



12-2014

Novel Insights into the Role of the Smoothened Cysteine Rich Domain in Hedgehog Signalling

Rajashree Rana

University of Tennessee Health Science Center

Follow this and additional works at: <https://dc.uthsc.edu/dissertations>



Part of the [Medical Biochemistry Commons](#), and the [Medical Cell Biology Commons](#)

Recommended Citation

Rana, Rajashree , "Novel Insights into the Role of the Smoothened Cysteine Rich Domain in Hedgehog Signalling" (2014). *Theses and Dissertations (ETD)*. Paper 207. <http://dx.doi.org/10.21007/etd.cghs.2014.0257>.

This Dissertation is brought to you for free and open access by the College of Graduate Health Sciences at UTHSC Digital Commons. It has been accepted for inclusion in Theses and Dissertations (ETD) by an authorized administrator of UTHSC Digital Commons. For more information, please contact jwelch30@uthsc.edu.

Novel Insights into the Role of the Smoothened Cysteine Rich Domain in Hedgehog Signalling

Document Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Program

Biomedical Sciences

Track

Cell Biology and Biochemistry

Research Advisor

Jie J. Zheng, Ph.D.

Committee

Martha M. Howe, Ph.D. David R. Nelson, Ph.D. Stacey K. Ogden, Ph.D. Stephen White, DPhil.

DOI

10.21007/etd.cghs.2014.0257

Comments

Six month embargo expired June 2015

**NOVEL INSIGHTS INTO THE ROLE OF THE SMOOTHENED CYSTEINE
RICH DOMAIN IN HEDGEHOG SIGNALLING**

A Dissertation
Presented for
The Graduate Studies Council
The University of Tennessee
Health Science Center

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy
From The University of Tennessee

By
Rajashree Rana
December 2014

Chapter 3 and Chapter 4 © 2013 by Nature Publishing Group.
All other material © 2014 by Rajashree Rana.
All rights reserved.

DEDICATION

This dissertation is dedicated to my parents, Mrs. Rinku Rana and Mr. Santanu Rana for their constant love and support.

ACKNOWLEDGEMENTS

Cardinally, I would sincerely like to thank my mentor and dissertation advisor Dr. Jie Zheng for introducing me to the world of science. One of Dr. Zheng's favorite quotes is "*time flies*", time indeed does fly and it is already five years that I have been studying under his guidance. One of the most important things that I have learnt from him is to respect time and make the most of each opportunity. He has been an amazingly supportive mentor and without his guidance and encouragement the completion of this dissertation would not have been possible. He has encouraged me to think critically and solve problems. Dr. Zheng is also a kind and generous human being. He is social and personally I have enjoyed the numerous lunchtime discussions we have had over the past years. At times Dr. Zheng has been a friend and walked alongside through the uneven roads of science, sometimes he has been a philosopher, while at other times he has been a guide. To sum it up all Dr. Zheng has been a wonderful mentor.

I would also take this opportunity to thank my committee members: Dr. Martha Howe, Dr. David Nelson, Dr. Stacey Ogden and Dr. Stephen White for their advice, time and support over the past five years. I have had an amazing committee and over the years each one has helped me in their own unique way. I would like to thank Dr. Howe for many reasons, but most of all for helping me improve my writing skills. From the early years of graduate school I have been fascinated by Dr. Nelson's biochemistry lectures. He has been an amazing teacher and the secret behind my grades. Hedgehog was just a spiny animal for me up until Dr. Ogden explained and introduced me to the world of Hedgehog signalling. Dr. White is obviously an amazing scientist and has provided numerous important scientific suggestions over the years, but most importantly his jovial comments lightened up the mood and helped survival through the not so glorious days of graduate school.

Additionally, I am grateful to all the Zheng lab members for all they have taught me. In particular I would like to thank Cristina Guibao for teaching me the various techniques in the lab and creating a congenial atmosphere to work. I am also grateful to Dr. Hojin Lee and Dr. Ju Bao for helping me with the docking experiments. I also appreciate Dr. Grace Royappa for teaching me the technique of NMR.

Furthermore, I would like to thank my parents for their unconditional love and the faith that they have in me. I also whole-heartedly appreciate my husband, Dr. Joydip Kundu, for his infinite love and constant support throughout my graduate study.

ABSTRACT

The Hedgehog (Hh) signal transduction pathway functions as one of the key developmental pathways and deranged Hh signalling is associated with numerous cancer and tumor conditions. The Smoothed (Smo) G protein coupled receptor (GPCR) functions as the signal transducer of the Hh pathway and is the most attractive drug target of the pathway. The structure of the Smo receptor includes seven membrane spanning domains, extracellular and intracellular loops connecting the membranous domains and the extracellular cysteine rich domain (CRD). The extracellular CRD of the Smo receptor is homologous to the Frizzled (FzD) CRD. The FzD CRD interacts with the physiological ligand of the FzD receptor: Wnt, but the function of the Smo CRD is not clearly elucidated.

In the present dissertation we determined the first structure of the *Drosophila* Smo CRD by solution Nuclear Magnetic Resonance (NMR) spectroscopy. The structure of the Smo CRD comprises primarily of four helices and is stabilized by disulfide bonds between the conserved cysteine residues. The tertiary structure of the Smo CRD is similar to the structure of the FzD CRD.

Additionally, we showed that both the *Drosophila* and the human Smo CRD interact with the glucocorticoid: Budesonide (Bud), albeit with different binding affinities. We identified the key residues in both species that are important for this interaction and generated the complex structure of the *Drosophila* Smo CRD with Bud using computational methodologies. The binding pocket in the Smo CRD can be targeted for the development of novel anti-cancer therapeutics.

The Smo receptor possesses a binding cavity located in the transmembrane domains and we determined that the Smo CRD also possesses ligand binding capabilities. This finding enabled us to propose a novel mechanism of Hh signalling. We speculate that certain Smo ligands bind to the Smo CRD and membrane spanning domains simultaneously to regulate intracellular signalling. Phosphatidylinositol-4-phosphate (PI(4)P) has been shown to regulate Smo activation. We determined that the fatty-acid side chains of the phosphatidylinositols interact with Smo CRD, and PI(4)P specifically enhances activator induced Hh signalling. Therefore, we speculated that PI(4)P may function to bridge the Smo extracellular and the membrane spanning domains, wherein the fatty-acid side chains of PI(4)P docks in the Smo CRD and the inositol head with the fourth position phosphate is necessary for anchoring in the cavity in the transmembrane domains; however, PI(4)P may not necessarily function as endogenous Smo ligand because in the cell-based luciferase reporter assays PI(4)P could only enhance signalling in the presence of an activator of Hh signalling. This suggests that PI(4)P binding to Smo is either a prerequisite and renders Smo active to further induce downstream signalling or PI(4)P acts as an allosteric activator of Hh signalling.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION	1
1.1. An Overview of the Components of the Hedgehog Signal Transduction Machinery	1
1.1.1. Hedgehog	1
1.1.2. Patched	2
1.1.3. Hedgehog Co-receptors	2
1.1.4. Smoothened	5
1.1.5. Hedgehog Signalling Complex	5
1.1.6. Ci/Gli Transcription Factors	6
1.2. The Smoothened Receptor	6
1.2.1. Smoothened Structure and Post-translational Modifications	6
1.2.2. Smoothened as a G Protein Coupled Receptor	9
1.2.3. Smoothened Signalling and Regulation by Small Molecules	10
1.3. The Hedgehog Signal Transduction Cascade	13
1.4. Implications of the Hedgehog Pathway in Cancer	15
1.4.1. Basal Cell Carcinoma	15
1.4.2. Medulloblastoma	16
1.4.3. Rhabdomyosarcoma	16
1.4.4. Targeting the Hedgehog Pathway to Treat Cancers	16
1.4.5. Promoting the Smoothened Cysteine Rich Domain as a Novel Target for Drug Development	17
CHAPTER 2. METHODS	18
2.1. Introduction	18
2.2. The Physics and Mathematics behind Nuclear Magnetic Resonance	19
2.3. The Nuclear Magnetic Resonance Instrumentation	23
2.3.1. Magnet	23
2.3.2. Lock System	25
2.3.3. Shim Coils	25
2.3.4. Sample Probe	25
2.3.5. Control Console and the Terminal Computer	26
2.4. Setting up a Nuclear Magnetic Resonance Experiment	26
2.5. Protein Structure Determination by Nuclear Magnetic Resonance	28
2.5.1. Synopsis	28
2.5.2. Nuclear Magnetic Resonance Experiments	29
2.5.3. Heteronuclear Backbone Assignment	30
2.5.4. Amino Acid Side Chain Assignment Strategy	31
2.5.5. Nuclear Overhauser Effect Correlation Spectroscopy Assignments and Distance Constraints	33
2.5.6. Structure Calculation and Refinement	35
2.5.7. Structure Validation	36
2.6. Nuclear Magnetic Resonance as a Tool to Study Protein-ligand Interactions	38

2.6.1. 1D Nuclear Magnetic Resonance Experiments to Screen and Identify Potential Ligands for Proteins.....	38
2.6.2. Chemical Shift Mapping to Determine the Ligand Binding Surface on the Protein.....	40
2.6.3. Binding Kinetics to Determine the Dynamics of Protein-ligand Interaction...	41
2.6.4. Dissociation Constant to Determine the Binding Affinity.....	41
2.6.5. Protein-ligand Complex Structure Determination by Nuclear Magnetic Resonance.....	44
CHAPTER 3. NUCLEAR MAGNETIC RESONANCE ASSIGNMENTS AND SOLUTION STRUCTURE OF THE SMOOTHENED CYSTEINE RICH DOMAIN	46
3.1. Introduction.....	46
3.2. Experimental Procedures	48
3.2.1. Vector Construction.....	48
3.2.2. Protein Expression and Purification.....	48
3.2.3. Circular Dichroism Spectroscopy	48
3.2.4. NMR Spectroscopy.....	49
3.3. Results and Discussion	49
3.3.1. Circular Dichroism Spectra.....	49
3.3.2. One Dimensional Proton NMR Spectra.....	49
3.3.3. Two Dimensional ^1H - ^{15}N and ^1H - ^{13}C Heteronuclear Single Quantum Coherence Spectra and Nuclear Magnetic Resonance Assignments.....	52
3.3.4. Nuclear Overhauser Effect Assignments Structure Calculation and Refinement.....	52
3.3.5. Solution Structure of the <i>Drosophila</i> Smo CRD	56
3.4. Discussion.....	61
CHAPTER 4. GLUCOCORTICOIDS MODULATE HEDGEHOG SIGNALLING.....	63
4.1. Introduction.....	63
4.2. Experimental Procedures	64
4.2.1. Protein Expression and Purification.....	64
4.2.2. Nuclear Magnetic Resonance Chemical Shift Perturbation Analysis.....	64
4.2.3. Structural Modelling of <i>Drosophila</i> Smoothened Cysteine Rich Domain in Complex with Budesonide.....	65
4.2.4. Bio-Layer Interferometry Assay	65
4.2.5. Cell-based Luciferase Reporter Assays	66
4.3. Results and Discussion	66
4.3.1. <i>Drosophila</i> Smo CRD Binds to the Synthetic Glucocorticoid Budesonide....	66
4.3.2. High Ambiguity Driven Biomolecular Docking of the <i>Drosophila</i> Smoothened Cysteine Rich Domain in Complex with Budesonide	68
4.3.3. Human Smoothened Cysteine Rich Domain Binds to the Synthetic Glucocorticoid Budesonide.....	71
4.3.4. Binding Affinity of the Glucocorticoid Budesonide for Smoothened Cysteine Rich Domain	71

4.3.5. Effect of Budesonide on <i>in-vitro</i> Hh Signalling	75
4.4. Discussion	79
4.4.1. Cysteine Rich Domains and Hydrophobic Ligands.....	79
4.4.2. Same yet Different <i>Drosophila</i> and Human Smoothened Cysteine Rich Domain.....	79
4.4.3. Ligand Binding Provides a Novel Insight into Hedgehog Signalling.....	81
CHAPTER 5. PHOSPHOLIPIDS AND HEDGEHOG SIGNALLING	82
5.1. Introduction.....	82
5.2. Experimental Procedures	83
5.2.1. Protein Expression and Purification.....	83
5.2.2. NMR Experiments	83
5.2.3. Cell-based Reporter Assays	84
5.3. Results and Discussion	84
5.3.1. Phosphatidylinositols Bind to the Human Smoothened Cysteine Rich Domain.....	84
5.3.2. The Phosphatidylinositol-4-phosphate Induces Backbone Chemical Shift Perturbations in Human Smoothened Cysteine Rich Domain.....	85
5.3.3. Phosphatidylinositols Exert Varied Effects on <i>in-vitro</i> Hedgehog Signalling.....	90
5.4. Discussion.....	90
CHAPTER 6. IMPLICATIONS AND FUTURE DIRECTIONS	95
6.1. The Structures of the Smoothened G Protein Coupled Receptor (Structure Determination Strategies follow the Rule of Divide and Conquer).....	95
6.2. Multiple Ligand Binding Sites on Smoothened Receptor Help Propose a Novel Mechanism for Hedgehog Signalling.....	97
6.3. The Unanswered Question: Does Smoothened Really Have an Endogenous Ligand?	100
6.4. Smoothened as a Template for Drug Discovery	101
LIST OF REFERENCES.....	105
APPENDIX A. <i>DROSOPHILA</i> SMO CRD ATOM ASSIGNMENTS.....	116
APPENDIX B. HUMAN SMO CRD ATOM ASSIGNMENTS	151
VITA.....	161

LIST OF TABLES

Table 3-1. NMR and refinement statistics used for calculating the structure of <i>Drosophila</i> Smo CRD.....	59
Table 4-1. The HADDOCK input and output parameters for docking Bud on <i>Drosophila</i> Smo CRD.....	70

LIST OF FIGURES

Figure 1-1. Schematic representation of the post-translational modifications of the Hh ligand.	3
Figure 1-2. Schematic representation of the Hh receptors.	4
Figure 1-3. Schematics of the Hedgehog Signalling Complex (HSC) and the domain architecture of the Gli transcription factor.	7
Figure 1-4. The structural domains of the Smo receptor and the disulfide bonds stabilizing the Smo extracellular linker and the extracellular loops.	8
Figure 1-5. Chemical structure of selected Smo modulators.	12
Figure 1-6. The Hedgehog Signal Transduction Cascade.	14
Figure 2-1. Behavior of nuclear spins in the absence and presence of an external magnetic field.	21
Figure 2-2. Magnetic field strength and NMR sensitivity.	22
Figure 2-3. Components of an NMR spectrometer.	24
Figure 2-4. Visualizing three dimensional NMR spectrum and generation of strips.	32
Figure 2-5. Flowchart for NMR based protein structure determination.	37
Figure 2-6. Transferred NOE experiments to determine protein-ligand interactions.	39
Figure 2-7. Representation of the different binding kinetics.	42
Figure 2-8. Example of chemical shift perturbation plot.	43
Figure 3-1. Primary sequence alignment of Smo CRD.	47
Figure 3-2. The CD spectra of the Smo CRD.	50
Figure 3-3. The 1D proton spectrum of <i>Drosophila</i> Smo CRD.	51
Figure 3-4. The 2D spectra of the <i>Drosophila</i> Smo CRD.	53
Figure 3-5. The NOE strip plot representing helical region of <i>Drosophila</i> Smo CRD.	54
Figure 3-6. Chemical structure depicting the atoms in an amino acid.	55
Figure 3-7. Structure refinement of the <i>Drosophila</i> Smo CRD.	57

Figure 3-8. Solution structure of the <i>Drosophila</i> Smo CRD.	58
Figure 3-9. Structure of the <i>Drosophila</i> Smo CRD and comparison with the mouse FzD8 CRD.	60
Figure 4-1. The <i>Drosophila</i> Smo CRD binds to the inhibitory glucocorticoid Bud.	67
Figure 4-2. The complex structure of <i>Drosophila</i> Smo CRD and the glucocorticoid Bud.	69
Figure 4-3. The human Smo CRD binds to the inhibitory glucocorticoid Bud.	72
Figure 4-4. Binding affinity measurement of <i>Drosophila</i> Smo CRD and Bud for residues F187, F188 and F191.	73
Figure 4-5. Binding affinity measurement of <i>Drosophila</i> Smo CRD and Bud for residues L127 and A132.	74
Figure 4-6. Binding affinity measurement of human Smo CRD and Bud.	76
Figure 4-7. BLI binding assays show that Bud binds to <i>Drosophila</i> Smo CRD with a lower affinity than human Smo CRD.	77
Figure 4-8. Bud attenuates SAG induced reporter gene activity.	78
Figure 5-1. The phosphatidylinositols with a single phosphate group bind to human Smo CRD.	86
Figure 5-2. The phosphatidylinositols with two phosphate groups bind to human Smo CRD.	87
Figure 5-3. The fatty-acid side chain of the phosphatidylinositols is essential for interaction with the human Smo CRD.	88
Figure 5-4. The phosphatidylinositol: PI(4)P, interacts with the human Smo CRD.	89
Figure 5-5. The effect of the phosphatidylinositols with one phosphate group on Hh signalling.	91
Figure 5-6. The effect of the phosphatidylinositols with two phosphate groups on Hh signalling.	92
Figure 5-7. The effect of the phosphatidylinositols with three phosphate groups on Hh signalling.	93
Figure 6-1. Superimposed structures of the <i>Drosophila</i> and the zebrafish Smo CRD.	96
Figure 6-2. A novel model for Smo allosteric regulation.	99

Figure 6-3. Design of Smoothened inhibitors.....	103
Figure 6-4. Design of Smo inhibitors.	104

LIST OF ABBREVIATIONS

ADC	Analog Digital Converter
AIR	Ambiguous Interaction Restraints
BCC	Basal Cell Carcinoma
BLI	Bio-Layer Interferometry
Boc	Brother of Cdo
Boi	Brother of Ihog
Bud	Budesonide
cAMP	cyclic Adenosine Mono Phosphate
CARA	Computer Aided Resonance Assignment
CD	Circular Dichroism
Cdo	Cell adhesion molecules Downregulated by Oncogenes
Ci	Cubitus interruptus
Cic	Ciclesonide
CK1	Casein Kinase 1
Cos2	Costal 2
COSY	Correlation Spectroscopy
CRD	Cysteine Rich Domain
CSP	Chemical Shift Perturbation
DHh	Desert Hedgehog
Disp	Dispatched
ECD	Extra Cellular Domain
ECL	Extra Cellular Loop
ECLD	Extra Cellular Linker Domain
FA	Fluocinolone Acetonide
FDA	Food and Drug Administration
FID	Free Induction Decay
FT	Fourier Transform
Fu	Fused
FzD	Frizzled
Gas	Growth Arrest Specific
Gli	Glioma associated genes
GPCR	G Protein Coupled Receptor
GRK	G Protein Coupled Receptor Kinase
GSK	Glycogen Synthase Kinase
HADDOCK	High Ambiguity Driven Biomolecular Docking
Hh	Hedgehog
HHIP	Hedgehog Interacting Protein
HSC	Hedgehog Signalling Complex
HSQC	Heteronuclear Single Quantum Coherence
ICL	Intra Cellular Loop
Ihh	Indian Hedgehog
IP3	Inositol-3,4,5- triphosphate
ITC	Isothermal Calorimetry

MB	Medulloblastoma
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect correlation Spectroscopy
PDB	Protein Data Bank
PI(3,4)P2	Phosphatidyl Inositol-3,4-bisphosphate
PI(3,4,5)P3	Phosphatidyl Inositol-3,4,5-triphosphate
PI(3,5)P2	Phosphatidyl Inositol-3,5-bisphosphate
PI(4)P	Phosphatidyl inositol-4-phosphate
PI(4,5)P2	Phosphatidyl inositol-4,5-bisphosphate
PI3P	Phosphatidylinositol-3-phosphate
PKA	Protein Kinase A
PPM	Parts Per Million
Ptc	Patched <i>Drosophila</i>
Ptch	Patched vertebrates
RCSB	Research Collaborator for Structural Bioinformatics
RDC	Residual Dipolar Coupling
RF	Radio Frequency
RMS	Rhabdomyosarcoma
RND	Resistance Nodulation cell Division
SHh	Sonic Hedgehog
SLMB	Slimb
Smo	Smoothed
SPR	Surface Plasmon Resonance
SSA	Super Streptavidin
SSD	Sterol Sensing Domain
STD	Saturation Transfer Difference
SuFu	Suppressor of Fused
TA	Triamcinolone Acetonide
TOCSY	Total Correlation Spectroscopy
Water-LOGSY	Water-ligand Observation with Gradient Spectroscopy
7TM	Seven Transmembrane Domains

CHAPTER 1. INTRODUCTION

1.1. An Overview of the Components of the Hedgehog Signal Transduction Machinery

The Hedgehog (Hh) signal transduction pathway plays an important role during embryogenesis controlling cell growth, proliferation and differentiation[1-3]. In adult tissues Hh signalling has been proposed to play a role in the proliferation and differentiation of adult stem cells[4]. Disruption of Hh signalling during development and the postnatal period leads to a plethora of developmental anomalies such as holoprosencephaly and Pallister-Hall Syndrome[5, 6]. Malfunctioning of the Hh pathway in adult life leads to tumor development and progression specifically in a subset of medulloblastoma (MB), basal cell carcinoma (BCC) and rhabdomyosarcoma (RMS)[7-10]. The core components of the Hh signal transduction cascade have been well conserved from *Drosophila* to humans, although some differences do exist[11, 12]. The G protein coupled receptor (GPCR), Smoothed (Smo), functions as the signal transducer of the Hh pathway and is the primary focus of the present dissertation[13-16]. This section will briefly describe the individual components of the Hh signal transduction cascade.

1.1.1. Hedgehog

The Hh pathway gets its name from the endogenous ligand of the pathway, Hh. The Hh gene was first identified as a segment polarity gene in *Drosophila melanogaster*[1]. The Hh mutations resulted in shorter embryos when compared to the wild-type embryos and exhibited a range of embryonic cuticular defects[1, 17]. The appearance of denticular or hairy larvae inspired the name hedgehog (Hh)[1]. The Hh gene is equally important in vertebrates wherein it has been described to play an important role in regulating cell fate and number in brain and spinal cord which in turn controls the patterning of limbs and internal organs[2, 11].

Hh binds to the Patched (Ptch) receptor and may act as a morphogen, a mitogen or a survival factor[2, 18]. Although the mechanism of action of Hh is conserved from *Drosophila* to vertebrates, *Drosophila* has only one type of Hh whereas vertebrates have three distinct types of Hh: Sonic Hedgehog (SHh), Indian Hedgehog (IHh) and Desert Hedgehog (DHh)[11, 19]. DHh is closely related to the fly Hh, and SHh and IHh are related to each other[20].

The Hh protein is synthesized as a precursor protein and undergoes significant post-translational modifications[11]. It has an amino (N) terminal signalling unit and a carboxyl (C) terminal protease domain[11]. The C-terminal region of the precursor protein is cleaved to release the signalling Hh protein (HhN) which in turn is dual-lipid modified[11]. One of the lipid modifications is palmitoylation and the other is the addition of cholesterol[21-23]. The palmitate modification is on a cysteine residue at the

N-terminus of the protein that is exposed after the signal peptide is cleaved off. The cholesterol modification is on the C-terminus of HhN following autocatalytic cleavage[11]. The schematics depicting the sequential lipid modifications on the Hh protein are shown in **Figure 1-1**. Lipid modifications of Hh have been suggested to aid in Hh oligomerization and form soluble protein complexes responsible for signalling[20, 24-26] Hh is released from the producing cell by the twelve transmembrane protein Dispatched (Disp)[27].

1.1.2. Patched

The patched (*ptc*) gene was also identified in *Drosophila* as a segment polarity gene[28]. In *Drosophila* there is only one type of Ptc whereas in zebrafish, mice and humans two homologs of Patched; Ptch1 and Ptch2 have been reported[2, 29]. Ptch functions as the receptor for the Hh ligand[28, 30, 31]. Unlike other signalling pathways in which receptors act as positive modulators of the pathway, Ptch acts as a negative regulator and inhibits signalling in the absence of the Hh ligand[32]. The binding affinity of Ptch for the Hh ligand is in the low nanomolar range[30, 33].

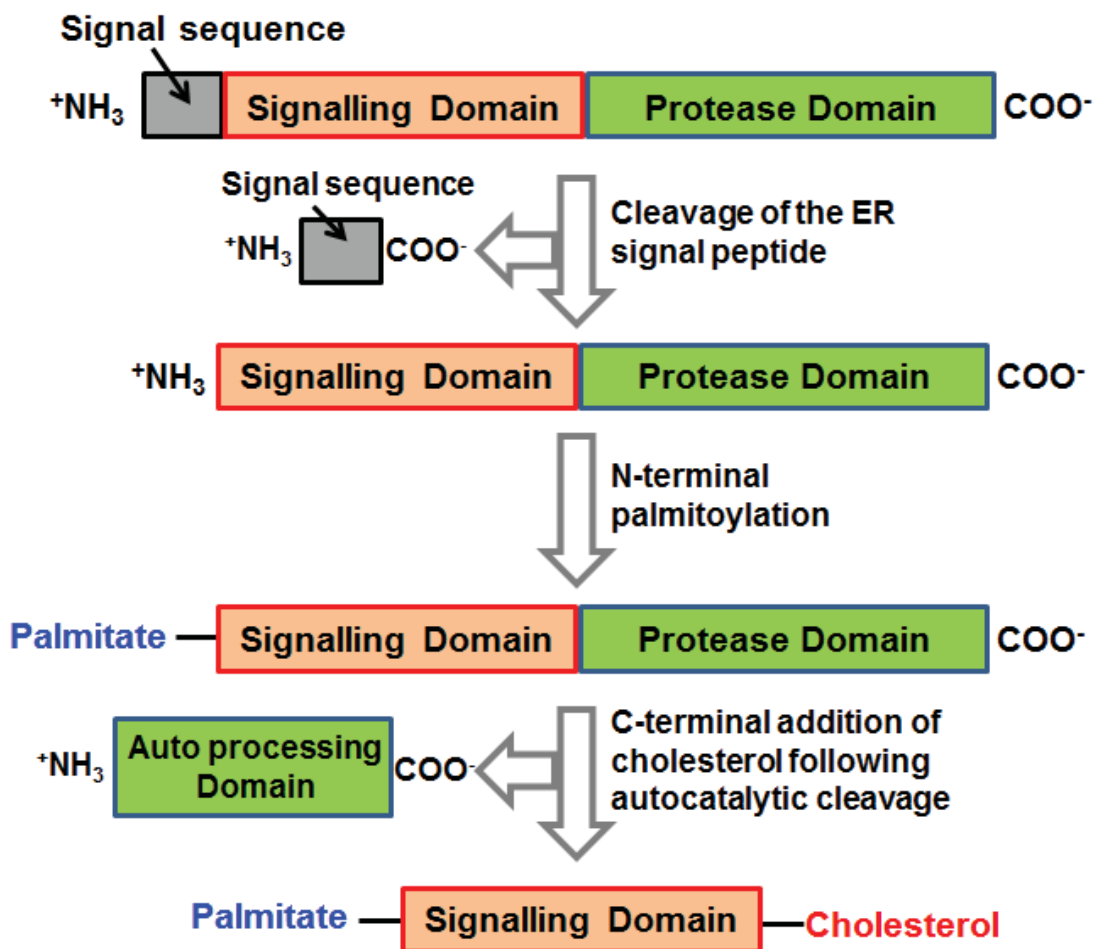
The Ptch receptor is predicted to possess twelve membrane spanning domains and is structurally related to the bacterial resistance-nodulation-cell division (RND) transporters[28, 30, 31]. Both the N- and C-terminus of the Ptch receptor are located in the cytoplasm. It has two large extracellular loops which bind to the Hh ligand[33]. The Ptch receptor structure also includes a sterol sensing domain (SSD) comprised of 180 amino acids and correspond to five consecutive transmembrane helical domains[34]. **Figure 1-2A** illustrates the different domains of the Ptch receptor.

In the absence of the Hh ligand, Ptch inhibits Smoothened (Smo) from functioning and transducing downstream signalling[2, 3]; however the functional details of how Ptch inhibits Smo are still subject to debate. Ptch does not inhibit Smo stoichiometrically because even in the excess of Smo, Ptch can inhibit Smo function[35]. Additionally, coimmunoprecipitation and cell imaging experiments have confirmed that Ptch does not physically associate or colocalize with Smo[36]. Ptch is known to control the subcellular distribution of Smo and thereby control its activity[36, 37]. When Hh binds to Ptch, it loses its constitutive repression on Smo, facilitating Smo translocation to the cell membrane in *Drosophila* and primary cilium in vertebrates to transduce signalling[36, 38].

1.1.3. Hedgehog Co-receptors

The Hh ligand has been reported to bind to other proteins which function as Hh co-receptors. Ihog and brother of Ihog (Boi) were amongst the first identified Hh co-receptors in *Drosophila*[39, 40]. Ihog receptors bind to the ligand Hh and the receptor Ptch[11, 18]. Ihog receptors possess a single transmembrane domain, an N-terminal extracellular and a C-terminal intracellular domain as shown in **Figure 1-2B**[41].

SHh precursor protein



Processed SHh-N protein

Figure 1-1. Schematic representation of the post-translational modifications of the Hh ligand.

The Hh ligand is secreted as an inactive precursor protein. The N-terminal signal sequence is cleaved following the addition of the palmitate moiety to the N-terminus and the cholesterol moiety to the C-terminus to produce the active Hh ligand. Source: Heal WP, Jovanovic B, Bessin S, Wright MH, Magee AI, Tate EW. Bioorthogonal chemical tagging of protein cholesterylation in living cells. *Chemical Communications* 2011,47:4081-4083.

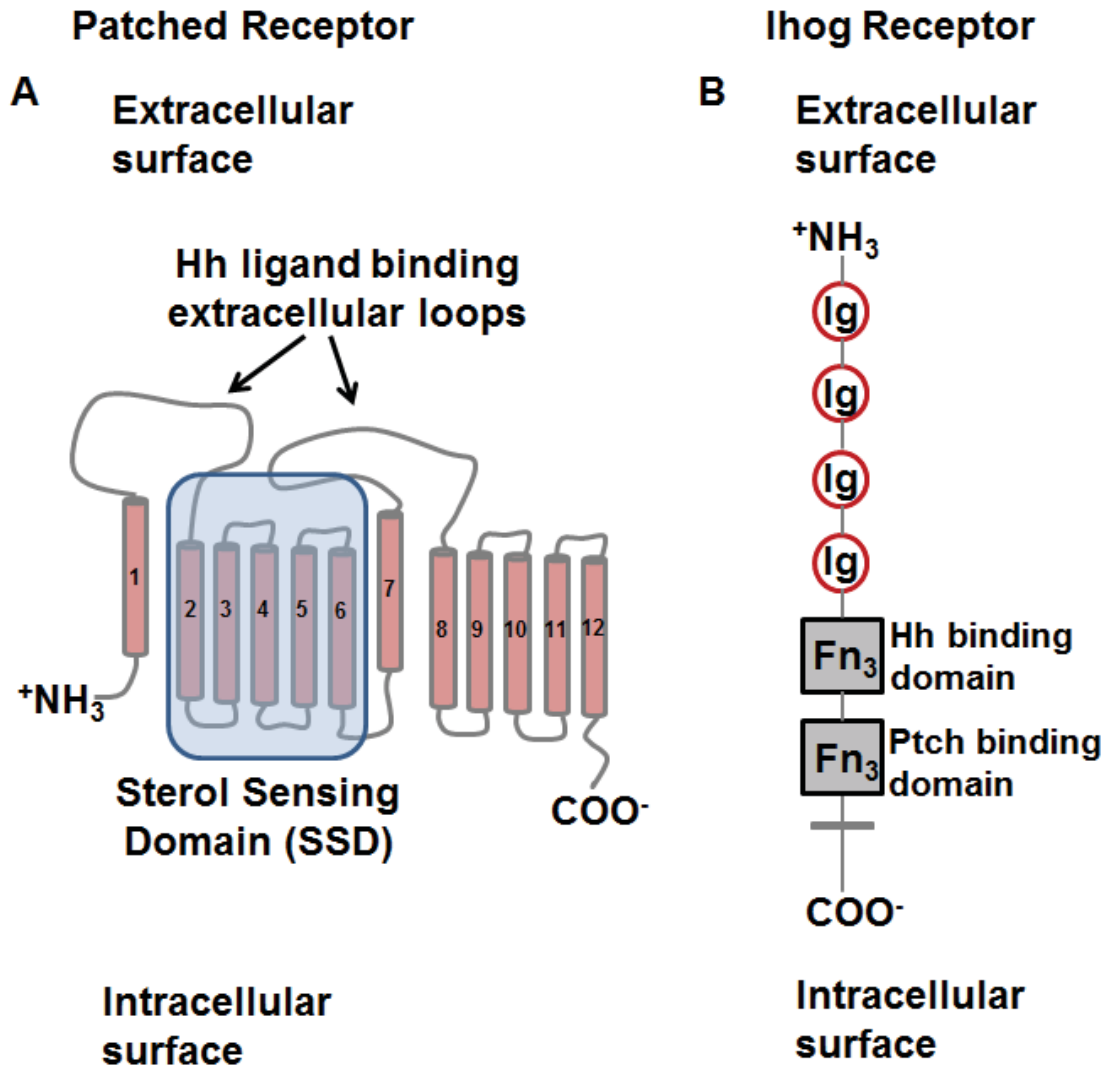


Figure 1-2. Schematic representation of the Hh receptors.

(A) The Ptch receptor possesses twelve membrane spanning domains. The first and the fourth extracellular loops are large and serve as the Hh binding site. Ptch also maintains a sterol sensing domain (SSD) formed by the transmembrane helices two through six. (B) The Hh co-receptors such as Ihog are a single-pass membrane protein. The extra-cellular domain is characterized by four immunoglobulin folds (Ig) and two or three fibronectin domains (Fn3). The first Fn3 domain binds to the Hh ligand and the second Fn3 domain binds to the Ptch receptor.

The extracellular domain (ECD) is characterized by four immunoglobulin (Ig) like domains and two fibronectin type III (FN3) domains in *Drosophila* and three FN3 domains in vertebrates[41]. The domain architecture of the Ihog receptor is shown in **Figure 1-2B**. The first FN3 domain binds to Hh and the second FN3 domain binds to Ptch[41, 42]. The vertebrate orthologs of Ihog are cell adhesion molecules down-regulated by oncogenes (Cdo) and brother of Cdo (Boc)[43].

Heparin in flies and calcium in vertebrates play an important role in stabilizing the interaction between Hh co-receptors and the ligand Hh[44]. Growth arrest specific gene 1 (Gas1) has also been reported to function as an Hh co-receptor in vertebrates[18, 45]. The physiological significance of the Hh co-receptors is underscored by the fact that missense mutations in these receptors were identified in patients suffering from holoprosencephaly[46].

1.1.4. Smoothened

The Smoothened (Smo) receptor is encoded by the *smo* gene which was first identified as a segment polarity gene in *Drosophila*[1]. The Smo receptor is a member of the Frizzled (FzD) class of G protein coupled receptors (GPCR) and functions as the signal transducer of the Hh pathway[13-15]. The aims of this dissertation have been to study the structural and functional significance of the Smo receptor and are discussed elaborately in the subsequent chapters. A brief introduction to some of the intricate details of the Smo receptor will be described in section 1.2.

1.1.5. Hedgehog Signalling Complex

Downstream of Smo, Hh signal is transduced by the Hedgehog Signalling Complex (HSC) which is comprised of the kinesin-like protein Costal2 (Cos2) in *Drosophila* / Kif 7 in vertebrates, the serine/threonine kinase Fused (Fu), Suppressor of Fused (SuFu) and the transcription factor *Cubitus interruptus* (Ci) in *Drosophila* and Glioma associated genes (Gli) in vertebrates[11, 47, 48]. The HSC binds microtubules in an Hh-dependent manner, probably aided by Cos2. Cos2 is a scaffold protein which also recruits other serine/threonine kinases like protein kinase A (PKA), glycogen synthase kinase 3 (GSK3) and casein kinase 1 (CK1)[47, 49]. Sequential phosphorylation of the transcription factor by the above mentioned kinases triggers its ubiquitination and proteolysis to generate a transcriptional repressor[50]. Thus Cos2 functions as a negative regulator of the pathway[11, 47, 49, 50].

On the other hand Fu functions as a positive modulator of the pathway[49]. Translocation of the HSC to the plasma membrane in response to the Hh ligand induces phosphorylation of Cos2 by Fu[11, 18, 51]. Consequently, Cos2 dissociates from the HSC, leading to the accumulation of the active transcription factor[11, 18].

The activity of the transcription factor Ci, is further inhibited by SuFu which is the primary mode of restricting Ci activation when Ci proteolysis is blocked[18, 52, 53]. In *Drosophila* loss of SuFu does not lead to ectopic Hh signalling, suggesting that there are other Ci inhibitory mechanisms present in *Drosophila*[18]. In vertebrates loss of SuFu leads to ectopic Hh signalling suggesting that SuFu is the primary inhibitor of Gli[18, 53, 54]. **Figure 1-3A** summarizes the components constituting the HSC.

1.1.6. Ci/Gli Transcription Factors

The expression of the Hh target genes are controlled by the zinc finger transcription factors Ci and Gli in *Drosophila* and vertebrates respectively[11, 55-57]. In vertebrates there are three different forms of Gli: Gli 1, Gli 2 and Gli 3[57]. The Ci protein is related to the Gli 2 and Gli 3 proteins[11, 58].

The Gli transcription factors share five highly conserved tandem C2H2 zinc fingers and a conserved His-Cys linker between the zinc fingers[18, 59]. The Ci, Gli 2 and Gli 3 proteins include an N-terminal transcription repressor domain and a C-terminal transcription activator domain as shown in **Figure 1-3B**[56, 60, 61]. The full length Ci, Gli 2 and Gli 3 proteins function as transcription activators, but proteolytic degradation of the C-terminus allows these transcription factors to function as transcription repressors[11, 56, 61]. Gli 1 lacks the N-terminal repressor domain and functions only as a transcription activator[11, 18].

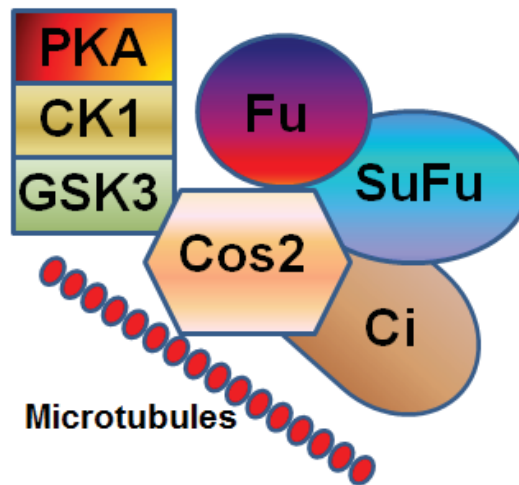
1.2. The Smoothed Receptor

The Smo receptor functions as the signal transducer in the Hh pathway[2, 11, 13, 14]. This section will elaborate on the structural and functional intricacies of the Smo receptor.

1.2.1. Smoothed Structure and Post-translational Modifications

Smo is a member of the FzD class of G protein coupled receptors (GPCRs) and functions as the signal transducer of the Hh pathway[15, 16, 62, 63]. Recently, the structure of the Smo seven transmembrane domains (7TM) and the extracellular linker domain (ECLD) were elucidated by X-ray crystallographic methods[64]. The structure showed that Smo possesses seven membrane spanning helical domains and a short eighth helix packed parallel to the membrane as shown in **Figure 1-4A**. The 7TM domains are connected by three intracellular loops (ICL) and three extracellular loops (ECL). The third extracellular loop (ECL3) is large and extends into the extracellular space[64]. Smo also possesses an N-terminal extracellular cysteine-rich domain (CRD) and a C-terminal intracellular domain that was not included in the crystal structure.

A Hedgehog Signaling Complex



B Gli Transcription Factor

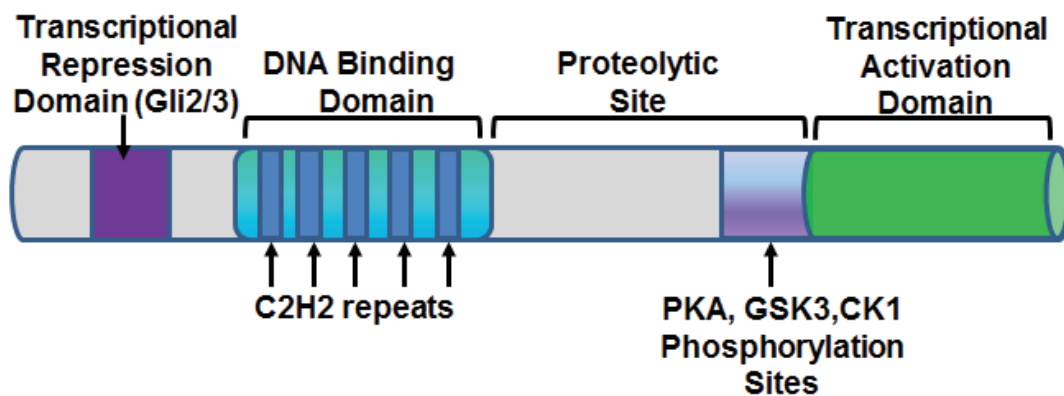


Figure 1-3. Schematics of the Hedgehog Signalling Complex (HSC) and the domain architecture of the Gli transcription factor.

(A) The key components of the HSC include the microtubule-associated protein Costal2 (Cos2), Fused (Fu), Suppressor of Fused (SuFu), the protein kinases: protein kinase A (PKA), casein kinase 1 (CK1), glycogen synthase kinase 3 (GSK3), and the transcription factor Ci/ Gli. (B) Domain architecture of the Gli transcription factor highlighting the main structural features. Source: Briscoe J, Therond PP. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature Reviews Molecular Cell Biology* 2013,14:416-429

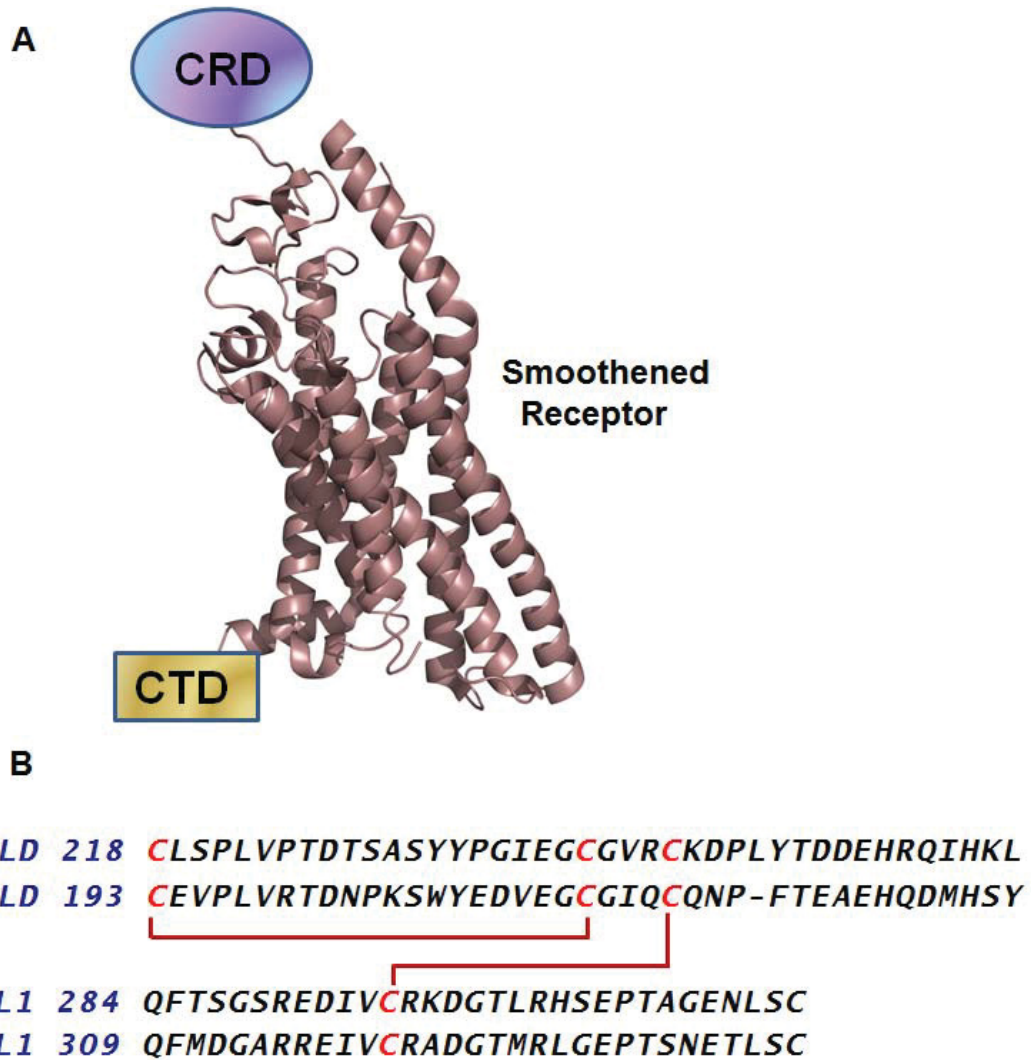


Figure 1-4. The structural domains of the Smo receptor and the disulfide bonds stabilizing the Smo extracellular linker and the extracellular loops.

(A) The structure of the seven transmembrane domains of the Smo receptor (pdb code 4JKV)[64]. The position of the extracellular cysteine rich domain (CRD) and the intracellular cytoplasmic domain (CTD) not included in the crystal structure are indicated. **(B)** Sequence alignment of the *Drosophila* (D) Smo extracellular linker domain (ECLD) and the extracellular loop 1 (ECL1) with the human (H) Smo ECLD and ECL1. The cysteines engaged in disulfide bond formation in human Smo are conserved in *Drosophila*. The red lines indicate the disulfide bond pattern.

The Smo N-terminal CRD is similar to the CRD of the FzD receptors and is connected to the first transmembrane domain by the ECLD[16, 65, 66]. The ECLD has three conserved cysteines; the first two cysteines form a disulfide bond within the ECLD and the third cysteine forms a disulfide bond with a conserved cysteine in the first extracellular loop as shown in **Figure 1-4B**[64]. These disulfide bonds stabilize the Smo extracellular structures close to the 7TM bundles. Additionally the residues in the first extracellular loop form hydrogen bonds with residues in the extracellular linker, further stabilizing the extracellular linker and the extracellular loops in proximity to one another[64].

The Smo cytoplasmic domain is about 241 amino acids longer in *Drosophila* than humans. Both *Drosophila* and vertebrate Smo C-termini contain putative phosphorylation sites and conserved clusters of basic amino acids[67, 68]. Phosphorylation neutralizes the charge on the basic residues and leads to a conformational switch which brings the cytoplasmic loop of the adjacent Smo receptor in close proximity and facilitates downstream signalling[68]. In the presence of Hh, *Drosophila* Smo is phosphorylated by PKA, CK1, CK2 and the G protein coupled receptor kinase (GRK2)[69-72]. The vertebrate protein kinases include CK1 α and GRK2[73].

On the intracellular surface the Smo receptor also possesses conserved tryptophan residues. Tryptophan 535 and W545 are of potential interest. Tryptophan 535 is conserved in the Smo family and is superimposable with the NPXXY motif of class A GPCRs[64]. Mutation of W535L leads to a constitutively active Smo receptor[74]. The W545 is homologous to the tryptophan of the KTXXXW motif of the FzD receptors[64]. In the FzD receptors the conserved KTXXXW motif of helix VIII is essential for the interaction of the FzD receptor and the Dishevelled protein[75]. This interaction is essential for the activation of the Wnt/beta catenin pathway[75, 76]. In Smo the conserved sequence is KATXXXW[64]. The hydroxyl group of the threonine and the indole nitrogen of the tryptophan form a hydrogen bond with residues from the intracellular tip of helix VII and the first intracellular loop respectively[64]. Thus the KATXXXW motif in Smo is essential for maintaining the tertiary fold of the protein; however, like the KTXXXW motif of the FzD receptor, the KATXXXW motif of the Smo receptor may directly bind to a downstream signalling protein or molecule to regulate signalling.

