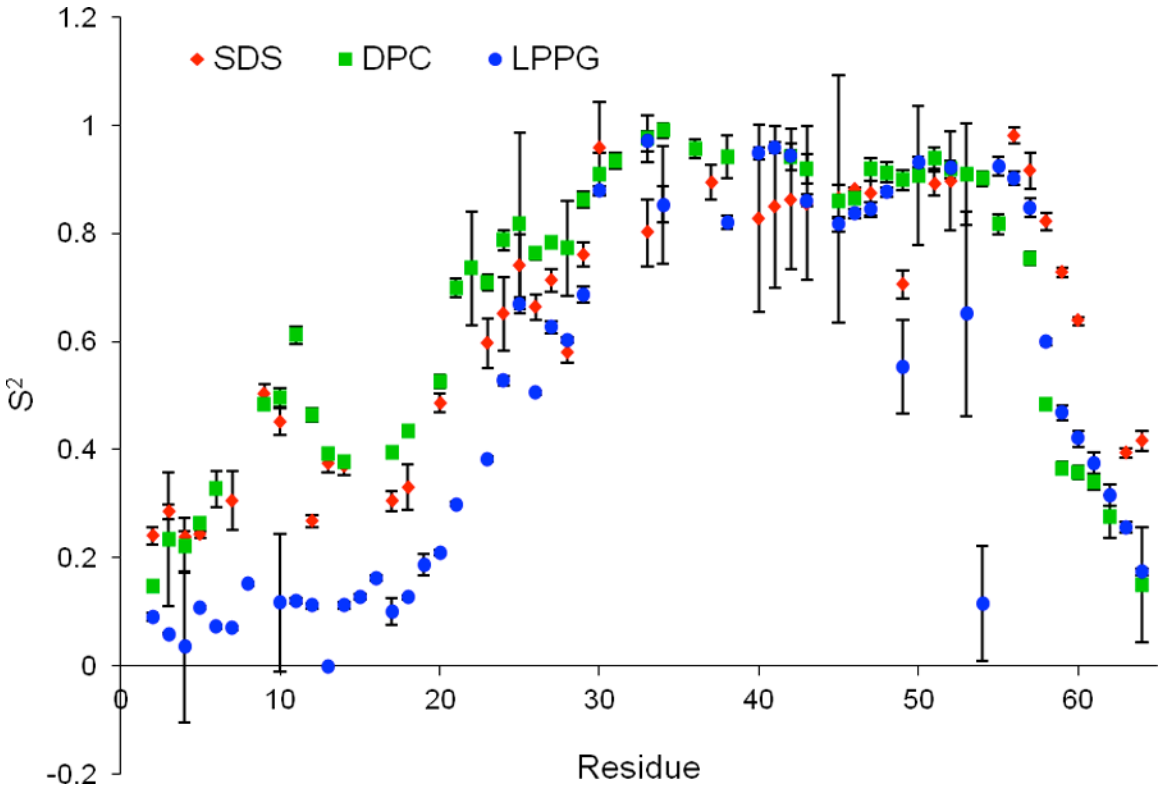

Figure 5.15: The per-residue calculated S^2 values of AR55 solubilized in SDS, DPC and LPPG micelles.

Chapter 6:

Analysis Of Kinks In Membrane Proteins (Based On My Contribution To Manuscript (4))

6.1. Introduction

In chapter 4 I discussed how residues present in the kink of AR55 are important for the function and/or folding of AR. By examining other membrane protein structures, it is clear that many membrane proteins have disruptions in their helices. It is thought that kinks in otherwise α-helical TM regions provide points that readily allow for conformational change and structural variability. This causes helical kinks to often be functionally important in proteins (126, 244, 245). For these reasons I decided to continue a project originally initiated by Michal Wieczorek (a former student of Dr. Rainey's) to computationally find and characterize kinks in protein helices. I then analyzed a large dataset of membrane protein structures in order to better understand the characteristics and causes of membrane protein kinks.

6.1.1. Previous Analyses Of Membrane Protein Helix Kinks

Despite the characteristics of an ideal α-helix being well defined, the features of disruptions in helices are poorly understood. Even though it is well recognized that membrane-spanning α-helices from mono- or polytopic membrane proteins may be

kinked (246, 247, 248, 249, 250, 251), there is disagreement in the literature concerning the nature, amino acid composition, and underlying causes of helical kinks (247, 249, 250). Compounding this issue, helical kinks are often neglected in the discussion and analysis of membrane protein structures. This neglect goes to the extent of regions of PDB (200) files for membrane proteins being annotated as helical even through regions of disruption.

Three analyses of kinking in datasets of α-helical membrane proteins have been carried out prior to this work. In 2001, Riek et al. examined 119 TM helices from 11 membrane proteins, with annotation of non-canonical helical character (*i.e.* π-helices or 3_{10}-helices) as well as kinks (247). This study identified 36 kinks in 31 helices, with ~26% of helices being kinked, and ~16% of these kinked helices having two kinks. A 2004 study by Yohannan et al. examined a set of 39 kinks from 10 membrane protein structures (249). This study had the unique aim of testing the hypothesis that kinks in TM helices can be traced back through evolution to ancestral proline residues and, as such, did not focus on kink detection and geometries. Finally, Hall et al. analyzed the positions of kinks in a larger subset of 405 TM helices (246), determining that 44% of TM helices were kinked, with only 35% of these kinks being caused by proline.

After the publication of this work two other related studies were published (250, 251). In both studies a manually annotated dataset was used. Kneissl et al. analyzed 818 helices and found 44% of them to be kinked. Meruelo et al. analyzed a smaller dataset of 41 non-homologous proteins but did not give statistics regarding the frequency of kinked helices. Both of these studies focus on helix kink prediction from protein sequence instead of kink characterization.

In this chapter I present an algorithm, implemented in a web-based Python application named MC-HELAN (Monte Carlo based HELix ANalysis). MC-HELAN uses heuristics to systematically detect and characterize helical kinks, regardless of whether the helix in question is an α-helix, π-helix or 3_{10} helix. I classify kinks (any change in helix axis) as either bends (a change in helix axis direction with all residues remaining helical) or disruptions (change in helix axis direction accompanied by a loss of helical character of the residues involved in the kink). I have used MC-HELAN to examine (as of Mar. 2010) all non-redundant α-helical membrane proteins contained in the PDB of Membrane Proteins (PDB_TM, (252, 253)). I then present general features of bends and disruptions in membrane proteins and assess the predictive nature of sequence determinants in the vicinity of kinks.

6.2. Materials And Methods

6.2.1. Dataset Preparation

The PDB_TM extensible markup language (XML) file and transformed PDB files for all structures identified as helical membrane-spanning proteins in the PDB_TM (252) were obtained (all depositions to Mar. 4, 2010). The XML file identifies TM polypeptide chains and the predicted N- and C-terminal boundaries of each TM span for each PDB entry. The PDB_TM transformed PDB files have atomic coordinates that are rotated and translated such that the putative membrane is in the xy-plane and centred at z=0 using the TMDET algorithm (254). The PISCES algorithm (255) was used to cull the set of PDB_TM entries, keeping only X-ray crystal structures with resolution less than 3.0 Å and an R-factor better than 0.75 (constraints are not placed on NMR derived structures;

only NMR and X-ray structures were used) and only the highest resolution polypeptide chain in cases of pairwise sequence identity ≥95%. For PDB entries consisting of ensembles of NMR structures, I have limited my analysis to the first model (*i.e.* that contained in the PDB_TM) rather than examining all ensemble members. MC-HELAN analysis was carried out on the data.

6.2.2. Monte Carlo Method For Helical Axis Definition

A residue at position i is denoted as a helix seed if it is not already classified as part of a helix and the following three empirically derived conditions are met: i) The dihedral angle pair (ϕ_i, ψ_i) lies within the 99.95% boundary of Lovell et al.'s statistically derived α- and 3_{10}-helical region (256); ii) The angles denoted by the triplet of points $\Theta_x = (C^{\alpha}_{i+x}, C^{\alpha}_{i}, C^{\alpha}_{i+1})$ for × = 2, 3, 4 lie within ranges of 35-50° for Θ_2, 60-80° for Θ_3 and 45-65°for Θ_4; and 3) The distances between C^{α}_{i} and C^{α}_{i+x} for × = 2, 3, 4 deviate from the distances observed in an ideal helix by at most 0.5 Å. Similar geometric criteria have been employed in other studies to locate helix kinks (247, 257). For each identified seed, a nascent helix including the backbone atoms of residues (i, i+1, … , i+4) is used to approximate a helical axis. The helix axis is denoted by the tuple (c,*v*), where c is a 3D point and *v* is a vector. The fit of the axis is minimized by a uniform random sampling in each of the components of c and *v*. An empirically determined objective function (E_h) is calculated, and the new parameters are kept only when E_h is lower. For the first 800 iterations, E_h is defined using:

$$E_h = \sum E_r \quad \text{and} \quad E_r = \begin{cases} (r-2)^2 & if\ r \geq 2 \\ 0 & if\ r < 2 \end{cases} \qquad 6.1$$

where r is the minimum distance between a given atom and the helix axis and the sum is carried out over all heavy backbone atoms of the helical segment. Next, 200 further iterations use the following modified E_r:

$$E_r = \left| r - \bar{r} \right|^2 \qquad 6.2$$

where $\bar{r}$ is the average distance of all of the atoms from the helical axis. This procedure was adequate for proper convergence of an accurate helical axis over as few as 5 residues, as is readily seen in examination of MC-HELAN output files. After definition of a helix axis, the nascent helix is extended to include a neighbouring residue if the residue has helical ϕ and ψ dihedral angles and the backbone atoms of the residue are on average < 3 Å away from the helix axis. The helix axis is then recomputed and the process repeated until no additional residues can be added. This procedure is applied from the N-terminus to the C-terminus of a peptide chain and then again from the C-terminus to the N-terminus.

Identification of disrupted TM helices is trivial based on the definition given above. Bends are identified in all instances where two helices are adjacent or overlapping. The precise location of the bend is determined as residue i using a voting procedure according to the following criteria: i) The residue with the largest deviation from ideal helical dihedral angles (ϕ=-62°, ψ=-41°, (258)) if the sum of the deviation is larger than 40°. ii) That residue i has helical Θ_x angles while i-1 and i-2 do not. iii) That the distance between C_i^α and C_{i+4}^α deviates from the distance expected in an ideal helix by more than 0.5 Å and either the C_{i-1}^α and C_{i+3}^α distance deviates by > 1 Å or the distance between C_{i-1}^α and C_{i+3}^α and C_{i-2}^α and C_{i+2}^α both deviate by > 0.5 Å from an ideal helix. In the case of ties in the voting process, the location of the bend was determined by identifying the residue i

in the overlapping segments that optimizes the fit of the helical axis for both incident helices.

6.2.3. Bend Prediction From Primary Structure

The ability to identify bends from sequence information alone was explored using a machine learning approach. The identification of a bend position was formulated as the classification for a window of 9 residues to have a bend site in position 5 or not. A dataset of 471 converged sequences of length 9 that contain a bend in the 5th (middle) position was compiled. Likewise, a set of 9067 sequences of equal length that do not contain a bend was compiled. In total, 10 classification replicates were performed using a 10-fold cross-validation on the 471 bent sequences combined with a random sample of negative examples of equal size. A feature vector for each character was defined with 1 or more numerical values from the set of possible features: hydropathy (259), membrane helical propensity (260), relative proportion of residue identity (Table 6.1), and a binary encoding for 6 physicochemical properties (hydrophobic, polar uncharged, cationic, anionic, glycine or proline). The final feature vector for one sequence instance is the concatenation of these features for each character in the input sequence. As a measure of baseline prediction considering only the presence of a proline residue, an alternative binary feature vector was used for each residue in the bend local sequence being assigned 1 for proline or 0 for non-proline. The performance of the classifiers was evaluated using precision, recall and the F-score. The classifier used was the support vector machine (SVM, (261)) using the Radial Basis Function kernel. For each classification experiment, optimal C and γ parameters for prediction were determined using a grid search.

6.3. Results And Discussion

6.3.1. Dataset Characteristics

After PISCES culling, the membrane protein dataset consisted of 840 TM helices from 205 polypeptide chains in 137 PDB files derived by both X-ray crystallography and NMR spectroscopy (Table 6.2). Highlighting the under-representation of polytopic membrane protein structures in the PDB, of the 205 TM polypeptide chains, only 129 (63%) are polytopic (Figure 5.1 (A)). The bitopic proteins were subdivided into monomeric vs. oligomeric biological assemblies, with 57/74 being structured as oligomers.

6.3.2. The MC-HELAN Server

For users wishing to analyze helical axes and kinks in a single protein (whether membrane-spanning or not), the Python program MC-HELAN is available in a web-mounted format (freely accessible at http://structbio.biochem.dal.ca/jrainey/MC-HELAN/). A sample MC-HELAN analysis of a protein, showing both the modified PDB file produced and the resulting membrane protein topology diagram, is provided in Figure 6.2. In this work, all reported statistics were derived using MC-HELAN in a batch script format.

6.3.3. Convergence Of MC-HELAN

Since the MC-HELAN algorithm relies on a heuristic approach, convergence of kinked residue identification is not guaranteed. In order to assess the degree of convergence, the

MC-HELAN analysis was completed 10 times for every protein, to evaluate the standard deviation values in kink angles and the consistency of kink locations. In this work, a kink is considered only if all 10 replicate predictions converged to a single solution.

Overall, 598 of the 616 kinks identified by MC-HELAN (>97%) converged to a unique solution. The presence of non-converged kinks was not due to failure of the algorithm in defining an accurate helix axis. Failure to converge was attributable in general to either a small kink (~5-10°) or to the location of a kink being assigned to pairs of neighbouring residues. The defined kink angles were well converged with the average value of the standard deviation of all converged kink angles being 0.043° and the largest standard deviation value being 4.2°. It should be noted that there were only 4 kinks (<0.7%) that had a standard deviation of > 1°. In these 4 cases, the kinks being analyzed had geometries that led to difficulty in exact helix axis definition by the MC-HELAN algorithm. Since both the convergence and the standard deviation for kink angle of a set of 10 predictions are a part of the standard output of MC-HELAN, lack of convergence or high standard deviation values, while generally very rare, should be employed by the user to highlight kinks where manual interpretation may be necessary.

6.3.4. Frequencies Of Kinks

Out of 810 TM segments with 100% convergence analyzed by MC-HELAN, 293 (36%) were an unperturbed helix, 462 (57%) contained a bend, 54 (7%) a disruption and 1 TM segment had no helical sections. The proportion of kinks detected (64%) is significantly higher than those detected previously in TM helices using other methods (26% (247) and 44% (246, 251)). MC-HELAN detects more kinks and, unlike

HELANAL, does not consider "curved" helices as a separate class. In addition, the extension over dynamic regions of a helix from both N- and C-termini in the MC-HELAN algorithm improves sensitivity to small deviations in the helical axis that may be missed by algorithms that rely strictly upon local parameters over fewer residues. The MC-HELAN algorithm also avoids the issue that least-squares fitting derived helical axes are dependent upon the number of residues analyzed (247). MC-HELAN analysis files of the full dataset are available for download at the MC-HELAN website.

The high prevalence of helical kinks in membrane proteins is likely reflective of both the degree of constraint and the uniqueness of the membrane environment. The physical presence of the lipid molecules and the anisotropy of a hydrophobic layer sandwiched on both sides by two potentially dramatically different hydrophilic environments limits the possible conformations of a protein. Due to this, the architectural toolkit of TM helices of membrane proteins may be hypothesized to require a high number of bends, disruptions and orientational diversity to provide functional uniqueness (246, 247, 248, 249). This is upheld by the fact that even non-kinked TM helices very rarely cross the membrane at an angle perpendicular to the membrane surface. Rather, they are observed to cross the membrane at a range of angles, with 10-30° relative to the membrane normal defined by TMDET being the most common (Figure 6.3).

6.3.5. Comparison Of MC-HELAN To HELANAL And DSSP

The membrane protein dataset was analyzed with both DSSP (262) and HELANAL (263) using the TM regions defined by the PDB_TM as boundaries for analysis (Table 6.2 and Figure 6.4). Although the ProKink (264) algorithm has been used in the past to analyze helix kinks (246), this algorithm is designed to calculate kink geometry given

helical axis information and a postulated kink position, rather than to directly detect or identify kinks *a priori*. Therefore, I did not compare ProKink directly to MC-HELAN. Although MC-HELAN, HELANAL and ProKink all have similar functions, the output data from each program is unique. HELANAL is able to determine which amino acids are kinked (at the user's discretion through parsing of the output data) but does not calculate a kink angle, while MC-HELAN provides both kinked residues and kink angle. Conversely, ProKink is optimized to determine kink geometry. ProKink may also be used to determine the kink location, but unlike HELANAL and MC-HELAN, this must be carried out in an iterative manner with intensive user analysis (*e.g.* (246)). Finally, MC-HELAN analyzes the TM helices and protein as a whole in the context of the plasma membrane, unlike either HELANAL or ProKink. Although MC-HELAN, HELANAL and ProKink have all been used to analyze helix kinks, these fundamental differences between the algorithms and output data must be kept in mind.

To test for differences in regional vs. localized kink detection, pairwise comparison between methods for kinks detected by each method at a given location was performed. For this comparison, additional parsing of output data was required. In the case of MC-HELAN, only converged kinks were considered valid for comparison. In the HELANAL algorithm, helices are classified as linear, curved and kinked. Because HELANAL does not determine the location of kinks, any residue that had a local helix axis bend angle > 20° (the cutoff used by HELANAL to categorize a helix as kinked) was considered kinked. Also, three residues are effectively lost at both the N- and C-termini of each helix by HELANAL due to the requirements of its helix axis calculation algorithm. For the DSSP analysis, any residue that did not have helical character was considered kinked. For

any of the 3 methods, an identified kink was only counted if there were at least 2 helical residues between the kink and the N- or C-terminus of the helix and there were at least 3 helical residues between it and the next nearest kink. Putative kinks located less than 3 residues apart were annotated as a single kink. Pairwise comparison was performed over windows of ±0-4 residues (Figures 6.5 and 6.6). Increasing the window size beyond ±1 residue led to only a small increase in the number of overlapping kinks observed (Figure 6.5). For consistency, all results are reported with a window size of ±4.