1.2.2. Smoothed as a G Protein Coupled Receptor

The Smo receptor has been classified as a member of the FzD class of GPCRs and shares less than 10% primary sequence identity with class A GPCRs[15, 16, 62, 63]. Smo retains the canonical 7TM helices characteristic of GPCRs and also shares some of the intracellular structural features with class A GPCRs such as the helical turn in the first intracellular loop and an intracellular helix VIII parallel to the cell membrane[62, 64]. On the extracellular surface the beta hairpin structure of the ECL2 is common for Smo and class A GPCRs. The canonical GPCR disulfide linkage between conserved cysteines in the ECL2 and TM3 is also found in Smo receptors[64, 77]. Furthermore Smo can be

phosphorylated by a GPCR kinase (GRK2) and is known to recruit β -arrestin[78]. Additionally, Smo has been shown to recruit G-proteins mainly $G\alpha_i$; further justifying the classification of Smo as a GPCR[15, 79].

Smo lacks some of the canonical class A GPCR features, such as the D/ERY motif in helix III[62, 64]. Conserved prolines which play an important role in receptor activation in helix V, VI, VII of class A GPCRs are absent in Smo. The Smo receptor has conserved glycines in helices V, VI and VII which may provide flexibility to the receptor allowing it to carry out structural rearrangements during receptor activation and signalling[64]. The extracellular loops of Smo are long compared to class A GPCRS and the extracellular CRD is unique to the FzD group of GPCRs[62, 64].

1.2.3. Smoothened Signalling and Regulation by Small Molecules

Smoothened activation is controlled by Ptch and involves three distinct phenomena: stabilization, conformational change and membrane localization[80, 81]. When Hh binds to Ptch it is internalized and degraded, leading to Smo stabilization and translocation from the endocytic vesicles to the cell membrane in *Drosophila* and primary cilium in vertebrates to activate downstream signalling[38, 68]. Thus Smo functions as the obligatory signal transducer of the Hh pathway located at the junction of the extracellular and intracellular events. Despite the central role played by Smo in mediating Hh signalling, the mechanism by which Smo activation is regulated and coupled to downstream signalling remains elusive. It has been suggested that Smo regulation may involve endogenous small molecules[80].

Various reports have shown that Smo activity can be regulated by a wide variety of small molecules[82, 83]. Cyclopamine, a plant alkaloid was the first described Smo inhibitor known to bind in the cavity of the 7TM domains. It is a natural product that can be isolated from corn lily and has anti-cancer effects[84-86]. Since the identification of cyclopamine, cell-based screens directed towards the Hh pathway have identified small molecules that directly bind Smo. The identified small molecules include agonists like 3-chloro-N-[4-(methylamino) cyclohexyl]-N-[3-(pyridin-4-yl) benzyl] benzo[b] thiopene-2-carboxamide (SAG) and purmorphamine and antagonists such as N-[(1E)-(3,5-dimethyl-1-phenyl-1H-pyrazol-4-yl) methylidene]-4-(phenylmethyl)-1-piperazinamine (SANT) and GDC-0449 molecules[83, 87-89]. All of these small molecules have been described to bind in the same site as cyclopamine[83, 85, 87, 90]. Presently, the only Food and Drug Administration (FDA)-approved Smo inhibitor is GDC-0449 which also docks in the 7TM domains[64, 91]. The recently elucidated crystal structure of the Smo 7TM domains in complex with an anti-tumor agent showed that the antitumor molecule docks in the long and narrow cavity within the 7TM helices making contacts with residues of helix I, II, V and VII[64]. The ECLD and the ECL form the orifice of the cavity and contact the antitumor agent[64].

In recent times oxysterols and glucocorticoids have been identified as two major groups of small molecules which bind to a site distinct from the cyclopamine binding site

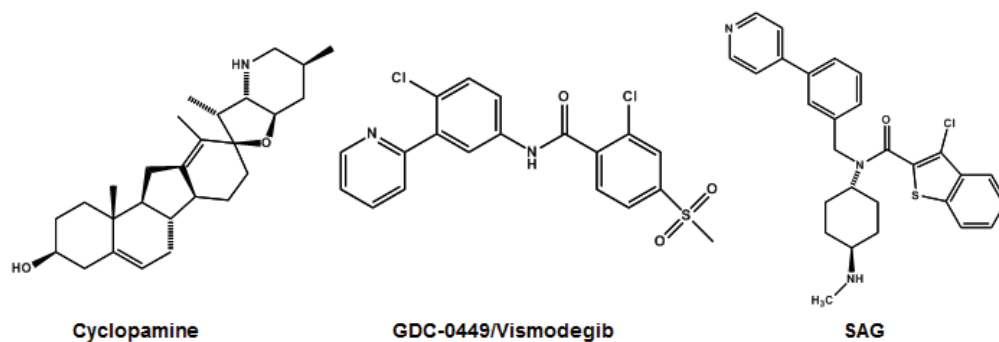
in the cavity of the 7TM domains[92-94]. Pharmacological and biochemical studies have shown that 20 (S)-hydroxyl cholesterol (20(S)-OHC) acts as an allosteric activator of vertebrate Hh signalling, functioning at the level of Smo[92]. Additional cholesterol derivatives like 7-keto-25-OHC and 7-keto-27-OHC have also been reported to regulate Smo activity[95]. The length of the carbon chain attached to the steroid moiety is reported to be six or eight carbon atoms long for maximal Hh pathway activation by oxysterols[92, 96]. Mutation experiments have suggested that an intact extracellular Smo CRD is required for oxysterols to bind to Smo[92, 95].

Glucocorticoids were identified as Smo modulators in high-content cell-based screening experiments[93]. They could be categorized into two groups: the activatory glucocorticoids and the inhibitory glucocorticoids. As the name suggests activatory glucocorticoids like fluocinolone acetonide (FA) and triamcinolone acetonide (TA) promoted Smo ciliary translocation and downstream Hh signalling[93]. The inhibitory glucocorticoids such as Budesonide (Bud) repressed Smo ciliary translocation and downstream Hh signalling[93]. Interestingly, although the activatory glucocorticoids were shown to bind in the cavity of the Smo 7TM domains by cyclopamine binding competition assays, the inhibitory glucocorticoids were suggested to bind to a site distinct from the cyclopamine binding site in the 7TM domains[93].

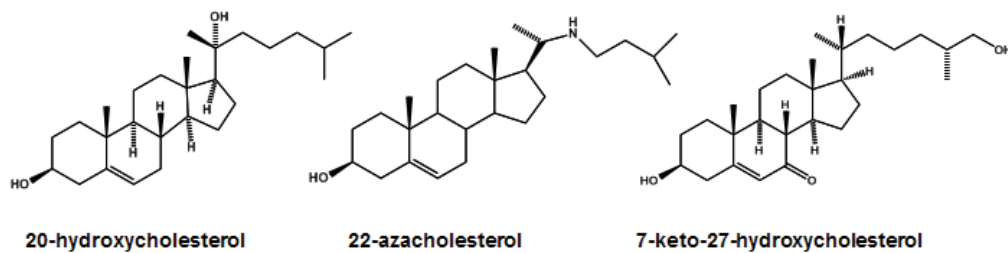
Most importantly the inhibitory glucocorticoid Bud acted on wild-type as well drug resistant Smo: SmoM2 and Smo D473H[93]. SmoM2 is a constitutively active form of Smo isolated from BCC patients[74]. There is a mutation, W535L, in the seventh transmembrane domain which may interfere with the interaction between the Smo receptor and the G-proteins. Smo D473H was isolated from MB patients treated with the FDA-approved drug GDC-0049[97]. This mutation is localized in the sixth transmembrane domain towards the extracellular surface and may play a role in small molecule binding. The occurrence of this mutation was accompanied by resistance to the drug[97]. Thus these Smo mutants were not responsive to previously identified Smo antagonists.

Most of the current literature asserts that mammalian Smo is responsive to 20-OHC and glucocorticoids whereas *Drosophila* Smo is cited as nonresponsive to these small molecules[92, 95, 96]; however, in *Drosophila*, the phospholipid, phosphatidylinositol-4 phosphate (PI4P), has been reported to regulate Smo activity[98]. The loss of the kinase (STT4) responsible for the synthesis of PI4P results in the Hh loss-of-function phenotypes. The loss of the phosphatase (Sac1) responsible for the degradation of PI4P results in the Hh gain-of-function phenotypes[98]. PI4P kinase depletion in a mouse fibroblast cell line also impairs activation of the Gli-luciferase reporter gene by SHh, implying that phospholipids play a role in regulating vertebrate Hh signalling[98].

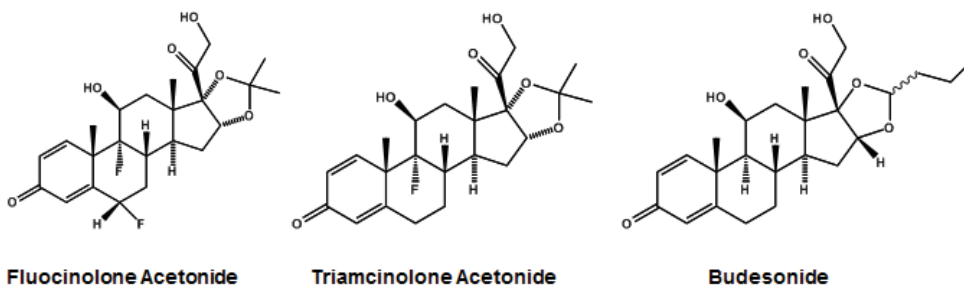
Clearly Smo activity is susceptible to regulation by more than one distinct class of molecules as shown in **Figure 1-5**. These small molecules may bind in the same or different site on Smo to regulate Hh signalling.



Cholesterol derivatives



Glucocorticoids



Phospholipids

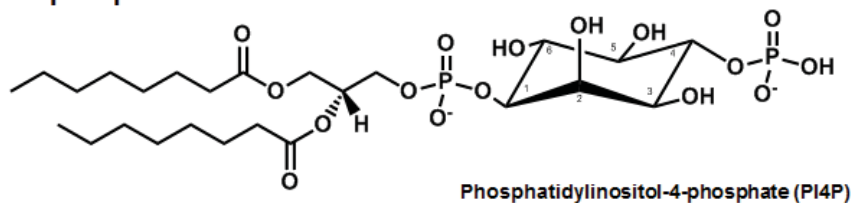


Figure 1-5. Chemical structure of selected Smo modulators.

Smo activity is regulated by small molecules such as cyclopamine, GDC-0449, SAG, cholesterol derivatives, glucocorticoids and phospholipids.

1.3. The Hedgehog Signal Transduction Cascade

The Hh signal transduction cascade is switched on when the endogenous Hh ligand binds to the large extracellular loops of the receptor protein Ptch assisted by the Hh co-receptors[28, 30, 33]. Hh binding to Ptch renders Ptch inactive and trapped in endocytic vesicles. Consequently Ptch loses its housekeeping inhibition on the Smo receptor[36]. The Smo receptor then translocates from the endocytic vesicles to the primary cilium in vertebrates or the cell membrane in *Drosophila*[36, 38, 99]. Thus Smo functions as the signal transducer of the pathway located at the nexus of the extracellular and intracellular events.

Smoothed in the plasma membrane is phosphorylated by the protein kinases PKA, CK1 and GPRK2[69-72]. Phosphorylation triggers a conformational switch in the cytoplasmic domains of the Smo receptor triggering downstream Hh signalling[68]. The Smo protein interacts with the kinesin-like protein Cos2[100, 101]. The Cos2 protein is one of the components of the HSC; the other members of the HSC include the serine/threonine kinase Fu and the transcription factor Ci/Gli[47-49]. The HSC is a cytoplasmic multi-protein complex responsible for transmitting the Hh signal from the Smo receptor to the Ci/Gli transcription factors[11, 18, 57]. The vertebrate analogue of Cos2 is Kif7[11, 12]. The Cos2 protein in turn is phosphorylated by the serine/threonine kinase Fu causing the release of the Ci/Gli transcription factor from the HSC[51, 102]. The Ci/Gli transcription factor accumulates in its full length form, and translocates to the nucleus to activate the transcription of the Hh target genes[11, 18]. Some of the Hh target genes include Cyclin D, Cyclin E, n-myc and patched[11, 55].

The Hh signal transduction is switched off in the absence of the Hh ligand as shown in **Figure 1-6A** and switched on in the presence of the Hh ligand as shown in **Figure 1-6B**. The Ptch receptor constitutively inhibits the Smo receptor. As a result the Smo receptor is trapped in the endocytic vesicles[36]. The downstream transcription factor Ci is associated with the scaffold protein Cos2[11, 18, 47, 48]. Costal 2 additionally recruits PKA, GSK3 and CK1[50]. These kinases phosphorylate the full length Ci/Gli transcription factors[11, 18]. The phosphorylated transcription factor serves as a signal to recruit the F-box protein Slimb (SLMB). SLMB catalyses C-terminal ubiquitination and successive proteolytic degradation of the transcription factor to generate a transcriptional repressor[103, 104]. The transcriptional repressor dissociates from the HSC, translocates to the nucleus and represses the transcription of the Hh target genes[11, 18].

The Hh signal is also known to be transduced by the Smo receptor via G-proteins[15, 79, 105]. Smoothed has specificity towards the $G\alpha_i$ protein. Recruitment of the $G\alpha_i$ protein by Smo in response to the Hh ligand decreases the intracellular concentrations of cyclic adenosine monophosphate (cAMP)[18]. Decreased cAMP leads to the inactivation of the protein kinase: PKA. The protein kinase, PKA, functions to phosphorylate and degrade the transcription factor to terminate Hh signalling. Thus inactivation of protein kinase A activates Hh signalling[18].

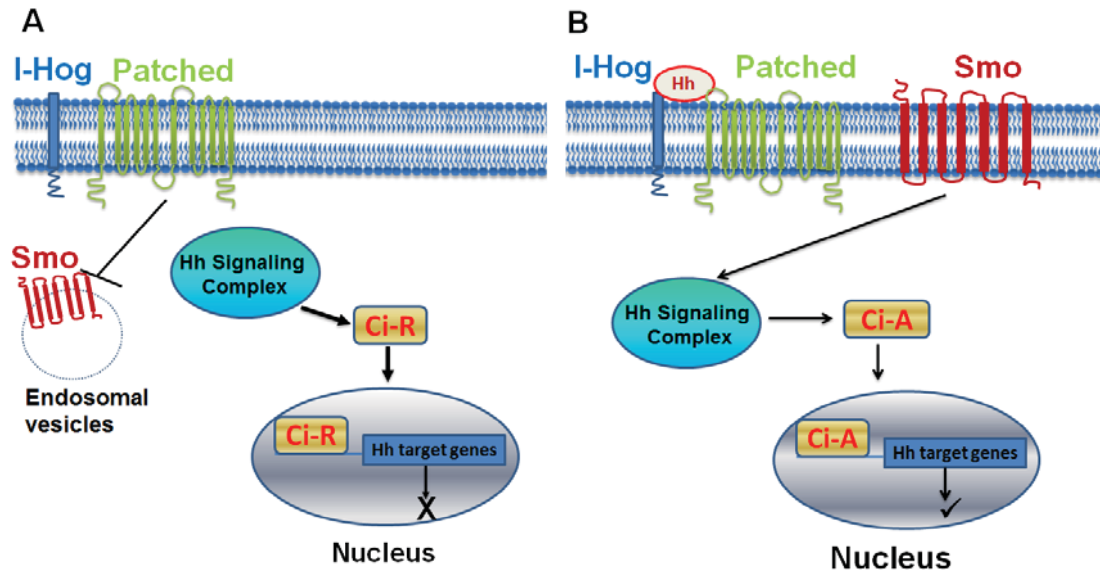


Figure 1-6. The Hedgehog Signal Transduction Cascade.

(A) In the absence of the Hh ligand, the Ptch receptor inhibits Smo; consequently Smo is trapped in the endocytic vesicles. The downstream Hh signalling complex (HSC) renders the transcription factor Ci inactive, which, although translocates to the nucleus, fails to induce the transcription of the Hh target genes. (B) In the presence of the Hh ligand, Ptch loses its constitutive repression on Smo; Smo translocates to the cell membrane in *Drosophila* or primary cilium in vertebrates. Thereby the HSC maintains Ci in its full length form which translocates to the nucleus to induce the expression of the Hh target genes.

Smoothened signalling is reported to be regulated by cytosolic β -arrestin proteins and G-protein-coupled receptor kinase, GRK[78]. The GRK proteins phosphorylate the C-terminal serine/threonine residues of the Smo receptor which serves as a signal to recruit β -arrestin[106]. Beta-arrestin bound to the Smo receptor leads to receptor desensitization and clathrin-mediated receptor internalization[78]. It has also been reported that β -arrestins together with the Kif proteins may aid the ciliary transportation of the Smo receptor to initiate downstream Hh signalling[107, 108].

1.4. Implications of the Hedgehog Pathway in Cancer

In adults Hh signalling is associated with maintenance of tissue homeostasis, cellular repair and regeneration, and continence of stem cells[2]. Abnormal Hh signalling in adults has been associated with tumorigenesis and cancer and was first reported in patients suffering from Gorlin's syndrome[10]. Gorlin's syndrome is caused due to a loss-of- function mutation in one of the *patched* genes and the phenotypes include BCC, MB and RMS[109, 110].

Depending on the mechanism of action the cancers associated with malfunctioning of the Hh pathway can be categorized under three major types[111, 112]. The first type (type I) is Hh ligand independent and results from inactivating mutations in Ptch (85%) or activating mutations in Smo (10%)[74, 113]. In the second type (Type II) of the Hh pathway cancers such as tumors of the lung, digestive tract and breast are not associated with oncogenic mutations in the pathway components rather they act in a ligand-dependent autocrine manner[10, 114]. The third type (Type III) of the Hh pathway- associated cancers are ligand dependent and act in a paracrine manner to activate signalling in the tumor microenvironment[115]. Ectopic Hh signalling in cancer may be associated with either the initiation of the tumor as in the case of BCC and MB, or the growth of the tumor as observed in the cancers of the digestive tract[10]. In a certain group of cancers such as lymphoma, although abnormal Hh signalling has been suggested, the exact role of Hh signalling has not been defined.

1.4.1. Basal Cell Carcinoma

Basal cell carcinoma (BCC) is the most common type of skin cancer in the United States. Ectopic Hh signalling plays a pivotal role in BCC and was first reported due to a mutation in the Ptch 1 gene located on chromosome 9[7, 109, 116]. In sporadic BCC 90% of cases are due to loss-of-function mutations in one of the *ptch* alleles, and in 10% of cases are due to gain- of-function mutations in *smo* gene[74, 110, 117]. Harmful UV radiations may have been the cause of the mutations.

In most cases BCC is not associated with metastasis, however if the tumor metastasizes it can be deadly[118]. BCC can be treated either by the surgical removal of the tumor or by the use of targeted therapeutics[118]. The natural product steroidal alkaloid cyclopamine was first used as a topical application, and since then several other

cyclopamine derivatives have been developed to enhance potency[119]. The Smo inhibitor GDC-0449 developed by Genentech-Curis is the first FDA-approved drug used for the treatment of BCC and acts on local as well as metastasized tumors[120].

1.4.2. Medulloblastoma

Medulloblastoma (MB) is a common childhood tumor of cerebellar origin[121]. Each year 400 children are diagnosed with MB in the United States, making it the most common pediatric brain tumor. Four distinct molecular sub-types of MB have been described on the basis of transcriptional profiles and aberrant Hh signalling has been reported in one of the subtypes, categorized as the Hh subtype of MB[122]. The first evidence of ectopic Hh signalling in MB was established when mutations in the *ptch* gene were identified in patients with Gorlin's syndrome[123].

Radio-therapy and chemical therapeutics have been commonly used for the treatment of MB; however such treatment is often associated with severe side effects including intellectual deterioration. The Smo antagonist GDC-0449 caused rapid regression of tumor and reduction of symptoms in MB patients; however, the effectiveness was limited due to the relapse of the tumor with a D473H resistant mutation in the sixth transmembrane helix of the Smo receptor[97, 124]. Therefore, second generation drugs are needed for the treatment of the Hh subtype of MB. These drugs should be effective against the resistant mutants and associated with minimal side-effects.

1.4.3. Rhabdomyosarcoma

Rhabdomyosarcoma (RMS) is the most common type of soft tissue sarcoma in children, accounting for 10% of solid pediatric malignancies[10]. RMS is mostly sporadic but is sometimes associated with Gorlin's syndrome[125]. Uncontrolled Hh signalling such as occurs after the loss of *Ptch* may lead to RMS[125]. Overexpression of Hh signalling also correlates with tumor proliferation in RMS. The metastatic property of the RMS cells can be suppressed by treatment with Hh pathway inhibitors[10].

1.4.4. Targeting the Hedgehog Pathway to Treat Cancers

Aberrant Hh signalling has been reported in a wide variety of cancer conditions and thus this pathway has emerged as a lucrative target for the development of drugs[7-10]. Abnormal Hh signalling reported in cancer may or may not be associated with oncogenic mutations[10]. For example, the type II and type III of the Hh-driven cancers do not harbor oncogenic mutations in any of its components. They are Hh ligand dependent and therefore can be treated by Hh antagonists[112]. In type I Hh tumors such as BCC, MB and RMS, which are reported to possess oncogenic mutations in *Ptch* or *Smo*, they can be treated with therapeutics acting at the level or downstream of the oncogenic mutations.

Theoretically each step of the Hh signal transduction cascade can be targeted to inhibit ectopic signalling; however Smo has emerged as the best defined target for drug development[126, 127]. Cyclopamine was the first identified small molecule which acts as an antagonist of Hh signalling by binding to the Smo receptor[84, 85, 119]. Subsequently a plethora of cyclopamine derivatives with enhanced bioavailability and potency have been reported to function as Hh pathway antagonists[128]. GDC-0449 is the only FDA-approved drug for the treatment of BCC[88, 89, 91]. Unfortunately resistance to this drug has already been reported and therefore the need arises to develop second generation drugs[97]. Smo inhibitors such as sonidegib (LDE225, Novartis), BMS-833923 and saridegib are also effective in the treatment of BCC and may be effective against the resistant mutants[126]. Smo inhibitors are also being tested on a variety of ligand-dependent tumors.

Targeting the Hh signalling pathway to treat cancers has been ongoing for years. Almost every step in the signal transduction has been targeted to inhibit ectopic signalling. Although Smo has emerged as the major drug target of the pathway, other candidates are also being analyzed as potential targets for the development of therapeutics. The monoclonal antibody 5E1 prevents the binding of Hh to Ptch and has demonstrated decent preclinical activity[112]. Hedgehog interacting protein (HHIP) mimics, SUFU mimics and Gli inhibitors like the GANT group of small molecules also function to ablate ectopic Hh signalling and are in various phases of development[129].

1.4.5. Promoting the Smoothened Cysteine Rich Domain as a Novel Target for Drug Development

Clearly Smo plays an important role in the Hh signal transduction cascade and therefore has emerged as a target for the development of drugs to inhibit ectopic Hh signalling associated with cancer conditions[126]. It is not surprising that Smo has been widely used as a target for drug discovery because GPCRs represent the core of modern medicine and Smo is a member of the FzD class of GPCRs[16, 62, 63]. Most of the characterized drugs such as GDC-0449 bind in the cavity of the 7TM domains[64]; however due to the occurrence of resistant mutations on treatment with such Smo inhibitors the need arises to develop new drugs[97]. The novel drugs may or may not be targeted towards the same binding site.

In such regards we hypothesized that the extracellular Smo CRD may function as a novel site for the development of next generation drugs effective against resistant mutants. The Smo CRD is homologous to the FzD CRD which binds to the endogenous Wnt ligand and activates downstream Wnt signalling[65, 66]. We speculated that like the FzD CRD the Smo CRD also possesses a well-defined structure and may play a role in binding to small molecules. Therefore to test our hypothesis we initiated structure-function studies of the Smo receptor, focusing primarily of the extracellular CRD. The following chapters discuss our findings with an emphasis on the structure of the Smo CRD and its role in binding small molecules to regulate downstream Hh signalling.

CHAPTER 2. METHODS

2.1. Introduction

Nuclear Magnetic Resonance (NMR) as the name indicates describes the ‘resonating’ behavior of certain ‘nucleus’ (possessing a property called spin) when placed in an external ‘magnetic’ field. NMR spectroscopy utilizes the phenomena of NMR to study physical and chemical properties of biomolecules. In the present day scenario NMR has a wide range of applications; it can be used to determine the structure of molecules, probe for interactions between various molecules and provide insights into the dynamic or the kinetic behavior of the molecules. The molecules resolved by NMR may be small organic molecules, natural products, peptides or proteins of several tens of kilo-daltons in molecular weight[130]. Using NMR the molecules can be studied in solution mimicking the physiological environment as well as in solid state. In this dissertation we will focus on solution NMR spectroscopy.

High resolution solution NMR spectroscopy provides insights into the structure of biomolecules at atomic resolution. The theoretical principle of NMR was first proposed by Pauli in 1924. He stated that certain atomic nuclei possess the property of spin and magnetic moment and when such nuclei are placed in an external magnetic field it leads to splitting of their energy levels. Historically the phenomenon of NMR was first described by Isidor Rabi in the year 1938. He measured the NMR in molecular beams and was awarded the Nobel Prize in Physics for his work in 1944. The groups of Felix Bloch from Stanford University and Edward Mills Purcell from the Harvard University advanced the technique of NMR for use in solids and liquids and jointly shared the Nobel Prize in 1952. Richard R. Ernst developed the Fourier transform (FT) NMR spectroscopy and received the Nobel Prize in 1991 for his phenomenal discovery. Kurt Wuthrich expanded the technique of NMR spectroscopy for determining the three dimensional structure of biomolecules such as proteins in solution and was bestowed with the Nobel Prize in 2002. Paul Lauterbur and Peter Mansfield were awarded the Nobel Prize for extending the use of NMR for biomedical imaging defined as Magnetic Resonance Imaging (MRI).

The technique of NMR provides molecular information at atomic resolution. The advantage of NMR over other complementary methods such as X-ray crystallography and mass spectrometry is that the samples used in NMR experiments is not destroyed and can be recovered after the measurement. Using NMR the molecules can be studied in solution mimicking the physiological environment as well as in solid state. Additionally NMR can also be used to study biological fluids such as blood and urine. In recent times the technique of NMR is also being utilized to analyze the quality of commercial commodities such as wine. In this dissertation solution NMR methodology was used to determine the structure of a protein domain and therefore will be discussed in detail.

Despite, the widespread use of the NMR methodology, there are some inherent limitations associated with this technique. The resolution derived from NMR experiments

is often considered to be lower than that derived from X-ray crystallography methods. Additionally there is a limitation to the size of the molecule that can be studied by solution NMR. As the size of the molecule increases it experiences a slower tumbling rate in the magnetic field leading to faster decay of signals which become difficult to detect. The data recorded in an NMR experiment is a mixed average of numerous individual signals from the biomolecular sample. Therefore with an increase in the size of the molecule the NMR signals overlap making unambiguous resonance assignments challenging[130, 131]. Consequently NMR derived data are convoluted and seek time and effort to interpret. Accordingly, the time taken to resolve a three dimensional structure by NMR is typically longer than the time taken to elucidate the structure by X-ray crystallography. The major benefit of NMR is that it provides information on biomolecules in solution as opposed to experimentally packed crystal conditions as is the case with X-ray crystallography methods.

2.2. The Physics and Mathematics behind Nuclear Magnetic Resonance

The samples studied by NMR are composed of molecules. The molecules in turn are composed of atoms. The atoms are comprised of electrons, protons and neutrons. The electrons occupy the electronic orbitals surrounding the nucleus of an atom whereas the protons and the neutrons are present in the nucleus of an atom. The subatomic particles comprising the electrons, protons and the neutrons are considered spinning on their axes. Spin is one of the fundamental properties of matter, appears in multiples of $\frac{1}{2}$ and may be positive or negative. Two or more particles with opposite spins may pair up to negate the observable effect of spin. In atoms the number of protons and neutrons may be even or odd. In case of atoms like ^{12}C and ^{16}O the number of protons and neutrons are both even such that they pair against each other and the nucleus of the atom have no net spin. Such nucleus cannot be detected by NMR. In case of atoms like ^1H , ^{15}N and ^{13}C the number of neutrons and/or the number of protons are odd such that the nucleus of the atom possesses a net half-integer spin which can be measured by NMR.

Atomic particles possessing the property of spin behave like small magnets with a north and south-pole and are described as spin vectors. The nuclei with the two spin states possess a magnetic moment (\hat{u}), which can be expressed as shown in **Equation 2-1**.

$$\hat{u} = \gamma h/4\pi \qquad \text{Equation 2-1}$$

Where h is the Planck's constant ($h=6.626 \times 10^{-34}$ Js), and γ is the gyromagnetic ratio of the particular nuclei.

Thus the magnitude of the magnetic momentum depends on the value of the gyromagnetic ratio. Since different nuclei have different gyromagnetic ratio they experience different magnetic moment. Magnetic moment is a vector quantity and defines the strength as well as the orientation of the nuclear magnet.

In the absence of any external magnetic field the small spin vectors with specific nuclear moments are randomly distributed negating each other's effect as represented in **Figure 2-1**. Consequently the net magnetic field is zero; however when placed in an external magnetic field of strength B_0 the spin vector aligns with the external field. The spin vectors may be aligned in the direction of the external magnetic field (N_α) or against the direction of the external magnetic field (N_β) giving rise to two different energy states. There is always an excess of nuclei in α (aligned with the external magnetic field, B_0) spin states than in the β (aligned against the external magnetic field, B_0) spin states, as elucidated in **Figure 2-1**. The ratio between the populations in the two spin states (N_α and N_β) can be described by the Boltzmann's equation as shown in **Equation 2-2**.

$$N_\alpha/N_\beta = e^{\Delta E/kT} \quad \text{Equation 2-2}$$

Where ΔE is the energy difference between the two spin states, T is the temperature (degree kelvin) and k ($k=1.3805 \times 10^{-23}$) is the Boltzmann's constant.

When placed in an external magnetic field of strength B_0 a particle with net spin can absorb a photon of frequency ν . **Equation 2-3** depicts the relationship between the frequency of the photon, the gyromagnetic ratio and the external magnetic field.

$$\nu = \gamma B_0 \quad \text{Equation 2-3}$$

Where γ is the gyromagnetic ratio of the particle and B_0 is the field strength of the external magnetic field.

Upon irradiation with the radio frequency (RF) radiation, energy is absorbed; causing the transition between α and β spin states. The spin vectors can undergo a transition between the two energy states by absorbing a photon of energy E as described in **Equation 2-4**.

$$E = h\nu \quad \text{Equation 2-4}$$

Therefore combining **Equation 2-3** and **Equation 2-4**, **Equation 2-5** is generated.

$$E = h\gamma B_0 \quad \text{Equation 2-5}$$

As stated earlier h is the Planck's constant ($h=6.626 \times 10^{-34}$ Js) and ν is the resonance frequency also known as the Larmor frequency of the spin.

Thus the amount of energy absorbed by a nucleus depends on several factors such as the strength of the external magnetic field. Stronger magnets induce larger α/β population differences as shown in **Figure 2-2**. The other factor that affects the α/β population differences is the gyromagnetic ratio of the nucleus in observation. Nucleus with greater gyromagnetic ratio absorbs and or emits more energy leading to intense absorbance signals.

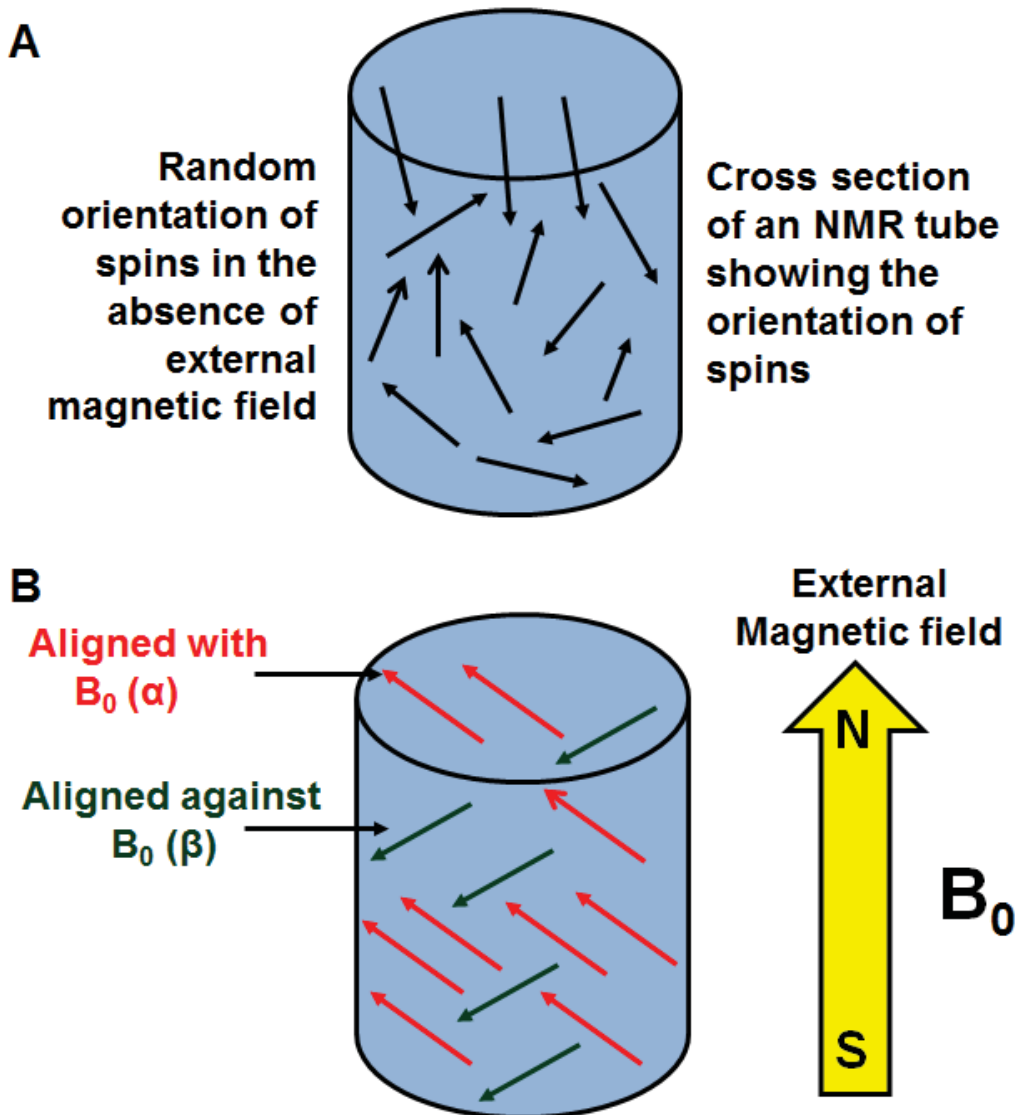


Figure 2-1. Behavior of nuclear spins in the absence and presence of an external magnetic field.

(A) In the absence of an external magnetic field the nuclear spins are randomly distributed and the resultant local magnetic field is zero. (B) In the presence of an external magnetic field the nuclear spins orient either in the direction of the magnetic field (α) or against the direction of the magnetic field (β). The population of the spins aligned in the direction of the magnetic field is usually in excess of the nuclear spins aligned against the external magnetic field.

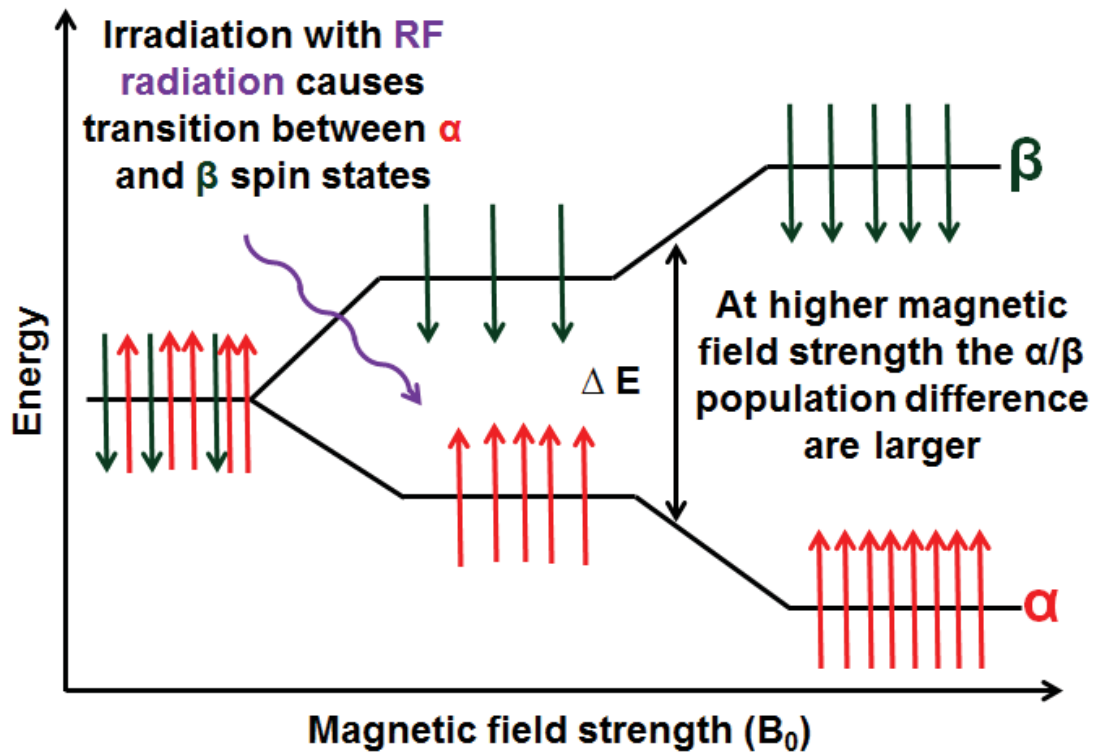


Figure 2-2. Magnetic field strength and NMR sensitivity.

Irradiation with RF radiation causes the nuclear spins to absorb energy and transit between the α and β spin states. Stronger magnets induce larger α/β population differences.

When the spins are irradiated by an orthogonal RF field, B_1 , close to the Larmor frequency of the nucleus under observation they are flipped down to the plane perpendicular to B_0 . The flipped vector precesses around the B_1 field while it relaxes back to its original position. During the process of relaxation the spinning vector generates electromagnetic radiation which induces a current in the coils located in the perpendicular plane. Plotting the induced current as a function of time generates a wave which decays with time. This is known as the free induction decay (FID). This time domain data can be converted to frequency domain data by FT.

The NMR data recorded is a set of peaks and each peak corresponds to a specific frequency. The resultant frequency depends on the strength of the external magnetic field. Since the external magnetic field varies with different spectrometers it would be difficult to compare NMR data obtained from different spectrometers. Therefore an internal reference such as tetramethylsilane is used to normalize the frequencies. The normalized frequency is defined as the chemical shift (δ). The formula for calculating the chemical shift is listed in **Equation 2-6**.

$$\delta \text{ (ppm)} = (\nu - \nu_{ref})/\nu_{ref} \times 10^6 \qquad \text{Equation 2-6}$$

Where ν is the frequency of the nuclei under observation and ν_{ref} is the frequency of the reference.

Thus as chemical shift (δ) is a ratio of both frequencies; the magnetic field dependence cancels out. This ratio specifies the position of the peak in a magnetic field independent way. The measurement of chemical shifts and assigning the chemical shifts to the specific nuclei within the molecular structure of the biomolecule forms the foundation of NMR spectroscopic methods.

2.3. The Nuclear Magnetic Resonance Instrumentation

The measurement of NMR spectroscopic data requires an NMR spectrometer. The NMR spectrometers are named based on the resonance frequency of the protons in the magnetic field generated by the given magnet. For example an NMR spectrometer named as 800MHz causes the protons in the sample to resonate at a frequency of 800MHz. The major components of the NMR spectrometer and their function are briefly described as follows. A schematic representation of the major components of the NMR spectrometer is illustrated in **Figure 2-3**.

2.3.1. Magnet

The NMR magnet is the most important and the most expensive part of the NMR spectrometer. The superconducting NMR magnets are capable of producing magnetic field strengths ranging from 6 to 23.5 Tesla.

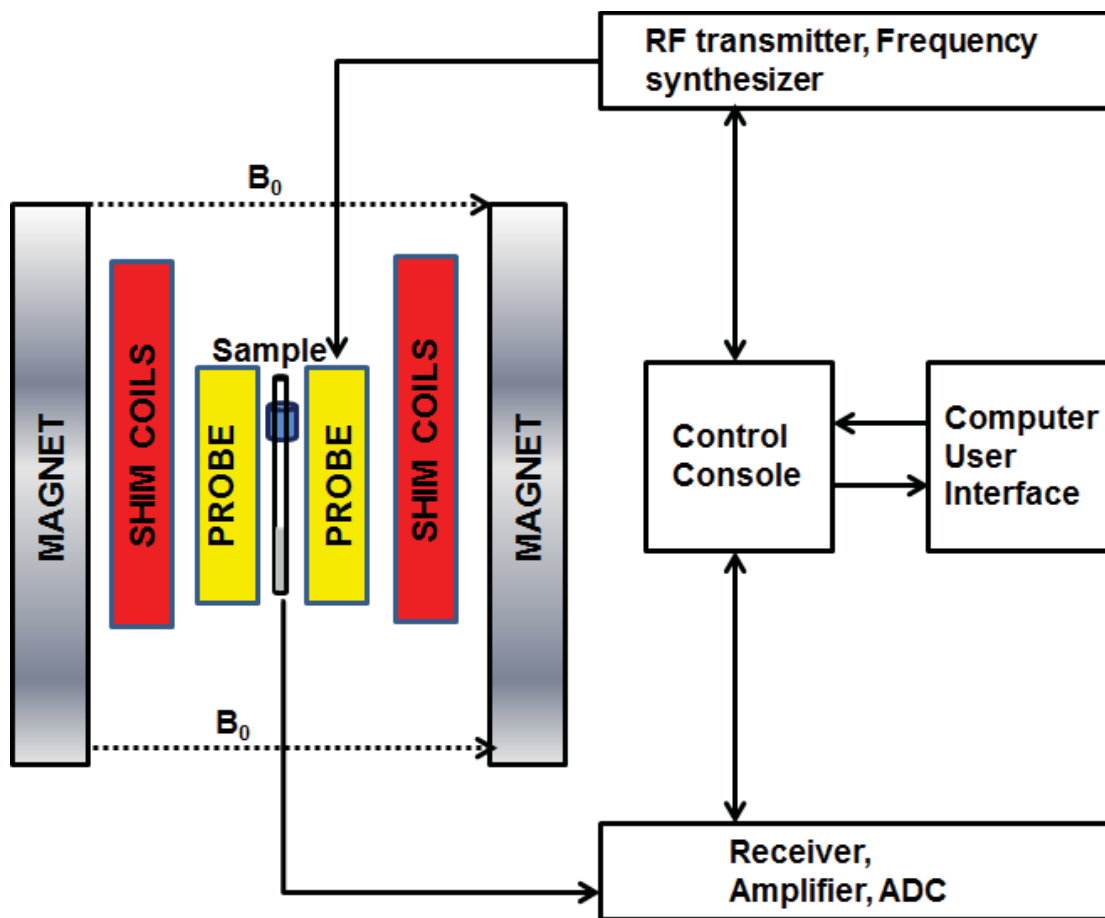


Figure 2-3. Components of an NMR spectrometer.
 The major components of the NMR hardware are illustrated.

The superconducting NMR magnet is an electromagnet (a magnet wherein the magnetic field is produced by electric current) made of superconducting wires. According to the principles of physics the superconducting wires have a resistance almost equal to zero at temperatures close to absolute zero (0 K). Such extreme temperatures are maintained by immersing the superconducting wires in liquid helium. The superconducting wires are a few miles in length and made of brittle niobium, tantalum, titanium and tin (NbTaTi)₃Sn material encapsulated in copper for strength. To accommodate the lengthy wire it is packed into a helical solenoid. The coil and the liquid helium are contained in large insulated Dewar. The Dewar in turn is placed in liquid nitrogen which acts as a buffer between the atmospheric air temperatures and the temperature of liquid helium.

2.3.2. Lock System

The lock system can be imagined as a separate spectrometer located within the main spectrometer. This spectrometer is customarily tuned to the frequency of a certain nuclei (typically deuterium that is added to the sample during sample preparation, 10% volume/volume). The lock system constantly invigilates the frequency of the deuterium nuclei. If an observable drift is reported in the deuterium frequency it is considered as a drift in the external magnetic field. This drift is then corrected by another coil in the main magnetic field. Thus the lock helps to monitor and correct any drift in the external magnetic field.

2.3.3. Shim Coils

The shim coils in a NMR spectrometer are responsible for maintaining a homogenous magnetic field. The inhomogeneities in the magnetic field may be caused by the design of the magnet, the materials used in the probe and ferromagnetic materials surrounding the magnet. The shim coils are capable of producing a local magnetic field that can oppose and cancel out the inhomogeneities in the external field. Since the external variations in the magnetic field may or may not be linear, the shim coils should be capable of producing a variety of local magnetic fields across all directions that is in the x, y and z axes. The magnetic field created by the shim coils depend on the calculated amount of current passed through each coil. The shim coils can be controlled by the user computer interface.

2.3.4. Sample Probe

The sample probe is the collective name used to describe the sample spinner, the temperature control circuit and the RF coils. The sample spinner rotates the NMR tube about its axis. When the sample is spun in the z axis the inhomogeneities in the x and y axes are averaged out resulting in narrower NMR line-width.

One of the major benefits associated with NMR spectroscopy is that the NMR data can be measured at a wide range of temperature. A thermocouple and electronic circuit monitors and maintains the sample temperature. Air or nitrogen which has been heated or cooled is passed over the sample to adjust the temperature as required.

The RF coils are responsible for producing the orthogonal B_1 magnetic field necessary to rotate the spins in the sample by 90^0 or 180^0 . Additionally the RF coils are also responsible for detecting the transverse magnetization of the sample spins as it precesses in the XY plane. The RF coils can resonate and has a bandwidth or specific range of frequencies at which it can resonate. The RF coils need to be tuned for the specific sample before the start of the experiment. Each sample has a different conductivity and dielectric constant which affect the resonance frequency of the RF coils. If the resonance frequency changes such that it does not match the resonance frequency of the nuclei under observation, the RF coil will decline to set up the effective B_1 field as well as fail to detect the signals from the sample. This in turn would lead to poor signal to noise ratio. Thus tuning the RF coils to the specific sample before the start of the NMR experiment is necessary.

2.3.5. Control Console and the Terminal Computer

The control console regulates the RF pulses through the RF transmitter and synthesizer. The signals detected from the sample by the RF coils are analog signals which are amplified by the RF amplifier and converted to digital signals by the analog digital converter (ADC). The digital signal is relayed to the terminal computer in the form of the FID. The FID represents the time domain data which is converted by the terminal computer to the frequency domain data by the mathematical FT. The terminal computer controls the spectrometer through RF pulse programs. The terminal computer also serves to display and store the detected digitized signals.

2.4. Setting up a Nuclear Magnetic Resonance Experiment

The following description summarizes the basic steps in setting up an NMR experiment.

- A pure biomolecular sample which in this case is the protein of interest is prepared in an appropriate aqueous buffer solution with 10% D_2O (volume/volume). The sample is carefully added to the NMR tube avoiding the introduction of air bubbles. The cap is secured on the NMR tube.
- The NMR tube is positioned in the spinner so that the sample is centered on the probe coil. The gauge can be used to check the center and symmetry of the sample before introducing into the spinner.