MC-HELAN was able to locate ~87% of kinks detected by either HELANAL or DSSP (Figure 6.5), as well as many additional kinks, demonstrating the effectiveness of the algorithm. Notably, although there are approximately the same number of regions of helical disruption located by DSSP and kinks located within TM helices by HELANAL (Table 6.2) they are largely non-overlapping, with only ~37% of kinks being detected by both methods. Finally, two kinks were detected by HELANAL with a lack of helical character as defined by DSSP that were detected at a level of < 100% convergence by MC-HELAN (Figure 6.4).

It is important to note that while MC-HELAN did locate more kinks than HELANAL, many of these kinked helices would have been classified as curved by HELANAL. In fact, out of the 840 helices analyzed, HELANAL only considers 68 helices to be linear while MC-HELAN identifies 293 converged linear helices. This large discrepancy arises from fundamental differences in the algorithms used to define helix axes, as HELANAL will fit some kinked helices to curves, while MC-HELAN will instead fit a kinked helix to multiple straight axes.

6.3.6. Architecture Of Helical Kinks

Examination of the frequency of kinks as a function of depth within the bilayer region (defined here relative to the center of the lipid membrane as determined by TMDET) demonstrates that kinks associated with the membrane surface are not overly prevalent since bend and disruption depths are highly variable (Figure 6.7). Helical disruptions appear equally probable at any depth, with a surprisingly high frequency of bends located near the center of the putative membrane bilayer, where the environment should be the most uniform, versus near the membrane surfaces. When considering bends that occur in the central 70% of the membrane, thus removing the headgroup associated bends, TM helix bends are distributed over a range of kink angles (Figure 6.8), with 10-20° being the most frequently observed. Disruptions show much greater variability in kink angle (Figure 6.8 (A)) and a reduced preference for small kink angles in comparison to bends. Surprisingly, some disruptions found by MC-HELAN are >90°. Close examination of these disruptions (Figure 6.9) indicates that they are not a failure of the algorithm. In all cases, the protein chain containing the large disruption was part of a larger multimeric complex, perhaps explaining how such a large deviation in helical character can be tolerated.

A reasonable *a priori* hypothesis would be that helical kinks in membrane proteins are more likely in situations where tertiary structuring could stabilize hydrogen bond donors and acceptors. Examination of the likelihood of bends or disruptions as a function of number of TM helices, however, clearly demonstrates that the frequency of helical kinking is independent of the number of TM helices in a membrane protein (Figure 6.1

(B)). In other words, regardless of the topology of an α-helical membrane protein, ~2/3 of its TM helices should be expected to be kinked and/or significantly disrupted.

6.3.7. Amino Acid Prevalence At Helical Kinks

The prevalence of amino acids at and in the region of helical bends and disruptions is a matter of significant literature debate. Proline residues are well recognized as inducing kinks in helices due to the disruption in hydrogen-bonding arising from the lack of a backbone amide proton (246, 247, 249). In the work of Yohannan et al., this was taken to the extreme of hypothesizing that proline residues are at the evolutionary root of all kinks in TM helices and this hypothesis was tested using a relatively small set of TM helices (249). Notably, the recent investigation of a much larger set of membrane proteins by Hall et al. (246) employing ProKink (264), found statistics somewhat at odds to those of Yohannan et al., with 19% of kinks being due to proline and 16% being associated with vestigial proline. Other studies have found ~50% (251) and 58% (250) of kinks to be associated with proline. Beyond proline, glycine is also typically assumed to be helix-destabilizing (265), although it should be noted that this may not necessarily be a valid assumption in non-polar environments (260). Since the MC-HELAN algorithm is entirely independent of amino acid sequence, it provides an ideal tool to investigate the induction of helical kinks by particular residues or motifs.

Examining the residues identified as being at the bend and within one helical turn of the 471 converged bends detected in the centre 70% of TM helices, with direct comparison to the likelihood of finding the same amino acid in a TM helix in general (Figure 6.10 and Table 6.1), several trends become apparent. First, proline is over 500% more frequent at both the i+2 and i+3 positions C-terminal to a bend (where the bend

occurs at position i) than anywhere else in a helix (Figure 6.10 (A)). However, proline is less likely to be at the position of the bend itself or N-terminal to it than it is to be in the rest of the helix. These results mirror previous studies (246, 250, 251) that find proline to be frequent two residues away from a kink, although the trend detected by MC-HELAN is much more apparent. Although proline is more frequent near a kink, only 33% of TM bends actually have a proline within one helical turn (Figure 6.11). Also, only 71% of TM helix proline residues are found within one turn of a bend or disruption, although many others appear at the termini of helices. This is a similar trend to Hall et al., who found 35% of bends to be due to proline (246), but lower than the ~50% of helical kinks due to proline found in other studies (247, 250, 251). Interestingly, the presence of a proline residue clearly shifts the distribution of bend angles towards higher values (Figure 6.8 (B)). Previous studies indicated that proline induced kinks in TM helices (247) vs. globular proteins (247, 258) have a consistent angle of ~ 23-26°, which was not upheld in this study.

Beyond proline, other residues also show increased frequency in the vicinity of TM helical bends. Most notably, aspartic acid is over 3 times more frequent at the i-4 position and lysine is over twice as frequent at the i+1 position. Decreased frequencies are also clear for a number of residues (Asn, Asp, Cys, Glu, His, Lys and Tyr) at specific positions surrounding a bend in a TM helix (Figure 6.8 (B)). Considering helical propensities in a non-polar environment (260), the observed trends of increased and decreased prevalence do not show a straightforward, direct relationship.

Disruptions in TM helices show a similar trend in amino acid frequencies to those at bends (Figure 6.12; not plotted as a function of position due to the relatively low number

of residues involved), with aspartic acid, asparagine, glutamic acid and proline all being significantly more frequent in a membrane disruption. All of these residues, notably, have relatively low helical propensities in a non-polar environment (260). Other studies show acidic and basic amino acids to be associated with kinks, but the specific nature of the association varies with the study (44, 250, 251). Finally, mirroring its relatively good helical propensity in non-polar environment (260), glycine is no more likely to be found in the vicinity of a TM helical kink than elsewhere in the helix. This glycine distribution is confirmed by Meruelo et al. (250) but not by Kneissl et al. (251). Overall, the polar or charged amino acids that show greatly increased frequency near kinks (Figures 6.11 and 6.12) may be capable of forming hydrogen-bonds to the backbone of the helix (266). This would break the typical hydrogen-bonding pattern of a α-helix, destabilizing it and allow a kink to form. It is important to note that the charged residues that are in proximity to kinks are themselves quite rare (Table 6.1), meaning that the observed discrepancies in frequencies may in some part arise from the limited dataset.

6.3.8. Prediction Of Kinks From Sequence Information

Although proline has been advanced as being essential for kink formation (249), this study and others (246, 247) find instances of proline residues that do not induce a kink and many kinks in the absence of a proline. With this in mind, I used machine learning algorithms to attempt to predict kinks in TM helices from the protein sequence. Through iteration, it was determined that the relative frequency of a residue in the vicinity of a bend (Table 6.1) for residues i+1 to i+4 was on its own the best feature to train a SVM classifier (F-score 0.59), versus training sets including side-chain identity or physicochemical properties (Table 6.3). Predicting from proline residues alone, in a

position-sensitive manner, achieves a maximum F-score of 0.50. Predicting kinks solely on the presence of proline provides the highest accuracy at 74%, but consequently the recall of bend prediction suffers greatly (38%) since most experimentally observed kinks do not involve a proline at all.

Since the classifier based on machine learning is able to accurately predict bends to some extent, primary structure definitely plays some role in creating kinks. However, even the best F-scores over all training sets and window sizes are relatively low (<0.6), indicating either that sequence information in the vicinity of a kink is only one factor in inducing bends or that the currently available dataset is still much too small. Other factors, such as tertiary structure, may therefore play large roles in kink formation. Recall, however, that bitopic and polytopic TM proteins show an equal prevalence of kinks (Figure 6.1 (B)), implying that tertiary structure is not exclusively responsible.

More recent studies using a neural network (250) or string kernels for SVM (251) have had greater success predicting helix kinks from the protein sequence. However, both studies are unable to predict all kinks successfully and conclude that primary structure is not the only factor that creates a helix kink. Kneissl et al. suggest that proline induced kinks may be due to local sequence while non-proline kinks are due to either the sequence of the helix as a whole, or the tertiary packing of the full length protein. This suggestion is in line with the results I present here especially considering the strong association of proline with kinks (Figure 6.10).

6.4. Summary

At this point in time, the exact factors causing TM helices to kink remain elusive. Sequence determinants certainly play some role but as of yet are not sufficiently well defined to accurately predict the presence of a kink. What is clear, however, is that the standard topology diagram inference that a helical membrane is made up of a bundle of canonical helices is far from the truth. In reality, topologies should be expected in which ~ 2/3 of TM helices are kinked and the membrane-traversal angles of the helical segments are highly variable (*e.g.* Figure 6.2 (C)).

Table 6.1: The relative frequencies, with respect to the frequency of the given residue in all TM helices (n = 12,079 residues), of each residue type at or within 4 residues of a TM helix bend. Only residues and bends located in the central 70% by depth of the putative membrane are counted. Frequencies of each residue in the analyzed region are given in brackets.

Residue	**Location of Residue relative to bend**								
(frequency)	**I-4**	**I-3**	**I-2**	**I-1**	**I**	**I+1**	**I+2**	**I+3**	**I+4**
Ala (12%)	1.0	1.0	0.8	1.0	0.9	0.8	0.7	0.9	1.0
Arg (1%)	0.7	1.9	0.5	1.5	1.1	0.8	1.8	0.5	0.8
Asn (1%)	0.6	1.6	1.4	1.1	1.0	1.3	0.6	0.3	0.6
Asp (1%)	3.9	1.5	1.4	1.2	0.3	0.6	0.8	0.6	2.0
Cys (1%)	0.5	1.1	1.0	1.4	1.3	0.4	0.2	1.2	0.8
Gln (1%)	1.3	0.9	1.6	0.4	1.1	0.6	0.7	0.4	1.3
Glu (1%)	1.1	0.7	1.4	0.7	0.7	0.2	0.7	1.8	1.3
Gly (10%)	0.6	0.9	1.0	1.1	0.9	1.0	0.9	1.4	0.7
His (1%)	1.4	0.3	0.6	0.9	1.1	0.8	0.5	0.5	1.4
Ile (11%)	0.9	1.0	1.0	0.7	1.3	0.9	1.0	0.7	1.0
Leu (17%)	1.2	1.1	1.1	1.1	0.8	1.1	0.9	0.7	0.9
Lys (1%)	2.3	1.6	0.8	1.1	2.2	2.8	0.0	2.0	0.9
Met (4%)	0.9	0.7	1.1	1.2	1.3	0.7	0.9	0.8	0.9
Phe (9%)	1.1	0.9	1.1	1.1	1.1	1.1	0.8	1.2	1.0
Pro (2%)	0.4	0.8	0.6	0.3	0.6	1.5	6.5	5.7	1.5
Ser (5%)	0.9	0.9	1.0	1.8	1.0	1.0	0.7	1.1	1.4
Thr (5%)	0.6	1.1	1.1	0.9	1.0	1.0	1.1	1.0	1.1
Trp (2%)	1.7	1.3	0.9	0.9	1.0	0.5	1.1	0.6	0.5
Tyr (3%)	1.5	0.8	0.7	0.3	1.4	1.7	1.2	1.1	1.1
Val (11%)	0.9	1.0	1.0	0.9	1.0	1.1	1.0	0.7	1.1

Table 6.2: TM helix characteristics (n=840) as classified by MC-HELAN, HELANAL and DSSP for the non-redundant set of currently solved protein structures (n = 205) containing at least one TM helix. (N/A is not applicable.)

	MC-HELAN[a]	HELANAL	DSSP
Linear helices	293	68	N/A
Curved helices	N/A	382	N/A
Kinked helices	516	387	N/A
Bent helices	462	N/A	N/A
Disrupted helices	54	N/A	142[c]
Bends	543	N/A	N/A
Disruptions	55	N/A	153[c]
Kinks used for pairwise comparison[b]	559	158	153

[a] The reported numbers are only for the converged helices found by MC-HELAN.
[b] Only kinks detected > 2 residues from the end of a helix. Multiple kinks within 3 residues are merged.
[c] Value obtained by parsing the files as in the pairwise analysis.

Table 6.3: The ability of classifiers trained using machine learning to predict helix bends from sequence information using training sets based either on relative frequency of a residue near a bend (Table 6.1) or on a binary proline/non-proline classification. These represent the best predictions of a wide variety of training sets and predictive window ranges that were employed.

Predictor	**Predictive residues**	**Precision**	**Recall**	**F-Score**
Relative frequency	i+1 to i+4	0.64	0.55	0.59
Relative frequency	i-4 to i+4	0.68	0.45	0.54
Proline	i-4 to i+4	0.74	0.38	0.50

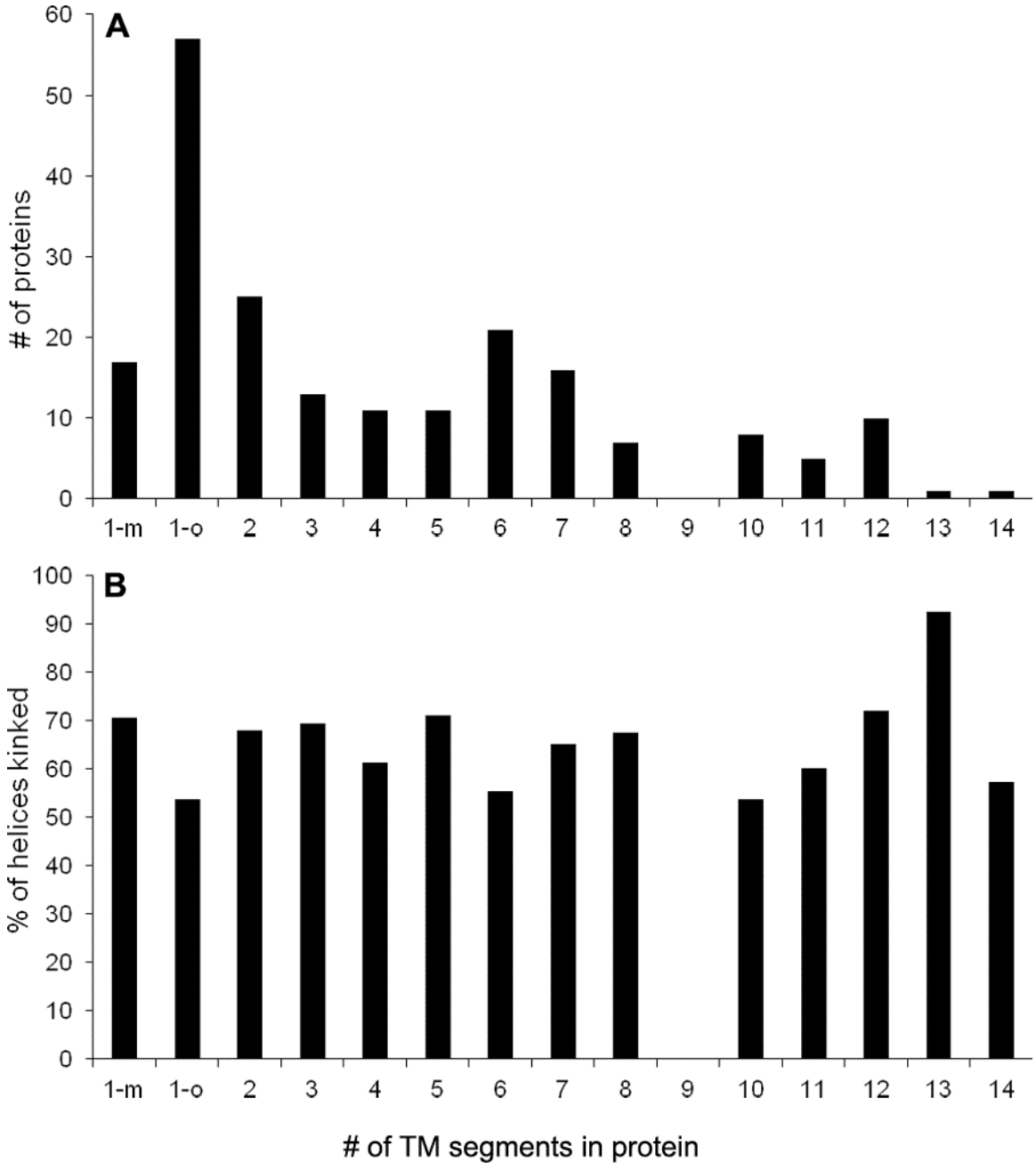

Figure 6.1: Composition of the culled helical membrane protein dataset. (A) Number of polypeptide chains (205 total) with indicated number of TM helices. (B) The proportion of kinked (*i.e.* bent or disrupted) helical segments identified by MC-HELAN in TM proteins of each size. Only TM segments that converged by MC-HELAN analysis (>97%) are considered in (B). The bitopic proteins have been subdivided into those that are in a monomeric state (1-m) and those that are part of an oligomeric complex (1-o).