- One of the benefits associated with NMR spectroscopy is NMR signals can be measured across a wide range of temperatures. Accordingly once the sample is placed in the center of the probe, the temperature can be adjusted using the command *edte*.
- The super conducting magnets used in high resolution NMR are prone to drift. To compensate for this drift and hold the magnetic field in a stable position it is needed to lock the magnetic field using an internal standard which in this case is deuterium oxide (D₂O). Typing the *lock* command displays a wide choice of internal standards, in our case we choose 90% H₂O+10% D₂O. Locking ensures temporal stability of the magnetic field. The lock power should be adjusted to maximum without saturation i.e. the increase in the signal must be linear with the increase of the power. The lock gain should be adjusted to maximize the lock signal. The lock phase is adjusted to obtain maximal signal at a given lock power and gain.
- The probe has a resonant circuit for each observed and decoupled nucleus. So there is a circuit for ¹H and for X, where X may be ¹⁵N, ¹³C, ¹⁹F, ³¹P. Each of these circuits is most sensitive at a particular frequency. Tuning is the process of adjusting the frequency until it coincides with the frequency of the pulses transmitted to the circuit. Matching is the process of adjusting the impedance (blockage/resistance) of the resonance circuit to the impedance of the resonance line connected to it. Thus at the start of each NMR experiment the frequency under observation which in this case is ¹H, ¹⁵N and ¹³C should be tuned and matched using the command *atmm* (or *atma* for automatic tuning and matching). The order of tuning the frequencies is ¹⁵N, ¹³C and ¹H. Practically on typing this command a curve is displayed on the screen, the minimum of the curve must be as low (matching) and as well as on the line (tuning) as possible.
- The next step in setting up an NMR experiment is shimming to ensure homogeneity of the magnetic field. The sample is surrounded with a set of shim coils. Each shim coil is capable of producing a small magnetic field with a particular spatial orientation and this can be utilized to cancel out the inhomogeneities in the main magnetic fields. The shim coils are labelled x, y and z corresponding to the spatial direction of the respective magnetic fields they generate. The orientation of the main magnetic field is vertical in the z direction and the axial z shim coils modify the field profile along the axis of the main field. The radial coils are in the x and y directions. The shim settings control the current through the shim coils which in turn controls the magnetic field gradients in the different directions. Since the z direction is parallel to the direction of the probe, the height of the sample in the NMR tube affects the z shims dramatically. For practical purposes a sample height of 5cm was maintained in all experiments and shimming was performed using the command *topshim*.
- Once the basic steps pertaining to all NMR experiments are performed the specific experiment (e.g. ¹H-¹⁵N Heteronuclear Single Quantum Coherence

(HSQC), ^1H - ^{13}C HSQC etc.) can be setup using the *rpar* command. This command provides a list of all the NMR experiments with a standard parameter set that is saved on the spectrometer. Note although an experiment is recalled from the manufacturer's list several parameters need to be modified depending on the specific protein of interest as described below.

- The proton offset is calibrated. The experiment without presaturation is phased to result in an absorptive signal. The offset is varied first in bigger than in smaller steps, until a minimum is reached. The experiment is started with command *zg* and processed with the command *efp*. This can also be achieved by acquiring in *gs* mode and making sure to get rid of high intensity slow component of the FID signal which is representative of water.
- Next the 90° proton hard pulse *p1* is calibrated. The *p1* is adjusted to result in the smallest signal possible. This can also be performed using the command *pulsecal*.
- The one dimensional (1D) experiment is selected from the set of NMR experiments. The *getprosol* command is performed with the determined proton hard pulse.
- The receiver gain can be set automatically using the command *rga*.
- The number of scans, spectral width, number of points to be collected in the indirect dimension, etc. need be determined based on the specific experiment and protein concentration of the sample. The experiment is started with the command *zg*. The FID is measured which is Fourier transformed using the command *efp* for a 1D experiment or *xfb* for a 2D experiment.

2.5. Protein Structure Determination by Nuclear Magnetic Resonance

2.5.1. Synopsis

Proteins are macromolecules of biological importance. The monomeric unit of the protein polymer is defined as an amino acid. There are twenty different amino acids which combine in different permutation and combinations to form the protein macromolecule. The individual amino acids in the protein are chemically held together by a peptide bond giving rise to the primary structure of the protein. The polypeptide in turn adopts local secondary conformations such as α - helices and β -sheets stabilized by hydrogen bonds. The polypeptide chain with its secondary structure folds in space giving rise to the tertiary structure of the protein. The tertiary structure of the protein is crucial and governs the function of the protein. Proteins are biologically significant macromolecules and play an important role in health and disease. Therefore it is an absolute necessary to study protein structure and function.

The method of NMR spectroscopy can be utilized to determine the three dimensional (3D) structure of proteins. In NMR a series of 1D, 2D and 3D experiments are performed in order to collect resonance/frequency information of the individual nuclei which make up the amino acids which form the building block of the protein molecules. The recorded NMR spectrum is a conglomeration of the superimposition of the NMR spectrum of the individual amino acids.

As stated earlier, the resonance frequency is converted to chemical shift data by referencing using an internal standard such as tetramethylsilane. The chemical shift measured by NMR is a reflection of the chemical environment of the nuclei. Some of the factors affecting the chemical shift of the nuclei are electron density and electronegativity of neighboring atoms. The chemical shift data need to be assigned to the specific nuclei within the molecular structure of the protein before any structural information can be obtained. The process of assigning each nuclear frequency within the molecular structure of the protein is known as NMR assignments and forms an important prerequisite for NMR based structure determination.

Essentially protein structures are made of hydrogen, nitrogen and carbon atoms. The respective NMR detectable isotopes of these atoms are ^1H , ^{15}N and ^{13}C . The natural abundance of ^1H is 99.9%, ^{15}N is 0.4% and ^{13}C is 1.1%. The natural abundance of proton is sufficient to be detected by NMR but the natural abundance of ^{15}N and ^{13}C is low and therefore isotope labeling of the protein is required. Typically for the structure determination of small proteins or protein domains it is ^{15}N and ^{13}C labelled by expressing the protein in minimal media containing isotope labelled nutrients. For determining the structure of larger proteins the protein could be triple labelled (^2H , ^{15}N and ^{13}C).

For best results and spectral compliances the different NMR experiments are performed under the same conditions of temperature and buffer. Regularly the protein sample used for acquiring multi-dimensional spectra and subsequent structure calculation by NMR methods should have a concentration close to one millimolar; however due to the challenging protein expression and purification methods associated with the protein described in this dissertation, the protein concentration attained was one fourth of one millimolar. Invariably, longer acquisition times and the use of high magnetic field and cryogenic probes helped overcome the problems associated with lower protein concentrations.

2.5.2. Nuclear Magnetic Resonance Experiments

Routinely the first spectrum recorded on the protein is the 1D proton spectrum. In a 1D proton spectrum thousands of protons resonate in a narrow spectral region. These protons need to be assigned within the chemical structure of the protein before obtaining any structural information. NMR assignments are made by establishing connectivity between protons and on the basis of the chemical properties of the amino acids.

In the 2D NMR spectrum two frequency dimensions are measured that allow resonances to be separated on the basis of their connectivity through bonds and through space. For example the 2D Correlation Spectroscopy (COSY) establishes connections or correlations between protons connected by three chemical bonds within a particular amino acid and the Total Correlation Spectroscopy (TOCSY) establishes connection between three or more chemical bonds within a particular amino acid. The Nuclear Overhauser effect correlation spectroscopy (NOESY) type of experiments determines correlation between protons through space and not through bonds. The 2D NOESY experiment shows a signal if the internuclear distance between the participating proton atoms is less than 5 Å. These protons may or may not be from the same amino acid. Protons which are far apart in the primary structure of the protein may fold close to one another in the tertiary structure of the protein and thus give rise to NOESY signal.

Each particular amino acid has a specific pattern in a COSY/TOCSY spectrum. Although the amino acid types can be identified in the correlation spectrums they cannot be assigned to specific amino acids within the protein sequence because the connectivity between the amino acids have not yet been established. The connectivity can be established using the 2D NOESY experiments. This methodology of homonuclear sequence specific resonance assignment is valuable for small proteins and when ^{15}N and ^{13}C labelled protein is not available; however, this method is challenging for proteins of bigger size wherein the process of heteronuclear backbone assignment is performed. The protein structure determined in this dissertation has a molecular weight of approximately fifteen kilodaltons, was expressed in bacteria as an isotope (^{15}N and ^{13}C) labelled protein and the NMR backbone assignments were performed using the heteronuclear backbone assignment strategy using the software Computer Aided Resonance Assignment (CARA) and is described briefly in the sections below[131].

2.5.3. Heteronuclear Backbone Assignment

For sequential assignment of the amide backbone a combination of 2D ^1H - ^{15}N HSQC and 3D backbone assignment experiments such as HNCA, HNCACAB and CBCA(CO)NH were used. These experiments can be imagined as an extension of the COSY type of experiments wherein the correlations are established between different nuclei types. In the 2D ^1H - ^{15}N HSQC experiment the correlation is established between the ^1H and ^{15}N nuclei. Amide proton resonances are well separated due to the dispersion by ^{15}N chemical shifts. Each amide group within a protein contributes to a single resonance peak, except for proline. The amino groups of the asparagine and glutamine side chains give rise to paired resonances depicting the same nitrogen shift but different proton shift. Thus, in theory each amino acid except prolines in the protein should contribute to a peak in the 2D ^1H - ^{15}N HSQC spectrum.

The 3D NMR experiments are performed to completely assign the backbone of the protein. The 3D NMR experiments are built from 2D NMR experiments by introducing an additional measurement in the third dimension as explained in

Figure 2-4[132]. The 3D HNCA spectrum shows two peaks in the carbon (third) dimension for each amide peak representing an amino acid. The stronger peak represents the corresponding $C\alpha$ resonance and the weaker peak represents the $C\alpha$ resonance of the preceding residue. In the 3D HNCACB spectrum four peaks are observed. The stronger peak in phase represents the self $C\alpha$ resonance and weaker peak in phase represents the preceding $C\alpha$ resonance. The other two observable peaks are out of phase, the stronger one represents the corresponding $C\beta$ resonance and the weaker one represents the $C\beta$ resonance of the preceding amino acid. In the 3D CBCA(CO)NH spectrum two peaks are observed corresponding to the $C\alpha$ and $C\beta$ resonances of the preceding amino acid.

By comparing the spins in the different spectra the peaks or spins of the self-resonances can be separated from the adjacent resonances. Thus from the simultaneous use of multiple 3D backbone assignment experiments, for each amide peak in the 2D ^1H - ^{15}N HSQC spectrum the $C\alpha$ and $C\beta$ of the self and the preceding amino acid can be identified and linked. Linking each amide peak to its predecessor helps in building a fragment. In theory the entire protein domain can be linked by this method from the N-terminus to the C-terminus. But in real situations the positioning of the proline residues disrupts the process of continuous linking giving rise to smaller fragments. This fragment is then assigned to the unique stretch of amino acids in the primary sequence (which is already known) of the protein. The heteronuclear assignment strategy primarily relies on the fact that the amino acids have a characteristic $C\alpha$ and $C\beta$ chemical shift distribution. Particularly some of the amino acids such as glycine, alanine, serine and threonine have unique $C\alpha$ and $C\beta$ chemical shifts which can be identified without any additional information. These amino acids serve to anchor the assignment within the primary structure of the protein. The longer the size of the linked fragment the higher is the chances of securing it to a unique stretch of amino acids in the protein primary sequence. The method described above is the manual method of protein backbone assignment. The process of manual assignment is generally tedious and time consuming.

The software CARA has an automated backbone assignment module described as 'Autolink' incorporated in the software[131]. The automatic linking algorithm uses the chemical shift information of the self and the preceding residues to calculate a probable assignment. The algorithm functions to increase the number of the linked residues, considers the secondary structure and consequently scores the assignment[131]. The Autolink module is fairly accurate; however it is mandatory to verify all automated assignments. In this dissertation the process of automated and manual assignments were used in combination. The automated assignments were verified manually and the gaps left in the automated assignments were filled by the process of manual assignment.

2.5.4. Amino Acid Side Chain Assignment Strategy

Each amino acid is characterized by a unique side chain. Some amino acids have a simple and short side chain such as alanine whereas others such as leucine and lysine are characterized by long side chains.

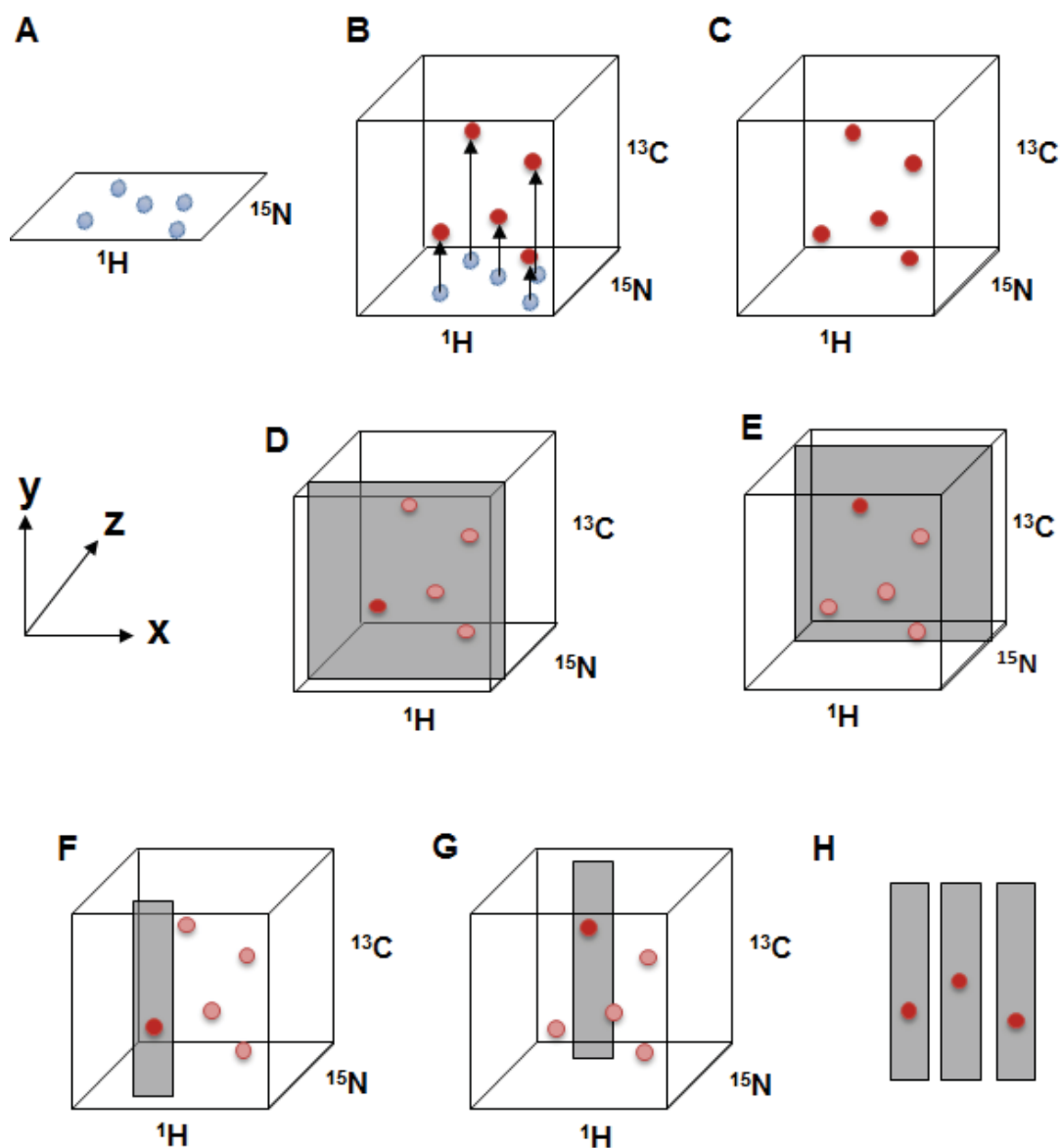


Figure 2-4. Visualizing three dimensional NMR spectrum and generation of strips. (A) The 2D ^1H - ^{15}N HSQC spectrum. The x and the y axes are ^1H and ^{15}N respectively. (B) and (C) The 2D ^1H - ^{15}N HSQC spectrum is extended in the third dimension which is the ^{13}C dimension. (D) and (E) The different planes of the 3D spectrum are represented. In this case the ^1H dimension forms the x-axis, the ^{13}C dimension forms the y-axis and the ^{15}N dimension is in the z-plane. (F) and (G) Each represents a particular position in the ^1H and ^{15}N dimensions but spans the entire width of the ^{13}C dimension. Thus each position generates a strip. (H) Each peak in the ^1H - ^{15}N HSQC generates a strip encompassing the ^{13}C dimension. Source: Higman P. Protein NMR. In. Bristol UK; November 2012.

The side chain protons include the likes of methyl protons and aromatic protons. Regularly side chain assignments are performed after the backbone assignment. Unlike the process of backbone assignment wherein automated assignments may be performed side chain assignments entirely rely on manual methods. Therefore the process of side chain assignments is exhausting and requires time and effort.

In the 2D ^1H - ^{13}C HSQC spectrum the correlation is established between the ^1H and the ^{13}C nuclei. Thus all the protons in the protein attached to a carbon atom in principle should be observed in the 2D ^1H - ^{13}C spectrum. The goal of the side chain assignment is to attribute each peak to the corresponding side chain proton in the protein. The 3D HBHA(CBCACO)NH and the (H)CCH-TOCSY experiments together with the 3D NOESY-HSQC experiments help in the assignment of the amino acid side chains. The 3D TOCSY provides correlation between all the methyl and methylene protons within a single amino acid. It is similar to the 2D-TOCSY as described earlier but is additionally resolved in the carbon dimension.

On completion of the backbone and side chain resonance assignments an atom list can be generated. The atom list enlists each amino acid in the protein and summarizes each of its hydrogen, nitrogen and carbon chemical shift.

2.5.5. Nuclear Overhauser Effect Correlation Spectroscopy Assignments and Distance Constraints

As stated earlier the NOESY experiments establish correlations through space. When the nuclei of two protons are spatially close to one another chemical exchange takes place meaning the nucleus exchanges between two or more different environment conditions. The chemical exchange between the proximal protons gives rise to the NOE “cross peak” in all the NOESY experiments. The intensity of the NOE “cross peak” is inversely related to the distance between the two participating nuclei as shown in **Equation 2-7**.

$$V \sim \frac{1}{d^6} \quad \text{Equation 2-7}$$

Where V is the volume of the NOESY cross peak and d is the spatial distance between the two atoms. The NOE peak intensity is inversely proportional to the 6th power dependence of the distance; therefore the distances measured by NOE are always less than 5 Å. The regular secondary structures of α -helices and β -sheets give rise to regular NOE pattern which helps in confirming the secondary structure of the protein. The NOESY cross peaks are assigned based on the correlations which have been established through the orchestrated analysis of TOCSY and NOESY spectra.

The intensity or the volumes of the NOESY cross peaks in a ^{15}N resolved ^1H - ^1H NOESY experiment is related to the distance between the participating protons. The NMR method of protein structure determination relies heavily on a dense network of such distance restraints. Some of the NOE peaks correspond to the protons within the

same amino acid; they are described as intra-residue NOEs. These residues do not provide important information regarding the protein secondary or tertiary structure. The other observable NOEs are for protons in different amino acids. These are described as inter-residue NOEs. The inter-residue NOEs in turn can be classified as sequential, medium or long range NOEs. The sequential NOEs are observed between the protons of the adjacent amino acids, the medium range NOEs are observed between protons of the amino acids which are less than equal to four amino acids apart. The sequential and medium range NOEs play a key role in defining the secondary structure of the protein. Alpha helices and β -sheets give rise to characteristic NOE pattern defining the local structure of the protein. The long range NOEs are observed between protons of the amino acids that are far apart in the primary sequence, at least four amino acids apart. The long range inter-residue NOEs are crucial for defining the tertiary structure of the protein. Thousands of NOEs are typically required to resolve a protein structure by NMR methods.

The practical challenge accompanied with this method is the associated degeneracy of the chemical shifts of some of the atoms. This means that some of the atoms may have the same chemical shift and thus the distance constraint measurements can become ambiguous. This can be controlled at least in part by the use of the ^{13}C -NOESY experiments. In which case the ^1H - ^1H cross peaks are now dispersed along the ^{13}C dimension reducing the chance of overlapping.

The intensities or volumes of the NOE cross peaks (V) can be converted into upper distance limits by using the following formula as shown in **Equation 2-8**.

$$V = \frac{c}{d^6} \qquad \text{Equation 2-8}$$

Where V is the volume of the NOESY peak, c is the calibration constant and d is the spatial distance between the two atoms. The calibration constant c is initially determined from NOEs between protons at fixed distance. Thus on completion of the NOE assignments a constraint list is generated which contains the upper distance limits of spatially neighboring atom pairs. The lower limit of the distance restraints is provided by the van der Waals radius between the two protons. The variation of the distance between the two protons between the lower limit (van der Waals forces $\sim 2\text{\AA}$) and the upper limit (obtained from the NOEs $\sim 5\text{\AA}$) allows a relatively large conformational space for the participating protons to arrange themselves relative to one another. In the structure calculation software, CYANA, the module CALIBA utilizes this principle and converts NOE peak intensities to distance constraints[133]. The NOE assignments may be performed manually or using a software program. In this dissertation the NOE assignments were performed using the program UNIO[134]. UNIO requires the ^{15}N and ^{13}C resolved ^1H - ^1H NOESY spectrum and the chemical shift list as an input for the automated NOE assignment. The chemical shift assignment list should be at the least 90% complete and the NOESY spectra should be calibrated to the proton list. As stated earlier all automated assignments were verified manually.

2.5.6. Structure Calculation and Refinement

The method of NMR based structure calculation utilizes experimentally derived constraints such as the distance restraints generated from NOESY experiments and angle constraints that can be generated from the chemical shift assignments to generate a structure that agrees with all the constraints and maintains a global minimum energy. Software programs such as CYANA utilize this information to resolve the tertiary structure of biomolecules from NMR derived data[133].

The dihedral angle constraints are used as input for structure calculation using CYANA. The dihedral angle constraint can be calculated by using the program TALOS+[135]. The TALOS+ program utilizes the chemical shift assignments ($H\alpha$, $C\alpha$, $C\beta$, CO and N) and the protein primary sequence as an input. It then compares the chemical shifts against a database of chemical shifts associated with high resolution structures. Comparison is based on triplet sequence of amino acids present in the database of the chemical shifts associated with high resolution structures. The program then generates the potential backbone dihedral angle (ϕ (ϕ) and ψ (ψ)) constraints[135]. This method relies on the fact that the secondary chemical shifts are correlated to the protein secondary structure. The use of the torsion angle restraints greatly reduces the conformational space to sample.

The software CYANA is based on the principle of simulated annealing molecular dynamics in torsion angle space[133]. In simple terms this means that the atoms are imitated to move under heating conditions and consequently cooled to a minimum energy level. During this process the only degrees of freedom of movement allowed is across the backbone torsion angles of the amino acids in the protein structure. In addition each of the generated structure should comply with the distance restraints that are used as an input for structure calculation.

In this method the protein primary structure is folded such that the atoms fulfill the observed distance constraints. The program uses the same restraints to calculate a bunch of structures or conformers. Typically hundred conformers are generated and the twenty conformers with the lowest energy or target function are selected. As explained earlier, in NMR the distances between the protons cannot be determined definitively, what is obtained instead is a range for the distance defined by upper and lower limits. Consequently these restraints are utilized to calculate an ensemble of structures rather than one single structure.

The initial structures are generated from a random conformation which is optimized by simulated annealing molecular dynamics in torsion angle space. Following this convention one hundred structures are generated. The quality of the structures generated can be assessed by monitoring some key factors. Firstly, the convergence rate between the generated structures should be high, i.e. the root mean square deviation between the structures should be minimum. Secondly, the number of distance restraints violated should be low. Thirdly, the final energy or the target function should be lowest. During the final stages of resolving the structure, the protein structures can be refined by

adding additional constraints such as hydrogen bonds constraints across the protein secondary structure and disulfide bond constraints if appropriate. The hydrogen bonds can be estimated by measuring slow exchanging amide protons in deuterium oxide.

In reality NMR based structure calculation is an iterative process between NOE assignments and evaluation and structure calculation. The structure is considered final when each of the resonance peaks in the NMR spectrum has been attributed to the corresponding nuclei within the protein structure. The quality of the determined structure is directly dependent on the completeness and correctness of the resonance assignments. The root mean square deviation between the twenty structures with the lowest energy is less than an Angstrom and the energy of the structures have reached a global minimum.

2.5.7. Structure Validation

One of the important methods of validating the resolved protein structure is by contemplating the Ramachandran plot and was originally developed by G.N Ramachandran, C. Ramakrishnan, and V. Sasisekharan[136, 137]. In the protein backbone the main chain N-C α and C α -C bonds are relatively free to rotate. These bonds respectively form the phi and psi dihedral angles. The Ramachandran plot displays the backbone dihedral angle pairs (ϕ, ψ) of the polypeptide chain in a given protein structure and is subdivided into primarily three regions and typically colour coded as red, yellow and white. Essentially the red region corresponds to the “highly favorable region” and encompasses the residues demonstrating no steric clashes. The yellow regions corresponds to the “additionally allowed region” and include the residues wherein the van der Waals radii are shorter than usual, this in turn allows the atoms to come close together but are still acceptable. The white disallowed region corresponds to conformations wherein the atoms in the polypeptide come closer than the sum of their van der Waals radii. This in turn would lead to steric clashes and are therefore disallowed for all amino acids except glycine because glycine does not possess a side chain. Needless to say that a well resolved structure should include maximum amino acids in the highly favorable region of the Ramachandran plot and theoretically none in the disallowed region[137].

Thus the Ramachandran plot is an important tool for validating the protein structure. Steric hindrance from amino acid side chains allows only certain angles for a folded protein. Deviation from the allowed values are called outliers, if the Ramachandran plot depicts an outlier it is usually because of a wrong assignment and therefore can be corrected. An extremely small percentage of the amino acids may still have torsion angles in the disallowed region because they may form part of a flexible loop. In the present dissertation the Ramachandran plot was used during the final steps of structure calculation to correct wrong assignments and reduce violations and global energy. The main steps of protein structure determination by NMR are summarized in **Figure 2-5**.

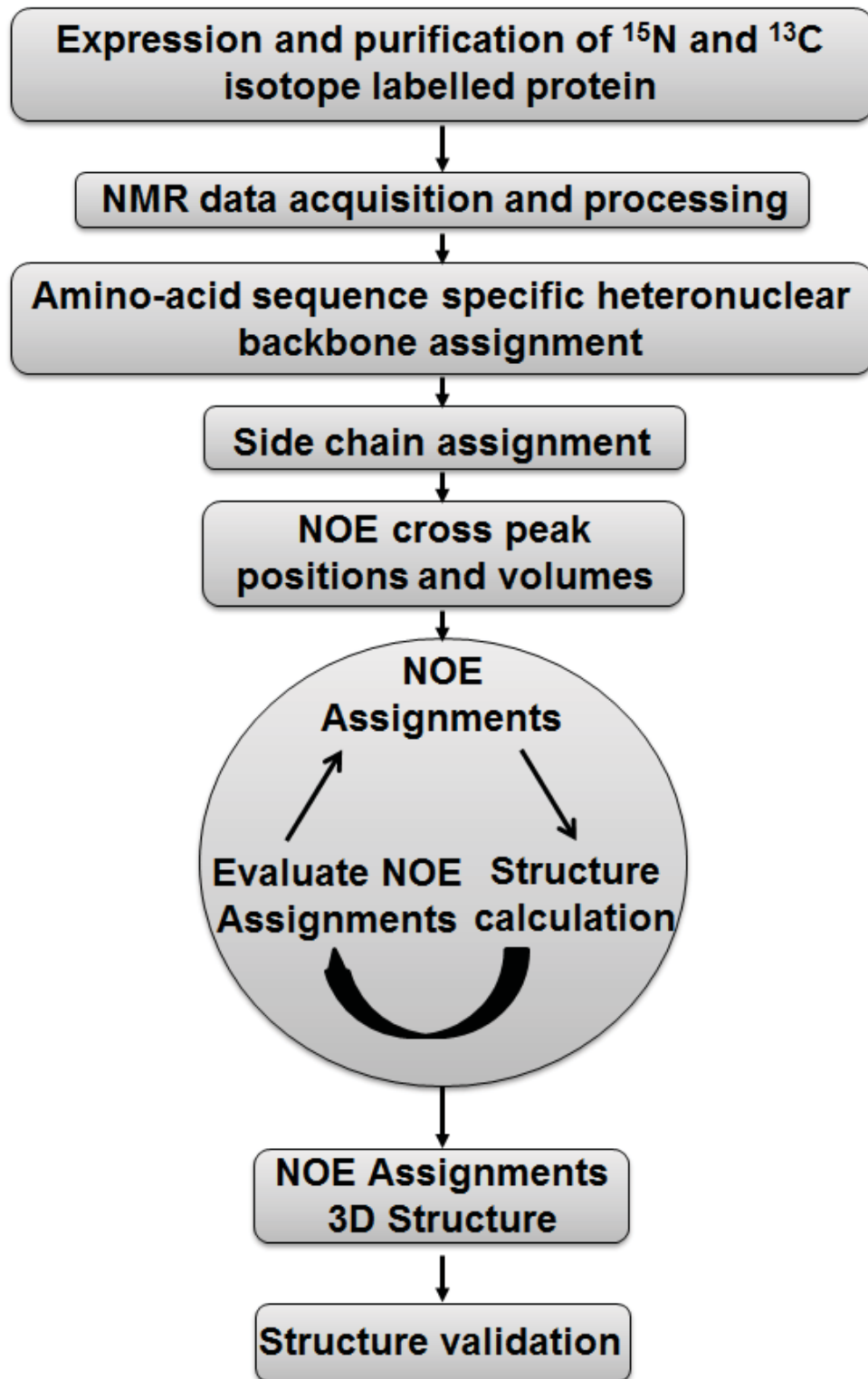


Figure 2-5. Flowchart for NMR based protein structure determination.
The major steps for determining the solution structure from NMR data are shown.

2.6. Nuclear Magnetic Resonance as a Tool to Study Protein-ligand Interactions

Apart from determining atomic resolution biomolecular structures, NMR being a powerful and sophisticated technique has widespread applications in scientific research, medicinal chemistry and pharmaceutical industry. The technique of NMR provides important information on biomolecular dynamics, protein folding, metabolite and chemical analysis and intermolecular interactions. In the pharmaceutical industry the technique of NMR is routinely used to support structure-based drug discovery[138]. NMR has also been employed to study drug intoxication[139]. In analytical chemistry NMR is used as a quality control technique for determining the purity of a sample. Interestingly NMR is also being used to profile commercial commodities like wine[140]. Amongst the various uses of NMR, in this dissertation NMR was widely used to study protein-ligand interactions and therefore will be discussed in detail.

2.6.1. 1D Nuclear Magnetic Resonance Experiments to Screen and Identify Potential Ligands for Proteins

Recurrently proteins interact with ligands of endogenous or synthetic origin and upon such interaction the structure of both the ligand as well as the protein are perturbed. The structural perturbations both on the ligand as well as the protein surface can be detected by NMR. The 1D proton spectrum of the ligand can be recorded in the absence and presence of an increasing concentration of the protein. Each peak in the 1D spectrum of the ligand corresponds to the protons of the ligand. The identity of each peak can be established by assigning the atoms of the ligand using the homonuclear assignment strategy. If the ligand interacts with the protein, the protons participating in the protein-ligand interaction witnesses a change in the chemical environment. Therefore the peaks shift. The magnitude of the peak shift is dependent on the strength of the protein-ligand interaction. Thus, even a simple 1D NMR experiment can be used to obtain important information about protein-ligand interactions.

Additional methods based on the principle of intermolecular NOE transfer are also used to detect protein-ligand interaction. If the protein-ligand interaction is undergoing fast exchange specific NMR methods based on the principle of transferred NOE can be used to study the behavior of the ligand in the protein-ligand complex. These experiments are described as saturation transfer difference (STD) NMR experiments and do not require isotope labeling of the protein or the ligand. A well separated protein signal is selectively irradiated with an RF pulse. The saturation then spreads to all the protons of the protein by spin diffusion. Intermolecular NOE transfers the saturation to the ligand protons at the binding interface[141]. This method is widely used to screen for drugs and identify a positive hit from a mixture through positive NOE signals. Additionally epitope mapping enables the identification of the part of the ligand that is in contact with the protein. The principle of the STD experiment is illustrated in **Figure 2-6A**.

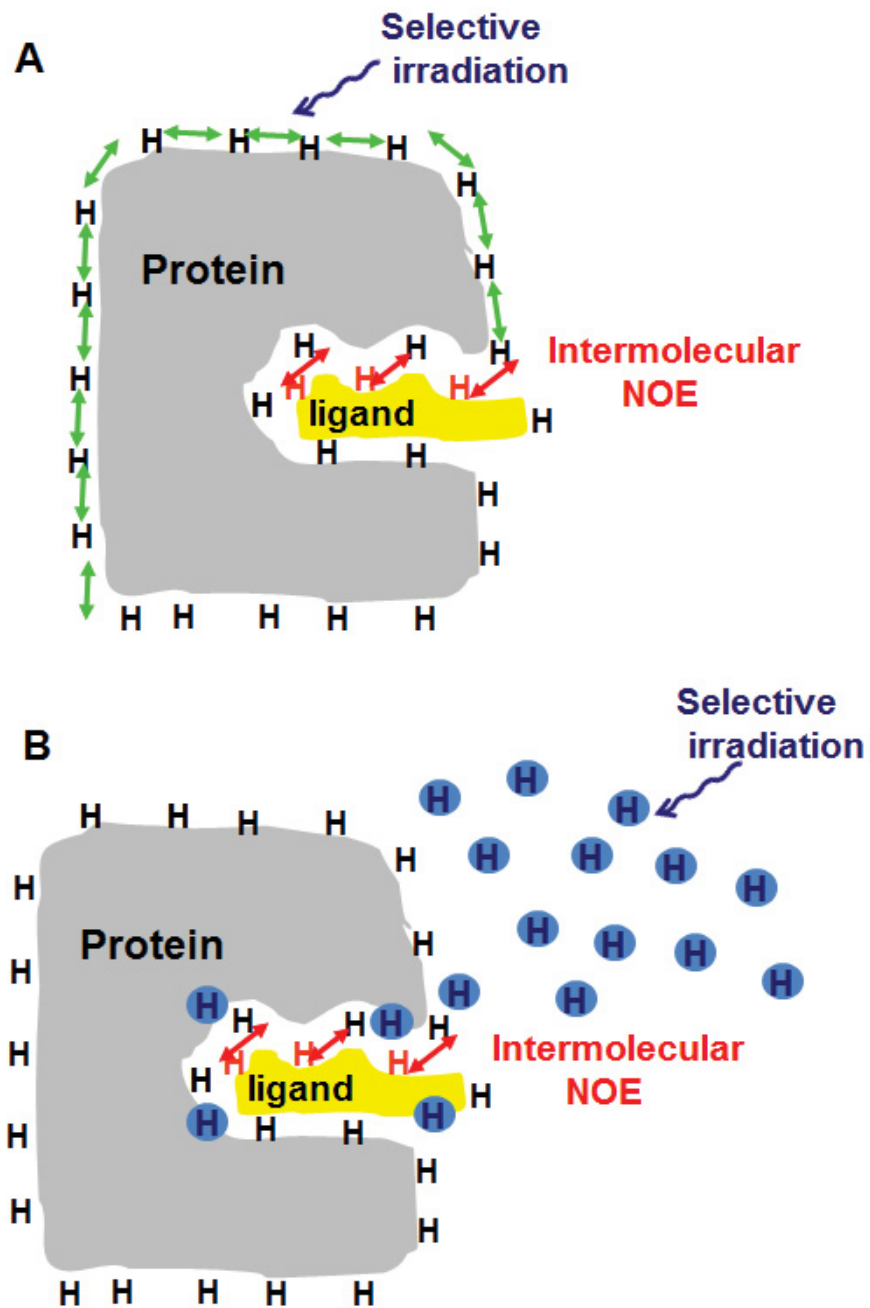


Figure 2-6. Transferred NOE experiments to determine protein-ligand interactions. (A) The principal of the Saturation Transfer Difference (STD) NMR experiment (B) The principal of the Water LOGSY NMR experiment.

Water-Ligand Observation with Gradient Spectroscopy (Water-LOGSY) is a widely used NMR experiment for the detection of protein-ligand interaction. Like the STD experiments, Water-LOGSY experiments are also based on the principle of intermolecular NOE transfer. Magnetization is transferred via intermolecular NOEs and spin diffusion. In this case, the bulk water is irradiated selectively with the RF pulse. The magnetization transfers to other protons of the protein by NOEs and from the protein-protons to the ligand-protons by spin diffusion. The NOE between the water and the protein have an opposite sign from the NOE between the water and the small molecules. The ligand concentration is twenty times higher than the protein concentration therefore the protein signals are not observed in this experiment. Positive and negative signals in the water-LOGSY experiments distinguish between the interacting and the non-interacting molecules[142]. The principle of Water-LOGSY is explained in **Figure 2-6B**. In this dissertation the technique of Water-LOGSY was widely used to screen and identify small molecules that interact with the protein of interest.

2.6.2. Chemical Shift Mapping to Determine the Ligand Binding Surface on the Protein

Chemical shift perturbation (CSP) mapping is a simple and unique technique to monitor protein-ligand interactions[143]. Regularly the unlabelled ligand is titrated to the ^{15}N -labelled protein and a ^1H - ^{15}N HSQC is recorded at each titration point. In a ^1H - ^{15}N HSQC spectrum each peak corresponds to an amide pair of the protein backbone. Following the technique of heteronuclear backbone assignment strategy as described above, each peak in the spectrum can be assigned to the specific amino acid of the protein[131]. If the ligand interacts with the protein, there will be a change in the chemical environment of the participating amino acid of the protein which will lead to a shift in the peak position in the ^1H - ^{15}N HSQC spectrum. The shift of the peaks can be measured at each titration point, as the protein is titrated with an increasing concentration of the ligand. The peaks depicting the maximum CSP are likely to form the ligand binding surface of the protein. Thus if the backbone assignment and the tertiary structure of the protein are known, CSPs can be used to map the ligand binding surface on the protein.

The CSP of each residue can be plotted against the concentration of the ligand to generate a titration curve. The titration curve can be fitted to obtain the dissociation constant, K_d . Thus the method of CSP is extremely useful and unique because using this method information about the binding surface of the protein as well as the binding affinity of the ligand can be obtained simultaneously[143]. Additionally the measurement of CSPs by NMR is generally fast. Typically for a sample of protein concentration of one-half of a millimolar NMR acquisition time is approximately 30 minutes. In the present dissertation for titration analysis with different ligands the concentration of the proteins varied widely, furthermore some of the ligands tested in the present dissertation were such that led to a drop in the intensity of the NMR signals. Therefore, the acquisition time for different CSP measurements in this dissertation varied widely ranging from 45 minutes to overnight measurements.

The experimental conditions during the CSP measurement should be kept constant. The external conditions of temperature and buffer should be maintained uniformly. Typically the same buffer is used for the protein and the ligand. A small change in pH or salt concentration may alter protein signals and complicate analysis. Additionally if the ligand is soluble only in organic solvents such as dimethyl sulfoxide a control experiment should be performed with the solvent to rule out the solvent effect.

2.6.3. Binding Kinetics to Determine the Dynamics of Protein-ligand Interaction

In addition to determining the ligand-binding surface on the protein, NMR provides important information about the kinetics of binding. The protein-ligand interactions can be categorized as fast exchange, intermediate exchange or slow exchange in the NMR timescale as summarized in **Figure 2-7**. In case of weak binding, the ligand associates and dissociates from the protein rapidly in sub-milliseconds. The lifetime of the complex is short. This form of exchange is described as fast exchange. The protein exists in equilibrium between the free and the ligand-bound form. The peaks corresponding to the free form of the protein are distinct from the ligand bound form; however since the free and the ligand bound protein states are undergoing fast chemical exchange the peaks average and one single peak is observed at a position between the two component peaks. If the protein-ligand interaction is strong the ligand remains bound to the protein. The time taken for the ligand to dissociate from the protein is more than one second. Therefore two distinct peaks should be observed simultaneously in NMR. The ratio of the two states in equilibrium determines the ratio of the peak integral. This phenomenon is described as slow exchange in NMR and is suggestive of tight binding. Protein-ligand interactions with intermediate dissociation rate exhibit a broad peak at the average position and are described as intermediate exchange in the NMR timescale. In the present dissertation all the observed protein-ligand interactions can be described as fast exchange on the NMR time scale and are representative of weak binding.

2.6.4. Dissociation Constant to Determine the Binding Affinity

NMR can also be used to determine the binding affinity or the dissociation constant (K_d) of the protein-ligand complex. In case of weak binding or fast exchange, titration of the protein with an increasing concentration of the ligand generates a series of moving peaks until the peak reaches an ultimate point corresponding to saturation of the protein by the ligand[143]. The titration curve is generated by plotting the observed CSP against the ligand concentration as shown in **Figure 2-8A**. The CSP for each residue in the protein on interaction with the ligand can be calculated as shown in **Equation 2-9**.

$$\Delta\delta = \sqrt{(\Delta\delta H)^2 + (\Delta\delta N / 5)^2} \quad \text{Equation 2-9}$$

Where, $\Delta\delta$ is the composite chemical shift difference in parts per million (ppm), $\Delta\delta H$ is the chemical shift difference in the proton dimension and $\Delta\delta N$ is the chemical shift difference in the nitrogen dimension.

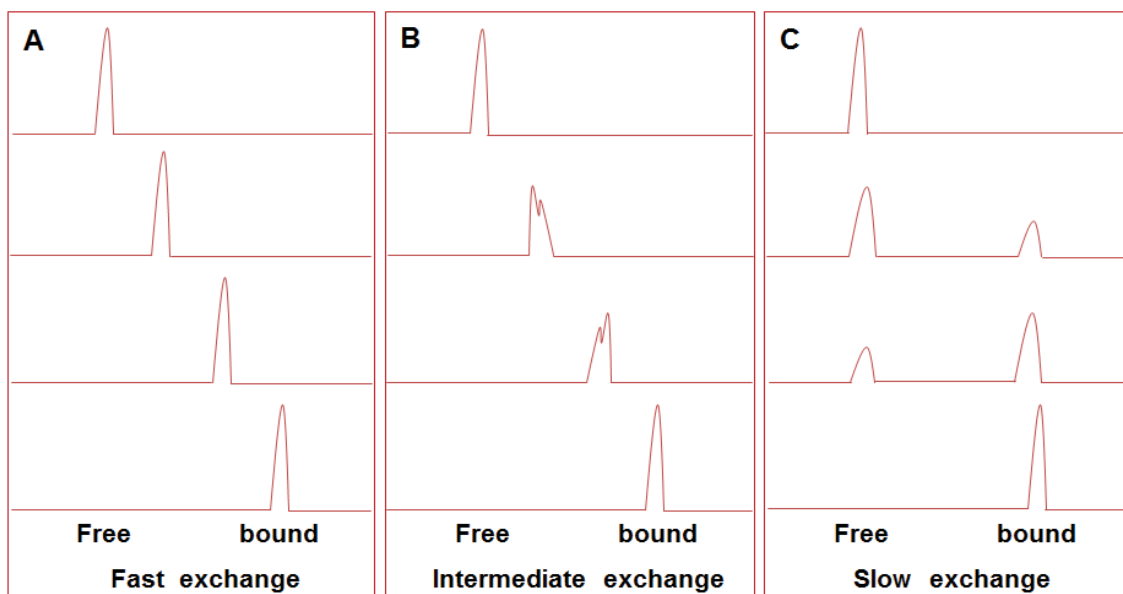


Figure 2-7. Representation of the different binding kinetics.

The different binding kinetics of the protein-ligand interactions can be categorized as fast exchange, intermediate exchange or slow exchange in the NMR timescale.

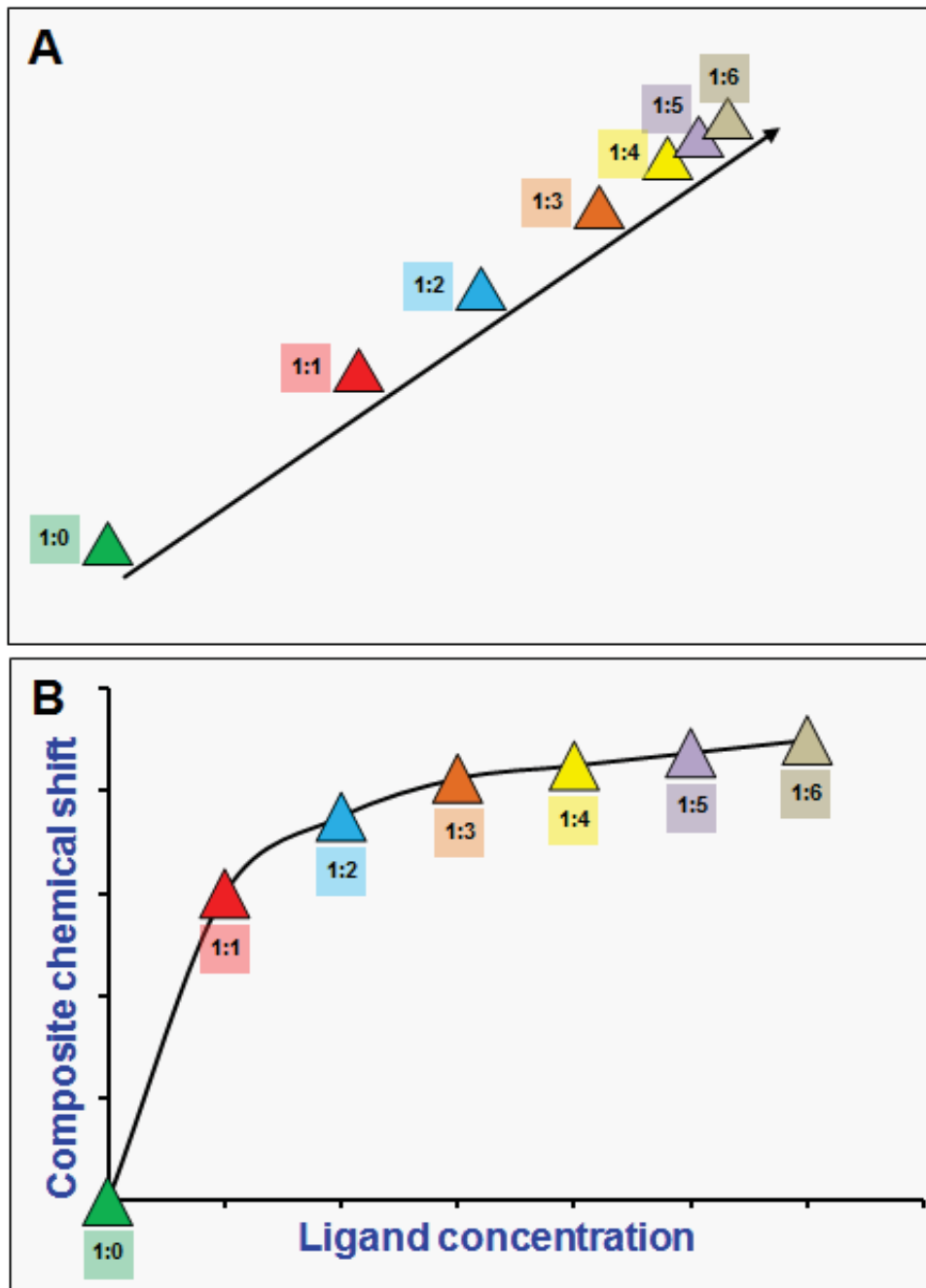


Figure 2-8. Example of chemical shift perturbation plot.

(A) In fast exchange the peak continues to move until saturation is reached on titration with higher concentrations of the ligand. (B) The composite chemical shift perturbation can be plotted against increasing ligand concentration to obtain the titration curve. The titration curve reaches a plateau as saturation is reached with higher concentrations of the ligand. The dissociation constant can be calculated from the titration curve by a nonlinear least-squares analysis.