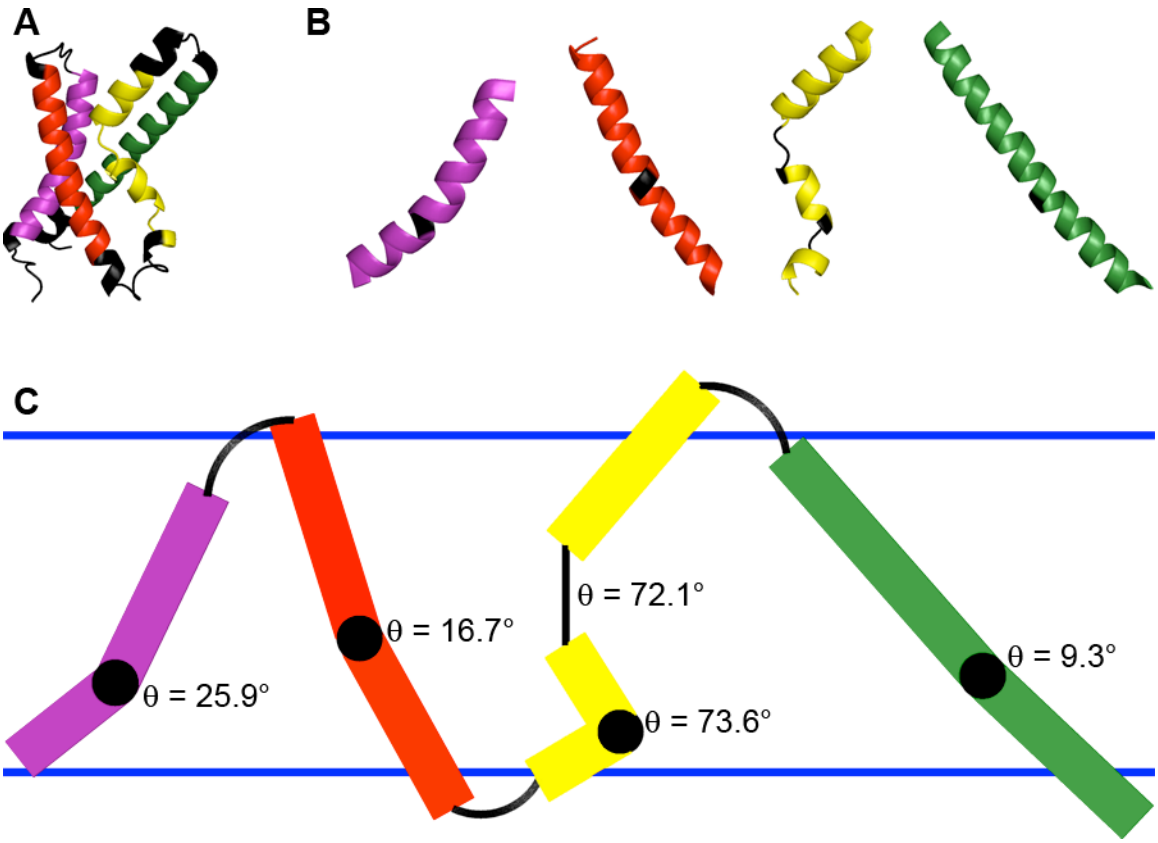

Figure 6.2: MC-HELAN analysis of the voltage sensor domain from a potassium channel (PDB entry 1ORS; (267)) with four helical TM domains. (A) Cartoon representation of the entire TM domain of 1ORS, with the residues identified by TMDET (254) as being membrane-spanning coloured with respect to TM domain. (B) MC-HELAN output specifically demonstrating the kinks observed in each of the four TM domains of 1ORS. Black residues indicate a membrane disruption or bend found by MC-HELAN. (C) The topology diagram generated by MC-HELAN for all TM helices of 1ORS with calculated bend or disruption angles indicated. The orientations and positions of each TM domain with respect to the putative membrane normal calculated by TMDET alongside kink characteristics determined by MC-HELAN are represented in the topology diagram.

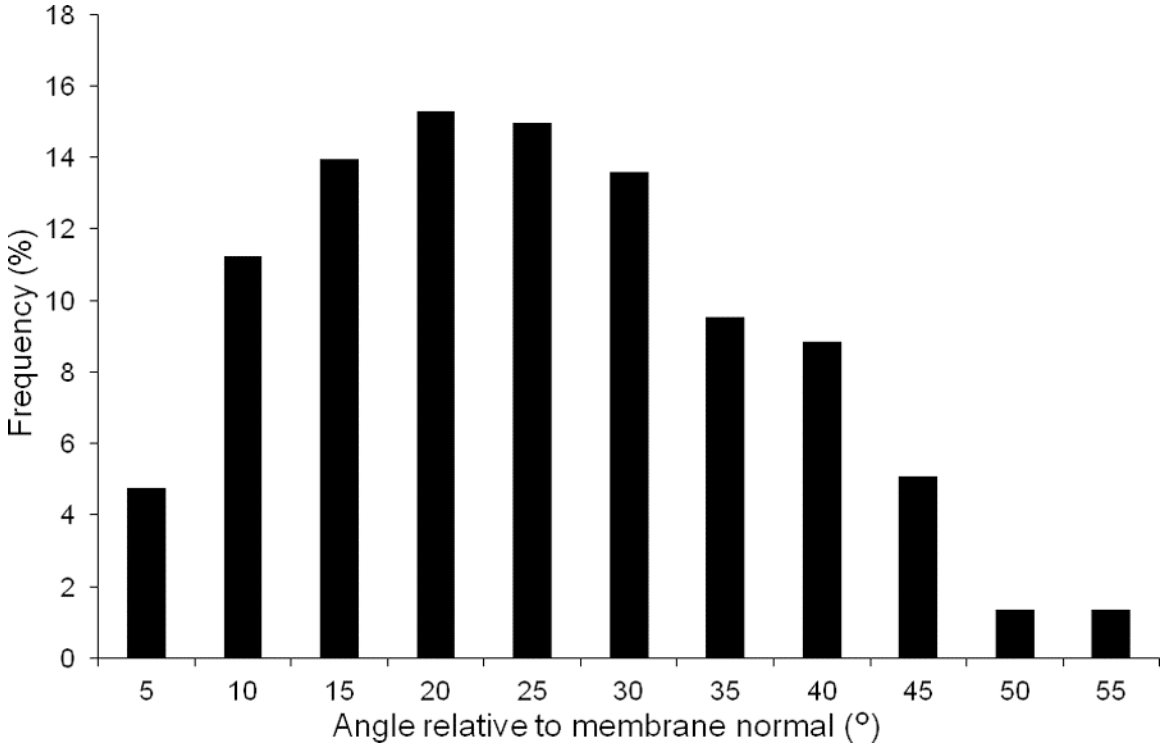

Figure 6.3: Observed distribution of the frequency of helix angles to the membrane normal for linear TM helices (n = 293) identified by MC-HELAN.

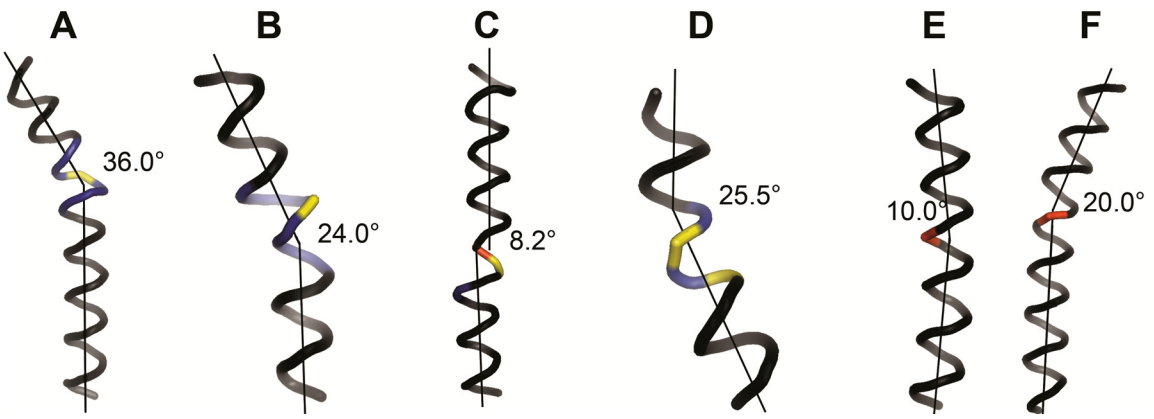

Figure 6.4: Examples of TM helices classified as kinked. Colours indicate the kinked residues detected by HELANAL (263) or DSSP (262) (blue); by MC-HELAN (red); or by MC-HELAN and either HELANAL or DSSP (yellow). Kinks are shown that were (A,B) detected by all three algorithms; (C,D) detected by HELANAL and DSSP, but did not have a converged position over 10 replicates of MC-HELAN analysis; and (E,F) by MC-HELAN only. In the illustrated cases, an overlap window of ±4 residues was used to define a kink between two different algorithms as being overlapped. In (C,D) the kinks shown by MC-HELAN were not converged, causing the kink to be 'missed' by MC-HELAN.

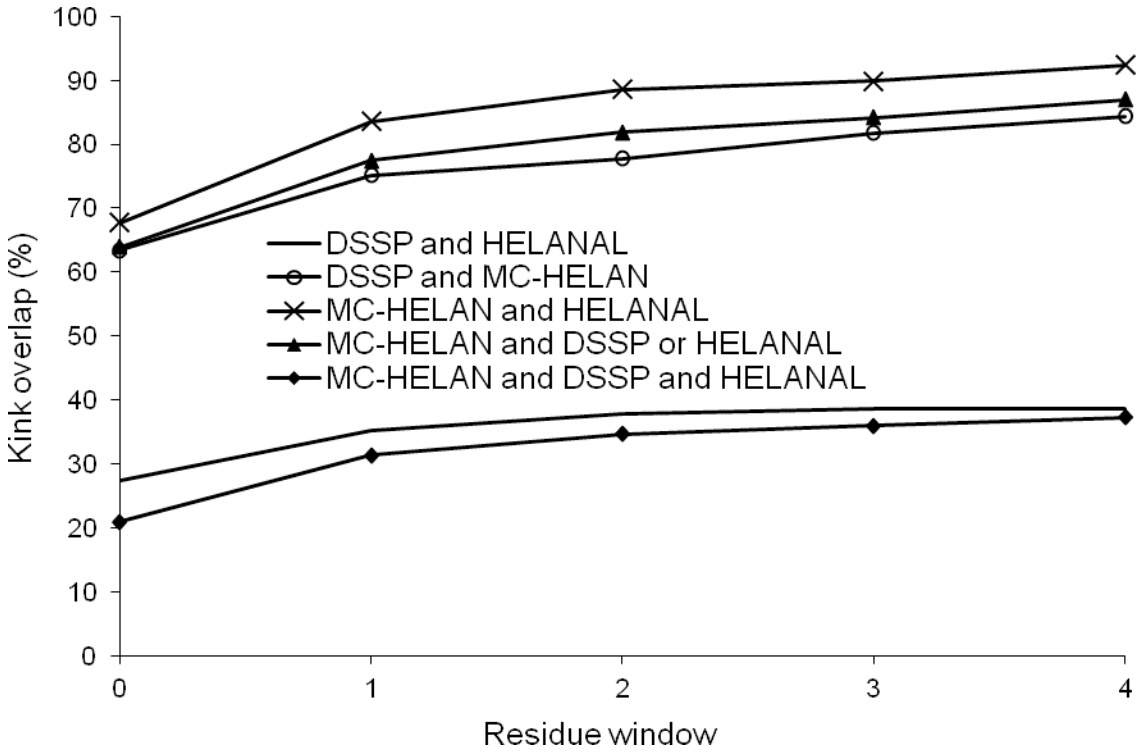

Figure 6.5: Comparison of kink detection in TM helices by MC-HELAN, HELANAL and DSSP. The percentage of overlapping bends for each category is shown for residue windows of 0-4, representing the overlap leniency. Leniency is defined such that, *e.g.*, if residue 30 were detected as kinked by MC-HELAN and residue 33 were detected as kinked by HELANAL, the kink is counted as overlapped only if the overlap leniency is ≥3.

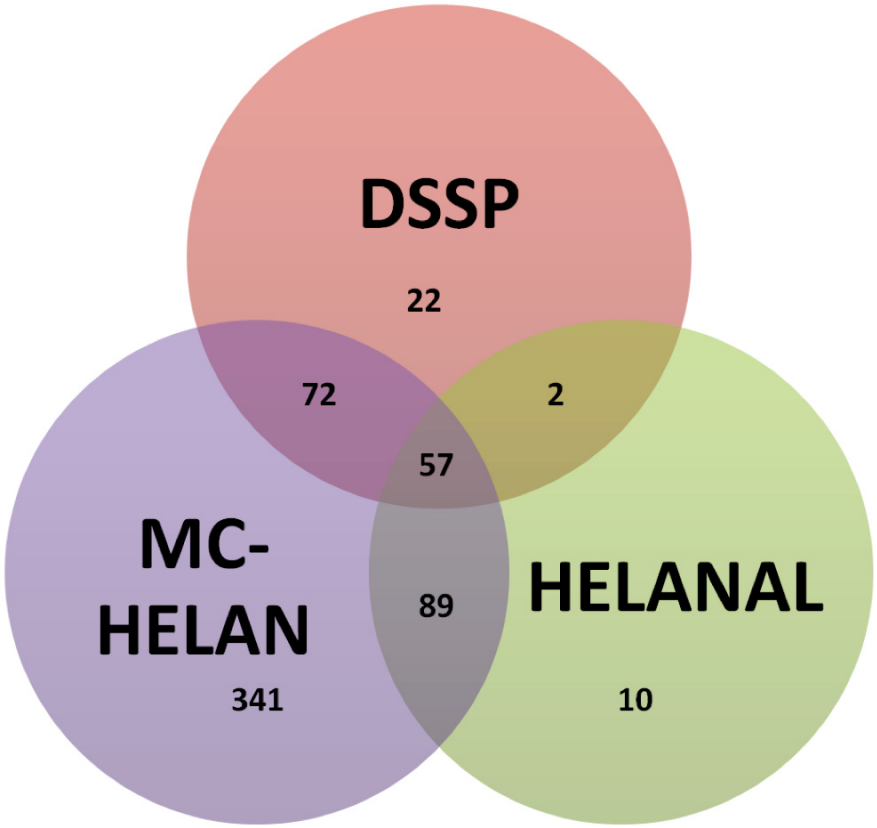

Figure 6.6: Comparison of kink detection in TM helices by MC-HELAN, HELANAL and DSSP. The residue leniency for this Venn diagram is 4, meaning that if 2 algorithms each detect a kink, so long as the detected kinks are within 4 residues of each other it is considered to be the same. The numbers in the circles represent the number of kinks in a particular class. For example, MC-HELAN detects 341 kinks that the other two algorithms do not.

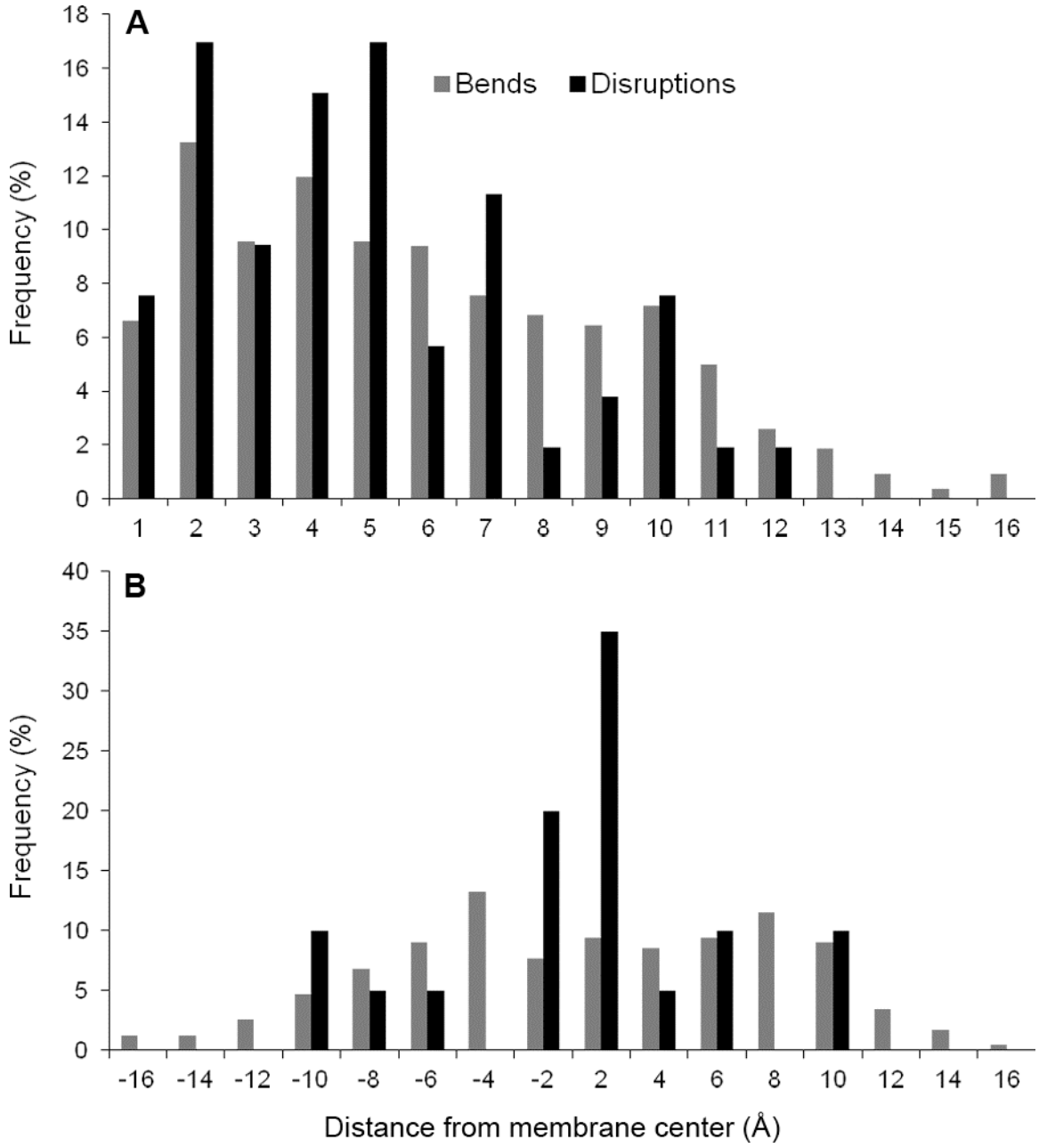

Figure 6.7: The percentage of bends (n = 471) and disruptions (n = 53) in TM helices located at different distances from the putative centre of the membrane as predicted by the TMDET algorithm (254). In (A), no distinction has been made for the side of the membrane that a kink is located. In (B), the TOPDB (268) was used to assign the bends (n=234) and disruptions (n=20) as either being on the outside (positive values) or the inside (negative values) of the membrane. The inside face is defined as being the cytoplasmic face of the membrane using the TOPDB convention. The shown distances represent the upper limit of histogram bins.