The titration curve reaches a plateau as saturation is reached with higher concentrations of the ligand as shown in **Figure 2-8B**. The dissociation constant can be calculated from the titration curve by a nonlinear least-squares analysis using **Equation 2-10**.

$$\Delta\delta = \Delta\delta_{\max} \left(([L] + [P] + K_d) - \sqrt{([L] + [P] + K_d)^2 - 4[P][L]} \right) / 2[P] \quad \text{Equation 2-10}$$

where [L] is the concentration of the ligand, [P] is the concentration of the protein, $\Delta\delta$ is the observed chemical shift change and $\Delta\delta_{\max}$ is the normalized chemical shift change at saturation.

The K_d values determined by NMR are often received with scepticism because NMR derived K_d values are typically higher suggestive of weaker binding when compared to K_d values determined by other orthogonal methods such as Surface Plasmon Resonance (SPR) or Isothermal Calorimetry (ITC). The probable cause of the error in K_d measurements by NMR could be attributed to the averaging effect during NMR experiments. The K_d values determined by NMR are therefore treated as the upper limit of interaction. Nevertheless, the K_d values obtained for different ligands by NMR can be compared. The NMR methodology has an additional advantage when compared to other methods because using this method the specific residues involved in the protein-ligand interaction can be determined.

2.6.5. Protein-ligand Complex Structure Determination by Nuclear Magnetic Resonance

The complex structure of the protein with its ligand which may be of physiological origin or may be a synthetic drug by far provides the maximum information about protein-ligand interaction. The interaction between the protein and the cognate ligand forms the basis of cellular signal transduction mechanisms. The complex structure of the ligand with the protein receptor provides a better understanding of the molecular details of signalling pathways. Furthermore, the ligand binding site on the receptors is frequently targeted for drug development to combat ectopic signalling associated with disease conditions. Therefore it is important to determine protein-ligand complex structures. The predominant method for determining the protein-ligand complex structure is X-ray crystallography; however sometimes it is challenging to attain the ideal crystallization conditions for the protein and the ligand especially if the interaction is weak. In such cases NMR provides an excellent alternative method for determining the complex structure of the protein and the ligand.

There are two methods of complex structure determination by NMR. Firstly, the complete complex structure can be determined based on intermolecular NOEs between the protein and the ligand. For complete complex structure determination by NMR, measuring the intermolecular NOEs between the protein and the ligand is an important prerequisite. If the exchange is slow indicating strong interaction $^{15}\text{N}/^{13}\text{C}$ - filtered NOESY experiments are often used to measure the intermolecular NOEs. If the exchange

is fast suggesting weak interaction a transfer NOE experiment is used to measure the intermolecular NOEs. Additionally Residual Dipolar Coupling (RDC) experiments may be performed to determine the orientation of the ligand with respect to the protein[144].

Secondly, NMR constraints can be used to dock or model the ligand on the protein using software programs such as High Ambiguity Driven Biomolecular Docking (HADDOCK)[145]. HADDOCK uses mutagenesis as well as biophysical and biochemical data to model the structure of the ligand on the receptor protein. Prior to docking the structure of the individual components, i.e. the ligand and the receptor should be determined. Additionally there should be no major change in conformation upon protein- ligand interaction. The binding surface can be determined by the above described CSP strategy or by mutagenesis. The binding surface input information such as the CSPs are defined is as Ambiguous Interaction Restraints (AIRs) in HADDOCK. The binding surface residues can be categorized as active or passive residues. The active residues include the amino acids which depict the maximum CSP with a solvent accessibility greater than 50%. The passive residues include the other amino acids which depicted smaller CSP or are near the active residues with solvent accessibility greater than 50%.

The software HADDOCK integrates the known structural information of the individual components with the binding surface information to generate the complex structure of the receptor and the ligand. In most docking protocols one of the binding partners is rotated or translated with respect to the other to generate different binding poses. Each pose is scored using an empirical scoring function. Empirical scoring functions use several terms such as polar interactions, hydrogen bonding, entropy and aromatic interactions which are critical for defining protein-ligand interactions. The problem associated with such docking protocols is that the conformational sampling space becomes very large with increase in the size of the target. The software HADDOCK proves to be advantageous over other docking protocols because along with the above described scoring function it also incorporates experimental binding information such as CSP data and mutagenesis data to maximize the efficiency and accuracy of docking[145].

CHAPTER 3. NUCLEAR MAGNETIC RESONANCE ASSIGNMENTS AND SOLUTION STRUCTURE OF THE SMOOTHENED CYSTEINE RICH DOMAIN*

3.1. Introduction

Smoothened is the signal transducer of the Hh pathway and is a canonical member of the FzD class of GPCRs[13-16, 62, 63]. Accordingly Smo retains seven membrane spanning helical domains linked through extracellular and intracellular loops. Despite the perpetual interest in the Smo receptor there was no structural evidence on this protein till last year. In 2013, the GPCR network elucidated the first crystal structure of the 7TM domains of Smo along with the conformation of the extra-cellular loops (ECL) and an extracellular linker domain (ECLD), both of which were demonstrated to be stabilized by disulfide bonds between conserved cysteine residues[64, 77]; however, the major portion of the Smo extracellular domain (ECD) is the cysteine rich domain (CRD), and this domain was not included in the reported crystal structure of the Smo receptor.

The CRDs of the FzD family GPCRs possesses modest homology and the cysteines in this domain are conserved as shown in **Figure 3-1**[65, 146]. While the CRD of FzD plays an essential role in Wnt ligand binding and receptor dimerization, the role of the Smo CRD is not clear; however, it has been shown that the Smo CRD is indispensable for Hh signalling in *Drosophila* and required for high levels of signalling in vertebrates[65, 66, 80, 81]. *Drosophila* and vertebrate Smo lacking the CRD fail to cycle to the plasma membrane and primary cilium respectively to induce maximum signalling[80, 81]. Furthermore, in *Drosophila* CRD deletion mutants fail to dimerize, suggesting that the CRD may govern Smo dimerization[68].

Based upon its functional importance, and the observation that the related FzD CRD binds to Wnt, it has been speculated that the Smo CRD may facilitate binding of an as yet unidentified endogenous small molecule that modulates Smo signalling activity in the presence of Hh[80]. To test this hypothesis and gain structural and functional understanding of this domain, we determined the structure of the *Drosophila* Smo CRD by solution NMR spectroscopy. Our findings show that Smo CRD is a structured domain predominated by alpha helical secondary structures stabilized by disulfide bonds and maintains the tertiary fold similar to the FzD CRD[147].

*Chapter 3 adapted with permission. Rana R, Carroll CE, Lee H-J, Bao J, Marada S, Grace CRR, et al. Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nat Commun* 2013,4.

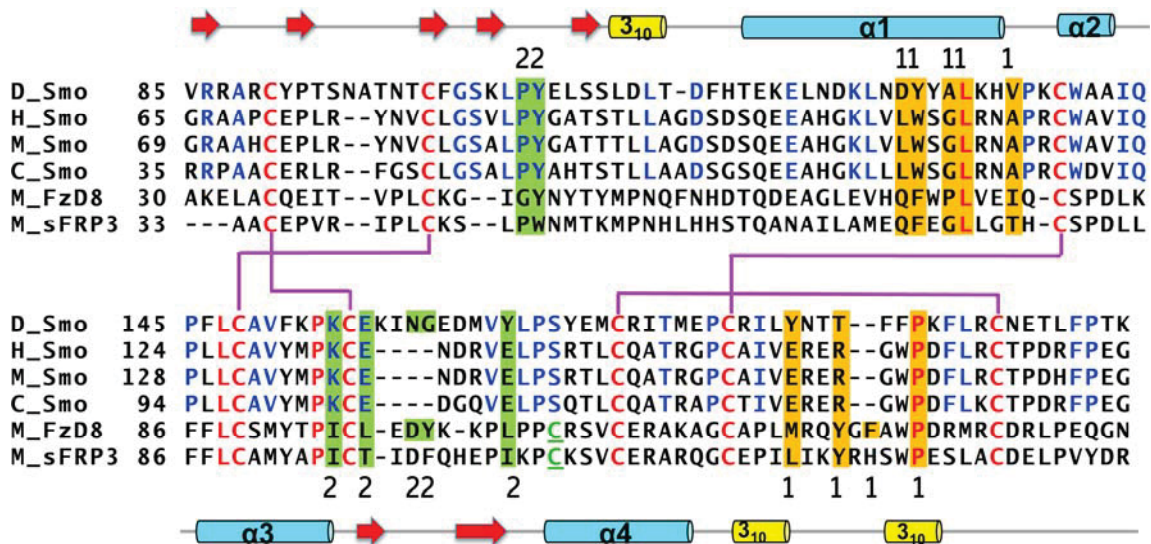


Figure 3-1. Primary sequence alignment of Smo CRD.

The primary sequence alignment of the Smo CRD and the FzD CRD. (D: *Drosophila*, H: human, M: mouse, C: chicken). The residues conserved in the Smo family are shown in blue whereas the residues colored in red are conserved both in the Smo and the FzD family. The purple lines indicate the disulfide bond pattern. The residues highlighted in yellow correspond to the “site 1” and the residues highlighted in green correspond to the “site 2” binding site in FzD-Wnt interaction. The secondary structure elements are represented above and below the primary sequence.

3.2. Experimental Procedures

3.2.1. Vector Construction

The *Drosophila* Smo CRD (V85-K202) was amplified from *pET-SmoN*[14]. The forward and the reverse primers used were 5'- ATA TAT CAT ATG GTC CGA CGT GCC CGT TGC- 3' and 5'- ATA TAC TCG AGT TAT TTC GTC GGA AAG AGT G- 3' respectively. The PCR amplified product was cloned in *pET-28a (+)* and validated by nucleotide sequencing. The human Smo CRD (G65-G177) was amplified from the *smo* cDNA purchased from Origene. The forward and the reverse primers used were 5'-TAT ATC ATA TGG GCC GGG CTG CCC CCT GC -3' and 5'- ATA TAT CTC GAG TTA GCC TTC AGG GAA GCG -3' respectively. The PCR amplified product was cloned in *pET-28a (+)* and validated by nucleotide sequencing.

3.2.2. Protein Expression and Purification

For the expression of both the *Drosophila* and human recombinant protein *pET28a-smoCRD* was transformed into BL21 (DE3) cells. The cells were grown in MOPS (3-(N-morpholino) propane sulfonic acid) minimal media supplemented with ¹⁵N ammonium chloride (1gm/liter) and ¹³C glucose (2gm/liter) as the sole source of nitrogen and carbon respectively. The media was supplemented with vitamin (10ml/liter of 10X Kao) and kanamycin (0.03gm/liter).

Cells were grown at 37⁰C to an optical density (OD₆₀₀) of 0.6 and expression was induced with 1mM IPTG (isopropyl β-D-1-thiogalactopyranoside), followed by overnight culturing at 37⁰C. Cells were resuspended in 20mM potassium phosphate, 500mM NaCl, 1% triton, pH 7.8, and lysed by sonication on ice. The protein of interest accumulated in the bacterial inclusion bodies, which were solubilized in 8M Urea and 10mM β-mercaptoethanol. The denatured protein solution was added drop-wise in an aqueous solution containing 1.25M NaCl, 50mM Tris, 1mM GSH, 2mM GSSG, 0.015% CHAPS, 1% glycerol, pH 9. Subsequently the diluted protein was concentrated using Amicon stirred cell concentrators and/or Pellicon XL devices (Millipore).

The concentrated protein was dialyzed in 5% acetic acid (volume/volume) and purified by reverse phase HPLC using a C18 column (Hitachi) followed by further purification by gel filtration (AKTA) on a Hiload 16/60 Superdex 200 preparative grade column (GE Healthcare). For the purpose of biophysical characterization the protein was dialyzed in 10 mM acetic-acid buffer, pH 5.

3.2.3. Circular Dichroism Spectroscopy

To gain understanding of the secondary structure of the *Drosophila* and human Smo CRD we recorded their Circular Dichroism (CD) spectra. The protein samples were

prepared in 10mM acetic-acid buffer, pH 5. The CD spectra were measured scanning in the far-UV wavelength of 240-190nm at 25⁰C with a 0.02cm path-length cell using an Aviv 62DS spectrometer (Aviv, Lakewood, NJ). A scan rate of 10nm/min was employed.

3.2.4. NMR Spectroscopy

A 250 μ M pure monomeric sample of the recombinant *Drosophila* (V85–K202) or 200 μ M pure monomeric sample of the recombinant human (G65-G177) Smo CRD was prepared in 10 mM deuterated acetic-acid buffer and 10% deuterium-oxide (volume/volume) for NMR experiments. All spectra were recorded using either ¹H, ¹⁵N labelled or ¹H, ¹⁵N, and ¹³C labelled protein on Bruker Avance 800 MHz or 600 MHz NMR spectrometers equipped with ¹H/¹⁵N /¹³C detect, triple resonance cryogenic probes at 25⁰C.

All spectra were processed using Topspin (Bruker Biospin) 3.0 NMR software and analysed using the program, CARA[131]. The backbone chemical shift assignment were obtained using the standard triple- resonance assignment strategy using the 2D ¹H-¹⁵N HSQC and 3D HNCA, HNCACB, CBCA(CO)NH, HN(CA)CO and HNCOC NMR experiments. The aliphatic side chain resonance assignments were obtained using HBHA(CBCACO)NH and (H)CCH-TOCSY. Distance constraints for the structure calculation were derived from ¹⁵N; ¹³C resolved ¹H-¹H NOESY with a mixing time of 100 ms.

3.3. Results and Discussion

3.3.1. Circular Dichroism Spectra

The CD spectra are widely used for the analysis of protein secondary structure. The shape, intensity and position of maxima and minima recorded for the *Drosophila* and human Smo CRD state that the Smo CRD fold is dominated by helices as shown in **Figure 3-2**.

3.3.2. One Dimensional Proton NMR Spectra

The 1D-¹H spectra of both *Drosophila* and human Smo CRD show that the proteins are folded. The 1D spectrum of the *Drosophila* Smo CRD is illustrated in **Figure 3-3**. The peaks are sharp and narrow. They are well dispersed covering a large range of chemical shifts (-0.5-10 ppm). The amide region encompasses chemical shifts spreading across 4 ppm, suggestive of a folded protein. Peaks are observed far out near negative chemical shifts suggestive of methyl groups. All these features are characteristic of folded protein spectra.

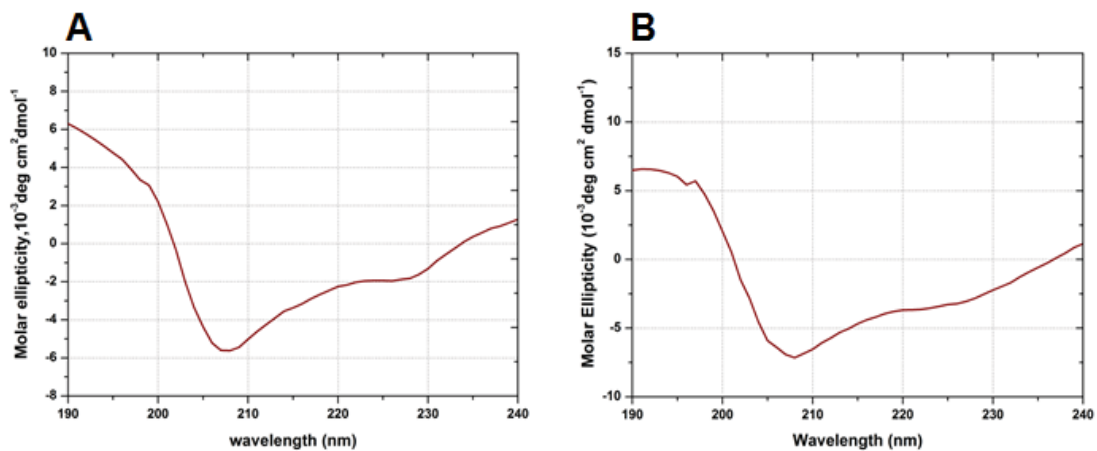


Figure 3-2. The CD spectra of the Smo CRD.

(A) The CD spectrum of the *Drosophila* Smo CRD at 25⁰C in 10mM acetic-acid buffer, pH 5. (B) The CD spectrum of the human Smo CRD at 25⁰ C in 10mM acetic-acid buffer, pH 5.

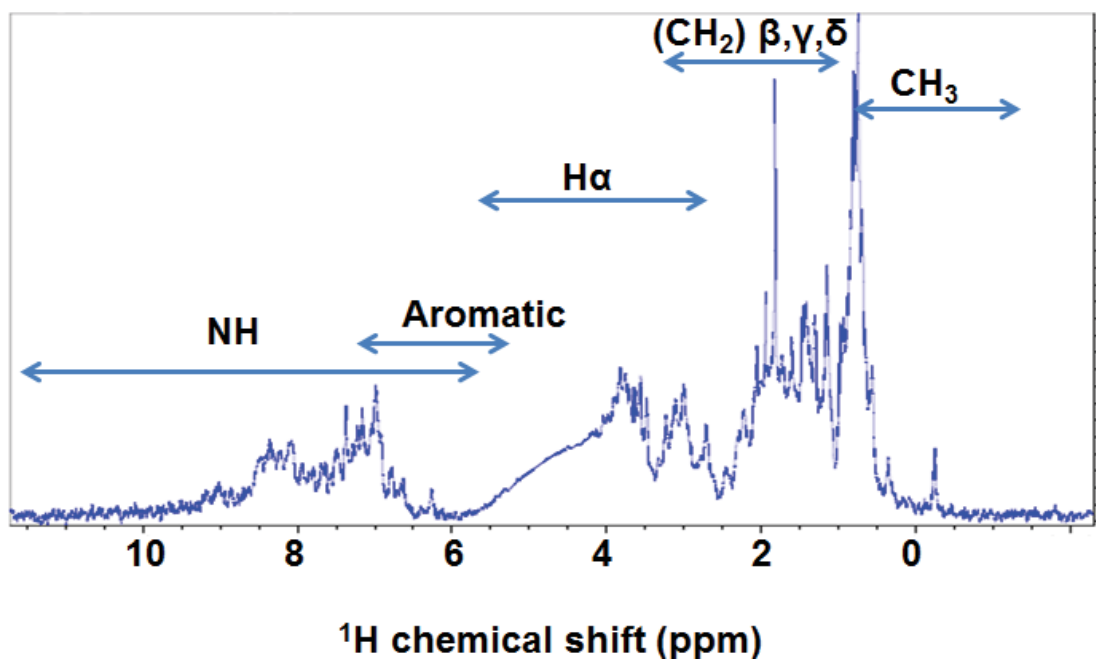


Figure 3-3. The 1D proton spectrum of *Drosophila* Smo CRD.

The 1D proton spectrum of Smo CRD shows widespread signals in the amide region that is indicative of folded protein. The signals arising from the different chemical environment in the protein are labelled on the spectrum.

3.3.3. Two Dimensional ^1H - ^{15}N and ^1H - ^{13}C Heteronuclear Single Quantum Coherence Spectra and Nuclear Magnetic Resonance Assignments

In the 2D ^1H - ^{15}N HSQC experiment the correlation between each amide proton and the connected amide nitrogen atom is measured. Except for proline, each amino acid in the protein has one amide pair in the backbone, which in principal should be detected in a 2D ^1H - ^{15}N HSQC experiment. Additionally the amide side chains of the asparagine and glutamine amino acids are also observed. The dispersion of the amide peaks for both the *Drosophila* and human Smo CRD in the 2D HSQC suggests that the Smo CRD is folded. The 2D ^1H - ^{15}N HSQC spectrum of the *Drosophila* Smo CRD is shown in **Figure 3-4A**.

Resonance assignments were performed using the standard triple resonance strategy for ^{13}C , ^{15}N labelled proteins[148]. In the 2D ^1H - ^{15}N HSQC spectrums, all the amide resonances of the 118 residues of the *Drosophila* Smo CRD (V85-K202) could be observed except D116. D116 is in the loop region preceding the first helix and hence could be undergoing motions in the intermediate time scale resulting in peak broadening for the backbone amide; nevertheless, all the side chain resonances of this residue were observed. The atom assignments of the *Drosophila* Smo CRD are listed in **Appendix A**. In the 2D ^1H - ^{15}N HSQC spectra of human Smo CRD (G65-G177), all the amide resonances of the 113 residues were assigned except two residues, L73 and R74. The atom assignments of the human Smo CRD are listed in **Appendix B**.

In the 2D ^1H - ^{13}C HSQC experiments the correlation between the aliphatic carbon and its attached proton is measured. The 2D ^1H - ^{13}C HSQC spectrum of the *Drosophila* Smo CRD is shown in **Figure 3-4B**. Assigning the peaks in the 2D ^1H - ^{13}C HSQC spectrum is also known as side chain assignment. Side chain assignments are a prerequisite for structure determination by NMR and were performed for the *Drosophila* Smo CRD. Approximately 94% of the peaks were assigned using HBHA(CBCACO)NH and (H)CCH-TOCSY NMR experiments.

3.3.4. Nuclear Overhauser Effect Assignments Structure Calculation and Refinement

The NOE or the distance constraint experiments were performed once the backbone and side chain assignments were completed. The NOE restraints were derived from ^{15}N ; ^{13}C resolved ^1H - ^1H NOESY with a mixing time of 100 ms. In the protein tertiary structure, due to protein folding, amino acids that may be far apart in the primary sequence are brought in close proximity. The NOE NMR experiments establish connectivity between two protons if they are within a distance of 5Å, irrespective of whether they are connected through bonds or just close together in space. The NOE experiments are very crucial for determining the 3D structure of the protein by NMR methods. An NOE strip plot for a helical domain of *Drosophila* Smo CRD encompassing residues N128-A132 is shown in **Figure 3-5**. The chemical structure of the representative amino acids with the detailed labelling of the atoms is shown in **Figure 3-6**.

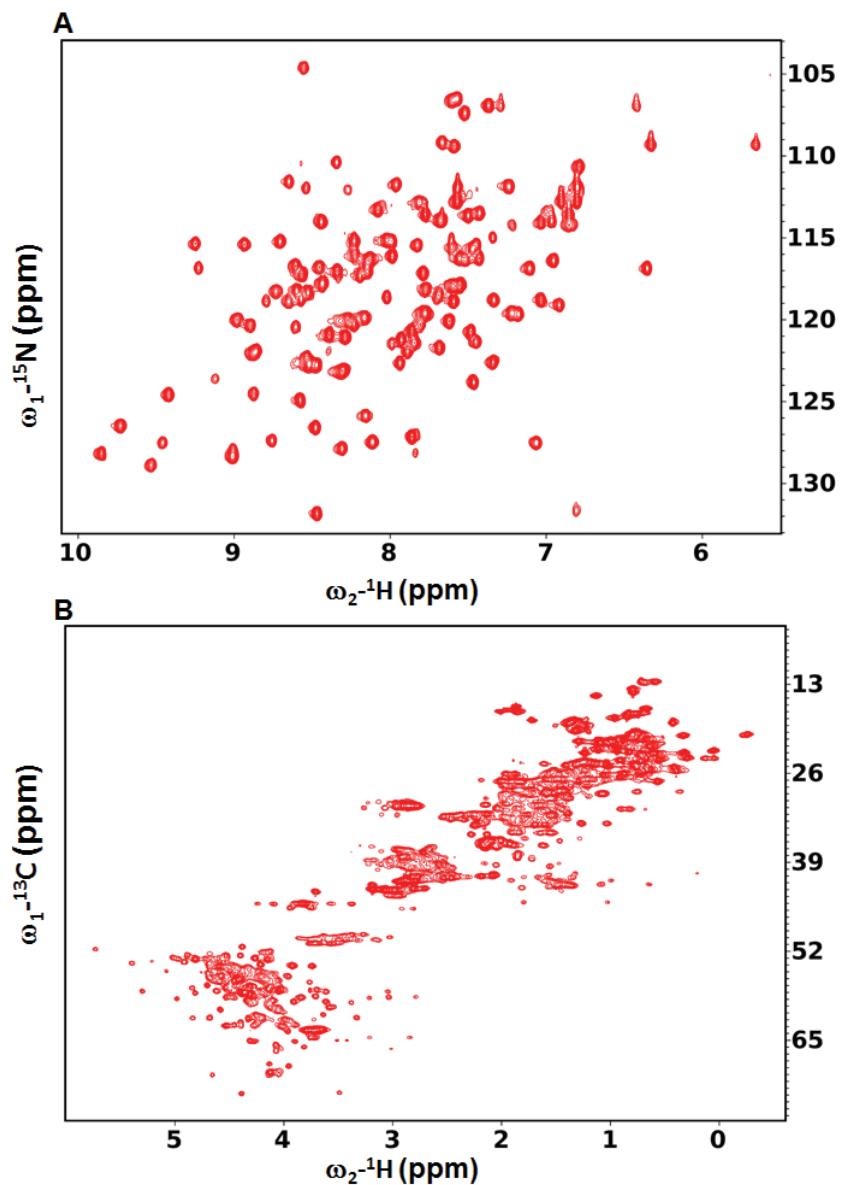


Figure 3-4. The 2D spectra of the *Drosophila* Smo CRD.

(A) The ^1H - ^{15}N HSQC spectrum of the *Drosophila* Smo CRD. The wide dispersion of the amide peaks in the Smo CRD is suggestive that the purified protein is folded (B) The ^1H - ^{13}C HSQC spectrum of the *Drosophila* Smo CRD.

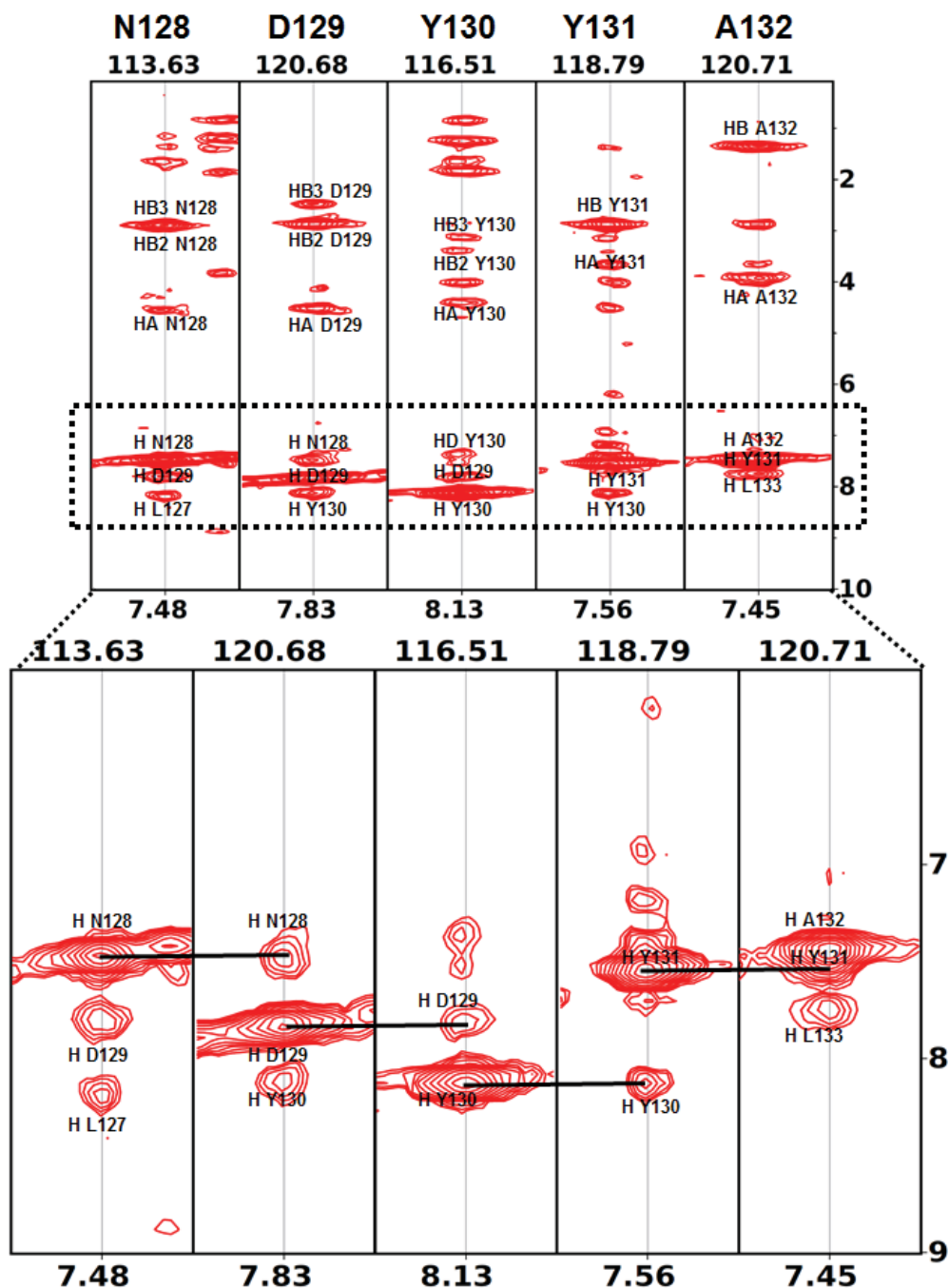


Figure 3-5. The NOE strip plot representing helical region of *Drosophila* Smo CRD. The NOE strip plot was obtained from ^{15}N resolved ^1H - ^1H NOESY spectrum. The strip represents five amino acids encompassing residues N128-A132. The peaks corresponding to the self and neighboring amino acids are labelled. The unlabelled peaks correspond to long range NOEs. In the lower inset the sequential connectivity of the amide protons characteristic of a helical region is shown in black lines.

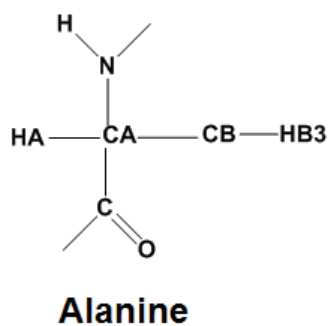
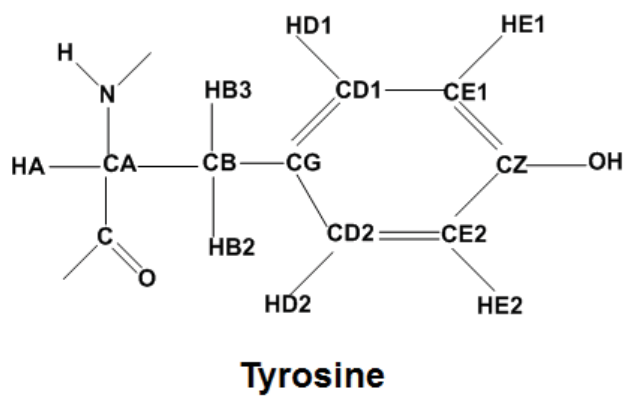
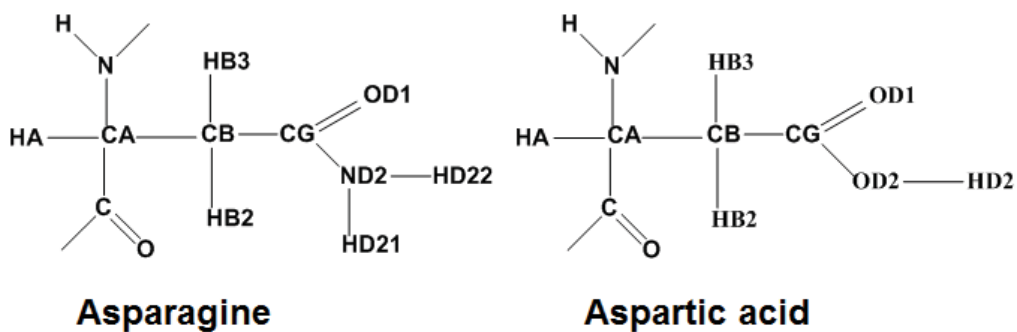


Figure 3-6. Chemical structure depicting the atoms in an amino acid.
 The structures of the amino acids represented in the NOE strip are illustrated.

An automated program UNIO was used for the NOE assignment and all these assignments were manually checked and modified[134]. Integrated NOE peaks were calibrated and converted to distance constraints using the program CALIBA[133]. The NOE restraints and 139 dihedral angle constraints were used as input for structure calculation using the program CYANA[133]. The dihedral angles were generated by the program TALOS+ which uses six kinds (HN, HA, CA, CB, CO, N) of chemical shift values for a given residue as input to give the ψ and ϕ angles[135]. A total of 100 conformers are initially generated and 20 conformers with the lowest energy are used to represent the 3-D NMR structure. Structure calculation using NMR methods is an iterative process. With the addition of an increasing number of correct NOEs, the root mean square deviation between the twenty structures with the lowest energy decreases as is shown in **Figure 3-7**. During the final structure calculation a total of 1421 meaningful NOE distance constraints were used. The structure was further refined by adding 22 hydrogen bond constraints and 4 disulfide bond constraints. The stereoview of the 20 structures with the lowest energy are shown in **Figure 3-8**. The ensemble of the final structures were checked and validated by PROCHECK. Structure figures for representation were generated using the software Pymol.

3.3.5. Solution Structure of the *Drosophila* Smo CRD

The determined solution structure of the *Drosophila* Smo CRD is of high precision. The statistical details for the NMR structure calculation are summarized in **Table 3-1**. The average pairwise root mean square deviation of the 20 structures from the lowest energy structure for the backbone and heavy atoms comprising the four helices (V85-E177) is 0.4Å and 1.28Å respectively. The NMR constraints, chemical shift data and the structure co-ordinates of the *Drosophila* Smo CRD have been deposited in the Protein Data Bank (PDB) and Research Collaborator for Structural Bioinformatics (RCSB) with accession codes 2MAH and RCSB103411 respectively.

The structure has no distance violation greater than 0.03 Å and no angle violations greater than 1°. The Ramachandran plot of the calculated Smo CRD structure enlisted 72.5% in the most favourable region, 25.7% in the additionally allowed region, 1.6% in the generously allowed region and 0.2% in the disallowed region. The residues in the disallowed region of the Ramachandran plot included a stretch of three amino acids serine, histidine and methionine, which are part of the histidine tag used to express the recombinant protein. The flexible histidine tag is positioned N-terminal to the Smo CRD and is not represented in any of the figures presented in this dissertation.

The Smo CRD possesses a well-defined tertiary structure predominated by alpha helical secondary structures. The structure with the lowest energy is used to describe the secondary structure elements as shown in **Figure 3-9A**. The globular fold of the Smo CRD is predominated by four helical bundles (α 1: E120-H135, α 2: C139-A142, α 3: F146-K154, α 4: S168-E177). The first helix is longest of the four helices, encompasses 16 residues and accommodates bulky aromatic residues. The second helix is the shortest owing to the positioning of P145.

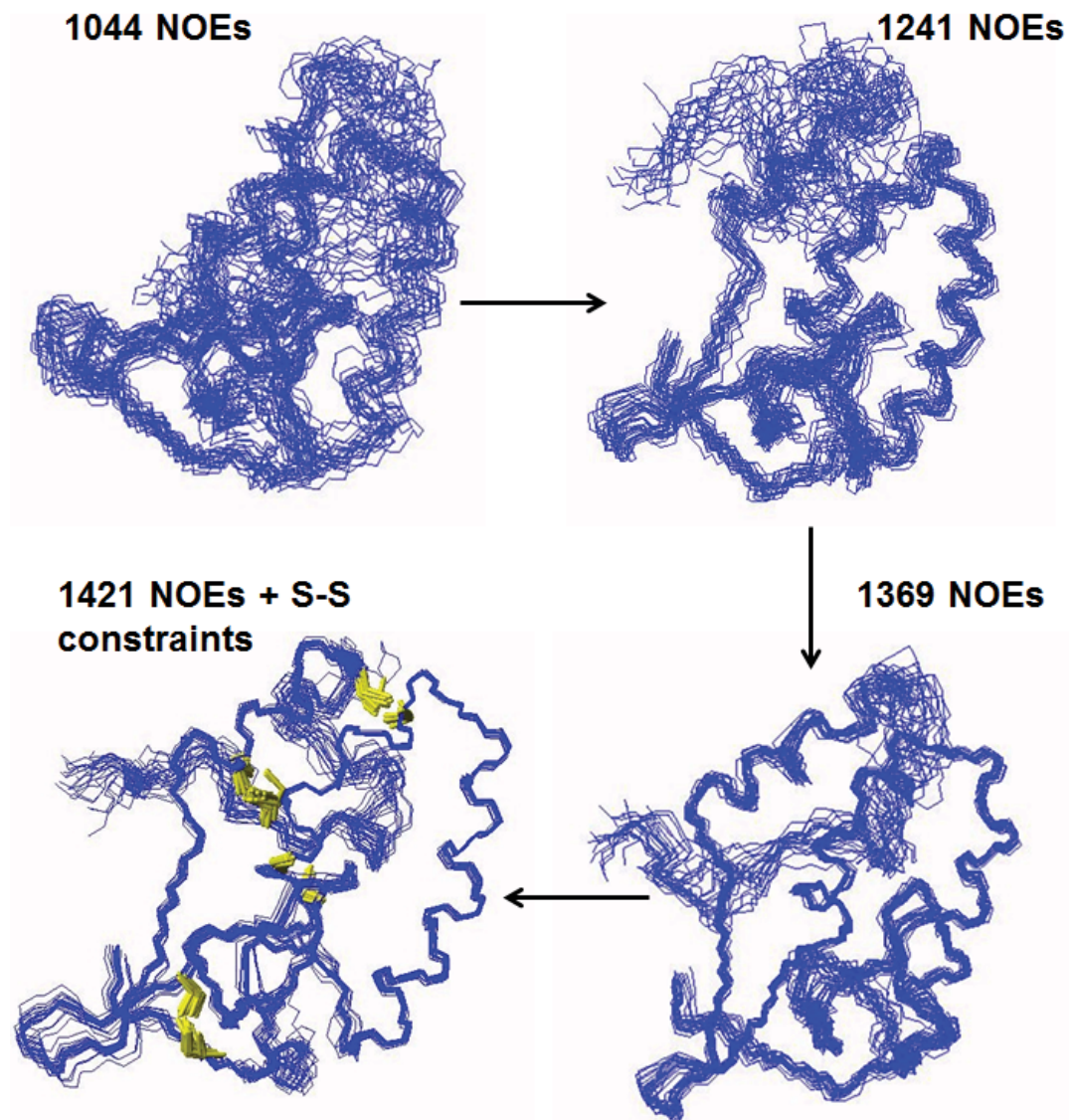


Figure 3-7. Structure refinement of the *Drosophila* Smo CRD.

Structures of the Smo CRD at various stages of NOESY assignment, structure calculation and refinement. The root mean square deviation between the 20 structures decreases with an increasing number of correct NOE distance constraints. The side chains of the cysteines forming the disulfide bonds are shown in yellow in the final structure. Each structure is an ensemble of the 20 structures with the lowest energy.

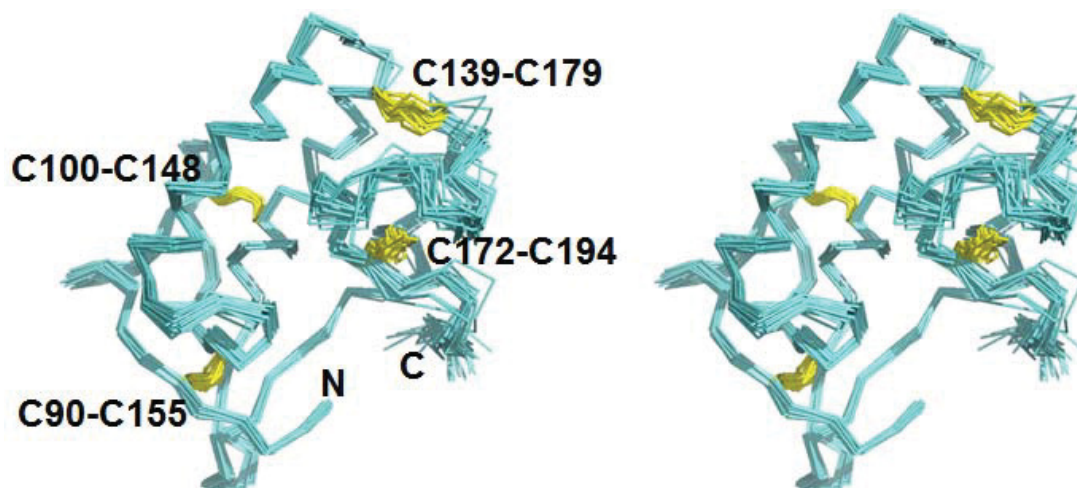


Figure 3-8. Solution structure of the *Drosophila* Smo CRD.
Stereo-view of the backbone atoms (N, C α , C') of the ensemble of the 20 structures with the lowest energy.

Table 3-1. NMR and refinement statistics used for calculating the structure of *Drosophila* Smo CRD.

NMR Distance and Dihedral Constraints	Values
Distance Constraints	
Total NOE	1421
Intra-residue	430
Inter-residue	
Sequential ($ i - j = 1$)	422
Medium-range ($ i - j < 4$)	248
Long-range ($ i - j > 5$)	321
Hydrogen Bonds	22
Total Dihedral Angle Restraints	
ϕ	71
ψ	68
Structure Statistics	
Violations (mean and s.d.)	
Distance Constraints (Å)	0.0176±0.0019
Dihedral Angle Constraints (°)	0.7607±0.0849
Max. Dihedral Angle Violation (°)	0.9547±0.0849
Max. Distance Constraint Violation (Å)	0.0238±0.0152
Average Pairwise r.m.s. Deviation (Å)	
(Val85-Lys202)	
Heavy	1.28±0.19
Backbone	0.7±0.15
Average Pairwise r.m.s. Deviation (Å)	
(Val85-Glu177)	
Heavy	1.08±0.19
Backbone	0.4±0.15

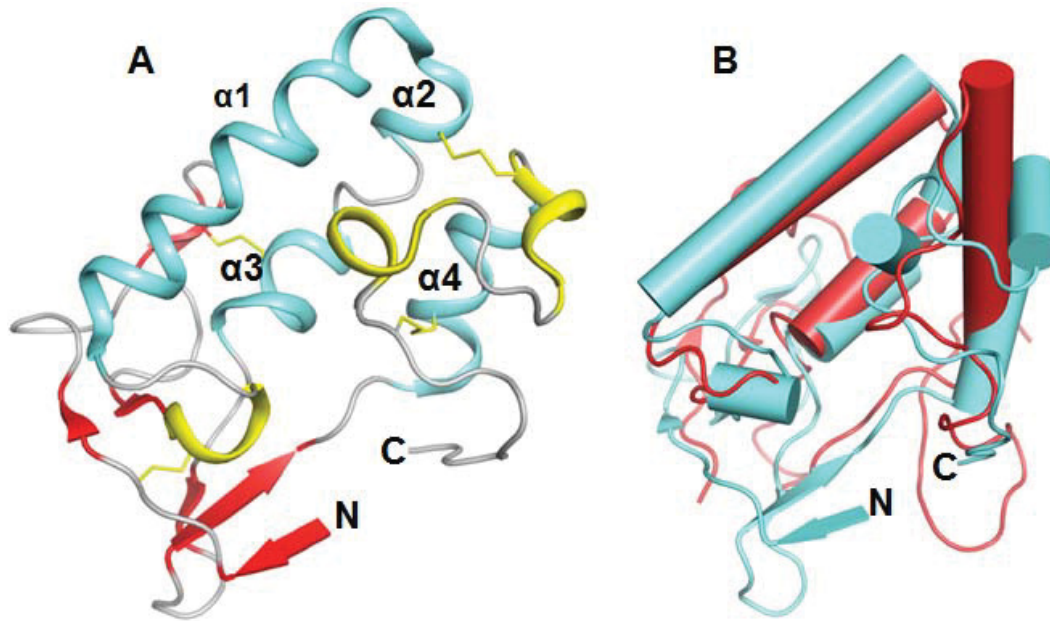


Figure 3-9. Structure of the *Drosophila* Smo CRD and comparison with the mouse FzD8 CRD.

(A) Ribbon representation of the Smo CRD. The alpha helices are labelled in order of their appearance in the primary structure. The disulfide bonds are indicated in yellow. (B) Superimposition of the Smo CRD with the FzD CRD. The Smo CRD is represented in cyan and the FzD CRD is coloured red. The helices in both proteins are represented as cylinders and the beta strands are represented as arrowheads.

The rigid backbone of the proline residue and the bulky hydrophobic side chain causes a break in the second helix. The third helix originates consecutive to P145 and includes 9 residues. The fourth helix is positioned between P167 and P178 and terminates in a 3_{10} helix. Although the Smo CRD is predominantly helical, it includes short stretches of N-terminal anti-parallel beta strands.

The Smo CRD contains eight conserved cysteines. NMR analysis depicted that the cysteine C-alpha chemical shifts and the cysteine C-beta chemical shifts are suggestive that all of the cysteines in the Smo CRD are oxidized[149]. The eight conserved cysteines form four disulfide bonds. The cysteine disulfide bond pairs were determined based on intermolecular NOEs observed between the β -protons of the cysteine residues and is as follows: C90-C155, C100-C148, C139-C179 and C172-C194. In addition to the eight conserved cysteines in the Smo CRD, the Smo extracellular domain contains five additional cysteines at *Drosophila* positions 84, 203, 218, 238, 242. In the 3D structure of the *Drosophila* Smo CRD the N- and C- termini are in close proximity, implying that C84 and C203 may form a disulfide bond. C218, 238 and 242 are predicted to be present in the ECLD connecting the Smo CRD to the first transmembrane domain. The cysteines in the Smo CRD play an important role in stabilizing the tertiary structure of this domain.

3.4. Discussion

The structure of the Smo CRD is predominated by alpha helices with distant relationship to the four-helix bundle. The tertiary structure of the *Drosophila* Smo CRD is similar to the tertiary structure of the FzD8 CRD as shown in **Figure 3-9B**[65, 66]. Despite a minimal similarity in the primary sequence the cysteines in the FzD and Smo CRD are conserved except for C107 in the FzD CRD, which is not conserved in the Smo CRD. The C107 in the FzD8 CRD forms a disulfide bond with C148[65, 66]. This represents the fifth disulfide bond in the FzD receptors, which is different in case of the Smo receptors. The FzD CRD contains ten conserved cysteines, which form five disulfide bonds[65, 66]. The *Drosophila* Smo CRD structure determined in this dissertation includes eight conserved cysteines, which form four disulfide bonds. The four disulfide bonds of Smo CRD determined here are also conserved in the FzD CRD[65, 66]. *Drosophila* C84 and C203 are located respectively at the N- and C-terminus of the structure determined in this study and may form a disulfide bond. Thus the CRDs of both the Smo and the FzD receptors may be stabilized by five disulfide bonds stabilizing the CRD to form the tight bundle of four helices.

Recently, the structure of the zebrafish Smo CRD was determined by crystallographic methods[150]. The construct (V29-Y212) used in the zebrafish study included all the thirteen cysteines present in the Smo ectodomain. The first ten cysteines form the canonical CRD and the last three cysteines form part of the extracellular linker connecting the extracellular domain structures and the first transmembrane domain. Although the zebrafish construct included all thirteen extracellular cysteine residues, the electron density was observed for the region including the ten cysteine residues

encompassing residues F41-Q158[150]. The structure of the human Smo extracellular linker was elucidated in an independent study focussing on the human Smo 7TM domains[64]. This structure showed that the first two cysteines of the linker form a disulfide bond and the last cysteine of the linker forms a disulfide bond with a cysteine in the ECL1 stabilizing the extracellular structures in proximity to the 7TM structures[64]. The tertiary structure of the determined *Drosophila* Smo CRD and the zebrafish Smo CRD is similar and the root mean square deviation between the two structures is 2.04 Å. The four disulfide bonds of the *Drosophila* Smo CRD are conserved in the zebrafish Smo CRD[150]. Additionally a fifth disulfide bond is formed by a conserved N-and C-terminal cysteine residue[150].

Thus the structure of the CRD has been conserved both in the FzD and Smo receptors and also across vertebrates (zebrafish Smo) and invertebrates (*Drosophila* Smo), suggesting that this domain may play an important evolutionarily conserved role. Given the fact that the FzD CRD plays a role in ligand binding we evaluated the ligand binding properties of the Smo CRD as described in the subsequent chapters.

CHAPTER 4. GLUCOCORTICOIDS MODULATE HEDGEHOG SIGNALLING[†]

4.1. Introduction

Glucocorticoids function as steroid hormones and are characterized by the classical steroid structure of four cycloalkane rings attached to each other[151]. The glucocorticoid nomenclature is derived from “glucose + cortex + steroids” because these groups of organic molecules play an important role in ‘glucose’ metabolism; are synthesized in the ‘adrenal cortex’ and retain the structure of the ‘steroids’[152]. Glucocorticoids are known to function by binding to the glucocorticoid receptor and play an important role in regulating immune response, metabolism, cell growth and proliferation, development and reproduction[153]. The anti-inflammatory properties of the glucocorticoids have led to their widespread use as therapeutics[154]. Glucocorticoids have also been used in the treatment of cancer mainly to combat the cancer-associated inflammation.

Recent reports have identified a variety of glucocorticoids which regulate Hh signalling by acting at the level of the Smo receptor[93]. The identified glucocorticoids were categorized as activatory or inhibitory glucocorticoids depending on their effect on Smo sub-cellular localization and transcription of Hh target genes. The activatory glucocorticoids such as flucinolone acetonide (FA) and triamcinolone acetonide (TA) facilitated Smo ciliary translocation and increased downstream Hh signalling[93]. In contrast the inhibitory glucocorticoids such as Budesonide (Bud) and ciclesonide (Cic) inhibited Smo ciliary translocation and suppressed downstream Hh signalling[93]. Additionally, the inhibitory glucocorticoids acted on wild-type as well as Smo mutants: SmoM2 and SmoD473H[93]. SmoM2 is a constitutively active form of Smo, first identified in BCC patients and is resistant to treatment with available Smo antagonists[74]. The D473H Smo mutant was identified in a MB patient being treated with the Smo antagonist GDC-0449 and was associated with tumor recurrence[97]. Furthermore, the inhibitory glucocorticoids were shown to modulate Smo function by binding to a site distinct from the orthosteric site present in the cavity of the 7TM core, previously demonstrated to bind the inhibitory small molecule cyclopamine[84, 85, 93].

In accordance we hypothesized that the Smo extracellular CRD may act as an allosteric binding site for glucocorticoids like Bud and additional Smo modulators. To test this hypothesis we analyzed the interaction between the glucocorticoid Bud and the *Drosophila* and human Smo CRD by NMR CSP experiments [143, 155]. Our results demonstrate that Bud binds to both *Drosophila* and human Smo CRD, albeit with different binding affinities[147].

[†] Chapter 4 adapted with permission. Rana R, Carroll CE, Lee H-J, Bao J, Marada S, Grace CRR, et al. Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nat Commun* 2013,4.

Furthermore, we generated the modelled structure of *Drosophila* Smo CRD in complex with the inhibitory glucocorticoid Bud, which showed that Bud docks in a conserved hydrophobic pocket in Smo CRD.

4.2. Experimental Procedures

4.2.1. Protein Expression and Purification

¹⁵N labelled *Drosophila* and human Smo CRD was expressed and purified as described in section 3.2.2.

4.2.2. Nuclear Magnetic Resonance Chemical Shift Perturbation Analysis

Chemical shift perturbation experiments were carried out using ¹⁵N labelled *Drosophila* or human Smo CRD prepared in 10 mM deuterated acetic-acid buffer and 10% deuterium-oxide (volume/volume). ¹H-¹⁵N HSQC spectra of the Smo CRD were recorded in the absence and presence of an increasing concentration of the glucocorticoid Bud at 25°C using Bruker Avance 600 MHz spectrometer equipped with ¹H/¹⁵N /¹³C detect, TCI triple resonance cryogenic probes. All data were processed using Topspin (Bruker Biospin) 3.0 NMR software and analysed using the program CARA[131].

A stock solution of 50 mM Bud (Sigma) was made in deuterated dimethyl sulfoxide for use in NMR titration experiments. The concentration of dimethyl sulfoxide in the NMR titration experiment was maintained at or below 1%. A control experiment was performed by titrating 1% dimethyl sulfoxide to *Drosophila* Smo CRD and no CSPs were observed.

The dissociation constant (K_d) values were calculated by a nonlinear least-squares analysis in the program Origin, using the **Equation 4-1**

$$\Delta\delta = \Delta\delta_{\max} \left(\frac{([L] + [P] + K_d) - \sqrt{([L] + [P] + K_d)^2 - 4[P][L]}}{2[P]} \right) \quad \text{Equation 4-1}$$

where [L] is the concentration of the glucocorticoid ligand, [P] is the concentration of the protein, $\Delta\delta$ is the observed chemical shift change and $\Delta\delta_{\max}$ is the normalized chemical shift change at saturation, calculated using **Equation 4-2**

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N / 5)^2} \quad \text{Equation 4-2}$$

Where, $\Delta\delta$ is the chemical shift in p.p.m.

4.2.3. Structural Modelling of *Drosophila* Smoothened Cysteine Rich Domain in Complex with Budesonide

The complex structure of the *Drosophila* Smo CRD and Bud was generated using the software High Ambiguity Driven protein-protein Docking (HADDOCK)[145]. The simulated annealing protocol in the program CNS was used to calculate the complex structure of the protein and the small molecule[145]. Topology and parameter files for Bud were generated using xplo2d and the charge parameters of Bud were calculated by using the antechamber module in the AMBER10 software package[156]. We used ambiguous restraints resulting from the CSPs of ^1H - ^{15}N Smo CRD on titration with Bud as input for the program HADDOCK[145]. To define the active and passive residues, the relative accessible surface area was calculated by the program NACCESS[157]. One thousand initial structures of the *Drosophila* Smo CRD in complex with Bud were generated. The structure with the lowest energy was selected and this was consistent with the CSPs observed in Smo CRD on titration with Bud.

4.2.4. Bio-Layer Interferometry Assay

The Bio-Layer Interferometry (BLI) assay (Octet RED instrument from FortéBio) was used to measure the interactions of the *Drosophila* and human Smo CRD with the glucocorticoid Bud. The purified protein was prepared in 0.1 M potassium phosphate buffer, pH 6.5 and was subsequently biotinylated. For the process of biotinylation 100 μl of 0.2 mM Smo CRD was incubated with 200 μl of 1.0 mg/ml EZ-Link NHS-LC-Biotin (Thermo Sci.) for 15 hours on a rocking platform at room temperature. Excess unreacted biotin was removed by dialysis in 0.1 M potassium phosphate buffer pH 6.5.

Super streptavidin (SSA) sensors were used to immobilize the biotinylated target protein on the optical sensors. Prior to the start of the assay the SSA sensors were pre-wet in 0.1M potassium phosphate buffer pH 6.5 with 0.01 % Trion-X100 for 10 minutes. For the binding affinity assay 3-5 μM biotinylated protein was immobilized on the SSA sensor by incubating for 30 minutes. Subsequently the free streptavidin sites were quenched by incubating with 0.5 mM biocytin (Thermo Sci.) for 10 minutes. Excess protein and biocytin were removed by washing with 0.1M potassium phosphate, pH 6.5 for 10 minutes.

To determine the binding affinity of the *Drosophila* and the human Smo CRD for Bud, kinetic experiments were carried out using three different concentrations of Bud (62, 41, and 31 μM). The association step for 50s was followed by the dissociation step for 100s. The assay buffer used was 0.1 M potassium phosphate, pH 6.5 with 0.5% dimethyl sulfoxide. Biocytin-loaded SSA sensor without loading biotinylated protein was used as a control to correct the systematic optical artefacts and baseline drifts. All steps were performed at 30 $^{\circ}\text{C}$ with 1000 rpm rotary shaking. The processed data were fitted locally with the integrated fitting function using the 1:1 binding model in FortéBio analysis software. The K_d values were calculated from curve fitting.

4.2.5. Cell-based Luciferase Reporter Assays

Cell-based Hh activity reporter assays were performed in mammalian Shh Light II cells[83, 86]. Shh Light II cells are derived from NIH-3T3 cells and co-transfected with Gli-responsive firefly luciferase reporter[58]. The stable clonal cell line was selected using the aminoglycoside antibiotic Geneticin (G418). The resulting cell line was transfected with pRL-TK constitutive Renilla-luciferase expression vector (Promega) and pVgRXX vector (Invitrogen) which codes for the ecdysone receptor and the glycopeptide antibiotic Zeocin resistance marker. Antibiotic resistance and cell cloning was used to generate Shh Light II cells. Presently the Shh Light 2 cell lines have been patented and we received the cells as a generous gift from the Ogden lab at St. Jude Children's Research Hospital.

For treatment with the drug Bud, Shh Light II cells were cultured to confluency in a 96-well plate and then treated with various concentrations of Bud in DMEM containing 0.5% (volume/volume) bovine calf serum. Bud was added to cellular media from concentrated stocks in dimethyl sulfoxide. After 48 hours, renilla and firefly luciferase levels were measured using Dual Luciferase Reporter Assay kit (Promega). Hh pathway activity was expressed as the ratio of firefly to renilla luciferase.

Experiments were performed three times in duplicate and error bars represent the standard error of the mean.

4.3. Results and Discussion

4.3.1. *Drosophila* Smo CRD Binds to the Synthetic Glucocorticoid Budesonide

Glucocorticoids have been reported to play a role in regulating Smo ciliary localization and downstream Hh signalling[93]. Budesonide has been described to function as an inhibitory glucocorticoid and stated to bind to a site different from the canonical binding site in the cavity of the Smo 7TM domains[93]. Therefore, we decided to test whether the *Drosophila* Smo CRD might bind to Bud. We used NMR CSPs to identify the residues that are involved in binding since this method allows us to determine precisely the amino acid residues involved in the interaction[143, 155].

In the 2D ^1H - ^{15}N HSQC spectra of the *Drosophila* Smo CRD in the absence and presence of an increasing concentration of the ligand Bud, CSPs induced by the binding of Bud were clearly observed indicating Bud bound to the CRD as shown in **Figure 4-1A**. The normalized maximum CSP for each residue of *Drosophila* Smo CRD in the presence of Bud is shown in **Figure 4-1B**. The interaction between the Smo CRD and Bud are guided by three phenylalanine residues, F188 depicted the maximum CSP, followed by F187. Although F191 demonstrated a smaller CSP when compared to F187 and F188 the CSPs on F191 were saturated on titration with an increasing concentration of Bud indicating tighter binding affinity.

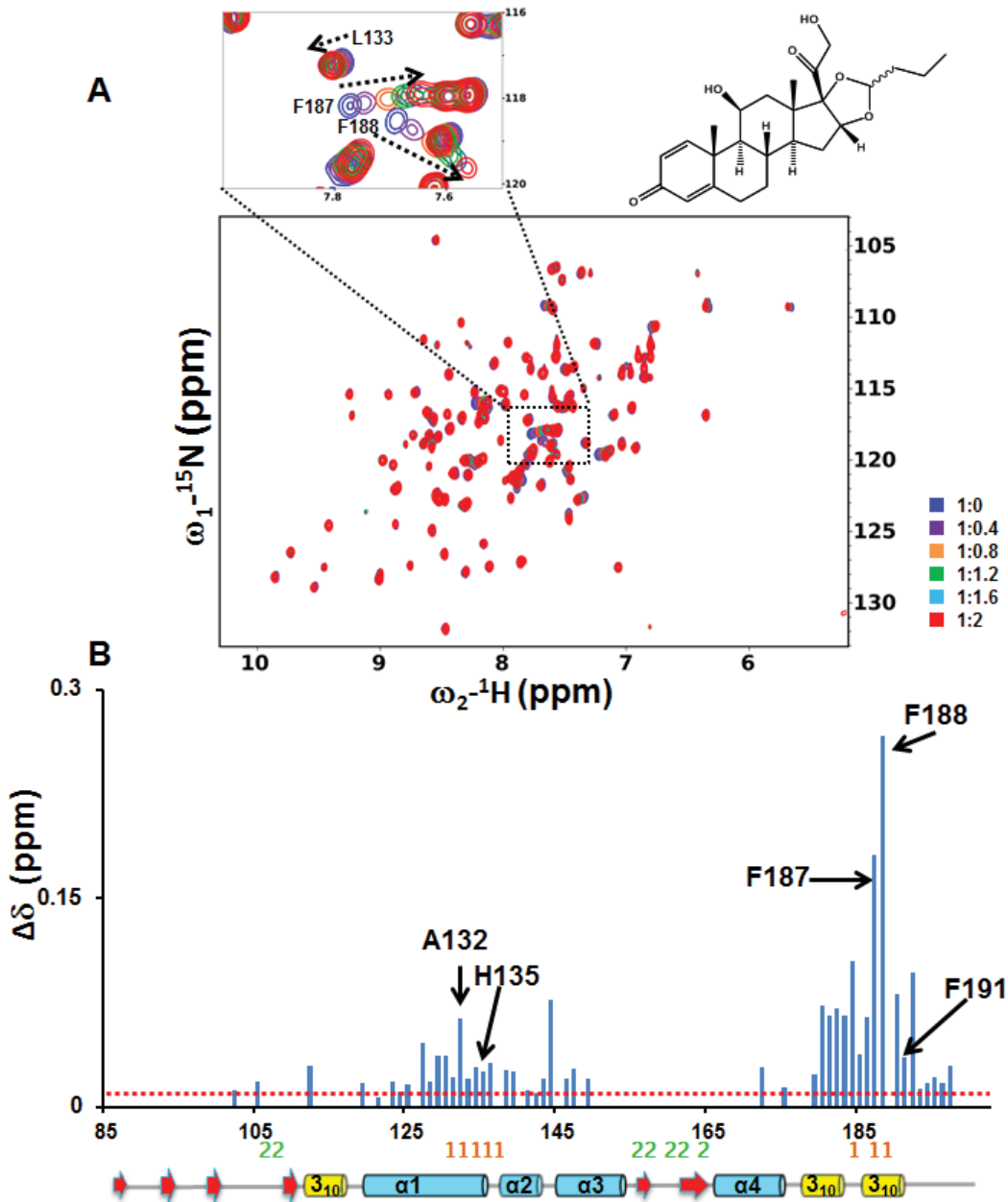


Figure 4-1. The *Drosophila* Smo CRD binds to the inhibitory glucocorticoid Bud. (A) ^1H - ^{15}N HSQC spectra depicting CSPs of Smo CRD on titration with an increasing concentration of the ligand Bud. The chemical structure of Bud is shown above the spectra. (B) The normalized maximum CSPs of Smo CRD upon addition of Bud are plotted versus residue numbers. The red line indicates CSPs greater than 0.01 ppm. The residues labelled in black form the Bud-binding pocket on Smo CRD as analyzed from the HADDOCK docking experiments. The mouse Fzd8-Wnt interacting “site 1” and “site 2” residues are shown in orange and green respectively. The corresponding secondary structure elements of the *Drosophila* Smo CRD are represented below the plot.

Phenylalanine 191 is conserved in the vertebrate Smo family and F188 is substituted by a tryptophan residue in the vertebrate Smo. The conservation of the Bud binding residues in the Smo family suggests that this pocket may have an important role in binding an as yet unidentified allosteric regulator of *Drosophila* Smo.

4.3.2. High Ambiguity Driven Biomolecular Docking of the *Drosophila* Smoothened Cysteine Rich Domain in Complex with Budesonide

The inhibitory glucocorticoid Bud binds in a hydrophobic pocket in the Smo CRD, however the details of the interaction remain obscure without a complex structure. We have solved the NMR solution structure of the *Drosophila* Smo CRD and precisely mapped the binding site of the glucocorticoid by NMR CSP titration data. Next we wanted to determine the structure of the Smo CRD in complex with Bud. Budesonide is a hydrophobic molecule and has limited solubility in an aqueous buffer. Therefore Bud cannot be added in excess amounts to the Smo CRD protein solution. This makes it difficult to obtain unambiguous distance restraints from NMR NOESY experiments required to determine the complex structure by NMR. Therefore we generated the complex structure of Smo CRD with Bud using the program HADDOCK[158].

The CSPs induced in Smo CRD on titration with Bud were used as input for complex structure calculation using HADDOCK[145]. The residues that have significant CSPs are further shown in the back-bone “ribbon” representation of the Smo CRD, wherein the thickness of the ribbon is proportional to the magnitude of CSP observed in Smo CRD on Bud binding. The backbone “ribbon” representation of Smo CRD along with the positioning of Bud is shown in **Figure 4-2A**. The results show that the residues of Smo CRD that have the largest CSPs induced by the bound Bud are located in the first helix and the C-terminal 3_{10} helical domain.

The CSP data on the *Drosophila* Smo CRD were used as ambiguous restraints for modelling the structure of the protein-ligand complex. The active residues included the amino acids which depicted the maximum CSP with a solvent accessibility greater than 50%. The passive residues include the other amino acids which depicted smaller CSPs and are near the active residues with solvent accessibility greater than 50%. One thousand binding poses were generated and the structure with the lowest energy in agreement with the CSP data was selected for representation. The surface of the binding pocket is shown in **Figure 4-2B**. The modelled structure showed that Bud docks into a hydrophobic pocket formed by the residues A132, H135, F187, F188 and F191 on the Smo CRD. The aromatic steroid rings of Bud interact with the aromatic phenylalanine rings on Smo CRD as shown in **Figure 4-2C**. The aliphatic carbon chain on Bud interacts with H135 and the aliphatic carbon chain of Bud with a free hydroxyl group interacts with A132. Thus Bud docks in the groove formed by the third and fourth helix and the 3_{10} helical domains as it folds back on the fourth helix. The details of the input and output parameters for HADDOCK determined *Drosophila* Smo CRD-Bud complex structure determination are summarized in **Table 4-1**.

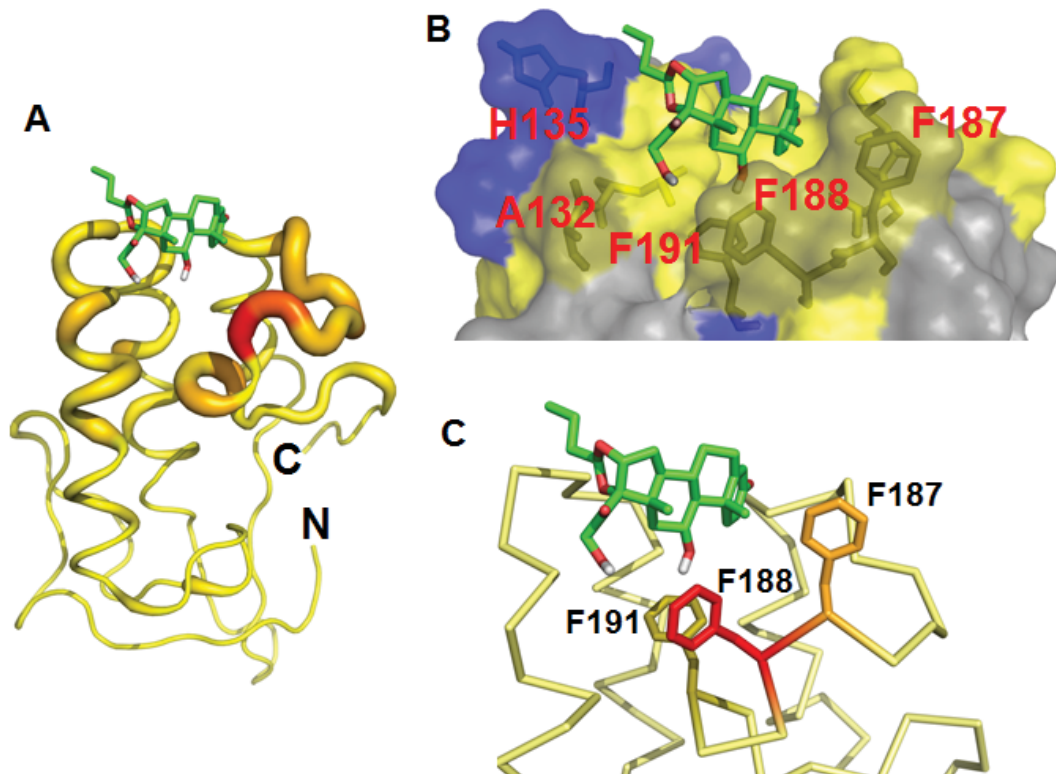


Figure 4-2. The complex structure of *Drosophila* Smo CRD and the glucocorticoid Bud.

(A) “Ribbon” representation of the Smo CRD. The backbone thickness of the ribbon is directly proportional to the weighted sum (in ppm) of the ^1H and ^{15}N CSPs upon binding to the ligand Bud. (B) Surface representation of the residues that interact with Bud. The aromatic residues are shown in yellow and the positively charged H135 is shown in blue. (C) The aromatic side chains of the Bud contacting Smo CRD residues are shown.

Table 4-1. The HADDOCK input and output parameters for docking Bud on *Drosophila* Smo CRD.

Parameters	Residue Number
Active Residues	D129, A132, F187, F188 and F191
Passive Residues	Y131, K134, K138, N184 and T186
Fully Flexible Region	F187-L192
Binding Energy	-2694 ± 37 kcal/mol
Buried Surface Area	575 ± 13 Å ²

4.3.3. Human Smoothened Cysteine Rich Domain Binds to the Synthetic Glucocorticoid Budesonide

The effect of glucocorticoids on Smo subcellular localization and subsequent Hh signalling has been studied in vertebrates[93]. In the vertebrate system it has been stated that Bud binds to Smo at a site distinct from the cavity located in the Smo 7TM domains[93]. Therefore we tested the binding of Bud to human Smo CRD.

The NMR titration experiments were performed to determine the specific residues involved in ligand binding. The 2D ^1H - ^{15}N HSQC spectra of human Smo CRD in the absence and presence of an increasing concentration of Bud showed that Bud binds to human Smo CRD as shown in **Figure 4-3A**. The normalized CSP for each residue of human Smo CRD in the presence of Bud is shown in **Figure 4-3B**. Comparing with *Drosophila* Smo CRD, Bud binds to human CRD in a similar region; however, there are some clear differences. For example, a smaller number of residues in human Smo CRD had CSPs induced by the binding of Bud, and the absolute values of CSPs observed in human Smo CRD are smaller. Furthermore, the CSP observed at the first point of titration is more than those observed at subsequent points of titration suggesting that at higher concentrations of Bud the interaction between the human Smo CRD and Bud is saturated.

The secondary and tertiary structure of the human Smo CRD can be predicted based on the structure of the *Drosophila* Smo CRD. Accordingly most of the observed CSPs localized in the flexible C-terminal loop. Arginine 161 demonstrated the maximum CSP. Some of the other residues depicting CSPs in the flexible C-terminus included E158, R159, G162 and W163. The other vicinity depicting distinct CSPs clustered in the region corresponding to the first helix encompassing residues L108, W109, G111 and L112. Although the two regions demonstrating CSPs are located faraway in the primary sequence of the protein, based on the structure of the *Drosophila* Smo CRD these regions should be held in close proximity in the tertiary structure of the protein.

4.3.4. Binding Affinity of the Glucocorticoid Budesonide for Smoothened Cysteine Rich Domain

The binding affinity of the glucocorticoid Bud for *Drosophila* and human Smo CRD was determined from NMR titration plots and BLI experiments. The residues which manifested the maximum CSPs on titration with Bud and formed the hydrophobic binding surface as inferred from HADDOCK modelling studies were selected to determine the binding affinity or K_d from NMR titration experiments. Namely the residues included were: F187, F188, F191, L127 and A132 as reported in **Figure 4-4** and **Figure 4-5**. The binding affinities of each of these residues were calculated and are mentioned in the figures. The binding affinity of the *Drosophila* Smo CRD for Bud determined from NMR experiments ranged from lower hundred to thousand micromolar, indicating that Bud binding induced backbone conformational stabilization in *Drosophila* CRD[143].

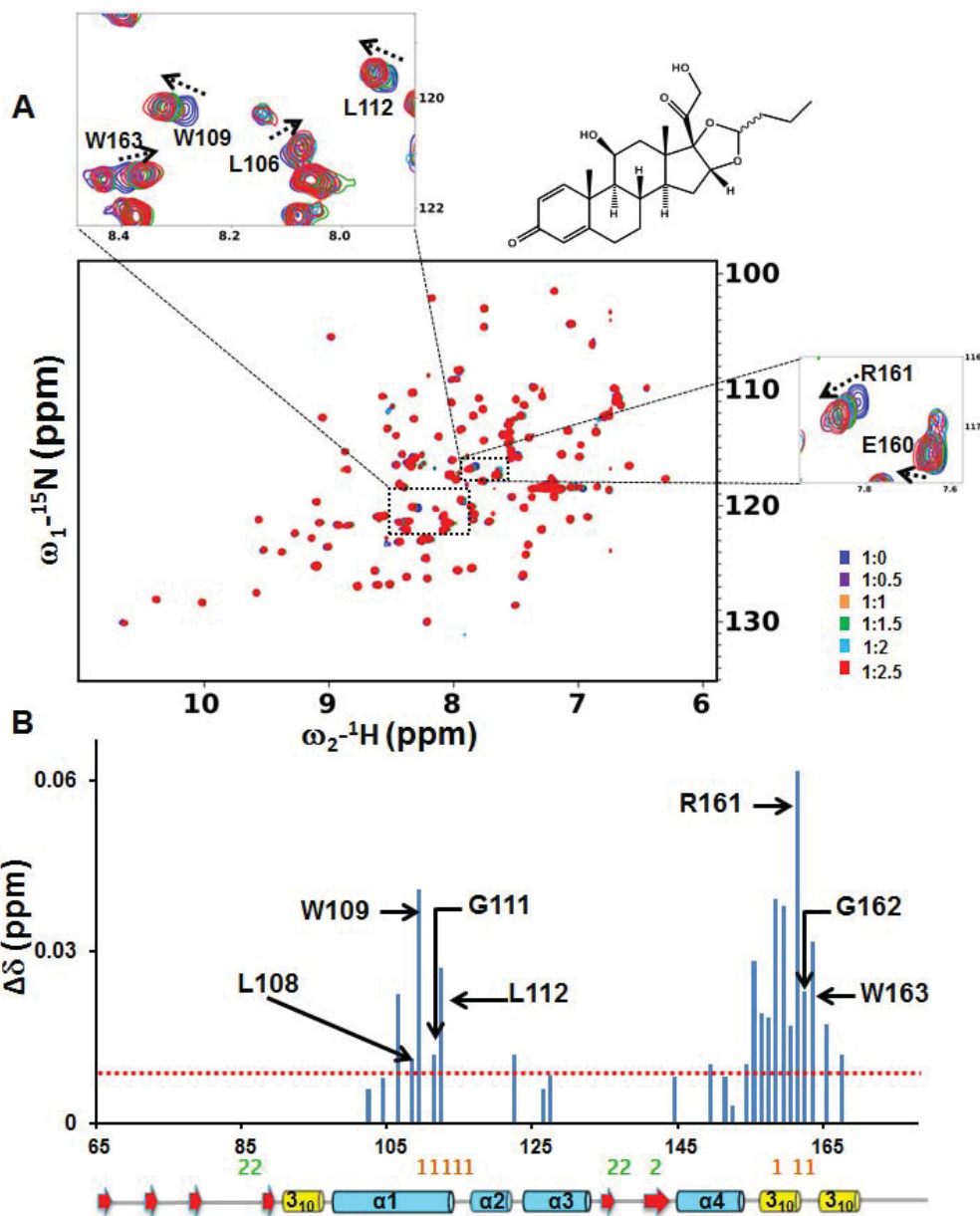


Figure 4-3. The human Smo CRD binds to the inhibitory glucocorticoid Bud. (A) ^1H - ^{15}N HSQC spectra depicting CSPs of human Smo CRD on titration with an increasing concentration of the ligand Bud. The chemical structure of Bud is shown above the spectra. (B) CSPs of human Smo CRD upon addition of Bud are plotted versus residue numbers. Dotted red line indicates CSPs greater than 0.01 ppm. Leucine 108, W109, G111, L112 and R161 are homologous to the mouse Smo residues that interact with 20-OHC. Glycine 162 is homologous to *Drosophila* F187 and W163 is homologous to *Drosophila* Smo F188 that interacts with Bud. All these residues map to the “site 1” of mouse FzD8-Wnt interaction. “Site 1” and “site 2” residues of mouse FzD8-Wnt interaction are shown in orange and green respectively. The secondary structure elements as in *Drosophila* Smo CRD structure are shown below the plot.

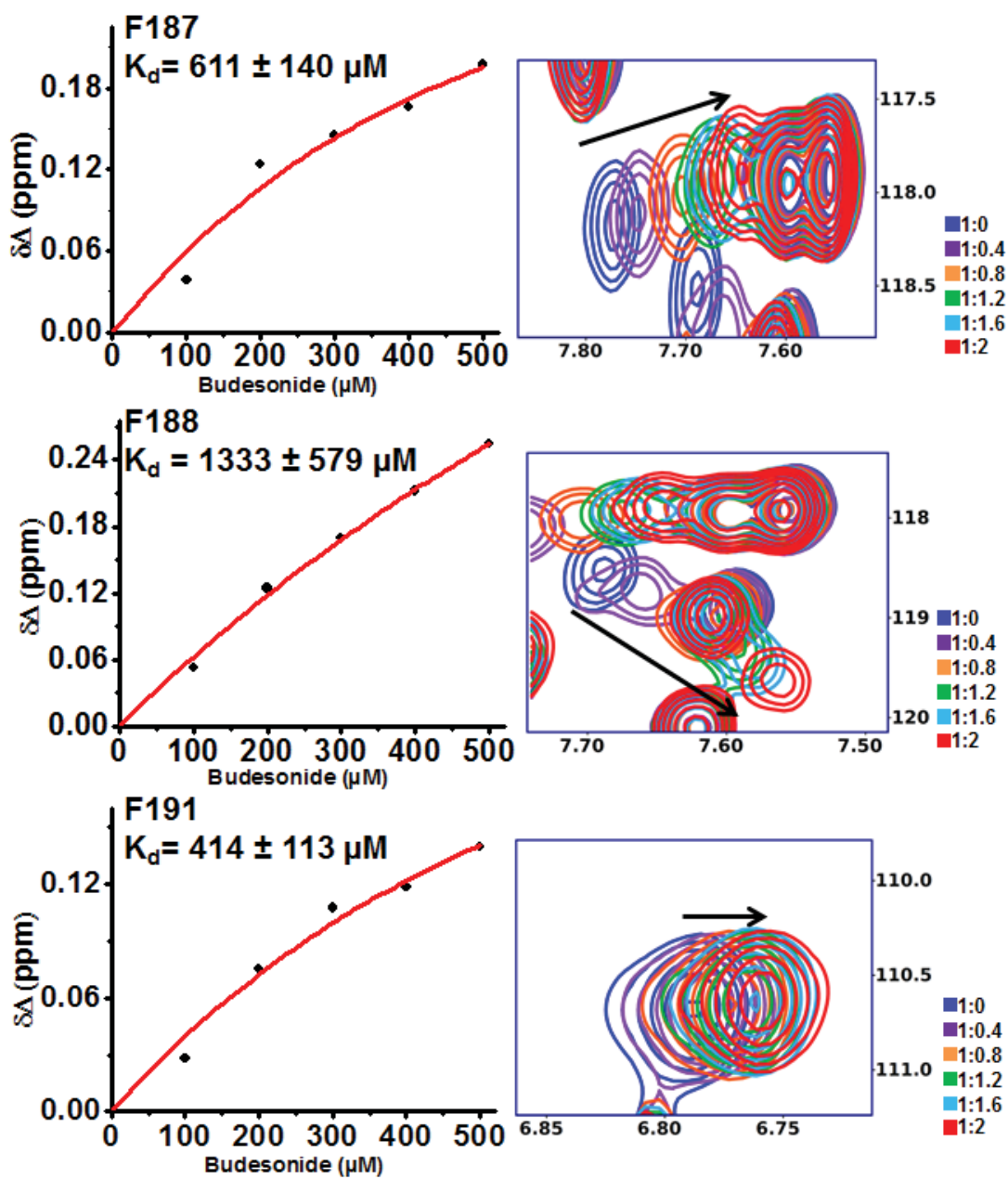


Figure 4-4. Binding affinity measurement of *Drosophila* Smo CRD and Bud for residues F187, F188 and F191.

Enlarged ^1H - ^{15}N HSQC spectra of F187, F188, and F191 during the titration of *Drosophila* Smo CRD and the ligand Bud. The spectra are colour coded according to the protein: Bud ratio. The chemical shift change is plotted against the concentration of Bud and the corresponding K_d values are shown.

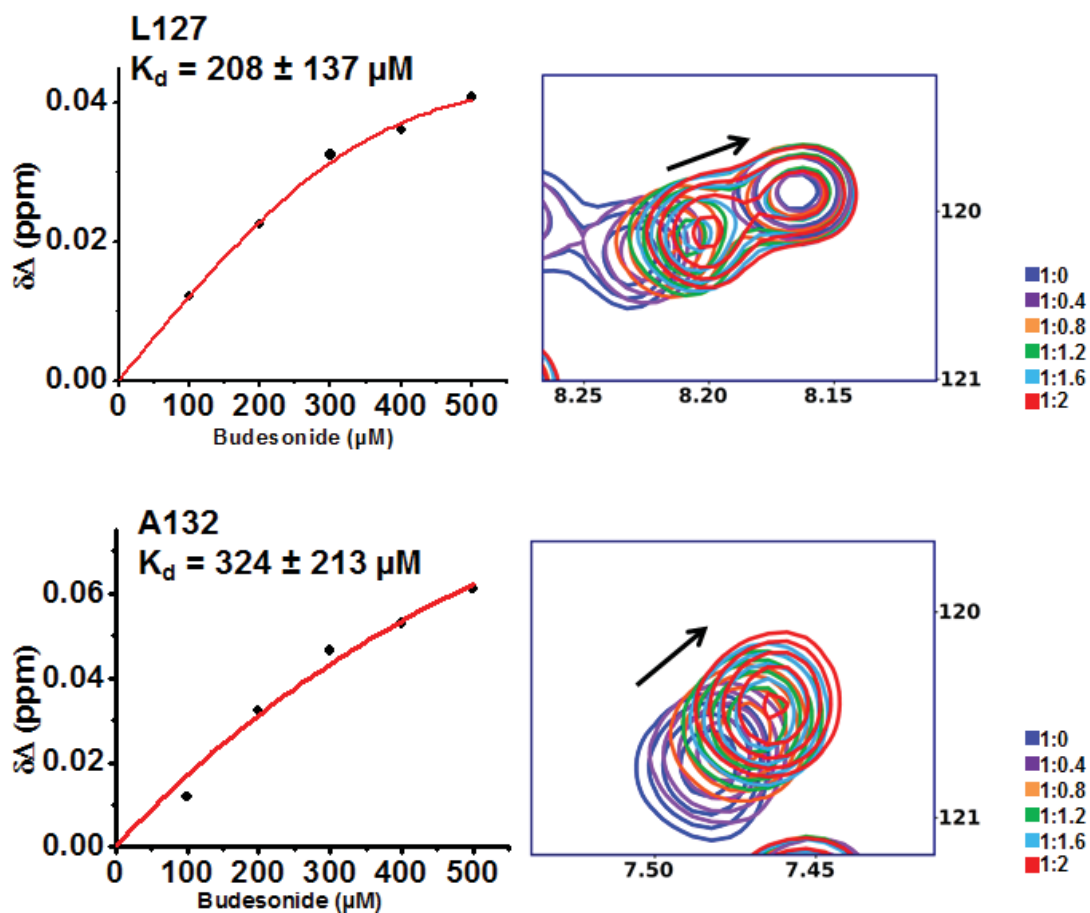


Figure 4-5. Binding affinity measurement of *Drosophila* Smo CRD and Bud for residues L127 and A132.

Enlarged ^1H - ^{15}N HSQC spectra of L127 and A132 during the titration of *Drosophila* Smo CRD and the ligand Bud. The spectra are colour coded according to the protein: Bud ratio. The chemical shift change is plotted against the concentration of Bud and the corresponding K_d values are shown.

For determining the NMR based binding affinity of Bud for human Smo CRD the residues included were W109, L112 and R161. The binding affinities measured from these residues in human Smo CRD clustered around 45 μM as shown in **Figure 4-6**. Judged from NMR titration data, it appeared that Bud binds to human Smo CRD with higher affinity.

Although NMR is a useful method for determining specific protein-ligand interactions in solution, it is not the ideal method for calculating binding affinity[159, 160]. The associated systemic error in calculating K_d values from NMR is likely caused by the averaging effect during NMR experiments[159]. The K_d values calculated by this method can only be treated as the upper limit for interaction. Therefore we validated the binding affinity measurements using additional BLI experiments.

We determined the binding affinity of Bud for *Drosophila* Smo CRD using BLI methods. Three different concentrations of Bud were tested: 62, 41 and 31 μM and the K_d values were determined for each concentration with the integrated fitting function by a 1:1 binding model. The 62 μM concentration of Bud was repeated once, whereas the 41 μM concentration of Bud was repeated thrice and the 31 μM concentration of Bud was repeated twice. The K_d value obtained by curve fitting at 62 μM Bud was 89 μM . The K_d values obtained for three different repeats at 41 μM Bud were 74 μM , 59 μM and 318 μM . The K_d values obtained for the two different repeats at 31 μM Bud were 93 μM and 85 μM . Thus the average K_d value of *Drosophila* Smo CRD for Bud was calculated to be $120 \pm 98 \mu\text{M}$ as reported in **Figure 4-7**.

To determine the binding affinity of Bud for human Smo CRD three different concentration of Bud were tested 62, 41 and 31 μM . The respective K_d values obtained by curve fitting were 74 μM (62 μM Bud), 37 μM (41 μM Bud), 44 μM (41 μM Bud,) and 54 μM (31 μM Bud). The average K_d value of human Smo CRD for Bud was $52 \pm 16 \mu\text{M}$ as noted in **Figure 4-7**. Both NMR and BLI data suggest that Bud binds to human Smo CRD with a higher affinity than *Drosophila* Smo CRD.

4.3.5. Effect of Budesonide on *in-vitro* Hh Signalling

Budesonide has been described to function as a negative regulator of Smo ciliary localization and downstream Hh signalling[93]. To reinvestigate the effect of the glucocorticoid Bud on *in-vitro* Hh signalling we performed cell-based luciferase assays. To determine whether Bud would attenuate signalling by the Smo receptor, Shh Light II cells were treated with increasing concentrations of Bud ranging from 0 to 25 μM . Hedgehog signalling was induced by adding the small molecule SAG(Enzo) which acts as a positive modulator of Hh signalling[83]. There was approximately a 50% inhibition of SAG induced *Gli*-luciferase activity on treatment with Bud as shown in **Figure 4-8**. Therefore we speculate that Bud acts as a weak synthetic mimic of an endogenous ligand for the Smo CRD, displacing it from this binding pocket.

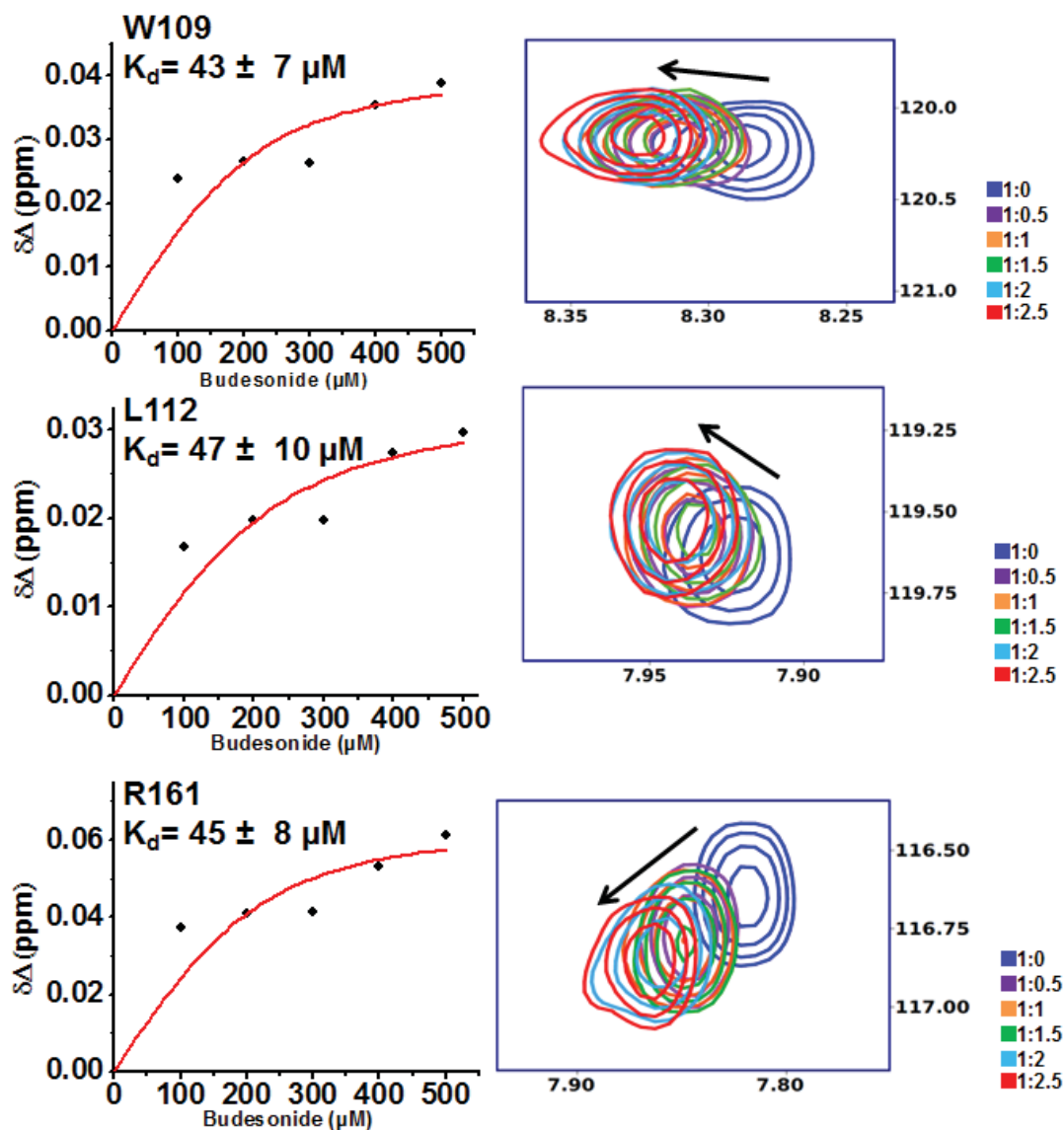


Figure 4-6. Binding affinity measurement of human Smo CRD and Bud.

Enlarged ^1H - ^{15}N HSQC spectra of W109, L112, and R161 during the titration of human Smo CRD and the ligand Bud. The spectra are colour coded according to the protein: Bud ratio. These are the same residues which interact with 20-OHC in vertebrate Smo. The chemical shift change is plotted against the concentration of Bud and the corresponding K_d values are shown.

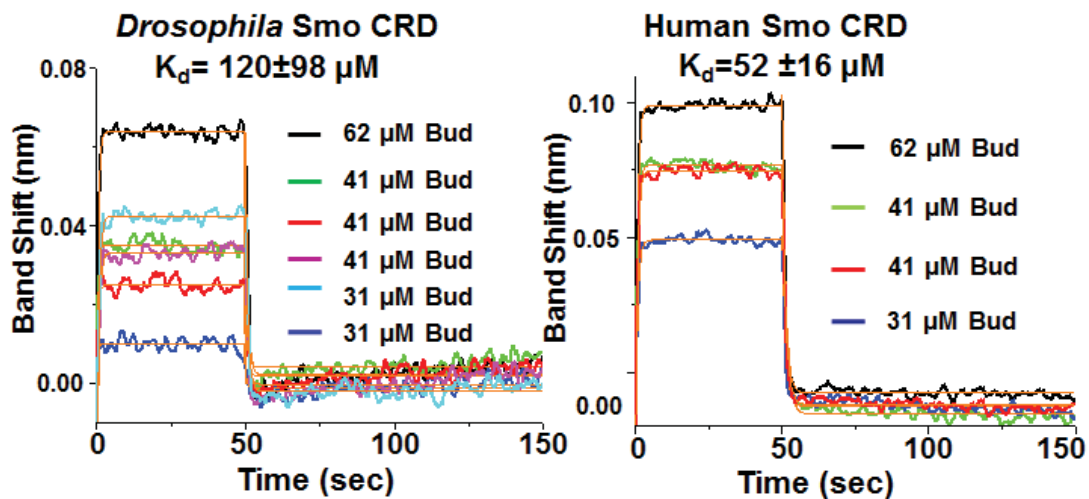


Figure 4-7. BLI binding assays show that Bud binds to *Drosophila* Smo CRD with a lower affinity than human Smo CRD.

The average K_d value of *Drosophila* Smo CRD for Bud was $120 \pm 98 \mu\text{M}$. The average K_d value of human Smo CRD for Bud was $52 \pm 16 \mu\text{M}$.

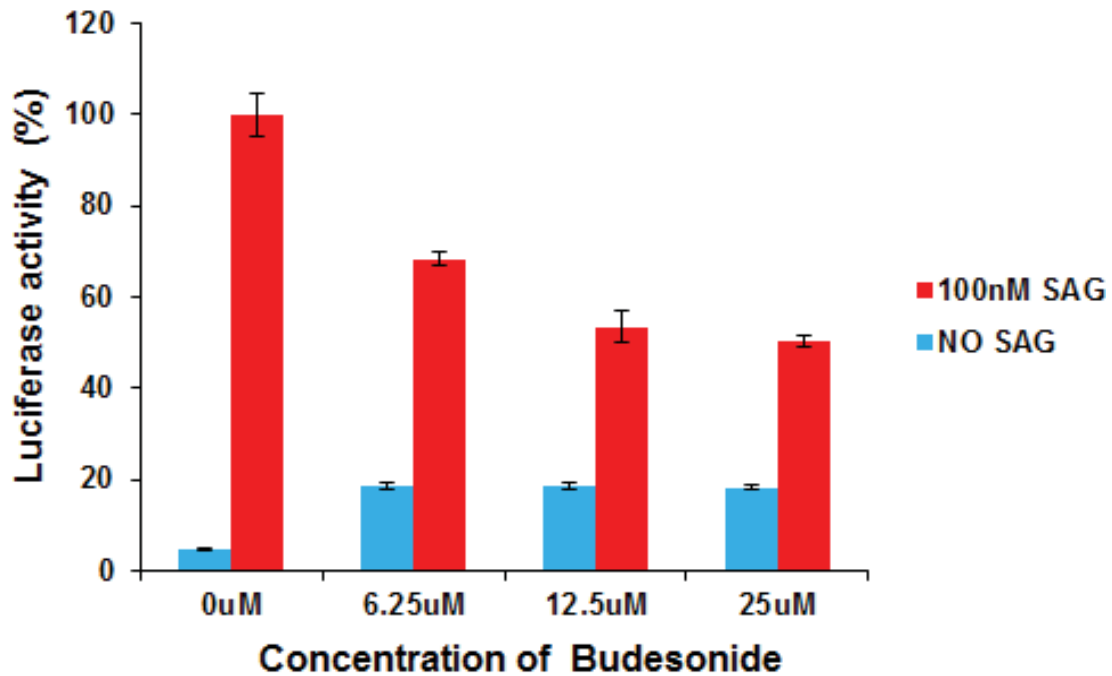


Figure 4-8. Bud attenuates SAG induced reporter gene activity.

Shh Light II cells were treated with increasing concentrations of Bud. Bud triggered a 40-50% reduction in SAG induced *Gli* luciferase activity. Experiments were performed three times in duplicate. Error bars indicate standard error of the mean (s.e.m)

4.4. Discussion

4.4.1. Cysteine Rich Domains and Hydrophobic Ligands

The Smo extracellular CRD is homologous to the extracellular CRD of the FzD receptors[65, 66, 146]. The FzD receptors function as the signal transducer in the Wnt signalling pathway[63, 65, 66]. The Wnt signal transduction cascade is an evolutionary conserved developmental pathway controlling cell growth and proliferation and abnormal Wnt signalling has been associated with cancer conditions[63]. The FzD receptors bind to the endogenous Wnt ligand through the N-terminal CRD[65, 66].