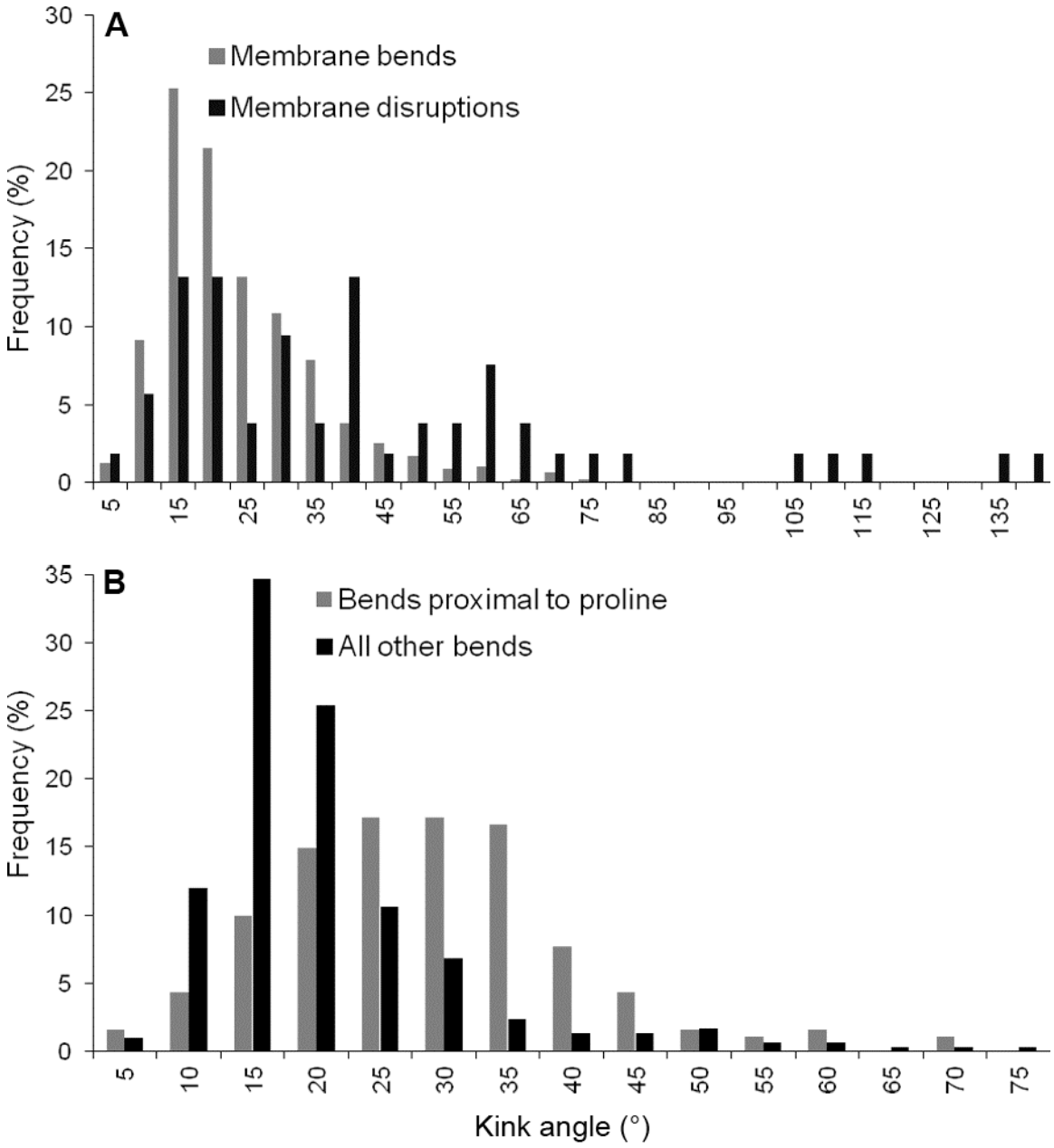

Figure 6.8: A) The frequency distribution of TM helix kink angles observed in bends (n = 471) and disruptions (n = 55) located in the center 70% by depth of the putative membrane. B) The bend angles in (A) subdivided into two distinct sets of bends with or without a proline within a range of ±4 residues.

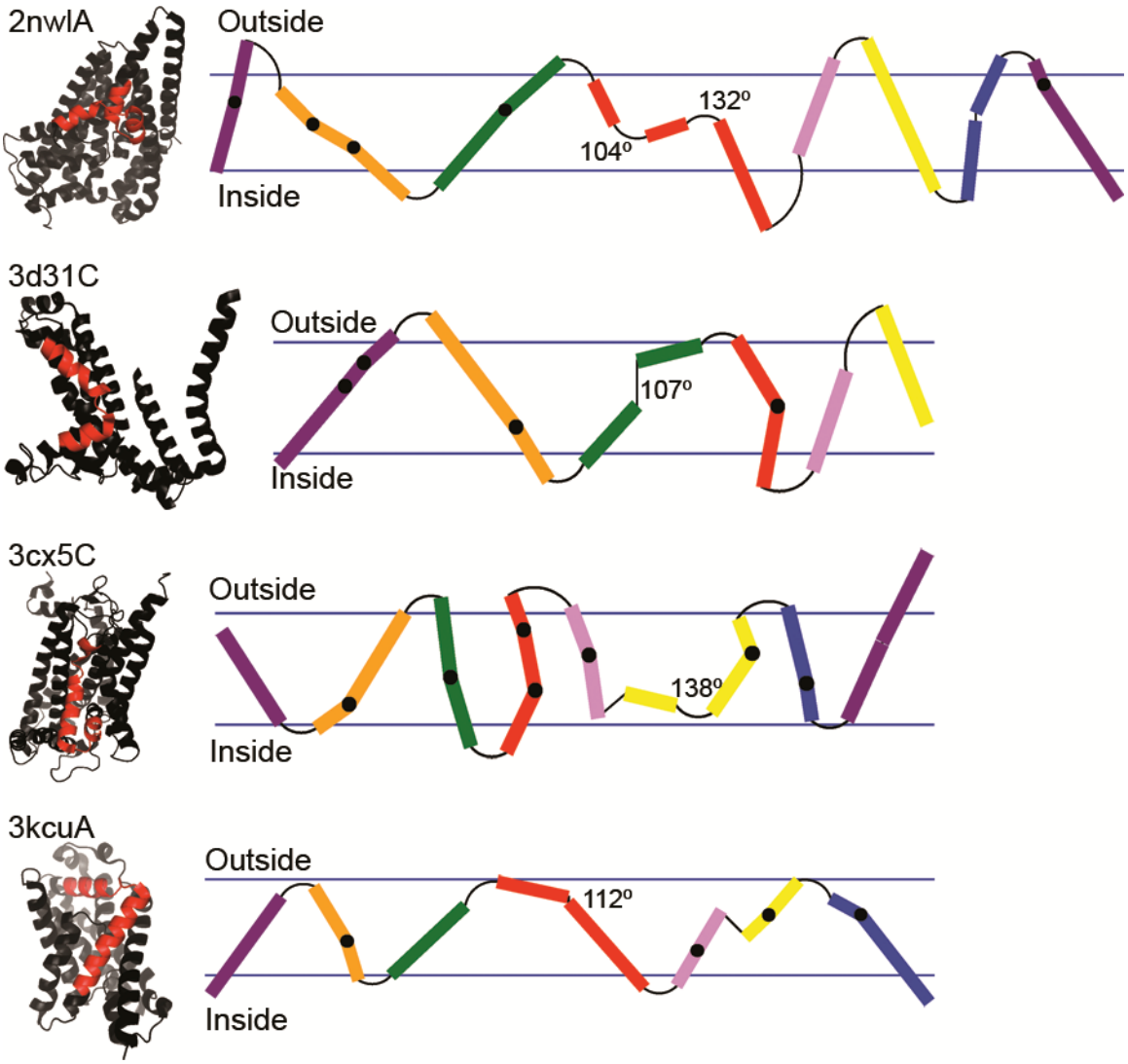

Figure 6.9: Structures and topology diagrams of all of disruptions found in the dataset with kink angle > 90º. On the left are images of single chains of the protein with the helix containing the highly-kinked disruption coloured red. On the right side of the figure are topology diagrams from MC-HELAN with the angle of the disruption indicated. Shown are the sodium-dependent aspartate transporter (2nwlA (269)), the molybdate/tungstate ABC transporter (3d31C (270)), cytochrome B (3cx5C (271)) and the formate transporter FocA (3kcuA (272)). The topology of the proteins are indicated with the cytoplasmic face of the membrane defined as inside according to the TOPDB (268) convention.

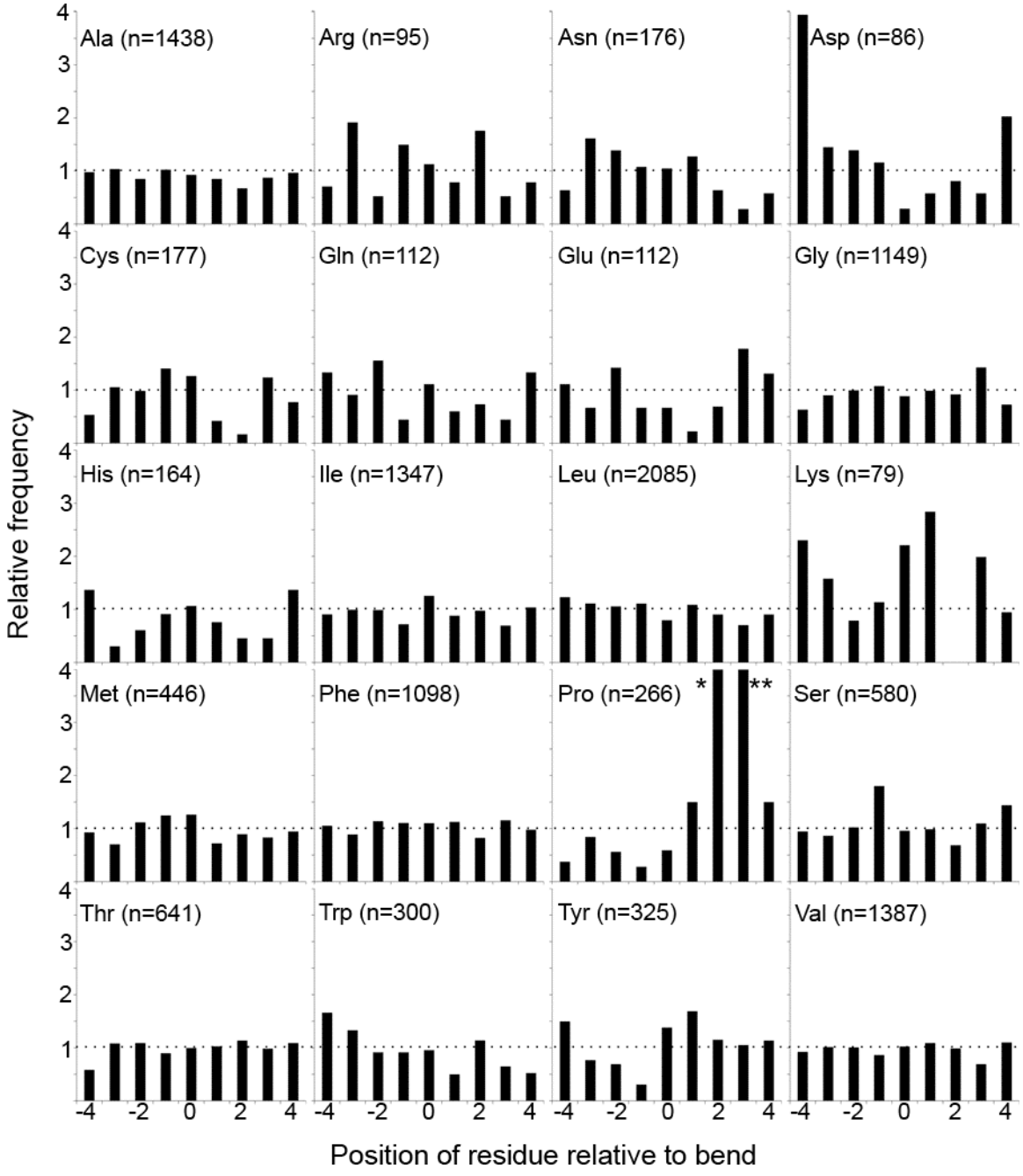

Figure 6.10: The relative frequencies of each amino acid type at or within 4 residues of a bend with respect to the frequency of the given residue in the central 70% of the putative membrane. Note that for proline the relative frequency values are beyond the scale of the plot (* =6.5 and ** = 6.7)

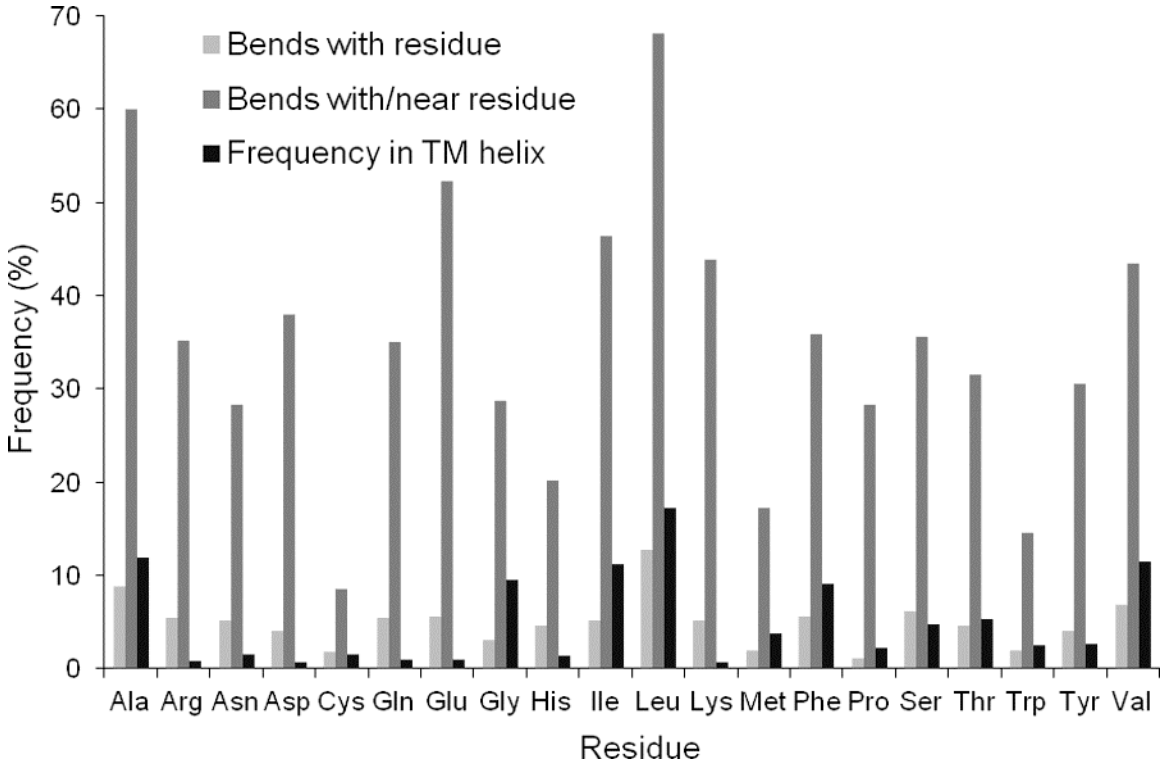

Figure 6.11: The proportion of TM helix bends (n=471) located within the central 70% by depth of the putative membrane that are found within one helical turn of each residue type. Multiple occurrences of a given residue at a given bend are only counted once and the relative frequencies of each residue in the central membrane region are shown.

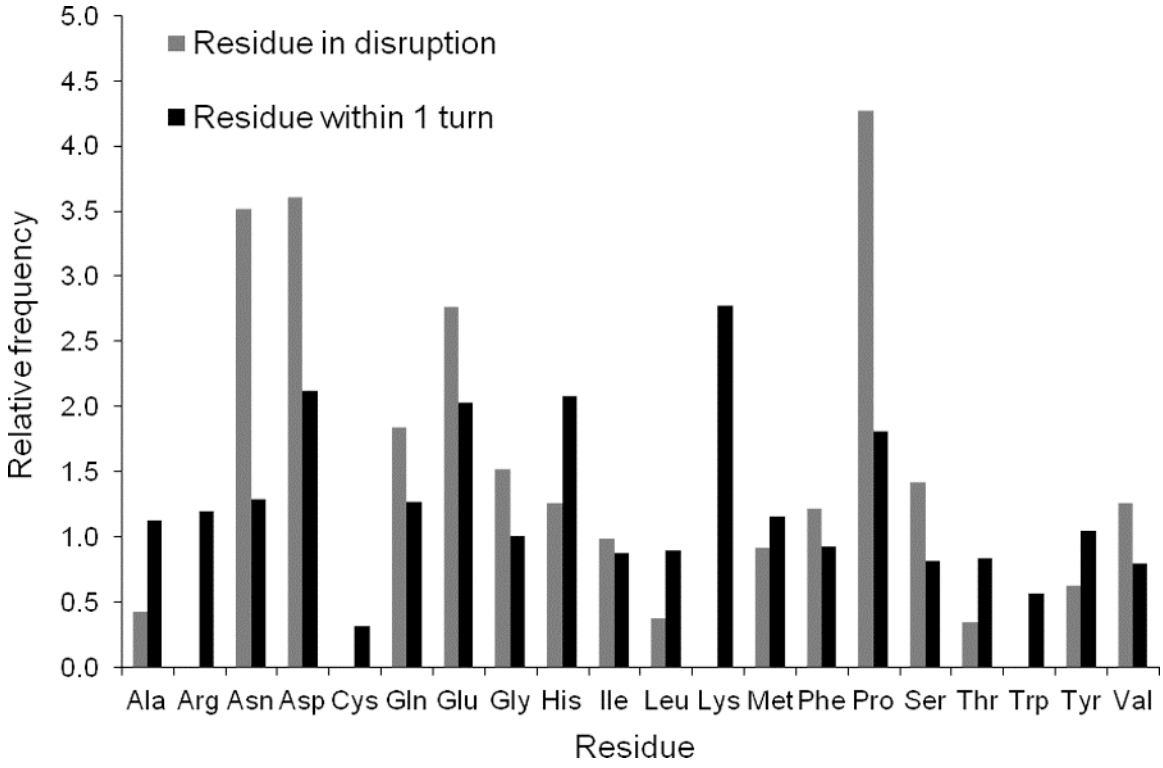

Figure 6.12: The relative frequencies of residues in or within one turn of a disruption, with respect to the frequency of the given residue within all examined TM helices. Data is shown for residues both in and within ~1 helical turn (*i.e.* 4 residues) of TM helix disruptions (n = 55) located in the central 70% of the putative membrane.