In the recently elucidated crystal structure of the Wnt8 ligand in complex with the FzD8 CRD, the Wnt molecule has been described as a “hand” and the palmitate modification on one of the cysteine residues of Wnt has been described as the “thumb” of the “hand” which protrudes out to bind in a hydrophobic groove in the FzD8 CRD[66]. This forms the first site of interaction between the FzD receptor and the Wnt ligand[66]. The second Wnt-FzD interaction site is mediated by protein-protein interactions. The second interacting surface on Wnt has been described as the “index finger” of the hand which grasps FzD CRD on a surface opposite to the first interaction site[66].

The first interaction site is dominated by hydrophobic interactions between the palmitate modification on Wnt and hydrophobic amino acid residues of the FzD CRD[66]. The nine residues of FzD CRD reported to interact with the palmitate modification are not strictly conserved in the Smo receptors; however Smo also possesses a similar hydrophobic surface in the equivalent position. In the present study, we show that the hydrophobic aromatic residues of Smo residing in a homologous region as the palmitate binding site on FzD CRD interact with the glucocorticoid Bud. This states that the Bud binding site identified in this study is similar to the lipid binding site of the FzD receptors[66]. This observation is not unexpected because it has been hypothesized that both Smo and FzD CRD have evolved from a common ancestor recognizing hydrophobic molecules[146, 161].

Glucocorticoids binding to the Smo CRD suggest that like the FzD CRD, the Smo CRD also possesses ligand binding capabilities; however Bud is a synthetic molecule which perhaps shares a similar chemical scaffold as that of the natural ligand of the Smo receptor, assuming that Smo indeed binds to a physiological ligand. In such respect contemporary research has shown that the hydrophobic pocket in the vertebrate Smo CRD binds to oxysterol derivatives[92, 95, 96].

4.4.2. Same yet Different *Drosophila* and Human Smoothened Cysteine Rich Domain

Despite a minimal primary sequence identity between the *Drosophila* and the human Smo CRD, the cysteines in this domain are conserved across species[146]. The

tertiary structure of the *Drosophila* Smo CRD determined in this study showed that the global fold of this domain is similar to the mouse FzD8 CRD structure[65, 66]. Additionally, the root mean square deviation of the *Drosophila* Smo CRD with the recently elucidated structure of the zebrafish Smo CRD is 2.04 Å suggesting that the structural fold of CRDs have been conserved across species[150].

Both the *Drosophila* and the human Smo CRD interact with the glucocorticoid Bud suggesting that the ligand binding properties of the CRD have been conserved across species. Likewise the vertebrate Smo CRD has been reported to interact with oxysterols like 20-hydroxyl cholesterol and oxysterol derivatives like azasterols, 7-keto-25-hydroxyl cholesterol and 7-keto-27-hydroxyl cholesterol[92, 95, 96]. In such studies the *Drosophila* Smo CRD has been reported insensitive to regulation by oxysterols[96]. The fact that the oxysterols may not bind to the *Drosophila* Smo CRD does not preclude the fact that the *Drosophila* Smo CRD may bind to a yet unidentified endogenous ligand. This argument is further supported by our observations that the synthetic glucocorticoid Bud binds to the *Drosophila* Smo CRD.

The CSPs observed in the NMR titration experiments both in the case of *Drosophila* and the human protein localized in the same region encompassing the residues of the first helix and the C-terminal loop which corresponds to the lipid binding site on the FzD CRD. In case of the *Drosophila* protein the aromatic phenylalanine residues played a key role in protein-ligand interaction. Of the three phenylalanines F187, F188 and F191 participating in this interaction, F187 is not conserved in the human Smo CRD and is instead a glycine residue, G162. The second phenylalanine F188 is W163 and the third phenylalanine, F191 is conserved in the human Smo CRD. Surprisingly the conserved phenylalanine in the human Smo CRD did not depict any observable CSP of the back-bone amide. The homologous tryptophan residue W163 depicted CSPs and saturation on titration with higher ratios of the glucocorticoid. The equivalent glycine residue, G162 also depicted CSPs suggesting that although the exact same residues in the *Drosophila* and the human protein may not be involved in ligand interactions, it is the same hydrophobic groove in both species that mediate protein-ligand interaction. In addition to the hydrophobic residues in the human Smo CRD, R161 showed significant CSPs on titration with the glucocorticoid Bud. In fact R161 of the human Smo CRD depicted the maximum CSP on titration with Bud suggesting that charged interactions may also play a role in modulating vertebrate Smo CRD ligand interaction. The human Smo CRD residues L108, W109, G111, L112 and R161 which depicted CSPs on titration with Bud are homologous to the mouse Smo residues that interact with the ligand 20-OHC[95, 96].

Although both the *Drosophila* and the human Smo CRD interacted with Bud, a noteworthy difference was observed between the two species. In case of the *Drosophila* Smo CRD the maximum value of the observed CSPs was four-five times higher than that observed with the human Smo CRD and also the CSPs observed in the human Smo CRD were saturated at higher concentrations indicating that the human Smo CRD possessed a higher binding affinity than the *Drosophila* Smo CRD.

4.4.3. Ligand Binding Provides a Novel Insight into Hedgehog Signalling

Smo is a canonical GPCR and likewise maintains a ligand binding pocket in the cavity of the 7TM domains[16, 62-64]. Cyclopamine, a natural product isolated from corn lily was the first identified Smo modulator known to dock in the cavity of the 7TM domains[84, 85]. Since the discovery of cyclopamine a plethora of small molecules have been identified that bind to Smo and regulate Hh signalling[82, 83, 87, 88, 162]. Although most of the identified small molecules were described to bind in the cavity of the 7TM domains there was no structural evidence till last year. Since the summer of May 2013 five structures of the Smo 7TM domains have been reported in complex with small molecules namely LY2940680, SANT-1, cyclopamine, SAG and Anta XV[64, 85, 90]. The complex structures revealed that the small molecules mostly dock into a cavity formed by the first, sixth and seventh transmembrane helices and make contacts with the residues in the extracellular loop 2[64, 85]. The different small molecules penetrate to different extents within the 7TM cavity. The small molecule SANT docks deeper into the interior of the 7TM domains whereas the other compounds partly dock in the membrane and partly projects outside the membrane layer, suggesting that these small molecules may interact with the Smo extracellular structures[64, 85].

In the present study we show that the inhibitory glucocorticoid Bud binds to both the *Drosophila* and the human Smo CRD. Therefore Smo has at least two distinct binding sites, one located in the cavity of the 7TM domains and the other located in the extracellular CRD. We speculate that some small molecules may bind to the Smo 7TM domains as well as the CRD.

G protein coupled receptor signalling is often associated with ligand binding to the extracellular domains which facilitates conformational rearrangements in the membrane domains to regulate signalling within the cell[163-165]. Likewise the Smo extracellular CRD may act as the binding site for an endogenous lipid molecule. The Smo CRD perhaps acts as “fishing bait” which serves to fetch the natural ligand. The bound ligand may then initiate conformational rearrangements which in turn would bridge the gap between the extracellular and the transmembrane structures to regulate subsequent Hh signalling within the cell.

CHAPTER 5. PHOSPHOLIPIDS AND HEDGEHOG SIGNALLING

5.1. Introduction

Phospholipids are a class of amphipathic lipids that form a major component of cellular membrane structures[166]. The chemical structures of phospholipids are comprised of two lipophilic fatty-acid chains esterified to a glycerol or sphingosine alcohol backbone and a head group attached to the hydrophilic phosphate group which in turn is attached to the alcohol backbone[166]. The fatty-acid side chains form the lipid bilayer of the membrane and the phosphate moiety faces the aqueous environment[166]. Phospholipids are divided into five major classes depending on the nature of the head group attached to the phosphate moiety. The five classes of phospholipids include: phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, phosphatidylserine and sphingomyelin[166]. In the present dissertation the major focus is on phosphatidylinositols and therefore will be discussed briefly.

Phosphatidylinositols are present in all cells and tissue types and are abundant in the brain tissues[167]. The phosphatidylinositols form a minor component of the membrane structures when compared to phosphatidylcholines and phosphatidylethanolamines[168]. The inositol head group of the phosphatidylinositols are susceptible to phosphorylation at positions 3, 4 and 5 to produce seven different types of phosphatidylinositols[167, 169]. The different phosphatidylinositols are functionally distinct and are enriched in different subcellular compartments. The specific lipid kinases and lipid phosphatases help in the inter-conversion of the phosphatidylinositols[167, 169].

Functionally, phosphatidylinositols and their derivatives play an important role in regulating cell migration, membrane trafficking and signal transduction[167, 169]. The phosphatidylinositols have also been suggested to play a role in regulating Hh signalling by acting at the level of the Smo receptor[98, 170]. Genetic experiments have shown that the loss of the kinase responsible for the synthesis of PI(4)P results in Hh loss-of-function phenotypes whereas loss of the phosphatase responsible for the degradation of PI(4)P results in Hh gain-of-function phenotypes[98].

The phosphatidylinositols possesses two fatty-acid side chains. These fatty-acid side chains are similar to the palmitate modification on Wnt that form the “site one” interaction site of the FzD-Wnt receptor ligand complex[66]. Extrapolating from the same we speculated that the fatty-acid side chains of the phosphatidylinositols may bind to Smo CRD to regulate Hh signalling. Additionally, as discussed in Chapter 4 we believe that certain Smo ligands can simultaneously engage the Smo extracellular and membrane spanning domains to regulate signalling. In such regards we hypothesized that the fatty-acid side chains of phosphatidylinositols bind to the Smo CRD and the inositol head anchors in the cavity of the 7TM domains to regulate signalling. Therefore we tested the ability of the different phosphatidylinositols to bind to human Smo CRD. We selected the synthetic phosphatidylinositols with fatty-acid side chains of eight carbon

atoms (octanoic-acid) so that it is readily soluble in an aqueous buffer and convenient for NMR experiments. Pertaining to our hypothesis we observed that the phosphatidylinositols irrespective of the number and the position of the phosphate groups interact with the human Smo CRD through the fatty-acid side chains.

Additionally we tested the effect of each of these phosphatidylinositols on *in-vitro* Hh signalling. The different phosphatidylinositols had varied effect on Hh signalling, suggesting that along with the fatty-acid side chains the inositol head with the phosphate modifications also play a role in regulating signalling. The phosphate group at the fourth position of the inositol head was particularly important for moderately enhancing SAG induced Hh signalling. Thereby, we speculated that the inositol head with the fourth position phosphate group is necessary for anchoring in the 7TM domains of the Smo receptor and the fatty-acid side chains bind to the Smo CRD. Perhaps the simultaneous engagement of extracellular and membrane bound structures by PI(4)P leads to an active conformation of Smo conducive for downstream Hh signalling.

5.2. Experimental Procedures

5.2.1. Protein Expression and Purification

Unlabelled and ^{15}N labelled human Smo CRD was expressed and purified as described in section 3.2.2.

5.2.2. NMR Experiments

The NMR Water-LOGSY experiments were performed to study the interaction of the human Smo receptor and the phosphatidylinositol ligands[142]. Water-LOGSY experiments were performed using a 10ms selective 180° pulse at the water signal frequency. The NOE mixing time was 2 seconds. The Water-LOGSY spectrum was recorded at 25°C with 2048 scans.

Six different phosphatidylinositols with different number and position of the phosphate groups were tested, namely: phosphatidyl inositol-4-phosphate (PI(4)P), phosphatidyl inositol-3-phosphate (PI(3)P), phosphatidyl inositol-3,4-bisphosphate (PI(3,4)P₂), phosphatidyl inositol-3,5-bisphosphate (PI(3,5)P₂), phosphatidyl inositol-4,5-bisphosphate (PI(4,5)P₂) and phosphatidyl inositol-3,4,5-triphosphate (PI(3,4,5)P₃). Each of these phosphatidylinositols possesses two octanoic-acid side chains. Inositol-3,4,5- triphosphate (IP₃) lacking the fatty-acid side chain was also tested. The different phosphatidylinositols tested in the present study are of synthetic origin and were purchased from Avanti Polar Lipids. These phosphatidylinositols are readily soluble in aqueous buffer and stock solutions were made by dissolving in 10mM deuterated acetic-acid buffer, pH 5.

The NMR Water-LOGSY spectra were recorded for 200 μ M of each of the free phosphatidylinositol ligands in 10mM deuterated acetic-acid buffer and 10% deuterium-oxide (volume/volume). Subsequently 10 μ M of the human Smo CRD was added to maintain a 1:20 (protein to ligand) ratio and the Water-LOGSY spectrum was recorded. The spectra were referenced using the internal standard, 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

The NMR CSP experiments were carried out using 15 N labelled human Smo CRD in 10 mM deuterated acetic-acid buffer and 10% deuterium-oxide (volume/volume). 1 H- 15 N HSQC spectra of the human Smo CRD were recorded in the absence and presence of the ligand phosphatidylinositol, PI(4)P. The 1 H- 15 N HSQC spectrum of the free protein was recorded with 32 scans. On addition of the PI(4)P ligand at a ratio of 1:0.5 (protein: ligand) there was an immediate drop in the intensity of the amide signals, therefore the 1 H- 15 N HSQC spectrum of the human Smo CRD with the PI(4)P ligand was recorded with 240scans.

All NMR spectra were recorded at 25 $^{\circ}$ C using Bruker Avance 600 MHz spectrometer equipped with 1 H/ 15 N / 13 C detect, TCI triple resonance cryogenic probes. All data were processed using Topspin (Bruker Biospin) 3.0 NMR software and analysed using the program CARRA[131].

5.2.3. Cell-based Reporter Assays

The cell-based Hh reporter assays were performed as described in section 4.2.5. Briefly, the Shh Light II cells were cultured to confluency in a 96-well plate. The cells were treated with various concentrations of the phosphatidylinositols ranging from 0-50 μ M, in DMEM containing 0.5% (volume/volume) bovine calf serum. The phosphatidylinositols were added to the media from concentrated stocks in Phosphate Buffer Solution (PBS). After 48 hours, renilla and firefly luciferase levels were measured using Dual Luciferase Reporter Assay kit (Promega). Hh pathway activity was expressed as the ratio of firefly to renilla luciferase. Experiments were performed at least three times in duplicate and error bars represent the standard error of the mean.

5.3. Results and Discussion

5.3.1. Phosphatidylinositols Bind to the Human Smoothened Cysteine Rich Domain

The results of the NMR Water-LOGSY experiments show that phosphatidylinositols irrespective of the number and the position of the phosphate groups bind to human Smo CRD. The phosphatidylinositols tested included: PI(4)P, PI(3)P, PI(3,4)P2, PI(4,5)P2, PI(3,5)P2 and PI(3,4,5)P3. Each of these synthetic phosphatidylinositols possesses an inositol head-group and two fatty-acid side chains, each of eight carbon atoms in length. The proton signals arising from the fatty-acid side

chains are well-pronounced in the aliphatic region of the ^1H -NMR spectrum of the phosphatidylinositols. The peak positions of the methyl and the methylene protons are well demarcated as shown in **Figure 5-1**. The six different phosphatidylinositols possess the same fatty-acid side chains and therefore give rise to the similar proton signals in ^1H -NMR spectrum.

The Water-LOGSY spectra of the phosphatidylinositols were recorded in the absence and the presence of the human Smo CRD. The positive Water-LOGSY signals of the ligands, in the presence of human Smo CRD demonstrate that all of the above mentioned phosphatidylinositols interact with the human Smo CRD through the fatty-acid side chains. Phosphatidylinositols with a single phosphate group bind to human Smo CRD as shown in **Figure 5-1A** and **Figure 5-1B**. The phosphatidylinositols bearing two phosphate modifications also interact with human Smo CRD as shown in **Figure 5-2A**, **Figure 5-2B** and **Figure 5-2C**. The phosphatidylinositol with three phosphate groups also interact with human Smo CRD as shown in **Figure 5-3A**. Inositol-3,4,5-triphosphate (IP3) possesses only the inositol head group with three phosphate modifications and lacks the fatty-acid side chains. The signals arising from the protons of the inositol head are close to the bulk water signal arising from the aqueous buffer and are therefore often hard to detect; however, two isolated proton signals of IP3 could be detected as shown in **Figure 5-3B**. The negative Water-LOGSY signal of the IP3 ligand on addition of human Smo CRD, as shown in **Figure 5-3B**, states that IP3 does not interact with human Smo CRD.

5.3.2. The Phosphatidylinositol-4-phosphate Induces Backbone Chemical Shift Perturbations in Human Smoothed Cysteine Rich Domain

Genetic experiments have shown that the activation of the *Drosophila* Smo receptor is dependent on the levels of PI(4)P [98, 170]. The loss of the kinase facilitating the synthesis of PI(4)P, results in Hh loss-of-function phenotypes whereas the loss of the phosphatase responsible for degrading PI(4)P results in Hh gain-of-function phenotypes [98, 170]. We have shown that the phosphatidylinositols interact with the human Smo CRD by Water-LOGSY experiments, alongside we wanted to identify the residues in the human Smo CRD that are important for this interaction, therefore we performed the NMR CSP experiments focussing on PI(4)P.

The 2D ^1H - ^{15}N HSQC spectra of the human Smo CRD was recorded in the absence and presence of the ligand, PI(4)P. There was an immediate drop in the intensity of the amide signals observed in the 2D ^1H - ^{15}N HSQC spectrum of the human Smo CRD upon addition of PI(4)P ligand at a ratio of 1:0.5 (protein: ligand). Therefore NMR titration experiments with an increasing concentration of the PI(4)P ligand could not be performed. Nevertheless, in the first titration point distinct CSPs corresponding to the amino acids E160 and R161 were observed as shown in **Figure 5-4**. These residues are homologous to the mouse Smo CRD residues which are important for binding to oxysterol ligands [95]. Arginine 161 also depicted the maximum CSP on titration with the glucocorticoid ligand Bud as described in Chapter 4.

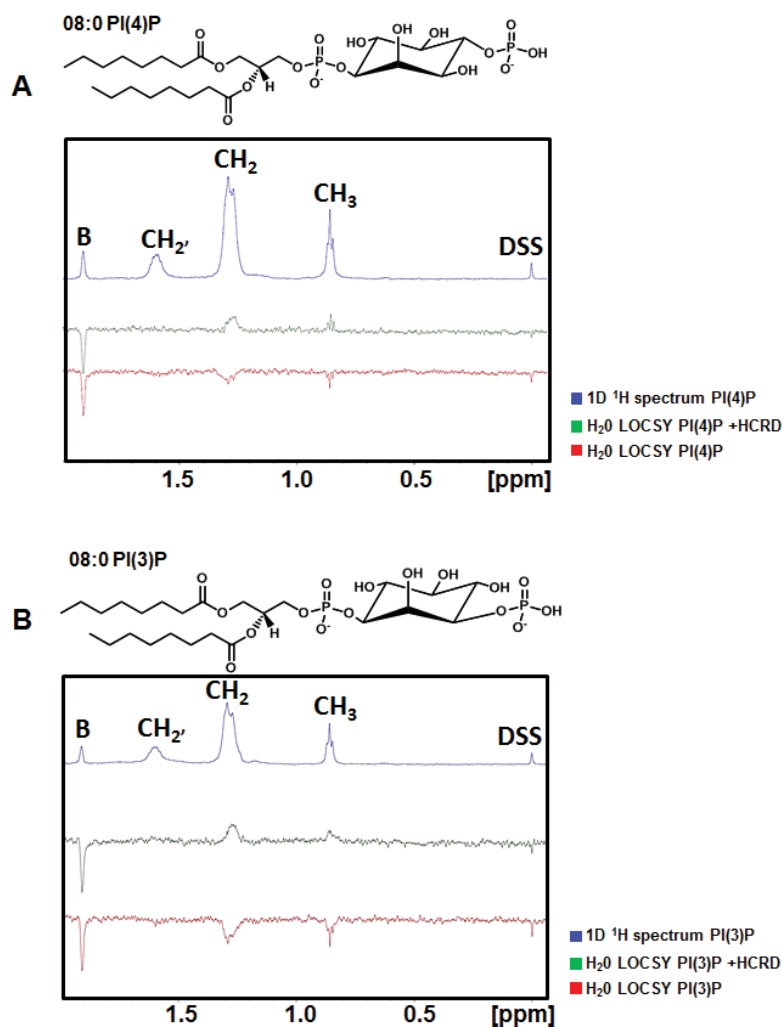


Figure 5-1. The phosphatidylinositols with a single phosphate group bind to human Smo CRD.

(A) The NMR Water-LOGSY spectrum shows that PI(4)P binds to human Smo CRD. **(B)** The NMR Water-LOGSY spectrum shows that PI(3)P binds to human Smo CRD. The blue spectrum represents the 1D proton spectrum of the ligand. The peaks corresponding to the methyl (CH₃), the methylene (CH₂) and the methylene protons adjacent to the ketone side chain (CH₂') are labelled in both spectra. The position of the acetic-acid buffer (B) peak and interanal standard DSS is also indicated. The positive ligand peaks on addition of the Smo CRD (green) as compared to the negative ligand peaks (red) in the absence of Smo CRD indicates that the ligand interacts with the protein.

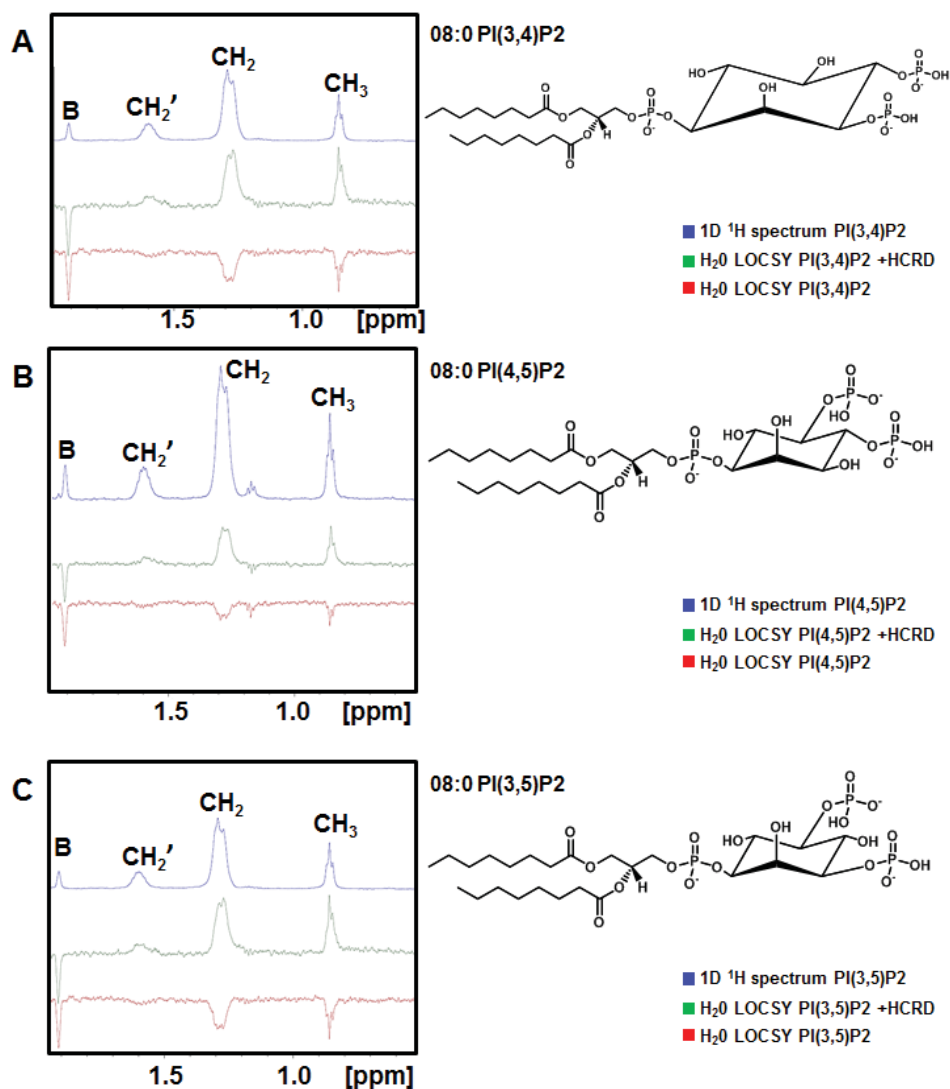


Figure 5-2. The phosphatidylinositols with two phosphate groups bind to human Smo CRD.

(A), (B) and (C) respectively show that PI(3,4)P2, PI(4,5)P2 and PI(3,5)P2 interact with Smo CRD. The peaks corresponding to the methyl (CH_3), the methylene (CH_2) and the methylene protons adjacent to the ketone side chain (CH_2') are labelled in all three spectra. The position of the acetic-acid buffer (B) peak is also indicated. The positive ligand peaks on addition of the Smo CRD (green) as compared to the negative ligand peaks (red) in the absence of Smo CRD indicates that the ligand interacts with the protein.

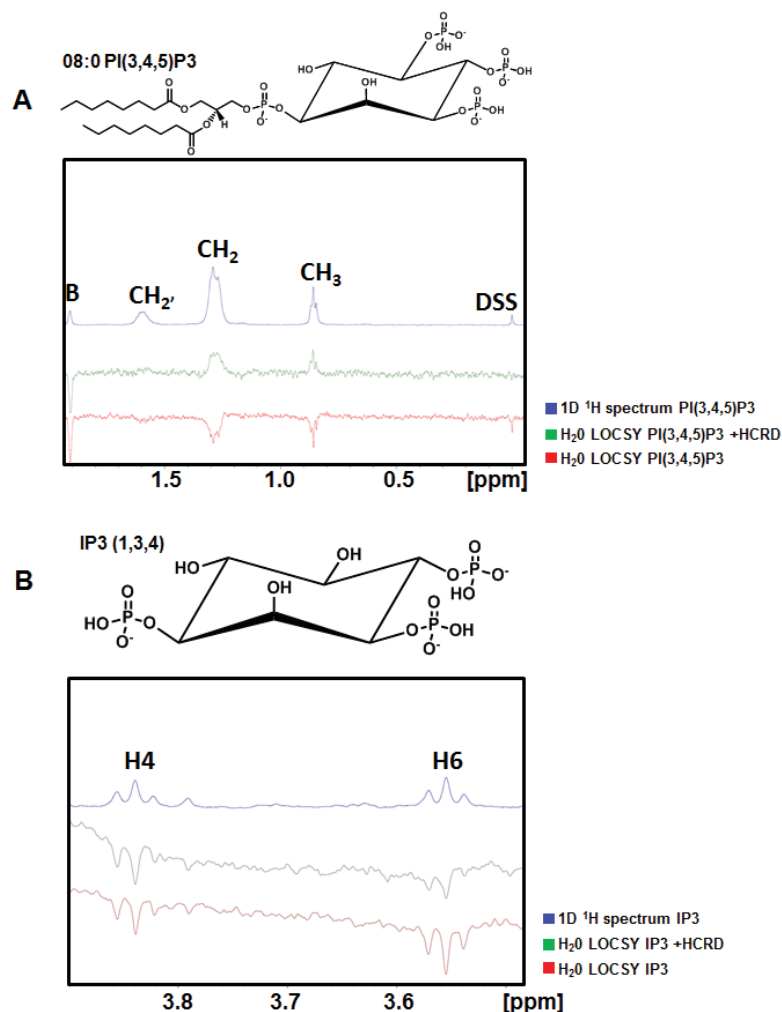


Figure 5-3. The fatty-acid side chain of the phosphatidylinositols is essential for interaction with the human Smo CRD.

(A) PI(3,4,5)P3 interacts with Smo CRD. The peaks corresponding to the methyl (CH₃), the methylene (CH₂) and the methylene protons adjacent to the ketone side chain (CH₂') are labelled in the spectra. The position of the acetic-acid buffer (B) peak and internal standard DSS is also indicated. The positive ligand peaks on addition of the Smo CRD (green) as compared to the negative ligand peaks (red) in the absence of Smo CRD indicates that the ligand interacts with the protein. (B) IP3 lacks the fatty-acid side chain and does not interact with the Smo CRD. The signals arising from the protons of IP3 are close to the bulk water signal arising from the aqueous buffer and are therefore hard to detect; however, two isolated proton signals (H4 and H6) of IP3 could be detected. The H4 and H6 protons are attached to the carbon 4 and carbon 6 respectively of the inositol ring. The negative ligand peaks in the presence (green) and absence (red) of protein indicates that the protein does not interact with the ligand.

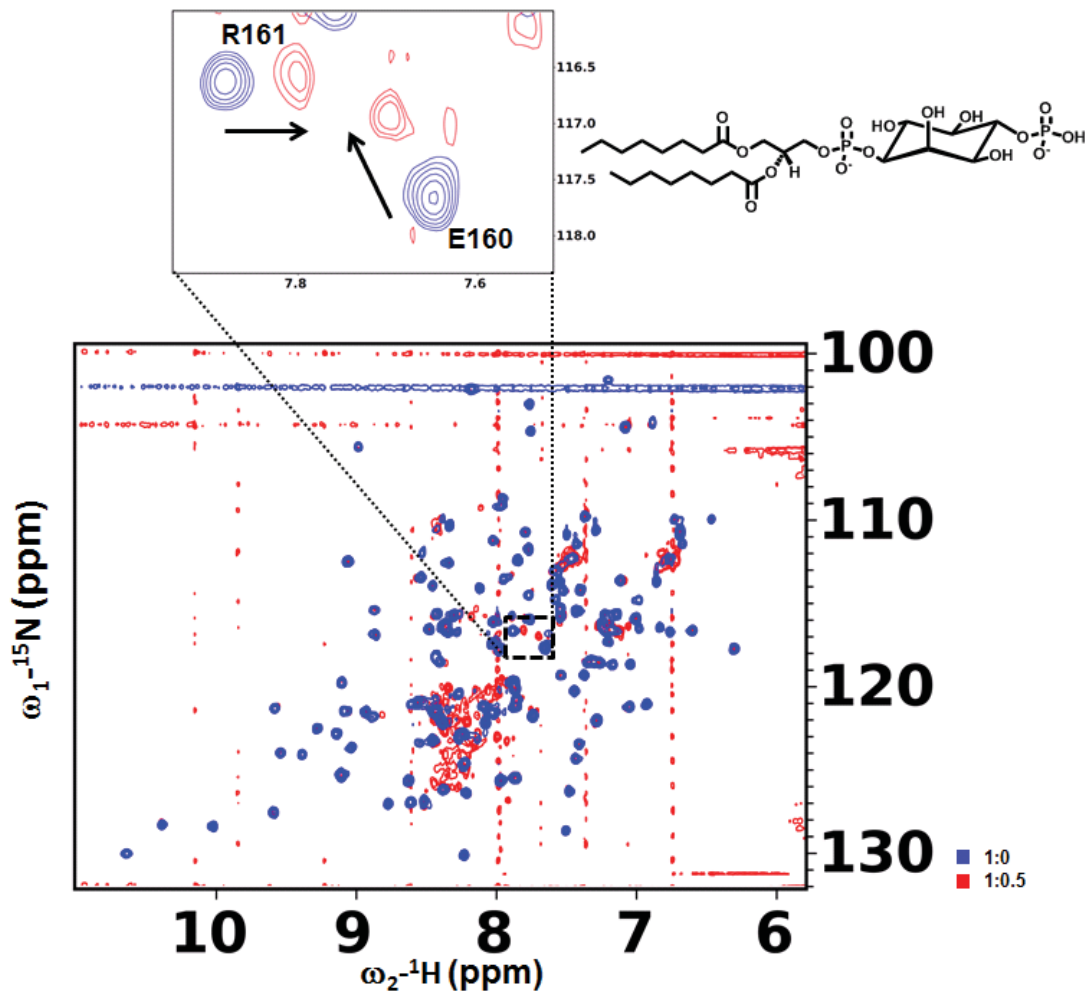


Figure 5-4. The phosphatidylinositol: PI(4)P, interacts with the human Smo CRD. ^1H - ^{15}N HSQC spectra of the human Smo CRD in the absence (blue) and presence (red) of the ligand PI(4)P. The ligand PI(4)P causes an immediate drop in the intensity of the amide signals. The upper inset depicts the specific chemical shift perturbations observed upon addition of PI(4)P. The chemical structure of PI(4)P is shown above the spectra.

5.3.3. Phosphatidylinositols Exert Varied Effects on *in-vitro* Hedgehog Signalling

To examine the effects of the different phosphatidylinositols on *in-vitro* vertebrate Hh signalling we performed cell-based luciferase reporter assays using the vertebrate NIH3T3 derived Shh Light 2 cells. The Shh Light 2 cells are stably transfected cell lines with Gli responsive firefly luciferase reporter and constitutive renilla luciferase expression vector. The Shh Light 2 cells were treated with an increasing concentration of the different phosphatidylinositols ranging from 0-50 μ M. Namely the different phosphatidylinositols tested included: PI(4)P, PI(3)P, PI(3,4)P₂, PI(4,5)P₂, PI(3,5)P₂, and PI(3,4,5)P₃. The inositol phosphate, IP₃ was also tested.

On treatment of the Shh Light 2 cells with PI(4)P, a 30-35% increase in SAG induced Hh signalling was observed as shown in **Figure 5-5A**. The small molecule SAG is an activator of Smo and is known to function by binding to the Smo transmembrane domains[83, 90]. In the present study SAG is used as a positive control to induce Hh signalling. In the absence of SAG, the synthetic PI(4)P activates signalling minimally which is basal when compared to SAG induced signalling suggesting that the synthetic PI(4)P with two octanoic-acid side chains, does not necessarily function as the endogenous ligand for Smo; however it acts in conjunction with SAG to further enhance Hh signalling. The other phosphatidylinositol with a single phosphate group, PI(3)P, did not affect Hh signalling either in the absence or presence of SAG as shown in **Figure 5-5B**, suggesting that the phosphate group at the fourth position in the inositol head group is important for signalling.

To study the effect of the phosphatidylinositols with two phosphate modifications, we treated Shh Light 2 cells with increasing concentrations of synthetic PI(4,5)P₂, PI(3,4)P₂ and PI(3,5)P₂. The synthetic phosphatidyl inositol PI(4,5)P₂, suppresses SAG induced Hh signalling by 15-20% as shown in **Figure 5-6A**. The synthetic phosphatidylinositols PI(3,4)P₂ and PI(3,5)P₂ did not have any observable effect on Hh signalling in the absence or presence of SAG as shown in **Figure 5-6B** and **Figure 5-6C**. Similarly, the phosphatidylinositol with three phosphate modifications, PI(3,4,5)P₃ with the octanoic-acid side chains and the inositol phosphate, IP₃, without the octanoic-acid side chains did not have any detectable effect on Hh signalling as shown in **Figure 5-7A** and **Figure 5-7B**.

5.4. Discussion

The Smo receptor functions as the signal transducer of the Hh pathway[13, 14]. Small molecules modulate Smo activity by anchoring either in the cavity of the 7TM domains or the extracellular CRD[64, 83-85, 90, 92, 95, 96, 150]. In Chapter 4 we introduced the novel hypothesis that perhaps certain Smo ligands engage both the Smo 7TM domains as well as the extracellular CRD leading to structural rearrangements in the Smo receptor to regulate downstream Hh signalling.

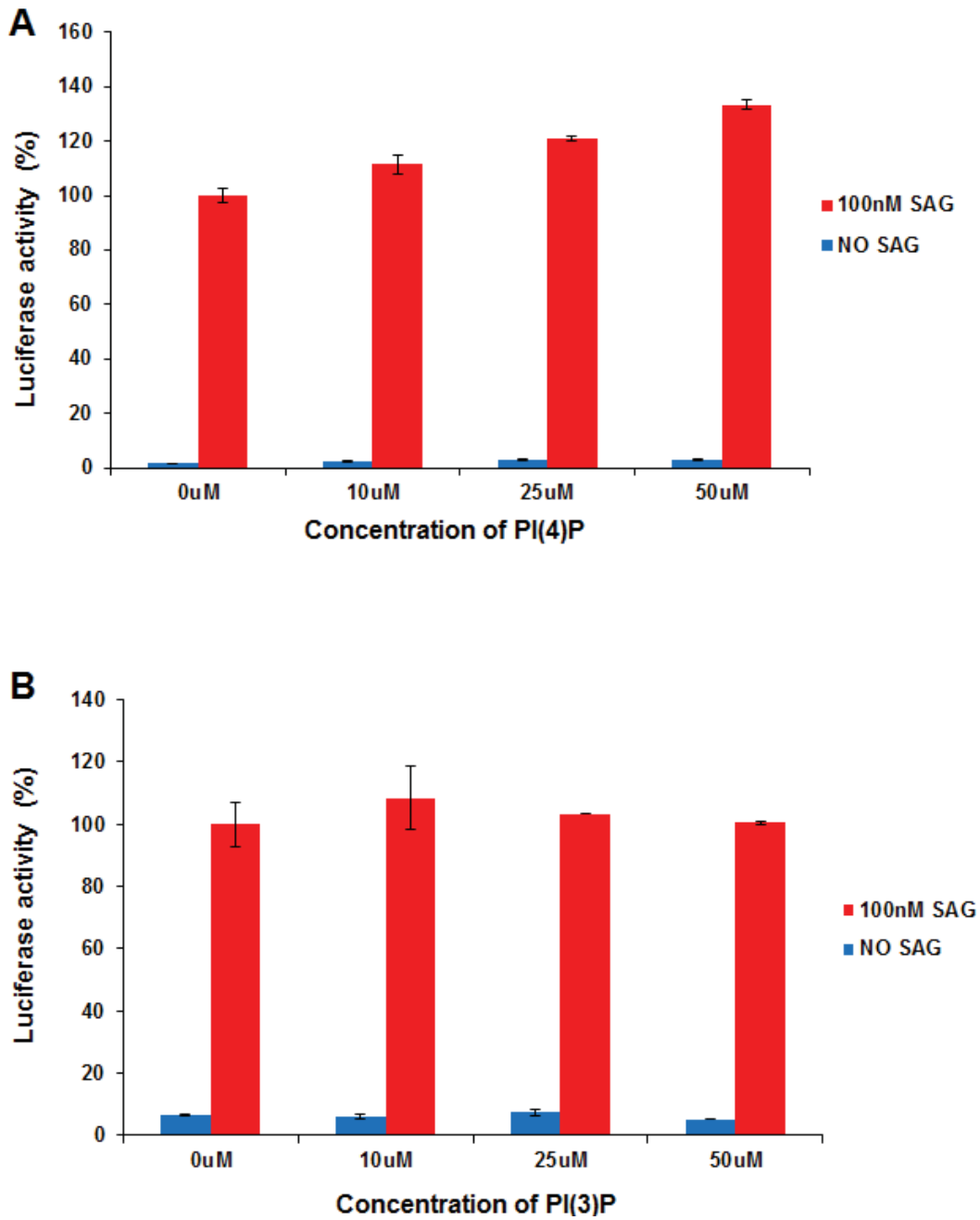


Figure 5-5. The effect of the phosphatidylinositols with one phosphate group on Hh signalling.
(A) PI(4)P moderately enhances SAG induced Hh signalling. **(B)** PI(3) does not effect Hh signalling.

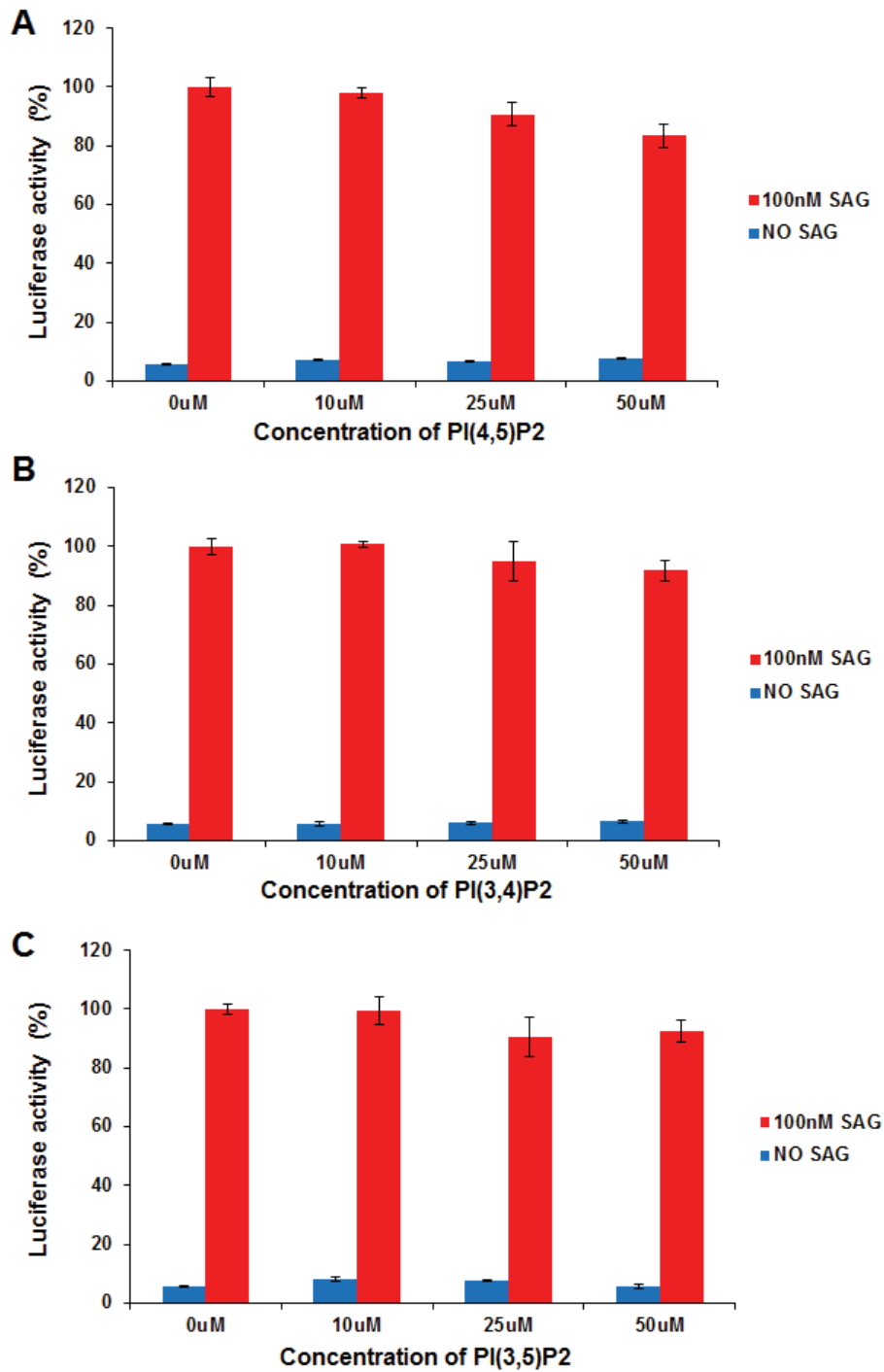


Figure 5-6. The effect of the phosphatidylinositols with two phosphate groups on Hh signalling.

(A) PI(4,5)P2 minimally decreases SAG induced Hh signalling. (B) and (C) PI(3,4)P2 and PI(3,5)P2 does not effect Hh signalling.

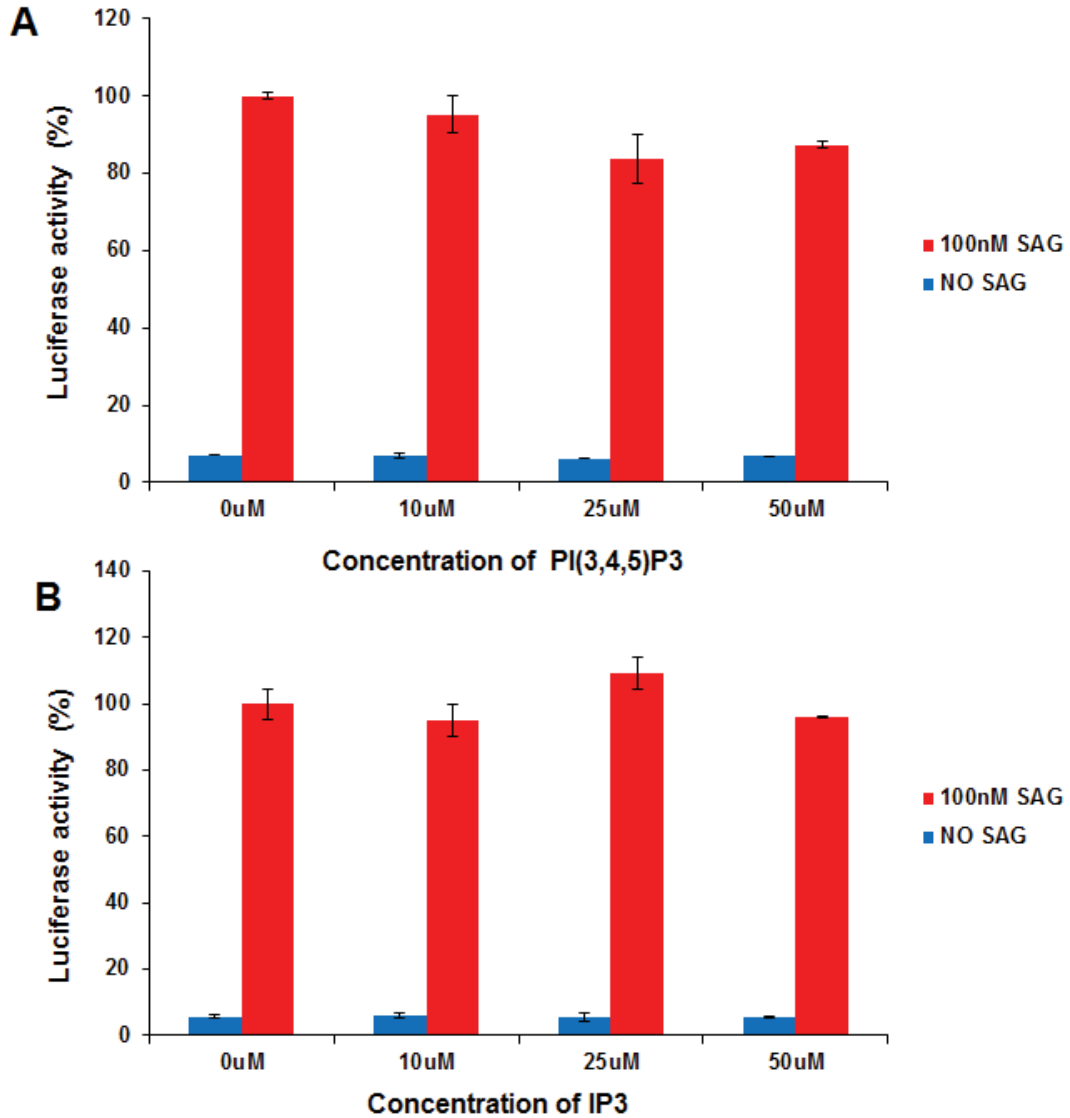


Figure 5-7. The effect of the phosphatidylinositols with three phosphate groups on Hh signalling.
(A) PI(3,4,5)P3 with the octanoic fatty-acid sidechains does not effect Hh signalling. **(B)** IP3 comprises the inositol head and lacks the fatty-acid sidechains. IP3 does not effect Hh signalling.

In such regards we speculated that phosphatidylinositols may be the potential Smo ligand which engages both the Smo membrane spanning and extracellular structures simultaneously. Such speculations were based on two major supporting factors. Firstly, genetic experiments have shown that phosphatidylinositols, specifically PI(4)P play a role in regulating Smo activation[98, 170]. Secondly, the FzD receptor which is homologous to the Smo receptor binds to its physiological ligand Wnt through a fatty-acid modification on Wnt[66]. Specifically, the palmitate fatty-acid modification on Wnt binds to the extracellular CRD of the FzD receptor[66]. Similarly we hypothesized that the fatty-acid sidechain of the phosphatidylinositols binds to the Smo CRD to regulate Hh signalling and perhaps the inositol head docks in the cavity of the membrane spanning domains.

In the present study we performed all experiments with commercially available synthetic phosphatidylinositols bearing two octanoic fatty-acid side chains. Pertaining to our hypothesis we observed that phosphatidylinositols irrespective of the number and position of the phosphate modifications engage the Smo CRD through the fatty-acid side chains. Furthermore, we showed that although the different phosphatidylinositols bind to the Smo CRD they have varied effect on *in-vitro* Hh signalling suggesting that the inositol head with different number and position of the phosphate groups regulate Hh signalling differently. Accordingly, we demonstrated that the synthetic PI(4)P enhances SAG induced Hh signalling by 30-35%, whereas PI(4,5)P₂ decreases SAG induced Hh signalling by 15-20%.

In conclusion we demonstrated that phosphatidylinositols bind to Smo CRD through the fatty-acid side chains and PI(4)P moderately enhances SAG induced Hh signalling; however the major reservation of the present study is that the synthetic phosphatidylinositols tested here did not have any significant effect on Hh signalling in the absence of the signal inducer SAG, suggesting that they may not necessarily function as endogenous Smo ligands. However, the observation that the synthetic PI(4)P moderately enhances SAG induced Hh signalling suggests that it might acts as a mimic of an allosteric activator of Smo. Furthermore, phosphatidylinositols are known to function as signalling molecules and regulate membrane trafficking of proteins[167, 169]. Similarly PI(4)P may promote the membrane trafficking of Smo receptor and thereby render Smo active to induce downstream signalling.

CHAPTER 6. IMPLICATIONS AND FUTURE DIRECTIONS

6.1. The Structures of the Smoothened G Protein Coupled Receptor (Structure Determination Strategies follow the Rule of Divide and Conquer)

The first seminal finding of the present dissertation is elucidating the structure of the extracellular CRD of the Smoothened receptor. The significance of the extracellular CRD to the Smo structure and function as well as the importance of the Smo receptor in Hh signalling and its implication in health and disease has been discussed in Chapter 1. As stated earlier Smo is a canonical member of the FzD class of GPCRs and likewise retains the characteristic extracellular CRD[16, 63, 65, 146]. Interestingly although the structure of the homologous FzD CRD has been known for over a decade now there was no structural information on any part of the Smo receptor till the summer of 2013[64-66, 90].

Despite continuous interest in studying the function of the Smo receptor, the lack of structural information up until recently has largely been attributed to the difficulties associated with the expression and purification of membrane proteins and specially GPCRs. To address this problem and provide a better understanding of the biology of human GPCRs, the GPCR network was formed as a large-scale collaboration to understand the structure and function of human GPCRs[171]. The major contribution of the GPCR network to the field of Hh signal transduction was the structure determination of the Smo 7TM domains[64]. Presently the complex structure of the Smo 7TM domains have been resolved with five different molecules and provide insight into the ligand binding pocket located within the 7TM domains[64, 85, 90]; however these structures lacked both the extracellular CRD and the cytoplasmic domains. Thus compiling the results of the present study about the structure of the Smo CRD with the GPCR network derived crystal structure of the Smo 7TM domains provides an essentially complete picture of the structure of the Smo receptor. The structure of the Smo cytoplasmic domains has not yet been determined. The length of the human Smo C-termini is about 240 residues and predicted to include the eighth helix which may pack parallel to the membrane. Resolving the structure of the Smo C-termini will further our understanding of how Smo functions to recruit downstream signalling components. The length of the *Drosophila* Smo C-termini is predicted to be 483 residues. The difference in the length of Smo C-termini between the two species may lead to different cytosolic engagements to regulate downstream Hh signalling.

Contemporary to our elucidation of the structure of the *Drosophila* Smo CRD, the structure of the zebrafish Smo CRD was determined by crystallographic methods[150]. The structure of the NMR derived solution structure of the *Drosophila* Smo CRD is similar to the crystal structure of the zebrafish Smo CRD with a root mean square deviation of 2.04 Å between the two structures as shown in **Figure 6-1**. The disulfide bond pattern between the two proteins is conserved. The structure of the zebrafish Smo CRD includes the N-and C- termini cysteines that were not included in the structure of the *Drosophila* Smo CRD.

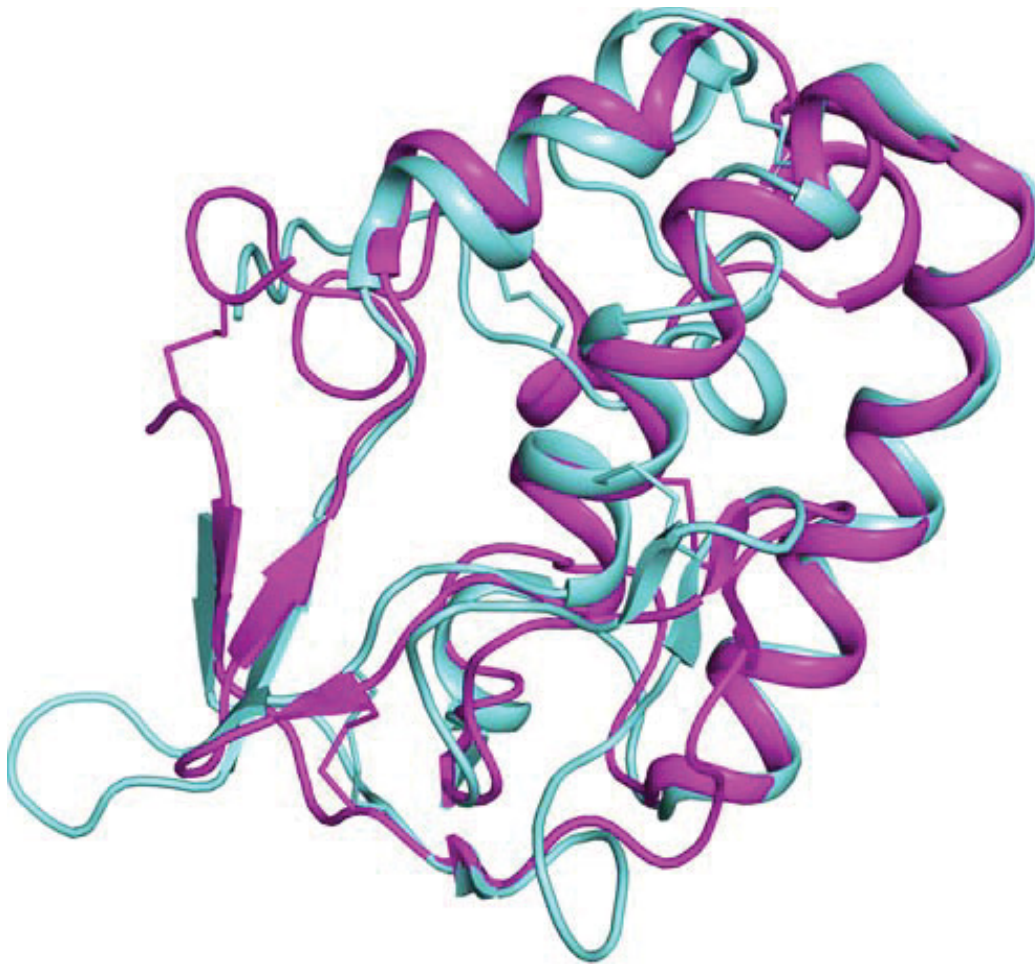


Figure 6-1. Superimposed structures of the *Drosophila* and the zebrafish Smo CRD. The *Drosophila* Smo CRD determined by NMR is represented in cyan and the zebrafish Smo CRD determined by crystallographic methods is shown in magenta. The sticks indicate the disulfide bonds in both structures.

These two cysteines form the fifth disulfide bond of the Smo CRD[150]. Smo CRD has been implicated to play a role in receptor dimerization and in the crystal structure of the zebrafish Smo CRD there were two Smo molecules in the crystallographic asymmetric unit[68, 150]. The interface between the two molecules buried a surface area of 490 Å² and the crystal contact was formed through a Zn ion which bonded to three different protein chains[150]. This suggested that the dimer observed in the crystal may not be of physiological relevance[150]. Additionally it was stated that the purified Smo CRD behaved as a monomer at a concentration of 5 μM; however this does not rule out the possibility that the Smo CRD may form a physiologically relevant dimer especially at higher concentrations.

6.2. Multiple Ligand Binding Sites on Smoothed Receptor Help Propose a Novel Mechanism for Hedgehog Signalling

The second important finding of the present study is that the Smo CRD possesses hydrophobic ligand binding capabilities. In the present work we showed that both the *Drosophila* and the human Smo CRD binds to the synthetic glucocorticoid Bud, albeit with different affinities as described in Chapter 4. Glucocorticoids regulate Hh signalling by acting at the level of the Smo receptor[93].

Recently, it has been shown that oxysterols and oxysterol derivatives also modulate Smo signalling by binding to the CRD[95, 96, 150]. It has been proposed that the Smo CRD perhaps acts as an allosteric binding site to regulate Hh signalling. The recent biochemistry experiments showed that the CRD is essential for binding to oxysterols derived on beads[96, 150]. In order to identify the specific residues involved in binding, individual or small stretches of amino acids in the Smo CRD were mutated and checked for their ability to bind oxysterols[95, 96, 150]. The choice of residues to mutate was based upon the complex structure of the FzD receptor and the Wnt ligand[66]. The Wnt protein interacts with FzD on two sites. The first site is mediated by a palmitate modification on Wnt which engages a hydrophobic groove in the CRD whereas the second site is dominated by protein-protein interactions between the receptor and the ligand protein[66]. The presumption is that the ligand binding site on Smo CRD would include the homologous residues as the palmitate binding site on FzD CRD. This approach for determining the key residues in Smo CRD required for ligand binding was solely biased on the complex structure of the FzD8 CRD and the Wnt8a ligand which revealed the intricate structural details of the interaction[66].

In the present dissertation we were successful in providing the first direct structural evidence of the residues in both *Drosophila* and human Smo CRD that are involved in binding to the hydrophobic glucocorticoid, Bud. NMR titration experiments provide an excellent choice of monitoring protein-ligand interaction and identifying unique residues involved in the interaction as described in Chapter 2[143, 155, 159, 160]. Using this methodology we were able to determine the specific residues both in the *Drosophila* and the human Smo CRD that are essential for ligand binding. The NMR based structural results of the present study were further supported by contemporary

biochemistry research. Necessarily the residues that depicted the maximum CSPs in the human CRD on titration with the glucocorticoid, Bud corresponded to the mouse residues which when mutated failed to bind the oxysterol ligands and consequently failed to induce downstream Hh signalling[95, 96]. Thus although the glucocorticoid and the oxysterol ligands are different the hydrophobic ligand binding pocket which serves to dock these molecules are similar in the mouse and the human Smo receptor. Thus in the recent times Smo CRD has emerged as the second ligand binding site on Smo[95, 96, 150]. This further strengthens the argument that the CRDs have evolved from an ancestral domain capable of recognizing hydrophobic molecules[146, 161].

The first ligand binding site on Smo was described a decade earlier and was identified to be located in the pocket of the 7TM domains[83, 84]. Cyclopamine a plant derived alkaloid is the first described small molecule modulator known to dock in the cavity of the 7TM domains[84, 85]. Since then a plethora of small molecules have been described which compete with cyclopamine for binding to the Smo receptor indicating that these synthetic small molecules dock in the cavity located in the 7TM domains[83, 93]. While some of these small molecules such as SAG and purmorphamine function as Smo agonists others such as cyclopamine and the SANT group of small molecules function as Smo antagonists[83, 84, 87]. The complex structures of the Smo receptor with the different ligands show that the ligands penetrate the binding-cavity to varying degree[64, 85, 90]. The antagonist SANT docks deep down in the cavity, whereas, cyclopamine stems outside the cavity of the 7TM domains and makes contacts with the extracellular loop structures[85].

Most of the described Smo modulators are synthetic. Oxysterols have emerged as the likely candidate as an endogenous ligand for Smo provided Smo activity is indeed regulated by a physiological ligand[92, 95, 96, 150]; however this is subject to debate as the cellular levels of oxysterols are low and there is no concrete evidence that Smo activity is indeed regulated by any endogenous ligand. Although the natural ligand of Smo is yet unknown, presently what is known clearly is that Smo has at least two definite binding sites, one located in the extracellular CRD and the other situated in the 7TM domains[64, 85, 150]. The extracellular region of Smo CRD may be flexible as shown in **Figure 6-2A**. Smo ligands may bind in the cavity of the 7TM domains and regulate signalling regardless of binding to the CRD. Similarly, small molecules may bind to Smo CRD and control Hh signalling irrespective of engaging the cavity in the 7TM domains, however; we speculate that perhaps there exists a certain class of small molecules which can bind to the Smo CRD and induce a conformational shift bridging the Smo extracellular and 7TM domain structures, which in turn would lead to conformational rearrangements in the membrane spanning domains to modulate downstream Hh signalling as shown in **Figure 6-2B**. The Smo CRD perhaps acts as “fishing bait”, which binds to the Smo ligand and induces conformational rearrangements between the Smo CRD and 7TM domains to regulate signalling.

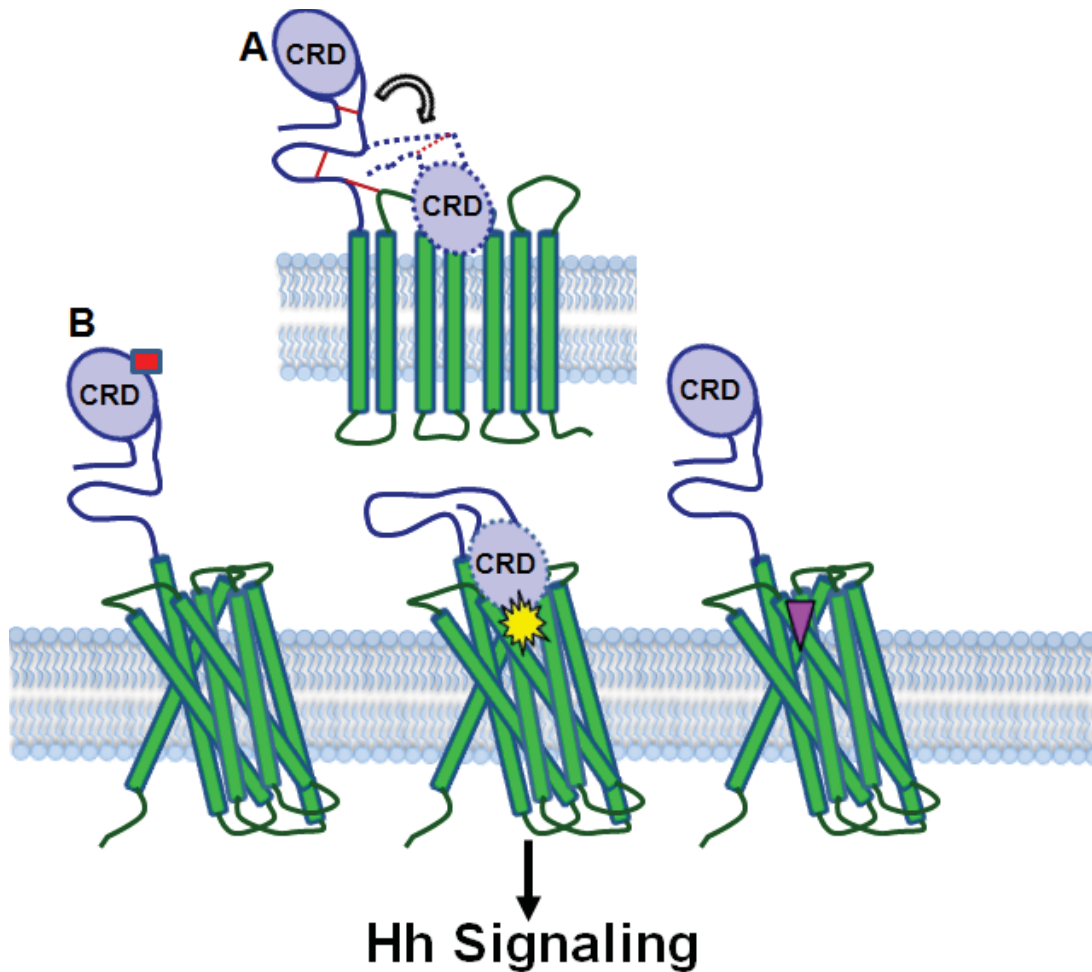


Figure 6-2. A novel model for Smo allosteric regulation.

(A) The Smo CRD may be flexible. The disulfide bonds stabilizing the N-terminal Smo extracellular linker domain are shown in red lines. (B) The Smo receptor has more than one binding site. Molecules like Bud and oxysterols bind to the Smo CRD (left, red rectangle) to regulate signalling. Cyclopamine and GDC-0449 are known to bind in the cavity located in the 7TM domains (right, pink inverted triangle) to control Hh signalling. We speculate that there is the possibility of a class of molecules (middle, yellow star) which would bind to the CRD and cause a conformational change of the Smo extracellular structures to bring it closer to the 7TM domains. This in turn would lead to structural rearrangements of the Smo 7TM domains to regulate signalling. Reprinted with permission. Rana R, Carroll CE, Lee H-J, Bao J, Marada S, Grace CRR, et al. Structural insights into the role of the Smoothed cysteine-rich domain in Hedgehog signalling. *Nat Commun* 2013,4.

6.3. The Unanswered Question: Does Smoothed Really Have an Endogenous Ligand?

For years now Smo activity has been demonstrated to be regulated by small molecules mostly of synthetic origin[83, 87, 88]. Presently there is also an FDA approved drug GDC-0449 for the treatment of BCC that functions by acting at the level of Smo[88, 89, 91]. Despite such major advancements in the field of Smo there is a dearth of knowledge regarding some of the basic aspects of Smo regulation and Hh signalling. One of the major unanswered questions is whether Smo activity is regulated by an endogenous ligand. The other major gap in the understanding of the Hh signal transduction cascade is how Ptch regulates Smo. The answers to both these questions may not be mutually exclusive that is Ptch perhaps regulates Smo activity by regulating the physiological ligand of Smo. Despite the current disparities regarding whether Smo activity is regulated by an endogenous ligand or not, we believe that Smo activity is regulated by a physiological ligand. Our presumption has originated from the following facts.

Firstly, the Smo CRD is similar to the CRDs related to FzD receptors and the Hh-interacting proteins[65, 66, 146]. These domains are suggested to have evolved from a common ancestral origin which play an important role in binding to hydrophobic ligands[146, 161]. Additionally the nearest member of the Smo CRD is the FzD CRD which indeed binds to the Wnt ligand through a palmitate modification suggesting that the FzD CRD retains the property of binding to hydrophobic ligands[66]. Similarly, the Smo CRD may pertain to its evolutionary history and like the other members may bind to a yet unidentified hydrophobic ligand.

Secondly, Smo is a GPCR and it is not unlikely for a GPCR to be regulated by physiological ligands[16, 62, 63, 163-165]. In canonical GPCR signalling ligand binding leads to conformational rearrangements in the membrane spanning domains to activate downstream signalling[16, 62, 63, 171]. Thus Smo may function as a typical GPCR and function by binding to an endogenous ligand. This is further supported by the fact that Smo possess the ability to bind to synthetic molecules to regulate signalling[64, 83, 87]. Perhaps these synthetic molecules act as a mimic of the yet unidentified natural ligand of the Smo receptor.

Thirdly, accepting the concept that Smo activity is regulated by a natural ligand may help answer the enigmatic question of how Ptch regulates Smo. Ptch has a sterol sensing domain which may help in the transport of the sterol which could function as the natural ligand of Smo[34]. The importance of the sterol sensing domain of Patched is emphasized by the fact that missense mutations in this domain of Patched abrogate its ability to inhibit the Smo receptor[172]. In the absence of the Hh ligand Ptch is most likely maintained in an inactive conformation which interferes with the functioning of the sterol sensing domains of the Ptch receptor and hinders the transport of the sterol which may function as the Smo ligand. In the presence of the Hh ligand the conformation of the Ptch receptor is such that it results in a functional sterol sensing domain. This in turn

leads the Smo receptor accessible to its natural ligand to regulate downstream Hh signalling.

Despite the facts supporting our presumption, the question still remains as to whether Smo activity is indeed regulated by an endogenous ligand. Smo in various respects is different from both the FzD receptors and the typical GPCRs and thereby may differ from them in its ability to bind to an endogenous ligand. Perhaps rather than a physiological ligand regulating Smo activity it is the sub-cellular distribution of Smo that regulates its activity. Vertebrate Smo is not functional when it is trapped in the endosomal vesicles. Trafficking of Smo to the primary cilium is an important prerequisite for Hh signalling[38]. Ptch through its sterol sensing domain perhaps maintains the lipid and sterol environment in the membrane conducive for Smo to reside and function.

In Chapter 5 we described that the phosphatidylinositol, PI(4)P enhances SAG induced Hh signalling. The fact that PI(4)P in the absence of SAG induces a minimal increase in Hh signalling which is basal when compared to SAG induced signalling suggests that PI(4)P may not necessarily function as the specific endogenous ligand for Smo; however PI(4)P may help in the formation of a lipid associated signalosome complex that help trafficking Smo to the primary cilium in vertebrates or plasma membrane in *Drosophila* to activate downstream signalling. This hypothesis is further supported by the fact that a similar observation was made in the Wnt pathway wherein phospholipids aided the formation of a signalosome complex required for cell surface localization of the receptors to regulate signalling[173].

6.4. Smoothed as a Template for Drug Discovery

Abnormal Hh signalling has been predominantly associated with BCC, MB, RMS and also been implicated in a variety of other cancers such as cancers of the digestive tract[7-10, 114]. The Ptch-Smo interaction is the most commonly disrupted step in cancers associated with ectopic Hh signalling. Although presently every step of the Hh signalling cascade is being studied for the development of anti-cancer therapeutics, the Smo GPCR has undoubtedly emerged as the most successful druggable candidate of the pathway[88, 89, 111].

The GPCR families of receptors collectively form the targets of most drugs and it is not uncommon that Smo being a GPCR has gained such intense success as a target for the development of anti-cancer therapeutics[171]. The teratogenic compound cyclopamine was the first identified Hh inhibitor described to dock in the cavity of the Smo transmembrane domains[84, 85]. Although cyclopamine was able to slow down tumor growth in animal models it was essentially not developed for therapeutic use. The soluble and more potent cyclopamine derivative saridegib also known as IPI-926 has entered a clinical trial for BCC[174]. Presently the only FDA approved Smo inhibitor is GDC-0449 or vismodegib[88, 89, 91]. Vismodegib is marketed as Erivedge and the recommended daily dose is 150mg administered orally for the treatment of adults with metastatic BCC or locally advanced BCC[174].

When GDC-0449 was administered for the treatment of MB it caused a rapid regression of tumor and reduction of symptoms in MB patients but there was relapse of the tumor with a D473H resistant mutation in the sixth transmembrane helix of the Smo receptor[97]. Additionally alanine scan mutagenesis of the Smo receptor identified novel mutations such as E518K that could confer resistance to GDC-0449[124]. Therefore second generation drugs that act on GDC-0449 resistant Smo mutants need to be developed. In the present work we have shown that the Smo CRD binds to the synthetic glucocorticoid, Bud, as elaborated in Chapter 4. Furthermore, Bud acts on wild-type as well GDC-0449 resistant Smo mutants[93]. Hence, we speculated that the Smo CRD could be evaluated for the development of Bud like molecules with higher potency that will represent a novel class of anti-cancer therapeutics.

Presently, the structures of the Smo transmembrane domains have been elucidated in complex with different antagonists namely LY2940680, SANT-1, cyclopamine and Anta XV[64, 85, 90]. We hypothesize that combining the structural scaffolds of the different Smo antagonists will help generate second generation Smo antagonists which will perhaps be more potent than the individual parent Smo antagonists. The Smo antagonist LY2940680 penetrates the cavity of the 7TM domains of Smo to a lesser extent when compared to that of SANT-1 which docks deeper inside the cavity[64]. Thus combining these two molecules as depicted in **Figure 6-3** may generate a Smo antagonist with higher efficacy.

Additionally, we speculate that there is indeed an endogenous ligand for Smo which functions by bridging the extracellular Smo CRD and the 7TM domains. Therefore synthesizing a drug that possesses the capability to recruit the Smo extracellular as well as the transmembrane structures simultaneously in theory may have a higher potency than small molecules occupying individual domains. In such respects we observed that the fatty-acid sidechain of the phosphatidylinositols binds to the Smo CRD as described in chapter 5. Therefore we speculated that another novel class of Smo antagonists can be synthesized by adding a fatty-acid sidechain to the known Smo antagonists such as cyclopamine and GDC-0449. The designs of such Smo inhibitors are depicted in **Figure 6-4**. As presumed if the fatty-acid sidechain of these antagonists bind to Smo CRD and the ring scaffolds dock in the cavity of the 7TM domains, these molecules may have a better chance of Smo regulation.

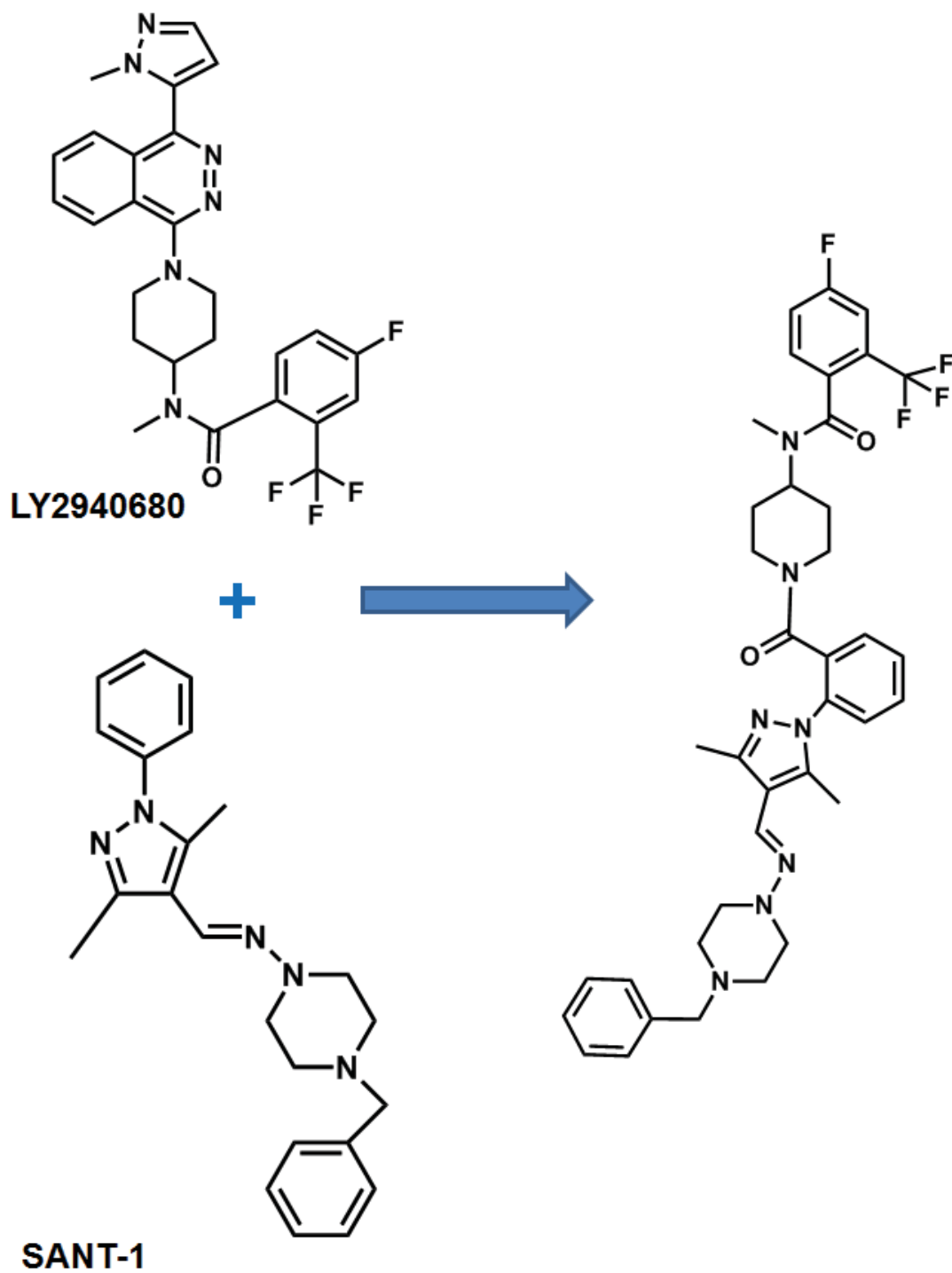


Figure 6-3. Design of Smoothened inhibitors.

Existing Smo inhibitors LY2940680 and SANT may be combined chemically to synthesize a novel Smo inhibitor.

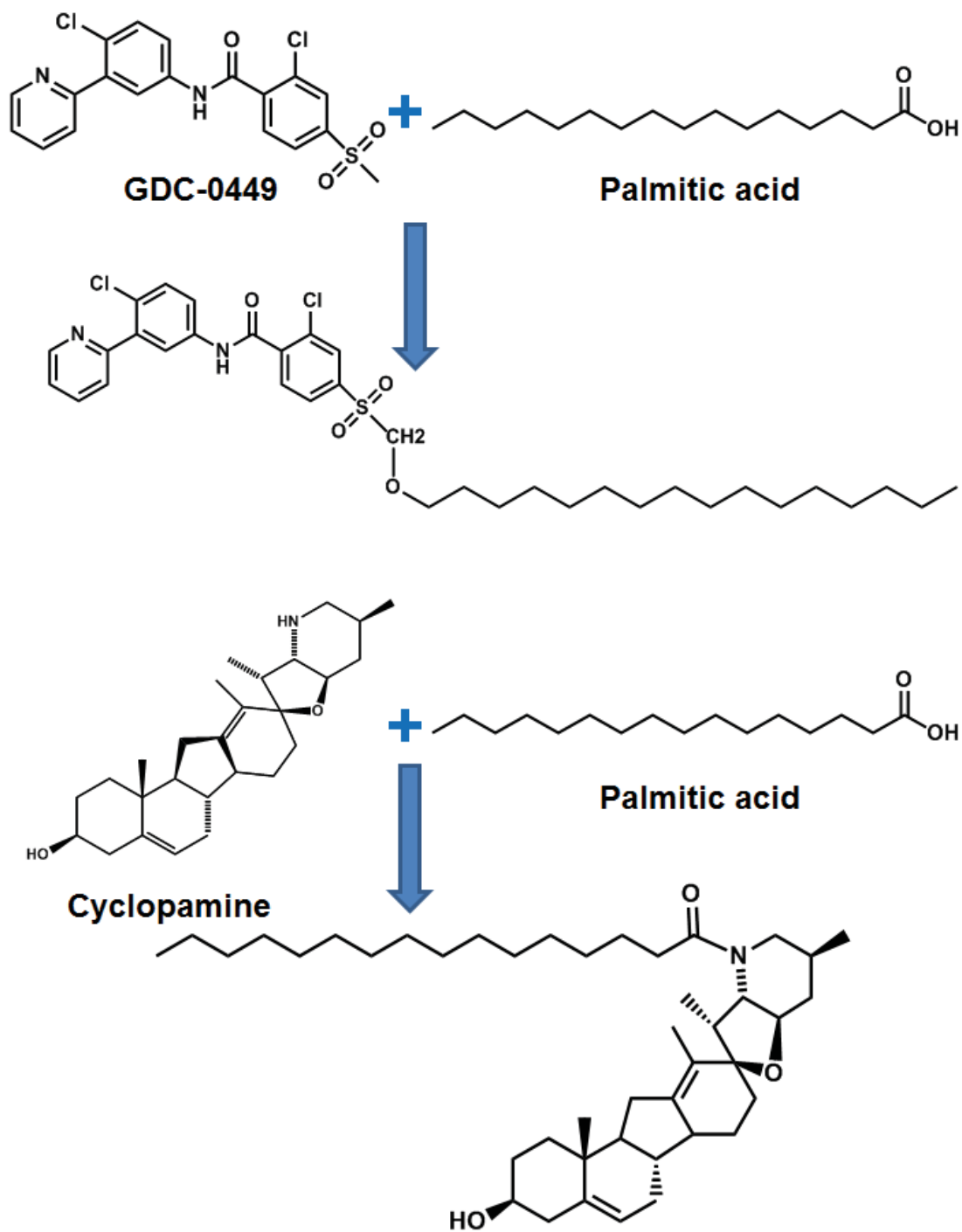


Figure 6-4. Design of Smo inhibitors.

A palmitic acid side chain may be linked chemically to the known Smo inhibitors such as GDC-0449 and cyclopamine to synthesize a novel class of Smo regulators.

LIST OF REFERENCES

1. Nussleinvohard C, Wieschaus E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* 1980,**287**:795-801.
2. Ingham PW, McMahon AP. Hedgehog signaling in animal development: paradigms and principles. *Genes & Development* 2001,**15**:3059-3087.
3. Jiang J, Hui C. Hedgehog Signaling in Development and Cancer. *Developmental Cell* 2008,**15**:801-812.
4. Beachy PA, Karhadkar SS, Berman DM. Tissue repair and stem cell renewal in carcinogenesis. *Nature* 2004,**432**:324-331.
5. Bale AE. Hedgehog signaling and human disease. *Annual Review of Genomics and Human Genetics* 2002,**3**:47-65.
6. Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, *et al.* Mutations in the human Sonic hedgehog gene cause holoprosencephaly. *Nature Genetics* 1996,**14**:357-360.
7. Hahn H, Wicking C, Zaphiropoulos PG, Gailani MR, Shanley S, Chidambaram A, *et al.* Mutations of the human homolog of *Drosophila* patched in the nevoid basal cell carcinoma syndrome. *Cell* 1996,**85**:841-851.
8. Zibat A, Missiaglia E, Rosenberger A, Pritchard-Jones K, Shipley J, Hahn H, *et al.* Activation of the hedgehog pathway confers a poor prognosis in embryonal and fusion gene-negative alveolar rhabdomyosarcoma. *Oncogene* 2010,**29**:6323-6330.
9. Gottardo NG, Gajjar A. Chemotherapy for Malignant Brain Tumors of Childhood. *Journal of Child Neurology* 2008,**23**:1149-1159.
10. Barakat MT, Humke EW, Scott MP. Learning from Jekyll to control Hyde: Hedgehog signaling in development and cancer. *Trends in Molecular Medicine* 2010,**16**:337-348.
11. Ingham PW, Nakano Y, Seger C. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nature Reviews Genetics* 2011,**12**:393-406.
12. Varjosalo M, Li SP, Taipale J. Divergence of hedgehog signal transduction mechanism between *Drosophila* and mammals. *Developmental Cell* 2006,**10**:177-186.
13. van den Heuvel M, Ingham PW. *smoothed* encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature* 1996,**382**:547-551.
14. Alcedo J, Ayzenzon M, VonOhlen T, Noll M, Hooper JE. The *Drosophila* *smoothed* gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell* 1996,**86**:221-232.
15. Ogden SK, Fei DL, Schilling NS, Ahmed YF, Hwa J, Robbins DJ. G protein G alpha(i) functions immediately downstream of *Smoothed* in Hedgehog signalling. *Nature* 2008,**456**:967-U976.
16. Schulte G. International Union of Basic and Clinical Pharmacology. LXXX. The Class Frizzled Receptors. *Pharmacological Reviews* 2010,**62**:632-667.
17. Mohler J. Requirements for Hedgehog, a segment polarity gene, in patterning larval and adult cuticle of *Drosophila*. *Genetics* 1988,**120**:1061-1072.

18. Briscoe J, Therond PP. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature Reviews Molecular Cell Biology* 2013,**14**:416-429.
19. Avaron F, Hoffman L, Guay D, Akimenko MA. Characterization of two new zebrafish members of the hedgehog family: Atypical expression of a zebrafish indian hedgehog gene in skeletal elements of both endochondral and dermal origins. *Developmental Dynamics* 2006,**235**:478-489.
20. Varjosalo M, Taipale J. Hedgehog: functions and mechanisms. *Genes & Development* 2008,**22**:2454-2472.
21. Porter JA, Ekker SC, Park WJ, vonKessler DP, Young KE, Chen CH, *et al.* Hedgehog patterning activity: Role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* 1996,**86**:21-34.
22. Porter JA, Young KE, Beachy PA. Cholesterol modification of hedgehog signaling proteins in animal development. *Science* 1996,**274**:255-259.
23. Heal WP, Jovanovic B, Bessin S, Wright MH, Magee AI, Tate EW. Bioorthogonal chemical tagging of protein cholesterylation in living cells. *Chemical Communications* 2011,**47**:4081-4083.
24. Zeng X, Goetz JA, Suber LM, Scott WJ, Schreiner CM, Robbins DJ. A freely diffusible form of Sonic hedgehog mediates long-range signalling. *Nature* 2001,**411**:716-720.
25. Guerrero I, Chiang C. A conserved mechanism of Hedgehog gradient formation by lipid modifications. *Trends in Cell Biology* 2007,**17**:1-5.
26. Whalen DM, Malinauskas T, Gilbert RJC, Siebold C. Structural insights into proteoglycan-shaped Hedgehog signaling. *Proceedings of the National Academy of Sciences of the United States of America* 2013,**110**:16420-16425.
27. Burke R, Nellen D, Bellotto M, Hafen E, Senti KA, Dickson BJ, *et al.* Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell* 1999,**99**:803-815.
28. Hooper JE, Scott MP. The Drosophila patched gene encodes a putative membrane-protein required for segmental patterning. *Cell* 1989,**59**:751-765.
29. Carpenter D, Stone DM, Brush J, Ryan A, Armanini M, Frantz G, *et al.* Characterization of two patched receptors for the vertebrate hedgehog protein family. *Proceedings of the National Academy of Sciences of the United States of America* 1998,**95**:13630-13634.
30. Stone DM, Hynes M, Armanini M, Swanson TA, Gu QM, Johnson RL, *et al.* The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* 1996,**384**:129-134.
31. Chen Y, Struhl G. Dual roles for patched in sequestering and transducing hedgehog. *Cell* 1996,**87**:553-563.
32. Ingham PW, Taylor AM, Nakano Y. Role of the Drosophila patched gene in positional signaling. *Nature* 1991,**353**:184-187.
33. Marigo V, Davey RA, Zuo Y, Cunningham JM, Tabin CJ. Biochemical evidence that Patched is the Hedgehog receptor. *Nature* 1996,**384**:176-179.
34. Kuwabara PE, Labouesse M. The sterol-sensing domain: multiple families, a unique role? *Trends in Genetics* 2002,**18**:193-201.

35. Taipale J, Cooper MK, Maiti T, Beachy PA. Patched acts catalytically to suppress the activity of Smoothed. *Nature* 2002,**418**:892-897.
36. Deneff N, Neubuser D, Perez L, Cohen SM. Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothed. *Cell* 2000,**102**:521-531.
37. Khaliullina H, Panakova D, Eugster C, Riedel F, Carvalho M, Eaton S. Patched regulates Smoothed trafficking using lipoprotein-derived lipids. *Development* 2009,**136**:4111-4121.
38. Rohatgi R, Milenkovic L, Scott MP. Patched1 regulates Hedgehog signaling at the primary cilium. *Science* 2007,**317**:372-376.
39. Beachy PA, Hymowitz SG, Lazarus RA, Leahy DJ, Siebold C. Interactions between Hedgehog proteins and their binding partners come into view. *Genes & Development* 2010,**24**:2001-2012.
40. Lum L, Yao SQ, Mozer B, Rovescalli A, Von Kessler D, Nirenberg M, *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* 2003,**299**:2039-2045.
41. Yao SQ, Lum L, Beachy P. The ihog cell-surface proteins bind hedgehog and mediate pathway activation. *Cell* 2006,**125**:343-357.
42. Zheng XY, Mann RK, Sever N, Beachy PA. Genetic and biochemical definition of the Hedgehog receptor. *Genes & Development* 2010,**24**:57-71.
43. Tenzen T, Allen BL, Cole F, Kang JS, Krauss RS, McMahon AP. The cell surface membrane proteins Cdo and Boc are components and targets of the hedgehog signaling pathway and feedback network in mice. *Developmental Cell* 2006,**10**:647-656.
44. McLellan JS, Zheng XY, Hauk G, Ghirlando R, Beachy PA, Leahy DJ. The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. *Nature* 2008,**455**:979-U962.
45. Izzi L, Levesque M, Morin S, Laniel D, Wilkes BC, Mille F, *et al.* Boc and Gas1 each form distinct Shh receptor complexes with Ptch1 and are required for Shh-mediated cell proliferation. *Developmental Cell* 2011,**20**:788-801.
46. Zhang W, Kang JS, Cole F, Yi MJ, Krauss RS. Cdo functions at multiple points in the sonic hedgehog pathway, and Cdo-deficient mice accurately model human holoprosencephaly. *Developmental cell* 2006,**10**:657-665.
47. Robbins DJ, Nybakken KE, Kobayashi R, Sisson JC, Bishop JM, Therond PP. Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein costal2. *Cell* 1997,**90**:225-234.
48. Sisson JC, Ho KS, Suyama K, Scott MP. Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway. *Cell* 1997,**90**:235-245.
49. Liu Y, Cao X, Jiang J, Jia J. Fused-Costal2 protein complex regulates Hedgehog-induced Smo phosphorylation and cell-surface accumulation. *Genes & Development* 2007,**21**:1949-1963.
50. Zhang WS, Zhao Y, Tong C, Wang GL, Wang B, Jia JH, *et al.* Hedgehog-regulated Costal2-kinase complexes control phosphorylation and proteolytic processing of Cubitus interruptus. *Developmental Cell* 2005,**8**:267-278.

51. Ruel L, Gallet A, Raisin S, Truchi A, Staccini-Lavenant L, Cervantes A, *et al.* Phosphorylation of the atypical kinesin Costal2 by the kinase Fused induces the partial disassembly of the Smoothened-Fused-Costal2-Cubitus interruptus complex in Hedgehog signalling. *Development* 2007,**134**:3677-3689.
52. Wang GL, Amanai K, Wang B, Jiang J. Interactions with Costal2 and Suppressor of fused regulate nuclear translocation and activity of Cubitus interruptus. *Genes & Development* 2000,**14**:2893-2905.
53. Methot N, Basler K. Suppressor of Fused opposes Hedgehog signal transduction by impeding nuclear accumulation of the activator form of Cubitus interruptus. *Development* 2000,**127**:4001-4010.
54. Svard J, Henricson KH, Persson-Lek M, Rozell B, Lauth M, Bergstrm A, *et al.* Genetic elimination of Suppressor of fused reveals an essential repressor function in the mammalian Hedgehog signaling pathway. *Developmental Cell* 2006,**10**:187-197.
55. Alexandre C, Jacinto A, Ingham PW. Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes & Development* 1996,**10**:2003-2013.
56. Sasaki H, Nishizaki Y, Hui CC, Nakafuku M, Kondoh H. Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling. *Development* 1999,**126**:3915-3924.
57. Hui C-c, Angers S. Gli Proteins in Development and Disease. In: *Annual Review of Cell and Developmental Biology, Vol 27*. Edited by Schekman R, Goldstein L, Lehmann R; 2011. pp. 513-537.
58. Sasaki H, Hui CC, Nakafuku M, Kondoh H. A binding site for Gli proteins is essential for HNF-3 beta floor plate enhancer activity in transgenics and can respond to Shh in vitro. *Development* 1997,**124**:1313-1322.
59. Pavletich NP, Pabo CO. Crystal Structure of a Five-Finger Gli-DNA complex: New Perspectives on Zinc Fingers. *Science* 1993,**261**:1701-1707.
60. Tsanev R, Tiigimaegi P, Michelson P, Metsis M, Osterlund T, Kogerman P. Identification of the gene transcription repressor domain of Gli3. *Febs Letters* 2009,**583**:224-228.
61. AzaBlanc P, RamirezWeber FA, Laget MP, Schwartz C, Kornberg TB. Proteolysis that is inhibited by Hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* 1997,**89**:1043-1053.
62. Katritch V, Cherezov V, Stevens RC. Structure-Function of the G Protein-Coupled Receptor Superfamily. In: *Annual Review of Pharmacology and Toxicology, Vol 53, 2013*. Edited by Insel PA; 2013. pp. 531-556.
63. Foord SM, Bonner TI, Neubig RR, Rosser EM, Pin JP, Davenport AP, *et al.* International Union of Pharmacology. XLVI. G protein-coupled receptor list. *Pharmacological Reviews* 2005,**57**:279-288.
64. Wang C, Wu H, Katritch V, Han GW, Huang X-P, Liu W, *et al.* Structure of the human smoothened receptor bound to an antitumour agent. *Nature* 2013,**497**:338-343.

65. Dann CE, Hsieh JC, Rattner A, Sharma D, Nathans J, Leahy DJ. Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. *Nature* 2001,**412**:86-90.
66. Janda CY, Waghray D, Levin AM, Thomas C, Garcia KC. Structural Basis of Wnt Recognition by Frizzled. *Science* 2012.
67. Chen Y, Jiang J. Decoding the phosphorylation code in Hedgehog signal transduction. *Cell Research* 2013,**23**:186-200.
68. Zhao Y, Tong C, Jiang J. Hedgehog regulates smoothed activity by inducing a conformational switch. *Nature* 2007,**450**:252-U254.
69. Chen Y, Li S, Tong C, Zhao Y, Wang B, Liu Y, *et al.* G protein-coupled receptor kinase 2 promotes high-level Hedgehog signaling by regulating the active state of Smo through kinase-dependent and kinase-independent mechanisms in *Drosophila*. *Genes & Development* 2010,**24**:2054-2067.
70. Jia H, Liu Y, Xia R, Tong C, Yue T, Jiang J, *et al.* Casein Kinase 2 Promotes Hedgehog Signaling by Regulating both Smoothened and Cubitus Interruptus. *Journal of Biological Chemistry* 2010,**285**:37218-37226.
71. Jia J, Tong C, Wang B, Luo L, Jiang J. Hedgehog signalling activity of Smoothened requires phosphorylation by protein kinase A and casein kinase I. *Nature (London)* 2004,**432**:1045-1050.
72. Maier D, Cheng S, Faubert D, Hipfner DR. A broadly conserved G-Protein-Coupled Receptor Kinase phosphorylation mechanism controls *Drosophila* Smoothened Activity. *PLoS Genet* 2014,**10**:e1004399.
73. Chen Y, Sasai N, Ma G, Yue T, Jia J, Briscoe J, *et al.* Sonic Hedgehog Dependent Phosphorylation by CK1 alpha and GRK2 Is Required for Ciliary Accumulation and Activation of Smoothened. *Plos Biology* 2011,**9**.
74. Xie JW, Murone M, Luoh SM, Ryan A, Gu QM, Zhang CH, *et al.* Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* 1998,**391**:90-92.
75. Wong HC, Bourdelas A, Krauss A, Lee HJ, Shao YM, Wu DQ, *et al.* Direct binding of the PDZ domain of Dishevelled to a conserved internal sequence in the C-terminal region of frizzled. *Molecular Cell* 2003,**12**:1251-1260.
76. Umbhauer M, Djiane A, Goisset C, Penzo-Mendez A, Riou JF, Boucaut JC, *et al.* The C-terminal cytoplasmic Lys-Thr-X-X-X-Trp motif in frizzled receptors mediates Wnt/beta-catenin signalling. *Embo Journal* 2000,**19**:4944-4954.
77. Carroll CE, Marada S, Stewart DP, Ouyang JX, Ogden SK. The extracellular loops of Smoothened play a regulatory role in control of Hedgehog pathway activation (vol 139, pg 612, 2012). *Development* 2012,**139**:827-827.
78. Chen W, Ren XR, Nelson CD, Barak LS, Chen JK, Beachy PA, *et al.* Activity-dependent internalization of smoothed mediated by beta-arrestin 2 and GRK2. *Science* 2004,**306**:2257-2260.
79. Riobo NA, Saucy B, DiLizio C, Manning DR. Activation of heterotrimeric G proteins by Smoothened. *Proceedings of the National Academy of Sciences of the United States of America* 2006,**103**:12607-12612.
80. Nakano Y, Nystedt S, Shivdasani AA, Strutt H, Thomas C, Ingham PW. Functional domains and sub-cellular distribution of the Hedgehog transducing protein Smoothened in *Drosophila*. *Mechanisms of Development* 2004,**121**:507-518.