Chapter 7:

Conclusion

In this work I have studied the structure of apelin both in solution and when bound to detergent micelles. I have also studied the structure of AR55 in 4 different membrane mimetics. Through the results I present in the previous chapters and in the context of other studies, a model of apelin and AR binding can be proposed which follows the membrane catalysis hypothesis.

7.1. A Model Of Apelin-AR Interactions

The results that I have obtained concerning apelin and AR can be fit to the membrane catalysis model of ligand-receptor interactions and are shown in Figure 7.1. An interesting trend that appears is that apelin seems to have two distinct structured regions. One region occurs in the middle of apelin-17 beginning at R6 and the other in the C-terminus of the peptide including F17. Both of these regions are required for the final binding and activation of AR to occur. However, these regions of apelin could have distinct functions in the steps leading up to receptor activation.

If apelin binds to AR following the membrane catalysis model there are several relevant states of apelin and AR to consider. When free in solution the central region of apelin from residues R6-L9 is only weakly structurally converged. However, after binding to micelles residues R6-K12 of apelin become more structurally converged. This is likely a direct result of residues R6 and R8 being in contact with the membrane surface. After interactions with the membrane, apelin would then be free to bind to AR.

My work also investigates the structure of apelin when bound to AR. The presence of two structurally converged regions in apelin indicates that these two regions could be acting independently of each other to allow AR binding and activation. An interesting observation is that E20 and D23 of AR form an anionic surface and R6 and R8 of apelin create a cationic surface on apelin. Given that mutation of these residues affects the binding of apelin to AR, it is a logical suggestion that these residues are partially responsible for AR activation. With R6 and R8 of apelin being associated with the extracellular loops of AR, the hydrophobic C-terminus of apelin may then be able to bind into the TM region of AR and cause signaling. Although from my data the assertion that the C-terminal phenylalanine binds to the ligand binding site of AR is speculative, other groups have produced molecular dynamics models where this occurs (128). In addition this two step mode of a ligand binding to a GPCR has been observed previously with the CXCR4 receptor (68).

The major portion of the model shown in Figure 7.1 that is not investigated by my work is the transition of apelin from a membrane-bound to a receptor-bound state. To analyze this transition is extremely difficult and I have been unable to find any solid experimental proof of this occurring with other systems. In order to analyze this transition, I believe that a single molecule technique would be helpful since convolution of signal from free, membrane bound and receptor bound states of the ligand would be a significant problem. Although single molecule techniques do exist, they do not have the resolution necessary to monitor the transition of apelin from the membrane to the receptor. Furthermore, characterizing the structural changes that apelin undergoes during this transition would be difficult. Further questions about how apelin transitions from the

membrane to AR can be postulated. For example, does apelin transition directly from the membrane to AR or does apelin first dissociate from the membrane, and then bind to AR? Overall, this portion of the model is very difficult to characterize and can likely only be investigated through a combination of experimental techniques.

7.2. Future Work

There are several tractable questions that arise out of my work that can be the subject of future research. Although I approximate the structure of apelin while bound to AR, more direct evidence of this structure would be helpful. One way to address this would be through crosslinking studies, where a chemical is used to crosslink apelin to nearby AR residues. These crosslinking studies could lend support to the hypothesis that R6 and R8 of apelin interact with E20 and D23 of AR.

Another approach to looking at the interaction of apelin with E20 and D23 would be mutagenesis of these residues. If this initial interaction between apelin and AR is largely electrostatic, mutation to other charged residues may conserve the activity of the system. Although mutagenesis may not directly determine the binding site of apelin to AR, it can provide context for future studies of AR.

I observe conformational sampling occurring from G13-17 in apelin-17 that is due to the P14 and P16 having a *cis*-*cis* or *trans*-*trans* conformation. This immediately raises the question of which conformer of apelin is actually responsible for binding to and/or activation of AR. The simplest way to address this problem would be peptide synthesis of apelin analogues which have modified proline residues that are locked into either a *cis* or

trans conformation. Completion of these studies would also be immediately application for drug design to AR.

The main limitation of my structural characterization of AR is that I have characterized only AR55 and therefore lose the tertiary structure context of AR55 within AR. Some of this structural context has been regained through molecular modeling studies but further structural studies of AR would be beneficial. A long term goal for the structural characterization of AR could be to solve several overlapping receptor fragments and then use these to model full length AR, similar to what has been completed for rhodopsin (61).

7.3. Impact

In addition providing details to the model of apelin-AR interactions detailed above my work has had much more tractable impact. The structural data of apelin that I have provided was the first high resolution structure of either apelin or AR ever generated. Indicative of the high level of interest in apelin as a possible therapeutic, in just two years after the publication of my apelin structure there were AR agonists and antagonists generated based on my data (273, 274, 275). Even if these agonists and antagonists do not end up being used as therapeutics, they are useful research tools for studying apelin and AR.

Currently my work concerning AR55 has not yet been published, but given the high level of interest concerning GPCR structures and the apelin-AR system this work will have significant impact as well. In particular my structure of AR55 puts the functionally essential E20 and D23 into a structural context. If it is indeed true that apelin interacts with E20 and D23 as well as with the TM helices of AR, this would suggest that targeting

drugs to either of these locations could influence AR signaling. In short, in addition to contributing to our understanding of peptide-GPCR interactions, my structural results will provide further information for drug design to AR.

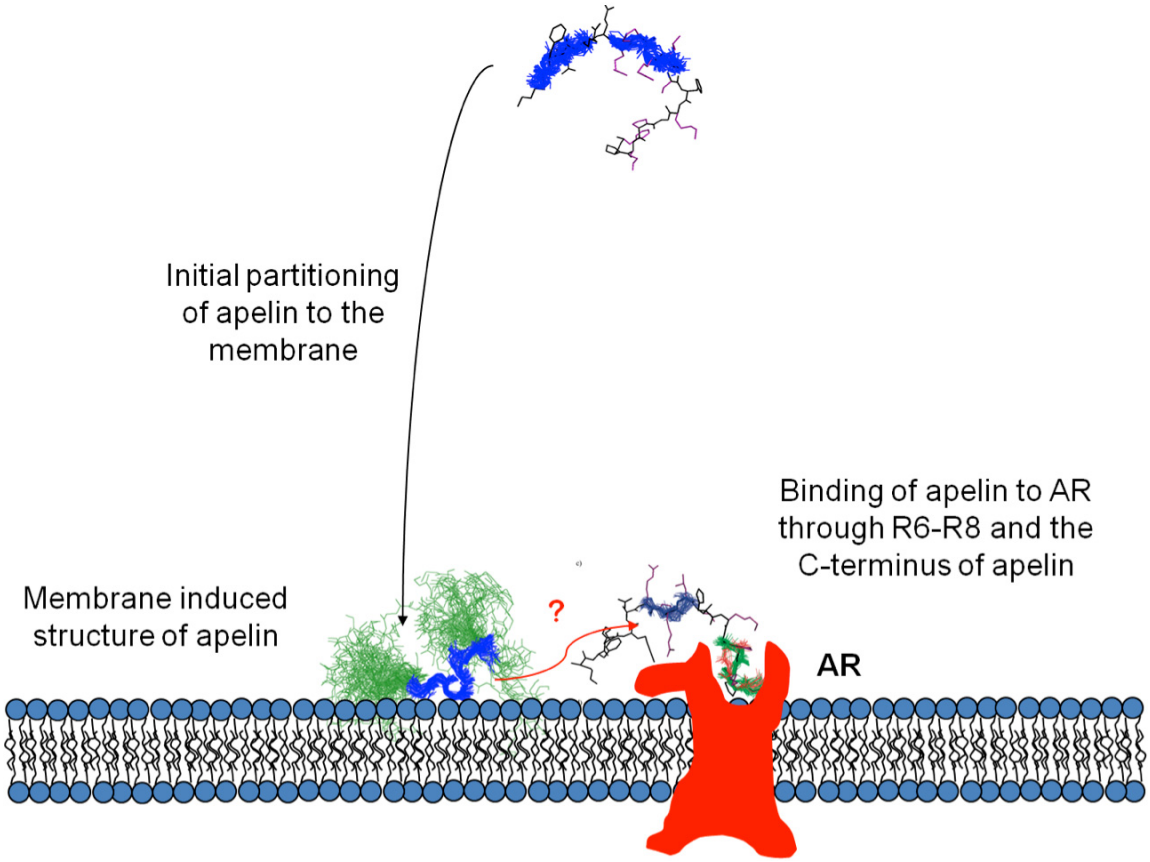

Figure 7.1: A model of apelin binding to AR. In this model apelin will first bind to the plasma membrane instead of AR. This results in increased local concentration of apelin and structural changes which may help with binding to AR. Apelin will then bind to AR through a mechanism not detailed in my work. Apelin may bind to AR through a two step process, with R6 and R8 of apelin interacting with D20 and E23 of AR followed by the C-terminal phenylalanine of apelin interacting with the TM helices of AR. When possible, I have shown the structure of apelin-17 in the various stages of this model. For clarity, I have omitted the structure of AR55 from this diagram.

Bibliography

1. Langelaan, D. N., Bebbington, E. M., Reddy, T., and Rainey, J. K. (2009) Structural Insight into G-Protein Coupled Receptor Binding by Apelin. *Biochemistry. 48*, 537-548.

2. Langelaan, D. N., and Rainey, J. K. (2009) Headgroup-Dependent Membrane Catalysis of Apelin-Receptor Interactions is Likely. *J. Phys. Chem. B. 113*, 10465-10471.

3. Langelaan, D. N., and Rainey, J. K. (2010) Membrane Catalysis of Peptide-Receptor Binding. *Biochem. Cell Biol. 88*, 203-210.

4. Langelaan, D. N., Wieczorek, M., Blouin, C., and Rainey, J. K. (2010) Improved Helix and Kink Characterization in Membrane Proteins Allows Evaluation of Kink Sequence Predictors. *J. Chem. Inf. Model. 50*, 2213-2220.

5. Vassilatis, D. K., Hohmann, J. G., Zeng, H., Li, F., Ranchalis, J. E., Mortrud, M. T., Brown, A., Rodriguez, S. S., Weller, J. R., Wright, A. C., Bergmann, J. E., and Gaitanaris, G. A. (2003) The G Protein-Coupled Receptor Repertoires of Human and Mouse. *Proc. Natl. Acad. Sci. U. S. A. 100*, 4903-4908.

6. Takeda, S., Kadowaki, S., Haga, T., Takaesu, H., and Mitaku, S. (2002) Identification of G Protein-Coupled Receptor Genes from the Human Genome Sequence. *FEBS Lett. 520*, 97-101.

7. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G., and Schioth, H. B. (2003) The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol. Pharmacol. 63*, 1256-1272.

8. Perez, D. M. (2003) The Evolutionarily Triumphant G-Protein-Coupled Receptor. *Mol. Pharmacol. 63*, 1202-1205.

9. Bleul, C. C., Farzan, M., Choe, H., Parolin, C., Clark-Lewis, I., Sodroski, J., and Springer, T. A. (1996) The Lymphocyte Chemoattractant SDF-1 is a Ligand for LESTR/fusin and Blocks HIV-1 Entry. *Nature. 382*, 829-833.

10. Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006) How Many Drug Targets are there? *Nat. Rev. Drug Discov. 5*, 993-996.

11. Kimple, A. J., Bosch, D. E., Giguere, P. M., and Siderovski, D. P. (2011) Regulators of G-Protein Signaling and their Galpha Substrates: Promises and Challenges in their use as Drug Discovery Targets. *Pharmacol. Rev. 63*, 728-749.

12. Clapham, D. E., and Neer, E. J. (1997) G Protein Beta Gamma Subunits. *Annu. Rev. Pharmacol. Toxicol. 37*, 167-203.

13. Kozasa, T., Jiang, X., Hart, M. J., Sternweis, P. M., Singer, W. D., Gilman, A. G., Bollag, G., and Sternweis, P. C. (1998) P115 RhoGEF, a GTPase Activating Protein for Galpha12 and Galpha13. *Science. 280*, 2109-2111.

14. Simonds, W. F. (1999) G Protein Regulation of Adenylate Cyclase. *Trends Pharmacol. Sci. 20*, 66-73.

15. Kammermeier, P. J., Ruiz-Velasco, V., and Ikeda, S. R. (2000) A Voltage-Independent Calcium Current Inhibitory Pathway Activated by Muscarinic Agonists in Rat Sympathetic Neurons Requires both $G_{\alpha q/11}$ and $G_{\beta\gamma}$. *J. Neurosci. 20*, 5623-5629.

16. Rhee, S. G. (2001) Regulation of Phosphoinositide-Specific Phospholipase C. *Annu. Rev. Biochem. 70*, 281-312.

17. Lutz, S., Shankaranarayanan, A., Coco, C., Ridilla, M., Nance, M. R., Vettel, C., Baltus, D., Evelyn, C. R., Neubig, R. R., Wieland, T., and Tesmer, J. J. (2007) Structure of Galphaq-p63RhoGEF-RhoA Complex Reveals a Pathway for the Activation of RhoA by GPCRs. *Science. 318*, 1923-1927.

18. Brzostowski, J. A., and Kimmel, A. R. (2001) Signaling at Zero G: G-Protein-Independent Functions for 7-TM Receptors. *Trends Biochem. Sci. 26*, 291-297.

19. Ma, L., and Pei, G. (2007) B-Arrestin Signaling and Regulation of Transcription. *Journal of Cell Science. 120*, 213-218.

20. Terrillon, S., and Bouvier, M. (2004) Roles of G-Protein-Coupled Receptor Dimerization. *EMBO Rep. 5*, 30-34.

21. Galvez, T., Duthey, B., Kniazeff, J., Blahos, J., Rovelli, G., Bettler, B., Prezeau, L., and Pin, J. (2001) Allosteric Interactions between GB1 and GB2 Subunits are Required for Optimal GABAB Receptor Function. *EMBO J. 20*, 2152-2159.

22. Nelson, G., Hoon, M. A., Chandrashekar, J., Zhang, Y., Ryba, N. J. P., and Zuker, C. S. (2001) Mammalian Sweet Taste Receptors. *Cell. 106*, 381-390.

23. Sun, X., Iida, S., Yoshikawa, A., Senbonmatsu, R., Imanaka, K., Maruyama, K., Nishimura, S., Inagami, T., and Senbonmatsu, T. (2011) Non-Activated APJ Suppresses the Angiotensin II Type 1 Receptor, Whereas Apelin-Activated APJ Acts Conversely. *Hypertens. Res. 34*, 701-706.

24. Violin, J. D., and Lefkowitz, R. J. (2007) B-Arrestin-Biased Ligands at Seven-Transmembrane Receptors. *Trends Pharmacol. Sci. 28*, 416-422.

25. O'Dowd, B. F., Heiber, M., Chan, A., Heng, H. H., Tsui, L. C., Kennedy, J. L., Shi, X., Petronis, A., George, S. R., and Nguyen, T. (1993) A Human Gene that shows Identity with the Gene Encoding the Angiotensin Receptor is Located on Chromosome 11. *Gene. 136*, 355-360.

26. Tatemoto, K., Hosoya, M., Habata, Y., Fujii, R., Kakegawa, T., Zou, M. X., Kawamata, Y., Fukusumi, S., Hinuma, S., Kitada, C., Kurokawa, T., Onda, H., and Fujino, M. (1998) Isolation and Characterization of a Novel Endogenous Peptide Ligand for the Human APJ Receptor. *Biochem. Biophys. Res. Commun. 251*, 471-476.

27. Alexandrov, A. I., Mileni, M., Chien, E. Y., Hanson, M. A., and Stevens, R. C. (2008) Microscale Fluorescent Thermal Stability Assay for Membrane Proteins. *Structure. 16*, 351-359.

28. Masri, B., Knibiehler, B., and Audigier, Y. (2005) Apelin Signalling: A Promising Pathway from Cloning to Pharmacology. *Cell. Signal. 17*, 415-426.

29. Wang, G., Anini, Y., Wei, W., Qi, X., O'Carroll, A., Mochizuki, T., Wang, H., Hellmich, M. R., Englander, E. W., and Greeley, G. H. (2004) Apelin, a New Enteric Peptide: Localization in the Gastrointestinal Tract, Ontogeny, and Stimulation of Gastric Cell Proliferation and of Cholecystokinin Secretion. *Endocrinology. 145*, 1342-1348.

30. Valle, A., Hoggard, N., Adams, A. C., Roca, P., and Speakman, J. R. (2008) Chronic Central Administration of Apelin-13 Over 10 Days Increases Food Intake, Body Weight, Locomotor Activity and Body Temperature in C57BL/6 Mice. *J. Neuroendocrinol. 20*, 79-84.

31. Vickers, C., Hales, P., Kaushik, V., Dick, L., Gavin, J., Tang, J., Godbout, K., Parsons, T., Baronas, E., Hsieh, F., Acton, S., Patane, M., Nichols, A., and Tummino, P. (2002) Hydrolysis of Biological Peptides by Human Angiotensin-Converting Enzyme-Related Carboxypeptidase. *J. Biol. Chem. 277*, 14838-14843.

32. Fan, X., Zhou, N., Zhang, X., Mukhtar, M., Lu, Z., Fang, J., DuBois, G. C., and Pomerantz, R. J. (2003) Structural and Functional Study of the Apelin-13 Peptide, an Endogenous Ligand of the HIV-1 Coreceptor, APJ. *Biochemistry. 42*, 10163-10168.

33. Medhurst, A. D., Jennings, C. A., Robbins, M. J., Davis, R. P., Ellis, C., Winborn, K. Y., Lawrie, K. W., Hervieu, G., Riley, G., Bolaky, J. E., Herrity, N. C., Murdock, P., and Darker, J. G. (2003) Pharmacological and Immunohistochemical Characterization of the APJ Receptor and its Endogenous Ligand Apelin. *J. Neurochem. 84*, 1162-1172.

34. Lee, D. K., Saldivia, V. R., Nguyen, T., Cheng, R., George, S. R., and O'Dowd, B. F. (2005) Modification of the Terminal Residue of Apelin-13 Antagonizes its Hypotensive Action. *Endocrinology. 146*, 231-236.

35. Bai, B., Tang, J., Liu, H., Chen, J., Li, Y., and Song, W. (2008) Apelin-13 Induces ERK1/2 but Not p38 MAPK Activation through Coupling of the Human Apelin Receptor to the Gi2 Pathway. *Acta Biochim. Biophys. Sin. (Shanghai). 40*, 311-318.

36. Falcao-Pires, I., Ladeiras-Lopes, R., and Leite-Moreira, A. F. (2010) The Apelinergic System: A Promising Therapeutic Target. *Expert Opin. Ther. Targets. 14*, 633-645.

37. Masri, B., Morin, N., Pedebernade, L., Knibiehler, B., and Audigier, Y. (2006) The Apelin Receptor is Coupled to Gi1 Or Gi2 Protein and is Differentially Desensitized by Apelin Fragments. *J. Biol. Chem. 281*, 18317-18326.

38. Li, Y., Chen, J., Bai, B., Du, H., Liu, Y., and Liu, H. (2012) Heterodimerization of Human Apelin and Kappa Opioid Receptors: Roles in Signal Transduction. *Cell. Signal. 24*, 991-1001.

39. Lee, D. K., Ferguson, S. S., George, S. R., and O'Dowd, B. F. (2010) The Fate of the Internalized Apelin Receptor is Determined by Different Isoforms of Apelin Mediating Differential Interaction with Beta-Arrestin. *Biochem. Biophys. Res. Commun. 395*, 185-189.

40. Szokodi, I., Tavi, P., Foldes, G., Voutilainen-Myllyla, S., Ilves, M., Tokola, H., Pikkarainen, S., Piuhola, J., Rysa, J., Toth, M., and Ruskoaho, H. (2002) Apelin, the Novel Endogenous Ligand of the Orphan Receptor APJ, Regulates Cardiac Contractility. *Circ. Res. 91*, 434-440.

41. Chandrasekaran, B., Dar, O., and McDonagh, T. (2008) The Role of Apelin in Cardiovascular Function and Heart Failure. *Eur. J. Heart Fail. 10*, 725-732.

42. Boucher, J., Masri, B., Daviaud, D., Gesta, S., Guigne, C., Mazzucotelli, A., Castan-Laurell, I., Tack, I., Knibiehler, B., Carpene, C., Audigier, Y., Saulnier-Blache, J. S., and Valet, P. (2005) Apelin, a Newly Identified Adipokine Up-Regulated by Insulin and Obesity. *Endocrinology. 146*, 1764-1771.

43. Heinonen, M. V., Purhonen, A. K., Miettinen, P., Paakkonen, M., Pirinen, E., Alhava, E., Akerman, K., and Herzig, K. H. (2005) Apelin, Orexin-A and Leptin Plasma Levels in Morbid Obesity and Effect of Gastric Banding. *Regul. Pept. 130*, 7-13.

44. Kleinz, M. J., and Davenport, A. P. (2005) Emerging Roles of Apelin in Biology and Medicine. *Pharmacol. Ther. 107*, 198-211.

45. Sorli, S. C., Le Gonidec, S., Knibiehler, B., and Audigier, Y. (2007) Apelin is a Potent Activator of Tumour Neoangiogenesis. *Oncogene. 26*, 7692-7699.

46. Kalin, R. E., Kretz, M. P., Meyer, A. M., Kispert, A., Heppner, F. L., and Brandli, A. W. (2007) Paracrine and Autocrine Mechanisms of Apelin Signaling Govern Embryonic and Tumor Angiogenesis. *Dev. Biol. 305*, 599-614.

47. Hashimoto, T., Kihara, M., Imai, N., Yoshida, S., Shimoyamada, H., Yasuzaki, H., Ishida, J., Toya, Y., Kiuchi, Y., Hirawa, N., Tamura, K., Yazawa, T., Kitamura, H., Fukamizu, A., and Umemura, S. (2007) Requirement of Apelin-Apelin Receptor System for Oxidative Stress-Linked Atherosclerosis. *Am. J. Pathol. 171*, 1705-1712.

48. Sorli, S. C., van den Berghe, L., Masri, B., Knibiehler, B., and Audigier, Y. (2006) Therapeutic Potential of Interfering with Apelin Signalling. *Drug Discov. Today. 11*, 1100-1106.

49. Principe, A., Melgar-Lesmes, P., Fernandez-Varo, G., del Arbol, L. R., Ros, J., Morales-Ruiz, M., Bernardi, M., Arroyo, V., and Jimenez, W. (2008) The Hepatic Apelin System: A New Therapeutic Target for Liver Disease. *Hepatology. 48*, 1193-1201.

50. Torres, J., Stevens, T. J., and Samso, M. (2003) Membrane Proteins: The 'Wild West' of Structural Biology. *Trends Biochem. Sci. 28*, 137-144.

51. Chien, E. Y., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V., and Stevens, R. C. (2010) Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. *Science. 330*, 1091-1095.

52. Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P., and Stevens, R. C. (2008) The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science. 322*, 1211-1217.

53. Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G., and Schertler, G. F. (2008) Structure of a beta1-Adrenergic G-Protein-Coupled Receptor. *Nature. 454*, 486-491.

54. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science. 289*, 739-745.

55. Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C., and Iwata, S. (2011) Structure of the Human Histamine H1 Receptor Complex with Doxepin. *Nature. 475*, 65-70.

56. Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I., and Kobilka, B. K. (2007) Crystal Structure of the Human beta2 Adrenergic G-Protein-Coupled Receptor. *Nature. 450*, 383-387.

57. Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V., and Stevens, R. C. (2010) Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science. 330*, 1066-1071.

58. Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T., and Kobayashi, T. (2012) Structure of the Human M2 Muscarinic Acetylcholine Receptor Bound to an Antagonist. *Nature. 482*, 547-551.

59. Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Yao, X. J., Weis, W. I., Stevens, R. C., and Kobilka, B. K. (2007) GPCR Engineering Yields High-Resolution Structural Insights into beta2-Adrenergic Receptor Function. *Science. 318*, 1266-1273.

60. Tikhonova, I. G., and Costanzi, S. (2009) Unraveling the Structure and Function of G Protein-Coupled Receptors through NMR Spectroscopy. *Curr. Pharm. Des. 15*, 4003-4016.

61. Katragadda, M., Alderfer, J. L., and Yeagle, P. L. (2001) Assembly of a Polytopic Membrane Protein Structure from the Solution Structures of Overlapping Peptide Fragments of Bacteriorhodopsin. *Biophys. J. 81*, 1029-1036.

62. Li, J., Edwards, P. C., Burghammer, M., Villa, C., and Schertler, G. F. (2004) Structure of Bovine Rhodopsin in a Trigonal Crystal Form. *J. Mol. Biol. 343*, 1409-1438.

63. Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P., and Nietlispach, D. (2010) Structure Determination of the Seven-Helix Transmembrane Receptor Sensory Rhodopsin II by Solution NMR Spectroscopy. *Nat. Struct. Mol. Biol. 17*, 768-774.

64. Park, S. H., Prytulla, S., De Angelis, A. A., Brown, J. M., Kiefer, H., and Opella, S. J. (2006) High-Resolution NMR Spectroscopy of a GPCR in Aligned Bicelles. *J. Am. Chem. Soc. 128*, 7402-7403.

65. Tian, C., Breyer, R. M., Kim, H. J., Karra, M. D., Friedman, D. B., Karpay, A., and Sanders, C. R. (2005) Solution NMR Spectroscopy of the Human Vasopressin V2 Receptor, a G Protein-Coupled Receptor. *J. Am. Chem. Soc. 127*, 8010-8011.

66. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K., and Stevens, R. C. (2007) High-Resolution Crystal Structure of an Engineered Human beta2-Adrenergic G Protein-Coupled Receptor. *Science. 318*, 1258-1265.

67. Clark-Lewis, I., Kim, K. S., Rajarathnam, K., Gong, J. H., Dewald, B., Moser, B., Baggiolini, M., and Sykes, B. D. (1995) Structure-Activity Relationships of Chemokines. *J. Leukoc. Biol. 57*, 703-711.

68. Kofuku, Y., Yoshiura, C., Ueda, T., Terasawa, H., Hirai, T., Tominaga, S., Hirose, M., Maeda, Y., Takahashi, H., Terashima, Y., Matsushima, K., and Shimada, I. (2009) Structural Basis of the Interaction between Chemokine Stromal Cell-Derived Factor-1/CXCL12 and its G-Protein-Coupled Receptor CXCR4. *J. Biol. Chem. 284*, 35240-35250.

69. Hoare, S. R. (2005) Mechanisms of Peptide and Nonpeptide Ligand Binding to Class B G-Protein-Coupled Receptors. *Drug Discov. Today. 10*, 417-427.

70. Parthier, C., Kleinschmidt, M., Neumann, P., Rudolph, R., Manhart, S., Schlenzig, D., Fanghanel, J., Rahfeld, J. U., Demuth, H. U., and Stubbs, M. T. (2007) Crystal Structure of the Incretin-Bound Extracellular Domain of a G Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. U. S. A. 104*, 13942-13947.

71. Grace, C. R., Perrin, M. H., DiGruccio, M. R., Miller, C. L., Rivier, J. E., Vale, W. W., and Riek, R. (2004) NMR Structure and Peptide Hormone Binding Site of the First Extracellular Domain of a Type B1 G Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. U. S. A. 101*, 12836-12841.

72. Rosenbaum, D. M., Rasmussen, S. G., and Kobilka, B. K. (2009) The Structure and Function of G-Protein-Coupled Receptors. *Nature. 459*, 356-363.

73. Rasmussen, S. G., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., Devree, B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S., Schnapp, A., Konetzki, I., Sunahara, R. K., Gellman, S. H., Pautsch, A., Steyaert, J., Weis, W. I., and Kobilka, B. K. (2011) Structure of a Nanobody-Stabilized Active State of the Beta(2) Adrenoceptor. *Nature. 469*, 175-180.

74. Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., and Kobilka, B. K. (2011) Crystal Structure of the beta2 Adrenergic Receptor-Gs Protein Complex. *Nature. 477*, 549-555.

75. Greenfield, N. J. (2004) Analysis of Circular Dichroism Data. *Methods Enzymol. 383*, 282-317.

76. Sreerama, N., and Woody, R. W. (2004) Computation and Analysis of Protein Circular Dichroism Spectra. *Methods Enzymol. 383*, 318-351.

77. Holzwarth, G., and Doty, P. (1965) The Ultraviolet Circular Dichroism of Polypeptides. *J. Am. Chem. Soc. 87*, 218-228.

78. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an Online Server for Protein Secondary Structure Analyses from Circular Dichroism Spectroscopic Data. *Nucleic Acids Research. 32*, W668-W673.

79. Levitt, M. H. (2008) *Spin Dynamics: Basics of Nuclear Magnetic Resonance.* John Wiley & Sons, Chichester, England; Hoboken, NJ.

80. Keeler, J. (2010) *Understanding NMR Spectroscopy.* John Wiley and Sons, Chichester, U.K.

81. Chary, K. V. R., and Govil, G. (2008) *NMR in Biological Systems from Molecules to Humans.* First ed., Springer, Dordrecht, the Netherlands.

82. Cavanagh, J.,. (2007) *Protein NMR Spectroscopy: Principles and Practice.* Academic Press, Amsterdam; Boston.

83. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *J. Magn. Reson. 89*, 496-514.

84. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (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*, 67-81.

85. Griesinger, C., Otting, G., Wuethrich, K., and Ernst, R. R. (1988) Clean TOCSY for Proton Spin System Identification in Macromolecules. *J. Am. Chem. Soc. 110*, 7870-7872.

86. Clubb, R. T., Thanabal, V., and Wagner, G. (1992) A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue 1HN, 15N, and 13C′ Chemical Shifts in 15N 13C-Labelled Proteins. *Journal of Magnetic Resonance (1969). 97*, 213-217.

87. Bax, A., and Ikura, M. (1991) An Efficient 3D NMR Technique for Correlating the Proton and 15N Backbone Amide Resonances with the Alpha-Carbon of the Preceding Residue in Uniformly 15N/13C Enriched Proteins. *J. Biomol. NMR. 1*, 99-104.

88. Grzesiek, S., and Bax, A. (1992) Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc. 114*, 6291-6293.

89. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR Molecular Structure Determination Package. *J. Magn. Reson. 160*, 65-73.

90. Schwieters, C. D., Kuszewski, J. J., and Marius Clore, G. (2006) Using Xplor–NIH for NMR Molecular Structure Determination. *Prog. Nucl. Magn. Reson. Spectrosc. 48*, 47-62.

91. Guntert, P. (2004) Automated NMR Structure Calculation with CYANA. *Methods Mol. Biol. 278*, 353-378.

92. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D Biol. Crystallogr. 54*, 905-921.

93. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983) Optimization by Simulated Annealing. *Science. 220*, pp. 671-680.

94. Aarts, E. H. L., and van Laarhoven, P. J. M. (1989) Simulated Annealing: An Introduction. *Statistica Neerlandica. 43*, 31-52.

95. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins. 59*, 687-696.

96. Zhang, H., Neal, S., and Wishart, D. S. (2003) RefDB: A Database of Uniformly Referenced Protein Chemical Shifts. *J. Biomol. NMR. 25*, 173-195.

97. Liu, G., Forouhar, F., Eletsky, A., Atreya, H. S., Aramini, J. M., Xiao, R., Huang, Y. J., Abashidze, M., Seetharaman, J., Liu, J., Rost, B., Acton, T., Montelione, G. T., Hunt, J. F., and Szyperski, T. (2009) NMR and X-RAY Structures of Human E2-Like Ubiquitin-Fold Modifier Conjugating Enzyme 1 (UFC1) Reveal Structural and Functional Conservation in the Metazoan UFM1-UBA5-UFC1 Ubiquination Pathway. *J. Struct. Funct. Genomics. 10*, 127-136.

98. Tatemoto, K., Takayama, K., Zou, M. X., Kumaki, I., Zhang, W., Kumano, K., and Fujimiya, M. (2001) The Novel Peptide Apelin Lowers Blood Pressure Via a Nitric Oxide-Dependent Mechanism. *Regul. Pept. 99*, 87-92.

99. De Mota, N., Lenkei, Z., and Llorens-Cortes, C. (2000) Cloning, Pharmacological Characterization and Brain Distribution of the Rat Apelin Receptor. *Neuroendocrinology. 72*, 400-407.

100. Zhou, N., Zhang, X., Fan, X., Argyris, E., Fang, J., Acheampong, E., DuBois, G. C., and Pomerantz, R. J. (2003) The N-Terminal Domain of APJ, a CNS-Based Coreceptor for HIV-1, is Essential for its Receptor Function and Coreceptor Activity. *Virology. 317*, 84-94.

101. Tyndall, J. D., Pfeiffer, B., Abbenante, G., and Fairlie, D. P. (2005) Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chem. Rev. 105*, 793-826.

102. Rath, A., Davidson, A. R., and Deber, C. M. (2005) The Structure of "Unstructured" Regions in Peptides and Proteins: Role of the Polyproline II Helix in Protein Folding and Recognition. *Biopolymers. 80*, 179-185.

103. Siligardi, G., and Drake, A. F. (1995) The Importance of Extended Conformations and, in Particular, the PII Conformation for the Molecular Recognition of Peptides. *Biopolymers. 37*, 281-292.

104. Rainey, J. K., Fliegel, L., and Sykes, B. D. (2006) Strategies for Dealing with Conformational Sampling in Structural Calculations of Flexible Or Kinked Transmembrane Peptides. *Biochem. Cell Biol. 84*, 918-929.

105. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of Peptide Fragments of Proteins in Aqueous Solution: Implications for Initiation of Protein Folding. *Biochemistry. 27*, 7167-7175.

106. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy. *Biochemistry. 31*, 1647-1651.

107. Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., and de Groot, B. L. (2008) Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science. 320*, 1471-1475.

108. Booth, V., Slupsky, C. M., Clark-Lewis, I., and Sykes, B. D. (2003) Unmasking Ligand Binding Motifs: Identification of a Chemokine Receptor Motif by NMR Studies of Antagonist Peptides. *J. Mol. Biol. 327*, 329-334.

109. Brown, J. E., and Klee, W. A. (1971) Helix-Coil Transition of the Isolated Amino Terminus of Ribonuclease. *Biochemistry. 10*, 470-476.

110. Dougherty, R. C. (1998) Temperature and Pressure Dependence of Hydrogen Bond Strength: A Perturbation Molecular Orbital Approach. *J. Chem. Phys. 109*, 7372-7378.

111. Slupsky, C. M., Sykes, D. B., Gay, G. L., and Sykes, B. D. (2001) The HoxB1 Hexapeptide is a Prefolded Domain: Implications for the Pbx1/Hox Interaction. *Protein Sci. 10*, 1244-1253.

112. Bray, B. L. (2003) Large-Scale Manufacture of Peptide Therapeutics by Chemical Synthesis. *Nat. Rev. Drug Discov. 2*, 587-593.

113. Benoiton, N. L. (2006) *Chemistry of Peptide Synthesis.* Taylor and Francis, Boca Raton, Fla.

114. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR. 6*, 277-293.

115. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR. 8*, 477-486.

116. Best, R. B., Buchete, N. V., and Hummer, G. (2008) Are Current Molecular Dynamics Force Fields Too Helical? *Biophys. J. 95*, L07-9.

117. Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF--a Program to Identify and Analyze Structural Motifs in Proteins. *Protein Sci. 5*, 212-220.

118. Collaborative Computational Project, Number 4. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D Biol. Crystallogr. 50*, 760-763.

119. Diamond, R. (1995) Coordinate-Based Cluster Analysis. *Acta Crystallogr. D Biol. Crystallogr. 51*, 127-135.

120. Maddison, D. R., Swofford, D. L., and Maddison, W. P. (1997) NEXUS: An Extensible File Format for Systematic Information. *Syst. Biol. 46*, 590-621.

121. Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., and Markley, J. L. (2008) BioMagResBank. *Nucleic Acids Res. 36*, D402-8.

122. Shen, Y., and Bax, A. (2010) Prediction of Xaa-Pro Peptide Bond Conformation from Sequence and Chemical Shifts. *J. Biomol. NMR. 46*, 199-204.

123. Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002) A Software Tool for the Prediction of Xaa-Pro Peptide Bond Conformations in Proteins Based on 13C Chemical Shift Statistics. *J. Biomol. NMR. 24*, 149-154.

124. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids.* Wiley, New York.

125. Hutchinson, E. G., and Thornton, J. M. (1994) A Revised Set of Potentials for Beta-Turn Formation in Proteins. *Protein Sci. 3*, 2207-2216.

126. Reddy, T., Ding, J., Li, X., Sykes, B. D., Rainey, J. K., and Fliegel, L. (2008) Structural and Functional Characterization of Transmembrane Segment IX of the NHE1 Isoform of the Na+/H+ Exchanger. *J. Biol. Chem. 283*, 22018-22030.

127. Vila, J. A., Baldoni, H. A., Ripoll, D. R., Ghosh, A., and Scheraga, H. A. (2004) Polyproline II Helix Conformation in a Proline-Rich Environment: A Theoretical Study. *Biophys. J. 86*, 731-742.

128. Iturrioz, X., Gerbier, R., Leroux, V., Alvear-Perez, R., Maigret, B., and Llorens-Cortes, C. (2010) By Interacting with the C-Terminal Phe of Apelin, Phe255 and Trp259 in Helix VI of the Apelin Receptor are Critical for Internalization. *J. Biol. Chem. 285*, 32627-32637.

129. Ding, J., Rainey, J. K., Xu, C., Sykes, B. D., and Fliegel, L. (2006) Structural and Functional Characterization of Transmembrane Segment VII of the Na+/H+ Exchanger Isoform 1. *J. Biol. Chem. 281*, 29817-29829.

130. Wishart, D. S., and Sykes, B. D. (1994) The 13C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure using 13C Chemical-Shift Data. *J. Biomol. NMR. 4*, 171-180.

131. Sargent, D. F., and Schwyzer, R. (1986) Membrane Lipid Phase as Catalyst for Peptide-Receptor Interactions. *Proc. Natl. Acad. Sci. U. S. A. 83*, 5774-5778.

132. Rich, A., Davidson, N. R., and Pauling, L.,. (1968) *Structural Chemistry and Molecular Biology.* W. H. Freeman, San Francisco.

133. Mavromoustakos, T., Theodoropoulou, E., Dimitriou, C., Matsoukas, J. M., Panagiotopoulos, D., and Makriyannis, A. (1996) Interactions of Angiotensin II with Membranes using a Combination of Differential Scanning Calorimetry and 31P NMR Spectroscopy. *Lett. Peptide Sci. 3*, 175-180.

134. Deber, C. M., and Behnam, B. A. (1984) Role of Membrane Lipids in Peptide Hormone Function: Binding of Enkephalins to Micelles. *Proc. Natl. Acad. Sci. U. S. A. 81*, 61-65.

135. Lopes, S. C., Fedorov, A., and Castanho, M. A. (2005) Lipidic Membranes are Potential "Catalysts" in the Ligand Activity of the Multifunctional Pentapeptide Neokyotorphin. *Chembiochem. 6*, 697-702.

136. Lopes, S. C., Fedorov, A., and Castanho, M. A. (2004) Cholesterol Modulates Maculosin's Orientation in Model Systems of Biological Membranes. Relevance Towards Putative Molecular Recognition. *Steroids. 69*, 825-830.

137. Young, J. K., Graham, W. H., Beard, D. J., and Hicks, R. P. (1992) The use of UV-Visible Spectroscopy for the Determination of Hydrophobic Interactions between Neuropeptides and Membrane Model Systems. *Biopolymers. 32*, 1061-1064.

138. Motta, A., Andreotti, G., Amodeo, P., Strazzullo, G., and Castiglione Morelli, M. A. (1998) Solution Structure of Human Calcitonin in Membrane-Mimetic Environment: The Role of the Amphipathic Helix. *Proteins. 32*, 314-323.

139. Contreras, L. M., de Almeida, R. F., Villalain, J., Fedorov, A., and Prieto, M. (2001) Interaction of Alpha-Melanocyte Stimulating Hormone with Binary Phospholipid Membranes: Structural Changes and Relevance of Phase Behavior. *Biophys. J. 80*, 2273-2283.

140. Backlund, B. M., Wikander, G., Peeters, T. L., and Graslund, A. (1994) Induction of Secondary Structure in the Peptide Hormone Motilin by Interaction with Phospholipid Vesicles. *Biochim. Biophys. Acta. 1190*, 337-344.

141. Warschawski, D. E., Arnold, A. A., Beaugrand, M., Gravel, A., Chartrand, É., and Marcotte, I. (2011) Choosing Membrane Mimetics for NMR Structural Studies of Transmembrane Proteins. *Biochim. Biophys. Acta. 1808*, 1957-1974.

142. Nyholm, T. K., Ozdirekcan, S., and Killian, J. A. (2007) How Protein Transmembrane Segments Sense the Lipid Environment. *Biochemistry. 46*, 1457-1465.

143. White, S. H., and Wimley, W. C. (1999) Membrane Protein Folding and Stability: Physical Principles. *Annu. Rev. Biophys. Biomol. Struct. 28*, 319-365.

144. Arshava, B., Taran, I., Xie, H., Becker, J. M., and Naider, F. (2002) High Resolution NMR Analysis of the Seven Transmembrane Domains of a Heptahelical Receptor in Organic-Aqueous Medium. *Biopolymers. 64*, 161-176.

145. Cohen, L. S., Arshava, B., Estephan, R., Englander, J., Kim, H., Hauser, M., Zerbe, O., Ceruso, M., Becker, J. M., and Naider, F. (2008) Expression and Biophysical Analysis of Two Double-Transmembrane Domain-Containing Fragments from a Yeast G Protein-Coupled Receptor. *Biopolymers. 90*, 117-130.

146. Cohen, L. S., Arshava, B., Neumoin, A., Becker, J. M., Guntert, P., Zerbe, O., and Naider, F. (2011) Comparative NMR Analysis of an 80-Residue G Protein-Coupled Receptor Fragment in Two Membrane Mimetic Environments. *Biochim. Biophys. Acta. 1808*, 2674-2684.

147. Estephan, R., Englander, J., Arshava, B., Samples, K. L., Becker, J. M., and Naider, F. (2005) Biosynthesis and NMR Analysis of a 73-Residue Domain of a Saccharomyces Cerevisiae G Protein-Coupled Receptor. *Biochemistry. 44*, 11795-11810.

148. Katragadda, M., Chopra, A., Bennett, M., Alderfer, J. L., Yeagle, P. L., and Albert, A. D. (2001) Structures of the Transmembrane Helices of the G-Protein Coupled Receptor, Rhodopsin. *J. Pept. Res. 58*, 79-89.

149. Lee, B. L., Sykes, B. D., and Fliegel, L. (2011) Structural Analysis of the Na+/H+ Exchanger Isoform 1 (NHE1) using the Divide and Conquer Approach. *Biochem. Cell Biol. 89*, 189-199.

150. Popot, J. L., Gerchman, S. E., and Engelman, D. M. (1987) Refolding of Bacteriorhodopsin in Lipid Bilayers. A Thermodynamically Controlled Two-Stage Process. *J. Mol. Biol. 198*, 655-676.

151. Shenkarev, Z. O., Balashova, T. A., Efremov, R. G., Yakimenko, Z. A., Ovchinnikova, T. V., Raap, J., and Arseniev, A. S. (2002) Spatial Structure of Zervamicin IIB Bound to DPC Micelles: Implications for Voltage-Gating. *Biophys. J. 82*, 762-771.

152. Xiao, S., Yang, L., and Li, F. (2012) The Structure and Assembly Model of the Third Transmembrane Domain of Slc11a1 in SDS Micelles Revealed by NMR Study of the Leu-Substituted Peptide. *J. Pept. Sci. 18*, 45-51.

153. Shenkarev, Z. O., Balandin, S. V., Trunov, K. I., Paramonov, A. S., Sukhanov, S. V., Barsukov, L. I., Arseniev, A. S., and Ovchinnikova, T. V. (2011) Molecular Mechanism of Action of Beta-Hairpin Antimicrobial Peptide Arenicin: Oligomeric Structure in Dodecylphosphocholine Micelles and Pore Formation in Planar Lipid Bilayers. *Biochemistry. 50*, 6255-6265.

154. Neumoin, A., Cohen, L. S., Arshava, B., Tantry, S., Becker, J. M., Zerbe, O., and Naider, F. (2009) Structure of a Double Transmembrane Fragment of a G-Protein-Coupled Receptor in Micelles. *Biophys. J. 96*, 3187-3196.

155. Morris, K. F., and Johnson, C. S. (1992) Diffusion-Ordered Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. *J. Am. Chem. Soc. 114*, 3139-3141.

156. Stilbs, P. (1982) Fourier Transform NMR Pulsed-Gradient spin—echo (FT-PGSE) Self-Diffusion Measurements of Solubilization Equilibria in SDS Solutions. *J. Colloid Interface Sci. 87*, 385-394.

157. Orfi, L., Lin, M., and Larive, C. K. (1998) Measurement of SDS Micelle-Peptide Association using (1)H NMR Chemical Shift Analysis and Pulsed-Field Gradient NMR Spectroscopy. *Anal. Chem. 70*, 1339-1345.

158. Jarvet, J., Danielsson, J., Damberg, P., Oleszczuk, M., and Graslund, A. (2007) Positioning of the Alzheimer Abeta(1-40) Peptide in SDS Micelles using NMR and Paramagnetic Probes. *J. Biomol. NMR. 39*, 63-72.

159. Szajdzinska-Pietek, E., Maldonado, R., Kevan, L., Jones, R. R. M., and Coleman, M. J. (1985) Electron Spin-Echo Modulation Studies of Doxylstearic Acid Spin Probes in Sodium and Tetramethylammonium Dodecyl Sulfate Micelles: Interaction of the Spin Probe with Deuterated Water and with Deuterated Terminal Methyl Groups in the Surfactant Molecules. *J. Am. Chem. Soc. 107*, 784-788.

160. Respondek, M., Madl, T., Gobl, C., Golser, R., and Zangger, K. (2007) Mapping the Orientation of Helices in Micelle-Bound Peptides by Paramagnetic Relaxation Waves. *J. Am. Chem. Soc. 129*, 5228-5234.

161. Wu, D. H., Chen, A. D., and Johnson, C. S. (1995) An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *J. Magn. Reson. 115*, 260-264.

162. Jönsson, B., Wennerström, H., Nilsson, P. G., and Linse, P. (1986) Self-Diffusion of Small Molecules in Colloidal Systems. *Coll. & Polym. Sci. 264*, 77-88.

163. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure. *J. Mol. Biol. 222*, 311-333.

164. Mehlis, B., Rueger, M., Becker, M., Bienert, M., Niedrich, H., and Oehme, P. (1980) Circular Dichroism Studies of Substance P and its C-Terminal Sequences. CD Spectra in Aqueous Solution and Effects of Hydrogen Ion Concentration. *Int. J. Pept. Prot. Res. 15*, 20-28.

165. Sreerama, N., Manning, M. C., Powers, M. E., Zhang, J. X., Goldenberg, D. P., and Woody, R. W. (1999) Tyrosine, Phenylalanine, and Disulfide Contributions to the Circular Dichroism of Proteins: Circular Dichroism Spectra of Wild-Type and Mutant Bovine Pancreatic Trypsin Inhibitor. *Biochemistry. 38*, 10814-10822.

166. Fasman, G. D., and Price, N. C. (1996) Circular Dichroism and the Conformational Analysis of Biomolecules. *Protein Sci. 5*, 2134.

167. Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) Turns in Peptides and Proteins. *Adv. Protein Chem. 37*, 1-109.

168. Deaton, K. R., Feyen, E. A., Nkulabi, H. J., and Morris, K. F. (2001) Pulsed-Field Gradient NMR Study of Sodium Dodecyl Sulfate micelle–peptide Association. *Magn. Reson. Chem. 39*, 276-282.

169. Chou, J. J., Baber, J. L., and Bax, A. (2004) Characterization of Phospholipid Mixed Micelles by Translational Diffusion. *J. Biomol. NMR. 29*, 299-308.

170. Cho, C. H., Urquidi, J., Singh, S., and Robinson, G. W. (1999) Thermal Offset Viscosities of Liquid H2O, D2O, and T2O. *J. Phys. Chem. B. 103*, 1991-1994.

171. Malliaris, A., Le Moigne, J., Sturm, J., and Zana, R. (1985) Temperature Dependence of the Micelle Aggregation Number and Rate of Intramicellar Excimer Formation in Aqueous Surfactant Solutions. *J. Phys. Chem. 89*, 2709-2713.

172. Paul, B. C., Islam, S. S., and Ismail, K. (1998) Effect of Acetate and Propionate Co-Ions on the Micellization of Sodium Dodecyl Sulfate in Water. *J Phys Chem B. 102*, 7807-7812.

173. Stafford, R. E., Fanni, T., and Dennis, E. A. (1989) Interfacial Properties and Critical Micelle Concentration of Lysophospholipids. *Biochemistry. 28*, 5113-5120.

174. Evans, J. N. S. (1995) *Biomolecular NMR Spectroscopy.* Oxford University Press, Oxford; New York.

175. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) The Solution Structure of Eglin c Based on Measurements of Many NOEs and Coupling Constants and its Comparison with X-Ray Structures. *Protein Sci. 1*, 736-751.

176. Almgren, M., Gimel, J. C., Wang, K., Karlsson, G., Edwards, K., Brown, W., and Mortensen, K. (1998) SDS Micelles at High Ionic Strength. A Light Scattering, Neutron Scattering, Fluorescence Quenching, and CryoTEM Investigation. *J. Colloid Interface Sci. 202*, 222-231.

177. Mascioni, A., Porcelli, F., Ilangovan, U., Ramamoorthy, A., and Veglia, G. (2003) Conformational Preferences of the Amylin Nucleation Site in SDS Micelles: An NMR Study. *Biopolymers. 69*, 29-41.

178. Lerch, M., Gafner, V., Bader, R., Christen, B., Folkers, G., and Zerbe, O. (2002) Bovine Pancreatic Polypeptide (bPP) Undergoes Significant Changes in Conformation and Dynamics upon Binding to DPC Micelles. *J. Mol. Biol. 322*, 1117-1133.

179. Braun, W., Wider, G., Lee, K. H., and Wuthrich, K. (1983) Conformation of Glucagon in a Lipid-Water Interphase by 1H Nuclear Magnetic Resonance. *J. Mol. Biol. 169*, 921-948.

180. Braun, W., Bosch, C., Brown, L. R., Go, N., and Wuthrich, K. (1981) Combined use of Proton-Proton Overhauser Enhancements and a Distance Geometry Algorithm for Determination of Polypeptide Conformations. Application to Micelle-Bound Glucagon. *Biochim. Biophys. Acta. 667*, 377-396.

181. Ying, J., Ahn, J. M., Jacobsen, N. E., Brown, M. F., and Hruby, V. J. (2003) NMR Solution Structure of the Glucagon Antagonist [desHis1, desPhe6, Glu9]Glucagon Amide in the Presence of Perdeuterated Dodecylphosphocholine Micelles. *Biochemistry. 42*, 2825-2835.

182. Lee, S., and Kim, Y. (1999) Solution Structure of Neuromedin B by (1)H Nuclear Magnetic Resonance Spectroscopy. *FEBS Lett. 460*, 263-269.

183. Dike, A., and Cowsik, S. M. (2006) Three-Dimensional Structure of Neuropeptide k Bound to Dodecylphosphocholine Micelles. *Biochemistry. 45*, 2994-3004.

184. Coutant, J., Curmi, P. A., Toma, F., and Monti, J. P. (2007) NMR Solution Structure of Neurotensin in Membrane-Mimetic Environments: Molecular Basis for Neurotensin Receptor Recognition. *Biochemistry. 46*, 5656-5663.

185. Inooka, H., Ohtaki, T., Kitahara, O., Ikegami, T., Endo, S., Kitada, C., Ogi, K., Onda, H., Fujino, M., and Shirakawa, M. (2001) Conformation of a Peptide Ligand Bound to its G-Protein Coupled Receptor. *Nat. Struct. Biol. 8*, 161-165.

186. Bader, R., Bettio, A., Beck-Sickinger, A. G., and Zerbe, O. (2001) Structure and Dynamics of Micelle-Bound Neuropeptide Y: Comparison with Unligated NPY and Implications for Receptor Selection. *J. Mol. Biol. 305*, 307-329.

187. Lerch, M., Mayrhofer, M., and Zerbe, O. (2004) Structural Similarities of Micelle-Bound Peptide YY (PYY) and Neuropeptide Y (NPY) are Related to their Affinity Profiles at the Y Receptors. *J. Mol. Biol. 339*, 1153-1168.

188. Lucyk, S., Taha, H., Yamamoto, H., Miskolzie, M., and Kotovych, G. (2006) NMR Conformational Analysis of Proadrenomedullin N-Terminal 20 Peptide, a Proangiogenic Factor Involved in Tumor Growth. *Biopolymers. 81*, 295-308.

189. Nanga, R. P., Brender, J. R., Xu, J., Hartman, K., Subramanian, V., and Ramamoorthy, A. (2009) Three-Dimensional Structure and Orientation of Rat Islet Amyloid Polypeptide Protein in a Membrane Environment by Solution NMR Spectroscopy. *J. Am. Chem. Soc. 131*, 8252-8261.

190. Engelman, D. M., Chen, Y., Chin, C. N., Curran, A. R., Dixon, A. M., Dupuy, A. D., Lee, A. S., Lehnert, U., Matthews, E. E., Reshetnyak, Y. K., Senes, A., and Popot, J. L. (2003) Membrane Protein Folding: Beyond the Two Stage Model. *FEBS Lett. 555*, 122-125.

191. Jacobs, R. E., and White, S. H. (1989) The Nature of the Hydrophobic Binding of Small Peptides at the Bilayer Interface: Implications for the Insertion of Transbilayer Helices. *Biochemistry. 28*, 3421-3437.

192. Popot, J. L., and Engelman, D. M. (1990) Membrane Protein Folding and Oligomerization: The Two-Stage Model. *Biochemistry. 29*, 4031-4037.

193. Bordag, N., and Keller, S. (2010) Alpha-Helical Transmembrane Peptides: A "Divide and Conquer" Approach to Membrane Proteins. *Chem. Phys. Lipids. 163*, 1-26.

194. Zou, C., Naider, F., and Zerbe, O. (2008) Biosynthesis and NMR-Studies of a Double Transmembrane Domain from the Y4 Receptor, a Human GPCR. *J. Biomol. NMR. 42*, 257-269.

195. Zheng, H., Zhao, J., Sheng, W., and Xie, X. Q. (2006) A Transmembrane Helix-Bundle from G-Protein Coupled Receptor CB2: Biosynthesis, Purification, and NMR Characterization. *Biopolymers. 83*, 46-61.

196. Neumoin, A., Arshava, B., Becker, J., Zerbe, O., and Naider, F. (2007) NMR Studies in Dodecylphosphocholine of a Fragment Containing the Seventh Transmembrane Helix of a G-Protein-Coupled Receptor from Saccharomyces Cerevisiae. *Biophys. J. 93*, 467-482.

197. Yeagle, P. L., Choi, G., and Albert, A. D. (2001) Studies on the Structure of the G-Protein-Coupled Receptor Rhodopsin Including the Putative G-Protein Binding Site in Unactivated and Activated Forms. *Biochemistry. 40*, 11932-11937.

198. Marley, J., Lu, M., and Bracken, C. (2001) A Method for Efficient Isotopic Labeling of Recombinant Proteins. *J. Biomol. NMR. 20*, 71-75.

199. Fan, Q. R., Garboczi, D. N., Winter, C. C., Wagtmann, N., Long, E. O., and Wiley, D. C. (1996) Direct Binding of a Soluble Natural Killer Cell Inhibitory Receptor to a Soluble Human Leukocyte Antigen-Cw4 Class I Major Histocompatibility Complex Molecule. *Proc. Natl. Acad. Sci. U. S. A. 93*, 7178-7183.

200. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., and Zardecki, C. (2002) The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr. 58*, 899-907.

201. Cheung, M. S., Maguire, M. L., Stevens, T. J., and Broadhurst, R. W. (2010) DANGLE: A Bayesian Inferential Method for Predicting Protein Backbone Dihedral Angles and Secondary Structure. *J. Magn. Reson. 202*, 223-233.

202. Drake, M. T., Shenoy, S. K., and Lefkowitz, R. J. (2006) Trafficking of G Protein–Coupled Receptors. *Circulation Research. 99*, 570-582.

203. Mirzadegan, T., Benko, G., Filipek, S., and Palczewski, K. (2003) Sequence Analyses of G-Protein-Coupled Receptors: Similarities to Rhodopsin. *Biochemistry. 42*, 2759-2767.

204. Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brunger, A. T., and Engelman, D. M. (1994) A Dimerization Motif for Transmembrane α-Helices. *Nat Struct Mol Biol. 1*, 157-163.

205. D'Aoust, J. P., and Tiberi, M. (2010) Role of the Extracellular Amino Terminus and First Membrane-Spanning Helix of Dopamine D1 and D5 Receptors in Shaping Ligand Selectivity and Efficacy. *Cell. Signal. 22*, 106-116.

206. Georgieva, E. R., Ramlall, T. F., Borbat, P. P., Freed, J. H., and Eliezer, D. (2008) Membrane-Bound Alpha-Synuclein Forms an Extended Helix: Long-Distance Pulsed ESR Measurements using Vesicles, Bicelles, and Rodlike Micelles. *J. Am. Chem. Soc. 130*, 12856-12857.

207. Bussell, R.,Jr, and Eliezer, D. (2003) A Structural and Functional Role for 11-Mer Repeats in Alpha-Synuclein and Other Exchangeable Lipid Binding Proteins. *J. Mol. Biol. 329*, 763-778.

208. Chandra, S., Chen, X., Rizo, J., Jahn, R., and Sudhof, T. C. (2003) A Broken Alpha -Helix in Folded Alpha -Synuclein. *J. Biol. Chem. 278*, 15313-15318.

209. Bussell, R.,Jr, Ramlall, T. F., and Eliezer, D. (2005) Helix Periodicity, Topology, and Dynamics of Membrane-Associated Alpha-Synuclein. *Protein Sci. 14*, 862-872.

210. Ulmer, T. S., Bax, A., Cole, N. B., and Nussbaum, R. L. (2005) Structure and Dynamics of Micelle-Bound Human Alpha-Synuclein. *J. Biol. Chem. 280*, 9595-9603.

211. Gao, X., and Wong, T. C. (2001) NMR Studies of Adrenocorticotropin Hormone Peptides in Sodium Dodecylsulfate and Dodecylphosphocholine Micelles: Proline Isomerism and Interactions of the Peptides with Micelles. *Biopolymers. 58*, 20-32.

212. Strandberg, E., Ozdirekcan, S., Rijkers, D. T., van der Wel, P. C., Koeppe, R. E.,2nd, Liskamp, R. M., and Killian, J. A. (2004) Tilt Angles of Transmembrane Model Peptides in Oriented and Non-Oriented Lipid Bilayers as Determined by 2H Solid-State NMR. *Biophys. J. 86*, 3709-3721.

213. Pervushin, K. V., Orekhov, V. Y., Popov, A. I., Musina, L. Y., and Arseniev, A. S. (1994) Three-Dimensional Structure of (1-71)Bacterioopsin Solubilized in methanol/chloroform and SDS Micelles Determined by 15N-1H Heteronuclear NMR Spectroscopy. *Eur. J. Biochem. 219*, 571-583.

214. Orekhov, V. Y., Pervushin, K. V., Korzhnev, D. M., and Arseniev, A. S. (1995) *Backbone Dynamics of (1–71)- and (1–36)Bacterioopsin Studied by Two-Dimensional ^{1}H-^{15}N NMR Spectroscopy.* Springer Netherlands, .

215. Dmitriev, O., Jones, P. C., Jiang, W., and Fillingame, R. H. (1999) Structure of the Membrane Domain of Subunit b of the Escherichia Coli F0F1 ATP Synthase. *J. Biol. Chem. 274*, 15598-15604.

216. Engler, A., Stangler, T., and Willbold, D. (2002) Structure of Human Immunodeficiency Virus Type 1 Vpr(34-51) Peptide in Micelle Containing Aqueous Solution. *Eur. J. Biochem. 269*, 3264-3269.

217. Schwaiger, M., Lebendiker, M., Yerushalmi, H., Coles, M., Groger, A., Schwarz, C., Schuldiner, S., and Kessler, H. (1998) NMR Investigation of the Multidrug Transporter EmrE, an Integral Membrane Protein. *Eur. J. Biochem. 254*, 610-619.

218. Mottamal, M., Shen, S., Guembe, C., and Krilov, G. (2007) Solvation of Transmembrane Proteins by Isotropic Membrane Mimetics: A Molecular Dynamics Study. *J Phys Chem B. 111*, 11285-11296.

219. Roccatano, D., Fioroni, M., Zacharias, M., and Colombo, G. (2005) Effect of Hexafluoroisopropanol Alcohol on the Structure of Melittin: A Molecular Dynamics Simulation Study. *Protein Sci. 14*, 2582-2589.

220. Hirota, N., Mizuno, K., and Goto, Y. (1997) Cooperative Alpha-Helix Formation of Beta-Lactoglobulin and Melittin Induced by Hexafluoroisopropanol. *Protein Sci. 6*, 416-421.

221. Fioroni, M., Diaz, M. D., Burger, K., and Berger, S. (2002) Solvation Phenomena of a Tetrapeptide in water/trifluoroethanol and water/ethanol Mixtures: A Diffusion NMR, Intermolecular NOE, and Molecular Dynamics Study. *J. Am. Chem. Soc. 124*, 7737-7744.

222. Roccatano, D., Colombo, G., Fioroni, M., and Mark, A. E. (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*, 12179-12184.

223. Reddy, T., and Rainey, J. K. (2010) Interpretation of Biomolecular NMR Spin Relaxation Parameters. *Biochem. Cell Biol. 88*, 131-142.

224. Jarymowycz, V. A., and Stone, M. J. (2006) Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences. *Chem. Rev. 106*, 1624-1671.

225. Spyracopoulos, L. (2005) Thermodynamic Interpretation of Protein Dynamics from NMR Relaxation Measurements. *Protein Pept. Lett. 12*, 235-240.

226. Kern, D., and Zuiderweg, E. R. (2003) The Role of Dynamics in Allosteric Regulation. *Curr. Opin. Struct. Biol. 13*, 748-757.

227. Abragam, A. (1961) *The Principles of Nuclear Magnetism.* Clarendon Press, Oxford.

228. Mandel, A. M., Akke, M., and Palmer, A. G.,3rd. (1995) Backbone Dynamics of Escherichia Coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol. 246*, 144-163.

229. Hiyama, Y., Niu, C. H., Silverton, J. V., Bavoso, A., and Torchia, D. A. (1988) Determination of 15N Chemical Shift Tensor Via 15N-2H Dipolar Coupling in Boc-Glycylglycyl[15N Glycine]Benzyl Ester. *J. Am. Chem. Soc. 110*, 2378-2383.

230. Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc. 104*, 4546-4559.

231. Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc. 104*, 4559-4570.

232. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J. Am. Chem. Soc. 112*, 4989-4991.

233. Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995) Rotational Diffusion Anisotropy of Human Ubiquitin from 15N NMR Relaxation. *J. Am. Chem. Soc. 117*, 12562-12566.

234. Clore, G. M., Gronenborn, A. M., Szabo, A., and Tjandra, N. (1998) Determining the Magnitude of the Fully Asymmetric Diffusion Tensor from Heteronuclear Relaxation Data in the Absence of Structural Information. *J. Am. Chem. Soc. 120*, 4889-4890.

235. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) Spectral Density Function Mapping using 15N Relaxation Data Exclusively. *J. Biomol. NMR. 6*, 153-162.

236. Diercks, T., Coles, M., and Kessler, H. (1999) An Efficient Strategy for Assignment of Cross-Peaks in 3D Heteronuclear NOESY Experiments. *J. Biomol. NMR. 15*, 177-180.

237. Kay, L., Keifer, P., and Saarinen, T. (1992) Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc. 114*, 10663-10665.

238. Spyracopoulos, L. (2006) A Suite of Mathematica Notebooks for the Analysis of Protein Main Chain 15N NMR Relaxation Data. *J. Biomol. NMR. 36*, 215-224.

239. d'Auvergne, E. J., and Gooley, P. R. (2003) The use of Model Selection in the Model-Free Analysis of Protein Dynamics. *J. Biomol. NMR. 25*, 25-39.

240. Sonnichsen, F. D., Van Eyk, J. E., Hodges, R. S., and Sykes, B. D. (1992) Effect of Trifluoroethanol on Protein Secondary Structure: An NMR and CD Study using a Synthetic Actin Peptide. *Biochemistry. 31*, 8790-8798.

241. Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A., and Doniach, S. (2007) Size and Shape of Detergent Micelles Determined by Small-Angle X-Ray Scattering. *J Phys Chem B. 111*, 12427-12438.

242. Hayter, J.,B., and Penfold, ,J. (1983) Determination of Micelle Structure and Charge by Neutron Small-Angle Scattering. *Coll. & Polym. Sci. 261*, 1022-1030.

243. Marrink, S. J., Tieleman, D. P., and Mark, A. E. (2000) Molecular Dynamics Simulation of the Kinetics of Spontaneous Micelle Formation. *J Phys Chem B. 104*, 12165-12173.

244. Harris, T., Graber, A. R., and Covarrubias, M. (2003) Allosteric Modulation of a Neuronal K+ Channel by 1-Alkanols is Linked to a Key Residue in the Activation Gate. *Am. J. Physiol. Cell. Physiol. 285*, C788-96.

245. Singh, R., Hurst, D. P., Barnett-Norris, J., Lynch, D. L., Reggio, P. H., and Guarnieri, F. (2002) Activation of the Cannabinoid CB1 Receptor may Involve a W6 48/F3 36 Rotamer Toggle Switch. *J. Pept. Res. 60*, 357-370.

246. Hall, S. E., Roberts, K., and Vaidehi, N. (2009) Position of Helical Kinks in Membrane Protein Crystal Structures and the Accuracy of Computational Prediction. *J. Mol. Graph. Model. 27*, 944-950.

247. Riek, R. P., Rigoutsos, I., Novotny, J., and Graham, R. M. (2001) Non-Alpha-Helical Elements Modulate Polytopic Membrane Protein Architecture. *J. Mol. Biol. 306*, 349-362.

248. von Heijne, G. (2006) Membrane-Protein Topology. *Nat. Rev. Mol. Cell Biol. 7*, 909-918.

249. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P., and Bowie, J. U. (2004) The Evolution of Transmembrane Helix Kinks and the Structural Diversity of G Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A. 101*, 959-963.

250. Meruelo, A. D., Samish, I., and Bowie, J. U. (2011) TMKink: A Method to Predict Transmembrane Helix Kinks. *Protein Sci. 20*, 1256-1264.

251. Kneissl, B., Mueller, S. C., Tautermann, C. S., and Hildebrandt, A. (2011) String Kernels and High-Quality Data Set for Improved Prediction of Kinked Helices in Alpha-Helical Membrane Proteins. *J. Chem. Inf. Model. 51*, 3017-3025.

252. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2005) PDB_TM: Selection and Membrane Localization of Transmembrane Proteins in the Protein Data Bank. *Nucleic Acids Res. 33*, D275-8.

253. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2004) Transmembrane Proteins in the Protein Data Bank: Identification and Classification. *Bioinformatics. 20*, 2964-2972.

254. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2005) TMDET: Web Server for Detecting Transmembrane Regions of Proteins by using their 3D Coordinates. *Bioinformatics. 21*, 1276-1277.

255. Wang, G., and Dunbrack, R. L.,Jr. (2003) PISCES: A Protein Sequence Culling Server. *Bioinformatics. 19*, 1589-1591.

256. Lovell, S. C., Davis, I. W., Arendall, W. B.,3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure Validation by Calpha Geometry: Phi,Psi and Cbeta Deviation. *Proteins. 50*, 437-450.

257. Mohapatra, P. K., Khamari, A., and Raval, M. K. (2004) A Method for Structural Analysis of Alpha-Helices of Membrane Proteins. *J. Mol. Model. 10*, 393-398.

258. Barlow, D. J., and Thornton, J. M. (1988) Helix Geometry in Proteins. *J. Mol. Biol. 201*, 601-619.

259. Eisenberg, D. (1984) Three-Dimensional Structure of Membrane and Surface Proteins. *Annu. Rev. Biochem. 53*, 595-623.

260. Liu, L. P., and Deber, C. M. (1998) Uncoupling Hydrophobicity and Helicity in Transmembrane Segments. Alpha-Helical Propensities of the Amino Acids in Non-Polar Environments. *J. Biol. Chem. 273*, 23645-23648.

261. Chang, C., and Lin, C. (2011) LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol. 2*, 27:1-27:27.

262. Kabsch, W., and Sander, C. (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers. 22*, 2577-2637.

263. Kumar, S., and Bansal, M. (1996) Structural and Sequence Characteristics of Long Alpha Helices in Globular Proteins. *Biophys. J. 71*, 1574-1586.

264. Visiers, I., Braunheim, B. B., and Weinstein, H. (2000) Prokink: A Protocol for Numerical Evaluation of Helix Distortions by Proline. *Protein Eng. 13*, 603-606.

265. O'Neil, K. T., and DeGrado, W. F. (1990) A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science. 250*, 646-651.

266. Creighton, T. E.,. (1993) *Proteins : Structures and Molecular Properties.* W.H. Freeman, New York.

267. Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R. (2003) X-Ray Structure of a Voltage-Dependent K+ Channel. *Nature. 423*, 33-41.

268. Tusnady, G. E., Kalmar, L., and Simon, I. (2008) TOPDB: Topology Data Bank of Transmembrane Proteins. *Nucleic Acids Res. 36*, D234-9.

269. Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K., and Gouaux, E. (2007) Coupling Substrate and Ion Binding to Extracellular Gate of a Sodium-Dependent Aspartate Transporter. *Nature. 445*, 387-393.

270. Gerber, S., Comellas-Bigler, M., Goetz, B. A., and Locher, K. P. (2008) Structural Basis of Trans-Inhibition in a molybdate/tungstate ABC Transporter. *Science. 321*, 246-250.

271. Solmaz, S. R., and Hunte, C. (2008) Structure of Complex III with Bound Cytochrome c in Reduced State and Definition of a Minimal Core Interface for Electron Transfer. *J. Biol. Chem. 283*, 17542-17549.

272. Wang, Y., Huang, Y., Wang, J., Cheng, C., Huang, W., Lu, P., Xu, Y. N., Wang, P., Yan, N., and Shi, Y. (2009) Structure of the Formate Transporter FocA Reveals a Pentameric Aquaporin-Like Channel. *Nature. 462*, 467-472.

273. Macaluso, N. J., and Glen, R. C. (2010) Exploring the 'RPRL' Motif of Apelin-13 through Molecular Simulation and Biological Evaluation of Cyclic Peptide Analogues. *ChemMedChem. 5*, 1247-1253.

274. Macaluso, N. J., Pitkin, S. L., Maguire, J. J., Davenport, A. P., and Glen, R. C. (2011) Discovery of a Competitive Apelin Receptor (APJ) Antagonist. *ChemMedChem. 6*, 1017-1023.

275. Murza, A., Parent, A., Besserer-Offroy, E., Tremblay, H., Karadereye, F., Beaudet, N., Leduc, R., Sarret, P., and Marsault, E. (2011) Elucidation of the Structure-Activity Relationships of Apelin: Influence of Unnatural Amino Acids on Binding, Signaling, and Plasma Stability. *ChemMedChem. 7*, 318-325.