81. Aanstad P, Santos N, Corbit KC, Scherz PJ, Trinh LA, Salvenmoser W, *et al.* The Extracellular Domain of Smoothed Regulates Ciliary Localization and Is Required for High-Level Hh Signaling. *Current Biology* 2009,**19**:1034-1039.
82. Heretsch P, Tzagkaroulaki L, Giannis A. Modulators of the hedgehog signaling pathway. *Bioorganic & Medicinal Chemistry* 2010,**18**:6613-6624.
83. Chen JK, Taipale J, Young KE, Maiti T, Beachy PA. Small molecule modulation of Smoothed activity. *Proceedings of the National Academy of Sciences of the United States of America* 2002,**99**:14071-14076.
84. Chen JK, Taipale J, Cooper MK, Beachy PA. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothed. *Genes & Development* 2002,**16**:2743-2748.
85. Weierstall U, James D, Wang C, White TA, Wang D, Liu W, *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature Communications* 2014,**5**.
86. Taipale J, Chen JK, Cooper MK, Wang BL, Mann RK, Milenkovic L, *et al.* Effects of oncogenic mutations in Smoothed and Patched can be reversed by cyclopamine. *Nature* 2000,**406**:1005-1009.
87. Sinha S, Chen JK. Purmorphamine activates the Hedgehog pathway by targeting Smoothed. *Nature Chemical Biology* 2006,**2**:29-30.
88. Robarge KD, Brunton SA, Castanedo GM, Cui Y, Dina MS, Goldsmith R, *et al.* GDC-0449-A potent inhibitor of the hedgehog pathway. *Bioorganic & Medicinal Chemistry Letters* 2009,**19**:5576-5581.
89. Rudin CM, Hann CL, Laterra J, Yauch RL, Callahan CA, Fu L, *et al.* Brief Report: Treatment of Medulloblastoma with Hedgehog Pathway Inhibitor GDC-0449. *New England Journal of Medicine* 2009,**361**:1173-1178.
90. Wang C, Wu H, Evron T, Vardy E, Han GW, Huang X-P, *et al.* Structural basis for Smoothed receptor modulation and chemoresistance to anticancer drugs. *Nat Commun* 2014,**5**.
91. De Smaele E, Ferretti E, Gulino A. Vismodegib, a small-molecule inhibitor of the Hedgehog pathway for the treatment of advanced cancers. *Current Opinion in Investigational Drugs* 2010,**11**:707-718.
92. Nachtergaele S, Mydock LK, Krishnan K, Rammohan J, Schlesinger PH, Covey DF, *et al.* Oxysterols are allosteric activators of the oncoprotein Smoothed. *Nature Chemical Biology* 2012,**8**:211-220.
93. Wang Y, Davidow L, Arvanites AC, Blanchard J, Lam K, Xu K, *et al.* Glucocorticoid Compounds Modify Smoothed Localization and Hedgehog Pathway Activity. *Chemistry & Biology* 2012,**19**:972-982.
94. Wang J, Lu J, Bond MC, Chen M, Ren X-R, Lyerly HK, *et al.* Identification of select glucocorticoids as Smoothed agonists: Potential utility for regenerative medicine. *Proceedings of the National Academy of Sciences of the United States of America* 2010,**107**:9323-9328.
95. Myers BR, Sever N, Chong YC, Kim J, Belani JD, Rychnovsky S, *et al.* Hedgehog Pathway Modulation by Multiple Lipid Binding Sites on the Smoothed Effector of Signal Response. *Developmental Cell* 2013,**26**:346-357.

96. Nedelcu D, Liu J, Xu Y, Jao C, Salic A. Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nature Chemical Biology* 2013,**9**:557-U554.
97. Yauch RL, Dijkgraaf GJP, Alicke B, Januario T, Ahn CP, Holcomb T, *et al.* Smoothened Mutation Confers Resistance to a Hedgehog Pathway Inhibitor in Medulloblastoma. *Science* 2009,**326**:572-574.
98. Yavari A, Nagaraj R, Owusu-Ansah E, Folick A, Ngo K, Hillman T, *et al.* Role of Lipid Metabolism in Smoothened Derepression in Hedgehog Signaling. *Developmental Cell* 2010,**19**:54-65.
99. Incardona JP, Gruenberg J, Roelink H. Sonic hedgehog induces the segregation of patched and smoothened in endosomes. *Current Biology* 2002,**12**:983-995.
100. Lum L, Zhang C, Oh S, Mann RK, von Kessler DP, Taipale J, *et al.* Hedgehog signal transduction via Smoothened association with a cytoplasmic complex scaffolded by the atypical kinesin, Costal-2. *Molecular Cell* 2003,**12**:1261-1274.
101. Ogden SK, Ascano M, Stegman MA, Suber LM, Hooper JE, Robbins DJ. Identification of a functional interaction between the transmembrane protein Smoothened and the Kinesin-related protein Costal2. *Current Biology* 2003,**13**:1998-2003.
102. Ranieri N, Ruel L, Gallet A, Raisin S, Therond PP. Distinct Phosphorylations on Kinesin Costal-2 Mediate Differential Hedgehog Signaling Strength. *Developmental Cell* 2012,**22**:279-294.
103. Smelkinson MG, Kalderon D. Processing of the Drosophila hedgehog signalling effector Ci-155 to receptor Ci-75 is mediated by direct binding to the SCF component slimb. *Current Biology* 2006,**16**:110-116.
104. Jia JH, Zhang L, Zhang Q, Tong C, Wang B, Hou FJ, *et al.* Phosphorylation by double-time/CKI epsilon and CKI alpha targets cubitus interruptus for slimb/beta-TRCP-mediated proteolytic processing. *Developmental Cell* 2005,**9**:819-830.
105. Ayers KL, Therond PP. Evaluating Smoothened as a G-protein-coupled receptor for Hedgehog signalling. *Trends in Cell Biology* 2010,**20**:287-298.
106. Meloni AR, Fralish GB, Kelly P, Salahpour A, Chen JK, Wechsler-Reya RJ, *et al.* Smoothened signal transduction is promoted by G protein-coupled receptor kinase 2. *Molecular and Cellular Biology* 2006,**26**:7550-7560.
107. Kovacs JJ, Whalen EJ, Liu R, Xiao K, Kim J, Chen M, *et al.* Beta-arrestin-mediated localization of Smoothened to the primary cilium. *Science* 2008,**320**:1777-1781.
108. Molla-Herman A, Boularan C, Ghossoub R, Scott MGH, Burtey A, Zarka M, *et al.* Targeting of beta-Arrestin2 to the Centrosome and Primary Cilium: Role in Cell Proliferation Control. *Plos One* 2008,**3**.
109. Johnson RL, Rothman AL, Xie J, Goodrich LV, Bare JW, Bonifas JM, *et al.* Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* 1996,**272**:1668-1671.
110. Gailani MR, StahleBackdahl M, Leffell DJ, Glynn M, Zaphiropoulos PG, Pressman C, *et al.* The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas. *Nature Genetics* 1996,**14**:78-81.
111. Rubin LL, de Sauvage FJ. Targeting the Hedgehog pathway in cancer. *Nature Reviews Drug Discovery* 2006,**5**:1026-1033.

112. Scales SJ, de Sauvage FJ. Mechanisms of Hedgehog pathway activation in cancer and implications for therapy. *Trends in Pharmacological Sciences* 2009,**30**:303-312.
113. Roewert-Huber J, Lange-Asschenfeldt B, Stockfleth E, Kerl H. Epidemiology and aetiology of basal cell carcinoma. *British Journal of Dermatology* 2007,**157**:47-51.
114. Berman DM, Karhadkar SS, Maitra A, de Oca RM, Gerstenblith MR, Briggs K, *et al.* Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature* 2003,**425**:846-851.
115. Yauch RL, Gould SE, Scales SJ, Tang T, Tian H, Ahn CP, *et al.* A paracrine requirement for hedgehog signalling in cancer. *Nature* 2008,**455**:406-410.
116. Gailani MR, Bale SJ, Leffell DJ, Digiovanna JJ, Peck GL, Poliak S, *et al.* Developmental defects in Gorlin Syndrome related to a putative tumor suppressor gene on chromosome-9. *Cell* 1992,**69**:111-117.
117. Epstein EH. Basal cell carcinomas: attack of the hedgehog. *Nature Reviews Cancer* 2008,**8**:743-754.
118. Cowey CL. Targeted therapy for advanced Basal-cell carcinoma: vismodegib and beyond. *Dermatology and therapy* 2013,**3**:17-31.
119. Tas S, Avci O. Induction of the differentiation and apoptosis of tumor cells in vivo with efficiency and selectivity. *European Journal of Dermatology* 2004,**14**:96-102.
120. Sekulic A, Migden MR, Oro AE, Dirix L, Lewis KD, Hainsworth JD, *et al.* Efficacy and safety of Vismodegib in advanced basal-cell carcinoma. *New England Journal of Medicine* 2012,**366**:2171-2179.
121. Kostaras X, Easaw JC. Management of recurrent medulloblastoma in adult patients: a systematic review and recommendations. *Journal of Neuro-Oncology* 2013,**115**:1-8.
122. Gibson P, Tong Y, Robinson G, Thompson MC, Currle DS, Eden C, *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* 2010,**468**:1095-1099.
123. Goodrich LV, Scott MP. Hedgehog and Patched in neural development and disease. *Neuron* 1998,**21**:1243-1257.
124. Dijkgraaf GJP, Alicke B, Weinmann L, Januario T, West K, Modrusan Z, *et al.* Small Molecule Inhibition of GDC-0449 Refractory Smoothened Mutants and Downstream Mechanisms of Drug Resistance. *Cancer Research* 2011,**71**:435-444.
125. Kappler R, Calzada-Wack J, Schnitzbauer U, Koleva M, Herwig A, Piontek G, *et al.* Molecular characterization of Patched-associated rhabdomyosarcoma. *Journal of Pathology* 2003,**200**:348-356.
126. Ruat M, Hoch L, Faure H, Rognan D. Targeting of Smoothened for therapeutic gain. *Trends in Pharmacological Sciences* 2014,**35**:237-246.
127. Gallinari P, Filocamo G, Jones P, Pazzaglia S, Steinkuhler C. Smoothened antagonists: a promising new class of antitumor agents. *Expert Opinion on Drug Discovery* 2009,**4**:525-544.

128. Jimeno A, Weiss GJ, Miller WH, Jr., Gettinger S, Eigel BJC, Chang ALS, *et al.* Phase I Study of the Hedgehog Pathway Inhibitor IPI-926 in Adult Patients with Solid Tumors. *Clinical Cancer Research* 2013,**19**:2766-2774.
129. Lauth M, Bergstrom A, Shimokawa T, Toftgard R. Inhibition of GLI-mediated transcription and tumor cell growth by small-molecule antagonists. *Proceedings of the National Academy of Sciences of the United States of America* 2007,**104**:8455-8460.
130. Keeler J. *Understanding NMR Spectroscopy*. West Sussex PO19 8SQ England: WILEY; 2005.
131. Keller RLJ. The Computer Aided Resonance Assignment Tutorial. . In. CANTINA Verlag, Goldau, Switzerland; 2004.
132. Higman P. Protein NMR. In. Bristol UK; November 2012.
133. Guntert P. Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, N.J.)* 2004,**278**:353-378.
134. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *Journal of Biomolecular NMR* 2002,**24**:171-189.
135. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* 1999,**13**:289-302.
136. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* 1963,**7**:95-&.
137. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. *Advances in protein chemistry* 1968,**23**:283-438.
138. Pellecchia M, Sem DS, Wuthrich K. NMR in drug discovery. *Nature Reviews Drug Discovery* 2002,**1**:211-219.
139. Komoroski EM, Komoroski RA, Valentine JL, Pearce JM, Kearns GL. The use of nuclear magnetic resonance spectroscopy in the detection of drug intoxication. *Journal of Analytical Toxicology* 2000,**24**:180-187.
140. Kosir IJ, Kidric J. Use of modern nuclear magnetic resonance spectroscopy in wine analysis: determination of minor compounds. *Analytica Chimica Acta* 2002,**458**:77-84.
141. Mayer M, Meyer B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society* 2001,**123**:6108-6117.
142. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *Journal of Biomolecular NMR* 2001,**21**:349-359.
143. Williamson MP. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* 2013,**73**:1-16.
144. Olejniczak ET, Meadows RP, Wang H, Cai ML, Nettlesheim DG, Fesik SW. Improved NMR structures of protein/ligand complexes using residual dipolar couplings. *Journal of the American Chemical Society* 1999,**121**:9249-9250.
145. Dominguez C, Boelens R, Bonvin A. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* 2003,**125**:1731-1737.

146. Pei J, Grishin NV. Cysteine-rich domains related to Frizzled receptors and Hedgehog-interacting proteins. *Protein Science* 2012,**21**:1172-1184.
147. Rana R, Carroll CE, Lee H-J, Bao J, Marada S, Grace CRR, *et al.* Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nat Commun* 2013,**4**.
148. Bax A, Grzesiek S. Methodological Advances in Protein Nmr. *Accounts of Chemical Research* 1993,**26**:131-138.
149. Sharma D, Rajarathnam K. C-13 NMR chemical shifts can predict disulfide bond formation. *Journal of Biomolecular Nmr* 2000,**18**:165-171.
150. Nachtergaele S, Whalen DM, Mydock LK, Zhao Z, Malinauskas T, Krishnan K, *et al.* Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *Elife* 2013,**2**.
151. Rousseau GG, Schmit JP. Structure-activity relationship for glucocorticoids.1. Determination of receptor binding and biological activity. *Journal of Steroid Biochemistry and Molecular Biology* 1977,**8**:911-919.
152. Olefsky JM, Kimmerling G. Effects of glucocorticoids on carbohydrate-metabolism. *American Journal of the Medical Sciences* 1976,**271**:202-210.
153. Rhen T, Cidlowski JA. Antiinflammatory action of glucocorticoids - New mechanisms for old drugs. *New England Journal of Medicine* 2005,**353**:1711-1723.
154. Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. *Molecular and Cellular Endocrinology* 2011,**335**:2-13.
155. Zheng J, Cahill SM, Lemmon MA, Fushman D, Schlessinger J, Cowburn D. Identification of the binding site for acidic phospholipids on the pH domain of dynamin: implications for stimulation of GTPase activity. *Journal of Molecular Biology* 1996,**255**:14-21.
156. D.A. Case, T.A. Darden, T.E. Cheatham I, C.L. Simmerling, J. Wang, R.E. Duke, *et al.* AMBER 10. *University of California, San Francisco*. 2008.
157. Hubbard SJ, Thornton JM. NACCESS. In. Department of Biochemistry and Molecular Biology, University College London.; 1993.
158. Dominguez C, Boelens R, Bonvin A. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* 2003,**125**:1731-1737.
159. Shan JF, Zheng J. Optimizing Dvl PDZ domain inhibitor by exploring chemical space. *Journal of Computer-Aided Molecular Design* 2009,**23**:37-47.
160. London TBC, Lee HJ, Shao YM, Zheng J. Interaction between the internal motif KTXXXI of Idax and mDvl PDZ domain. *Biochemical and Biophysical Research Communications* 2004,**322**:326-332.
161. Bazan JF, de Sauvage FJ. Structural Ties between Cholesterol Transport and Morphogen Signaling. *Cell* 2009,**138**:1055-1056.
162. Roberg-Larsen H, Strand MF, Krauss S, Wilson SR. Metabolites in vertebrate Hedgehog signaling. *Biochemical and Biophysical Research Communications* 2014,**446**:669-674.

163. Grace CRR, Perrin MH, DiGrucchio MR, Miller CL, Rivier JE, Vale WW, *et al.* NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor. *Proceedings of the National Academy of Sciences of the United States of America* 2004,**101**:12836-12841.
164. Koth CM, Murray JM, Mukund S, Madjidi A, Minn A, Clarke HJ, *et al.* Molecular basis for negative regulation of the glucagon receptor. *Proceedings of the National Academy of Sciences of the United States of America* 2012,**109**:14393-14398.
165. Dong MQ, Li ZJ, Pinon DI, Lybrand TP, Miller LJ. Spatial approximation between the amino terminus of a peptide agonist and the top of the sixth transmembrane segment of the secretin receptor. *Journal of Biological Chemistry* 2004,**279**:2894-2903.
166. Berg JM TJ, Stryer L. *Biochemistry*. New York: W H Freeman; 2002.
167. Tan J, Brill JA. Cinderella story: PI4P goes from precursor to key signaling molecule. *Critical Reviews in Biochemistry and Molecular Biology* 2014,**49**:33-58.
168. Lemmon MA. Membrane recognition by phospholipid-binding domains. *Nature Reviews Molecular Cell Biology* 2008,**9**:99-111.
169. Skwarek LC, Boulianne GL. Great Expectations for PIP: Phosphoinositides as Regulators of Signaling During Development and Disease. *Developmental Cell* 2009,**16**:12-20.
170. Yang Y, Lin X. Hedgehog Signaling Uses Lipid Metabolism to Tune Smoothed Activation. *Developmental Cell* 2010,**19**:3-4.
171. Stevens RC, Cherezov V, Katritch V, Abagyan R, Kuhn P, Rosen H, *et al.* The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nature Reviews Drug Discovery* 2013,**12**:25-34.
172. Strutt H, Thomas C, Nakano Y, Stark D, Neave B, Taylor AM, *et al.* Mutations in the sterol-sensing domain of Patched suggest a role for vesicular trafficking in Smoothed regulation. *Current Biology* 2001,**11**:608-613.
173. Kim I, Pan W, Jones SA, Zhang Y, Zhuang X, Wu D. Clathrin and AP2 are required for PtdIns(4,5)P-2-mediated formation of LRP6 signalosomes. *Journal of Cell Biology* 2013,**200**:419-428.
174. Lin TL, Matsui W. Hedgehog pathway as a drug target: Smoothed inhibitors in development. *Oncotargets and Therapy* 2012,**5**:47-58.

APPENDIX A. *DROSOPHILA* SMO CRD ATOM ASSIGNMENTS

Residue Number	Residue Name	Atom	Chemical Shift
85	VAL	C	175.218
85	VAL	CA	57.885
85	VAL	CB	35.895
85	VAL	CG1	18.515
85	VAL	CG2	20.273
85	VAL	H	7.805
85	VAL	HA	5.297
85	VAL	HB	1.461
85	VAL	QG1	0.422
85	VAL	QG2	-0.26
85	VAL	N	112.794
86	ARG	C	174.128
86	ARG	CA	54.192
86	ARG	CB	34.757
86	ARG	CG	26.085
86	ARG	CD	43.85
86	ARG	H	8.43
86	ARG	HA	4.578
86	ARG	HB2	1.638
86	ARG	HB3	1.534
86	ARG	QG	1.486
86	ARG	QD	2.977
86	ARG	N	117.854
87	ARG	C	176.265
87	ARG	CA	57.824
87	ARG	CB	29.913
87	ARG	CG	28.214
87	ARG	CD	43.181
87	ARG	H	8.569
87	ARG	HA	4.619
87	ARG	QB	1.723
87	ARG	QG	1.517
87	ARG	QD	3.168
87	ARG	HE	7.419
87	ARG	N	124.928

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
87	ARG	NE	117.269
88	ALA	C	174.173
88	ALA	CA	51.382
88	ALA	CB	23.05
88	ALA	H	8.082
88	ALA	HA	4.379
88	ALA	QB	1.236
88	ALA	N	127.39
89	ARG	C	177.847
89	ARG	CA	55.271
89	ARG	CB	31.215
89	ARG	CG	26.896
89	ARG	CD	43.086
89	ARG	H	8.133
89	ARG	HA	4.692
89	ARG	HB2	1.8
89	ARG	HB3	1.745
89	ARG	QG	1.587
89	ARG	QD	3.18
89	ARG	HE	7.313
89	ARG	N	116.82
89	ARG	NE	116.871
90	CYS	C	175.288
90	CYS	CA	51.785
90	CYS	CB	38.246
90	CYS	H	9.709
90	CYS	HA	5.732
90	CYS	HB2	3.014
90	CYS	HB3	2.514
90	CYS	N	126.567
91	TYR	CA	54.951
91	TYR	CB	39.807
91	TYR	CD1	134.091
91	TYR	CE1	116.951
91	TYR	H	8.87
91	TYR	HA	5.079
91	TYR	HB2	3.04

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
91	TYR	HB3	2.699
91	TYR	QD	6.785
91	TYR	QE	6.318
91	TYR	N	120.309
92	PRO	CA	61.773
92	PRO	CB	31.173
92	PRO	CG	27.009
92	PRO	CD	50.108
92	PRO	HA	4.836
92	PRO	HB2	2.03
92	PRO	HB3	1.936
92	PRO	HG2	2.185
92	PRO	HG3	1.994
92	PRO	HD2	3.862
92	PRO	HD3	3.712
93	THR	C	173.688
93	THR	CA	61.324
93	THR	CB	70.191
93	THR	CG2	21.227
93	THR	H	7.978
93	THR	HA	4.253
93	THR	HB	4.041
93	THR	QG2	0.885
93	THR	N	115.177
94	SER	C	175.079
94	SER	CA	57.057
94	SER	CB	66.02
94	SER	H	8.451
94	SER	HA	4.702
94	SER	HB2	4.058
94	SER	HB3	3.811
94	SER	N	116.822
95	ASN	C	175.705
95	ASN	CA	55.564
95	ASN	CB	37.637
95	ASN	H	8.802
95	ASN	HA	4.385

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
95	ASN	QB	2.715
95	ASN	HD21	7.564
95	ASN	HD22	6.893
95	ASN	N	118.803
95	ASN	ND2	112.721
96	ALA	C	179.229
96	ALA	CA	53.087
96	ALA	CB	19.563
96	ALA	H	8.004
96	ALA	HA	4.262
96	ALA	QB	1.299
96	ALA	N	118.571
97	THR	C	175.334
97	THR	CA	60.586
97	THR	CB	70.195
97	THR	CG2	21.698
97	THR	H	7.551
97	THR	HA	4.581
97	THR	HB	4.653
97	THR	QG2	0.945
97	THR	N	106.312
98	ASN	C	174.221
98	ASN	CA	52.79
98	ASN	CB	36.768
98	ASN	H	7.897
98	ASN	HA	4.528
98	ASN	HB2	3.194
98	ASN	HB3	2.633
98	ASN	HD21	7.285
98	ASN	HD22	6.414
98	ASN	N	122.599
98	ASN	ND2	106.804
99	THR	C	173.651
99	THR	CA	59.011
99	THR	CB	72.745
99	THR	CG2	20.913
99	THR	H	7.49

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
99	THR	HA	4.979
99	THR	HB	3.486
99	THR	QG2	0.841
99	THR	N	107.151
100	CYS	C	173.688
100	CYS	CA	55.634
100	CYS	CB	45.854
100	CYS	H	8.632
100	CYS	HA	4.095
100	CYS	HB2	2.919
100	CYS	HB3	2.802
100	CYS	N	118.809
101	PHE	C	176.261
101	PHE	CA	59.067
101	PHE	CB	36.576
101	PHE	CD1	132.752
101	PHE	CE1	130.73
101	PHE	CZ	131.44
101	PHE	H	9.018
101	PHE	HA	4.275
101	PHE	HB2	3.301
101	PHE	HB3	3.021
101	PHE	QD	6.883
101	PHE	QE	6.886
101	PHE	HZ	7.334
101	PHE	N	128.19
102	GLY	C	174.173
102	GLY	CA	45.112
102	GLY	H	8.566
102	GLY	HA2	4.239
102	GLY	HA3	3.561
102	GLY	N	104.504
103	SER	C	173.873
103	SER	CA	57.436
103	SER	CB	63.718
103	SER	H	7.533
103	SER	HA	4.529

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
103	SER	QB	3.73
103	SER	N	116.201
104	LYS	C	176.632
104	LYS	CA	56.409
104	LYS	CB	32.169
104	LYS	CG	24.365
104	LYS	CD	28.605
104	LYS	CE	41.628
104	LYS	H	8.768
104	LYS	HA	4.056
104	LYS	QB	1.538
104	LYS	QG	1.198
104	LYS	QD	1.417
104	LYS	QE	2.785
104	LYS	N	127.456
105	LEU	CA	52.01
105	LEU	CB	40.68
105	LEU	CG	26.621
105	LEU	CD1	25.453
105	LEU	CD2	22.219
105	LEU	H	8.304
105	LEU	HA	4.12
105	LEU	HB2	1.323
105	LEU	HB3	0.205
105	LEU	HG	1.239
105	LEU	QD1	0.63
105	LEU	QD2	0.626
105	LEU	N	128.016
106	PRO	CA	62.807
106	PRO	CB	29.625
106	PRO	CG	26.537
106	PRO	CD	49.961
106	PRO	HA	4.453
106	PRO	HB2	2.145
106	PRO	HB3	1.926
106	PRO	HG2	1.95
106	PRO	HG3	1.776

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
106	PRO	HD2	3.53
106	PRO	HD3	3.027
107	TYR	C	175.239
107	TYR	CA	53.718
107	TYR	CB	38.637
107	TYR	CD1	134.566
107	TYR	CE1	117.894
107	TYR	H	6.311
107	TYR	HA	4.689
107	TYR	HB2	3.173
107	TYR	HB3	3.026
107	TYR	QD	6.836
107	TYR	QE	6.525
107	TYR	N	116.871
108	GLU	C	176.282
108	GLU	CA	57.502
108	GLU	CB	31.105
108	GLU	CG	36.1
108	GLU	H	8.591
108	GLU	HA	4.138
108	GLU	QB	1.864
108	GLU	QG	2.135
108	GLU	N	118.221
109	LEU	C	175.42
109	LEU	CA	53.202
109	LEU	CB	43.587
109	LEU	CG	27.153
109	LEU	CD1	25.498
109	LEU	CD2	23.139
109	LEU	H	8.176
109	LEU	HA	4.974
109	LEU	HB2	2.217
109	LEU	HB3	0.963
109	LEU	HG	1.507
109	LEU	QD1	0.713
109	LEU	QD2	0.588
109	LEU	N	119.826

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
110	SER	C	173.833
110	SER	CA	53.006
110	SER	CB	64.691
110	SER	H	7.924
110	SER	HA	5.02
110	SER	HB2	3.208
110	SER	HB3	2.838
110	SER	N	111.561
111	SER	C	175.602
111	SER	CA	57.388
111	SER	CB	66.333
111	SER	H	8.628
111	SER	HA	4.841
111	SER	QB	3.016
111	SER	N	111.459
112	LEU	C	179.571
112	LEU	CA	56.032
112	LEU	CB	41.904
112	LEU	CG	26.938
112	LEU	CD1	25.003
112	LEU	CD2	24.013
112	LEU	H	9.519
112	LEU	HA	4.383
112	LEU	QB	1.551
112	LEU	HG	1.294
112	LEU	QD1	0.413
112	LEU	QD2	0.552
112	LEU	N	128.995
113	ASP	C	176.486
113	ASP	CA	58.352
113	ASP	CB	41.259
113	ASP	H	8.191
113	ASP	HA	4.168
113	ASP	HB2	2.754
113	ASP	HB3	2.418
113	ASP	N	115.013
114	LEU	C	175.556

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
114	LEU	CA	54.483
114	LEU	CB	39.219
114	LEU	CG	26.318
114	LEU	CD1	21.976
114	LEU	H	8.732
114	LEU	HA	4.329
114	LEU	HB2	2.141
114	LEU	HB3	1.47
114	LEU	HG	1.056
114	LEU	QQD	0.645
114	LEU	N	118.298
115	THR	C	172.812
115	THR	CA	58.759
115	THR	CB	71.985
115	THR	CG2	20.435
115	THR	H	7.574
115	THR	HA	2.784
115	THR	HB	2.647
115	THR	QG2	0.328
115	THR	N	106.427
116	ASP	CA	52.757
116	ASP	CB	40.737
116	ASP	HA	4.398
116	ASP	HB2	2.459
116	ASP	HB3	2.076
117	PHE	C	176.169
117	PHE	CA	57.477
117	PHE	CB	39.27
117	PHE	CD1	131.518
117	PHE	CE1	132.594
117	PHE	CZ	131.862
117	PHE	H	7.575
117	PHE	HA	4.498
117	PHE	HB2	3.109
117	PHE	HB3	2.648
117	PHE	QD	7.151
117	PHE	QE	7.407

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
117	PHE	HZ	6.935
117	PHE	N	117.929
118	HIS	C	175.08
118	HIS	CA	56.439
118	HIS	CB	31.104
118	HIS	CD2	119.685
118	HIS	CE1	137.643
118	HIS	H	9.23
118	HIS	HA	4.585
118	HIS	HB2	3.258
118	HIS	HB3	3.071
118	HIS	HD2	7.098
118	HIS	HE1	8.077
118	HIS	N	115.333
119	THR	C	173.696
119	THR	CA	58.725
119	THR	CB	72.851
119	THR	CG2	21.622
119	THR	H	7.343
119	THR	HA	4.528
119	THR	HB	4.387
119	THR	QG2	1.096
119	THR	N	106.891
120	GLU	C	178.417
120	GLU	CA	59.38
120	GLU	CB	29.873
120	GLU	CG	36.482
120	GLU	H	8.841
120	GLU	HA	3.611
120	GLU	HB2	1.797
120	GLU	HB3	1.528
120	GLU	HG2	1.89
120	GLU	HG3	1.617
120	GLU	N	121.798
121	LYS	C	177.745
121	LYS	CA	59.47
121	LYS	CB	32.13

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
121	LYS	CG	24.437
121	LYS	CD	28.386
121	LYS	CE	42.016
121	LYS	H	8.307
121	LYS	HA	3.903
121	LYS	QB	1.751
121	LYS	HG2	1.288
121	LYS	HG3	1.214
121	LYS	QD	1.599
121	LYS	QE	2.897
121	LYS	N	121.091
122	GLU	C	178.051
122	GLU	CA	59.094
122	GLU	CB	30.468
122	GLU	CG	36.919
122	GLU	H	7.537
122	GLU	HA	3.852
122	GLU	QB	1.948
122	GLU	HG2	2.143
122	GLU	HG3	2.039
122	GLU	N	117.79
123	LEU	C	178.07
123	LEU	CA	59.007
123	LEU	CB	42.75
123	LEU	CG	27.001
123	LEU	CD1	24.927
123	LEU	H	7.321
123	LEU	HA	3.211
123	LEU	HB2	1.81
123	LEU	HB3	1.569
123	LEU	HG	1.295
123	LEU	QQD	1.021
123	LEU	N	118.763
124	ASN	C	177.689
124	ASN	CA	57.31
124	ASN	CB	39.288
124	ASN	H	8.335

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
124	ASN	HA	4.127
124	ASN	HB2	2.714
124	ASN	HB3	2.529
124	ASN	HD21	7.498
124	ASN	HD22	6.842
124	ASN	N	116.987
124	ASN	ND2	112.59
125	ASP	C	178.868
125	ASP	CA	57.51
125	ASP	CB	39.596
125	ASP	H	8.376
125	ASP	HA	4.306
125	ASP	HB2	2.629
125	ASP	HB3	2.51
125	ASP	N	120.935
126	LYS	C	179.458
126	LYS	CA	55.698
126	LYS	CB	31.082
126	LYS	CG	24.521
126	LYS	CD	28.082
126	LYS	CE	41.937
126	LYS	H	7.748
126	LYS	HA	4.168
126	LYS	QB	1.78
126	LYS	QG	1.171
126	LYS	QD	1.726
126	LYS	QE	2.938
126	LYS	N	119.626
127	LEU	C	178.732
127	LEU	CA	57.722
127	LEU	CB	41.236
127	LEU	CG	25.852
127	LEU	CD1	25.602
127	LEU	CD2	22.655
127	LEU	H	8.221
127	LEU	HA	3.688
127	LEU	HB2	1.661

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
127	LEU	HB3	1.339
127	LEU	HG	1.13
127	LEU	QD1	0.392
127	LEU	QD2	0.046
127	LEU	N	120.255
128	ASN	C	178.394
128	ASN	CA	55.628
128	ASN	CB	38.269
128	ASN	H	7.482
128	ASN	HA	4.58
128	ASN	HB2	2.914
128	ASN	HB3	2.864
128	ASN	HD21	7.775
128	ASN	HD22	6.846
128	ASN	N	113.63
128	ASN	ND2	113.685
129	ASP	C	179.095
129	ASP	CA	56.623
129	ASP	CB	39.883
129	ASP	H	7.834
129	ASP	HA	4.482
129	ASP	HB2	2.866
129	ASP	HB3	2.455
129	ASP	N	120.683
130	TYR	C	176.554
130	TYR	CA	62.142
130	TYR	CB	37.436
130	TYR	CD1	133.432
130	TYR	CE1	118.373
130	TYR	H	8.132
130	TYR	HA	3.984
130	TYR	HB2	3.38
130	TYR	HB3	3.1
130	TYR	QD	7.33
130	TYR	QE	6.75
130	TYR	N	116.512
131	TYR	C	177.915

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
131	TYR	CA	61.395
131	TYR	CB	37.345
131	TYR	CD1	133.47
131	TYR	CE1	118.363
131	TYR	H	7.558
131	TYR	HA	3.632
131	TYR	QB	2.851
131	TYR	QD	7.174
131	TYR	QE	6.895
131	TYR	N	118.786
132	ALA	C	179.39
132	ALA	CA	54.142
132	ALA	CB	18.404
132	ALA	H	7.453
132	ALA	HA	3.935
132	ALA	QB	1.354
132	ALA	N	120.715
133	LEU	C	177.28
133	LEU	CA	56.634
133	LEU	CB	41.045
133	LEU	CG	28.196
133	LEU	CD1	26.081
133	LEU	CD2	23.773
133	LEU	H	7.764
133	LEU	HA	3.765
133	LEU	HB2	2.069
133	LEU	HB3	1.57
133	LEU	HG	1.63
133	LEU	QD1	0.868
133	LEU	QD2	0.67
133	LEU	N	117.212
134	LYS	C	177.121
134	LYS	CA	58.865
134	LYS	CB	31.247
134	LYS	CG	23.834
134	LYS	CD	28.957
134	LYS	CE	41.183

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
134	LYS	H	7.001
134	LYS	HA	3.039
134	LYS	HB2	0.896
134	LYS	HB3	0.818
134	LYS	HG2	0.127
134	LYS	HG3	0.047
134	LYS	QD	1.077
134	LYS	HE2	2.529
134	LYS	HE3	2.46
134	LYS	N	118.756
135	HIS	C	174.232
135	HIS	CA	57.245
135	HIS	CB	30.221
135	HIS	CD2	120.468
135	HIS	CE1	138.791
135	HIS	H	6.973
135	HIS	HA	4.304
135	HIS	HB2	3.129
135	HIS	HB3	2.868
135	HIS	HD2	7.035
135	HIS	HE1	7.623
135	HIS	N	113.331
136	VAL	CA	59.159
136	VAL	CB	30.604
136	VAL	CG1	22.649
136	VAL	CG2	20.108
136	VAL	H	7.821
136	VAL	HA	4.531
136	VAL	HB	2.15
136	VAL	QG1	0.935
136	VAL	QG2	0.77
136	VAL	N	121.486
137	PRO	CA	66.637
137	PRO	CB	32.212
137	PRO	CG	27.321
137	PRO	CD	51.022
137	PRO	HA	4.07

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
137	PRO	HB2	2.405
137	PRO	HB3	1.977
137	PRO	QG	1.998
137	PRO	HD2	3.75
137	PRO	HD3	3.568
138	LYS	C	180.048
138	LYS	CA	59.398
138	LYS	CB	31.296
138	LYS	CG	24.657
138	LYS	CD	27.876
138	LYS	CE	41.76
138	LYS	H	9.234
138	LYS	HA	4.099
138	LYS	QB	1.877
138	LYS	QG	1.597
138	LYS	QD	1.718
138	LYS	QE	2.931
138	LYS	N	116.832
139	CYS	C	175.012
139	CYS	CA	58.297
139	CYS	CB	38.803
139	CYS	H	7.496
139	CYS	HA	4.32
139	CYS	HB2	3.23
139	CYS	HB3	3.031
139	CYS	N	116.333
140	TRP	C	177.462
140	TRP	CA	58.821
140	TRP	CB	30.425
140	TRP	CD1	126.832
140	TRP	CE3	119.77
140	TRP	CZ2	113.921
140	TRP	CZ3	122.123
140	TRP	CH2	124.328
140	TRP	H	8.254
140	TRP	HA	4.452
140	TRP	HB2	3.234

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
140	TRP	HB3	2.964
140	TRP	HD1	7.061
140	TRP	HE1	9.844
140	TRP	HE3	6.825
140	TRP	HZ2	6.824
140	TRP	HZ3	6.175
140	TRP	HH2	5.215
140	TRP	N	119.987
140	TRP	NE1	128.198
141	ALA	C	178.845
141	ALA	CA	54.188
141	ALA	CB	18.323
141	ALA	H	7.968
141	ALA	HA	3.736
141	ALA	QB	1.324
141	ALA	N	116.093
142	ALA	C	178.414
142	ALA	CA	53.243
142	ALA	CB	19.832
142	ALA	H	6.929
142	ALA	HA	4.147
142	ALA	QB	1.321
142	ALA	N	116.351
143	ILE	C	176.35
143	ILE	CA	62.228
143	ILE	CB	39.375
143	ILE	CG1	31.092
143	ILE	CG2	17.874
143	ILE	CD1	14.59
143	ILE	H	8.695
143	ILE	HA	4.212
143	ILE	HB	2.045
143	ILE	HG12	1.697
143	ILE	QG2	0.958
143	ILE	QD1	1.128
143	ILE	N	115.15
144	GLN	C	173.47

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
144	GLN	CA	61.829
144	GLN	CB	25.973
144	GLN	CG	33.328
144	GLN	H	7.315
144	GLN	HA	3.332
144	GLN	QB	1.555
144	GLN	HG2	1.265
144	GLN	HG3	1.028
144	GLN	HE21	6.316
144	GLN	HE22	5.64
144	GLN	N	122.564
144	GLN	NE2	109.407
145	PRO	CA	65.894
145	PRO	CB	30.481
145	PRO	CG	27.833
145	PRO	CD	49.756
145	PRO	HA	4.076
145	PRO	QB	2.131
145	PRO	HG2	1.913
145	PRO	HD2	3.375
145	PRO	HD3	3.268
146	PHE	C	175.919
146	PHE	CA	59.712
146	PHE	CB	39.329
146	PHE	CD1	132.442
146	PHE	CE1	130.945
146	PHE	CZ	132.38
146	PHE	H	6.892
146	PHE	HA	4.289
146	PHE	HB2	3.23
146	PHE	HB3	2.868
146	PHE	QD	6.865
146	PHE	QE	7.009
146	PHE	HZ	7.309
146	PHE	N	119.137
147	LEU	C	180.297
147	LEU	CA	57.952

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
147	LEU	CB	41.645
147	LEU	CG	26.484
147	LEU	CD1	27.209
147	LEU	CD2	22.994
147	LEU	H	8.589
147	LEU	HA	3.611
147	LEU	HB2	1.918
147	LEU	HB3	1.004
147	LEU	HG	1.643
147	LEU	QD1	0.565
147	LEU	QD2	0.323
147	LEU	N	116.693
148	CYS	C	175.556
148	CYS	CA	57.678
148	CYS	CB	39.396
148	CYS	H	8.001
148	CYS	HA	4.127
148	CYS	HB2	2.898
148	CYS	HB3	2.706
148	CYS	N	115.085
149	ALA	C	177.745
149	ALA	CA	54.306
149	ALA	CB	18.568
149	ALA	H	7.671
149	ALA	HA	3.916
149	ALA	QB	1.354
149	ALA	N	121.743
150	VAL	C	175.806
150	VAL	CA	64.622
150	VAL	CB	32.936
150	VAL	CG1	21.66
150	VAL	CG2	23.716
150	VAL	H	7.46
150	VAL	HA	3.735
150	VAL	HB	1.511
150	VAL	QG1	0.609
150	VAL	QG2	0.296

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
150	VAL	N	115.667
151	PHE	C	176.192
151	PHE	CA	60.42
151	PHE	CB	40.182
151	PHE	CD1	132.71
151	PHE	CE1	129.398
151	PHE	CZ	130.249
151	PHE	H	8.93
151	PHE	HA	4.048
151	PHE	HB2	3.143
151	PHE	HB3	2.704
151	PHE	QD	7.374
151	PHE	QE	7.148
151	PHE	HZ	6.935
151	PHE	N	115.306
152	LYS	C	184.176
152	LYS	CA	53.428
152	LYS	CB	33.274
152	LYS	CG	25.54
152	LYS	CD	29.658
152	LYS	CE	42.81
152	LYS	H	7.802
152	LYS	HA	4.816
152	LYS	QB	1.785
152	LYS	QG	1.226
152	LYS	HD2	1.675
152	LYS	HD3	1.379
152	LYS	HE2	2.815
152	LYS	HE3	2.708
152	LYS	N	115.47
153	PRO	CA	62.477
153	PRO	CB	32.347
153	PRO	CG	27.718
153	PRO	CD	50.34
153	PRO	HA	4.245
153	PRO	QB	1.728
153	PRO	QG	1.547

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
153	PRO	HD2	3.449
153	PRO	HD3	3.147
154	LYS	C	173.356
154	LYS	CA	58.027
154	LYS	CB	31.285
154	LYS	CG	25.275
154	LYS	CD	28.73
154	LYS	CE	40.857
154	LYS	H	8.537
154	LYS	HA	3.039
154	LYS	HB2	0.687
154	LYS	HB3	-0.075
154	LYS	QG	0.411
154	LYS	HD2	0.749
154	LYS	HD3	0.354
154	LYS	HE2	2.183
154	LYS	HE3	2.081
154	LYS	N	122.391
155	CYS	C	171.655
155	CYS	CA	53.804
155	CYS	CB	41.219
155	CYS	H	7.037
155	CYS	HA	5.391
155	CYS	HB2	2.309
155	CYS	HB3	1.792
155	CYS	N	127.562
156	GLU	C	173.371
156	GLU	CA	54.874
156	GLU	CB	34.63
156	GLU	CG	35.918
156	GLU	H	9.401
156	GLU	HA	4.492
156	GLU	HB2	1.894
156	GLU	HB3	1.839
156	GLU	HG2	2.07
156	GLU	HG3	2.026
156	GLU	N	124.483

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
157	LYS	C	176.509
157	LYS	CA	55.856
157	LYS	CB	31.809
157	LYS	CG	24.612
157	LYS	CD	28.476
157	LYS	CE	41.177
157	LYS	H	8.526
157	LYS	HA	4.581
157	LYS	HB2	1.531
157	LYS	HB3	1.391
157	LYS	HG2	1.135
157	LYS	HG3	1.042
157	LYS	QD	1.433
157	LYS	HE2	2.822
157	LYS	HE3	2.767
157	LYS	N	122.694
158	ILE	C	176.337
158	ILE	CA	60.847
158	ILE	CB	39.625
158	ILE	CG1	27.17
158	ILE	CG2	16.511
158	ILE	CD1	12.6
158	ILE	H	9.001
158	ILE	HA	4.049
158	ILE	HB	1.556
158	ILE	HG12	1.365
158	ILE	HG13	1.294
158	ILE	QG2	0.666
158	ILE	QD1	0.58
158	ILE	N	128.382
159	ASN	C	175.496
159	ASN	CA	53.97
159	ASN	CB	37.238
159	ASN	H	9.474
159	ASN	HA	4.212
159	ASN	HB2	2.949
159	ASN	HB3	2.666

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
159	ASN	HD21	7.578
159	ASN	HD22	6.795
159	ASN	N	127.596
159	ASN	ND2	112.813
160	GLY	C	173.651
160	GLY	CA	45.167
160	GLY	H	8.476
160	GLY	HA2	4.094
160	GLY	HA3	3.373
160	GLY	N	101.656
161	GLU	C	174.422
161	GLU	CA	54.294
161	GLU	CB	32.934
161	GLU	CG	35.875
161	GLU	H	7.591
161	GLU	HA	4.587
161	GLU	QB	1.804
161	GLU	HG2	2.105
161	GLU	HG3	1.999
161	GLU	N	120.042
162	ASP	C	176.214
162	ASP	CA	55.408
162	ASP	CB	41.629
162	ASP	H	8.489
162	ASP	HA	4.34
162	ASP	HB2	2.908
162	ASP	HB3	2.555
162	ASP	N	126.603
163	MET	C	172.9
163	MET	CA	53.649
163	MET	CB	35.529
163	MET	CG	32.804
163	MET	CE	18.224
163	MET	H	8.853
163	MET	HA	4.911
163	MET	HB2	1.787
163	MET	HB3	1.278

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
163	MET	HG2	2.468
163	MET	HG3	2.342
163	MET	QE	1.723
163	MET	N	121.977
164	VAL	C	172.991
164	VAL	CA	58.203
164	VAL	CB	34.629
164	VAL	CG1	24.114
164	VAL	CG2	21.05
164	VAL	H	8.964
164	VAL	HA	4.814
164	VAL	HB	1.559
164	VAL	QG1	0.907
164	VAL	QG2	0.685
164	VAL	N	119.963
165	TYR	C	176.464
165	TYR	CA	57.549
165	TYR	CB	38.274
165	TYR	CD1	133.471
165	TYR	CE1	117.567
165	TYR	H	8.137
165	TYR	HA	4.444
165	TYR	HB2	2.789
165	TYR	HB3	2.581
165	TYR	QD	6.75
165	TYR	QE	6.348
165	TYR	N	125.849
166	LEU	CA	52.426
166	LEU	CB	42.038
166	LEU	CG	27.028
166	LEU	CD1	24.515
166	LEU	CD2	24.681
166	LEU	H	8.308
166	LEU	HA	4.683
166	LEU	QB	1.505
166	LEU	HG	1.582
166	LEU	QD1	0.769

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
166	LEU	QD2	0.701
166	LEU	N	123.151
167	PRO	CA	61.704
167	PRO	CB	32.02
167	PRO	CG	27.297
167	PRO	CD	50.256
167	PRO	HA	4.678
167	PRO	HB2	2.237
167	PRO	HB3	2.117
167	PRO	HG2	1.957
167	PRO	HG3	1.875
167	PRO	HD2	3.824
167	PRO	HD3	3.659
168	SER	C	175.473
168	SER	CA	56.324
168	SER	CB	65.193
168	SER	H	7.668
168	SER	HA	4.632
168	SER	HB2	4.265
168	SER	HB3	3.894
168	SER	N	113.955
169	TYR	C	178.116
169	TYR	CA	62.057
169	TYR	CB	37.925
169	TYR	CD1	133.485
169	TYR	CE1	118.18
169	TYR	H	8.492
169	TYR	HA	3.724
169	TYR	HB2	3.294
169	TYR	HB3	2.626
169	TYR	QD	6.95
169	TYR	QE	6.688
169	TYR	N	122.582
170	GLU	C	177.421
170	GLU	CA	58.867
170	GLU	CB	28.969
170	GLU	CG	36.427

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
170	GLU	H	8.503
170	GLU	HA	3.846
170	GLU	HB2	1.932
170	GLU	HB3	1.701
170	GLU	QG	2.15
170	GLU	N	118.362
171	MET	C	176.804
171	MET	CA	58.648
171	MET	CB	33.058
171	MET	CG	32.282
171	MET	CE	16.827
171	MET	H	7.409
171	MET	HA	3.708
171	MET	QB	2.139
171	MET	QG	2.456
171	MET	QE	2.006
171	MET	N	116.221
172	CYS	C	178.047
172	CYS	CA	58.737
172	CYS	CB	35.006
172	CYS	H	6.817
172	CYS	HA	3.857
172	CYS	HB2	2.278
172	CYS	HB3	2.131
172	CYS	N	114.131
173	ARG	C	179.435
173	ARG	CA	59.641
173	ARG	CB	29.273
173	ARG	CG	25.796
173	ARG	CD	43.003
173	ARG	H	8.885
173	ARG	HA	3.376
173	ARG	HB2	1.404
173	ARG	HB3	1.192
173	ARG	QG	1.381
173	ARG	QD	2.872
173	ARG	N	124.629

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
174	ILE	C	175.624
174	ILE	CA	63.811
174	ILE	CB	38.229
174	ILE	CG1	26.623
174	ILE	CG2	17.501
174	ILE	CD1	14.098
174	ILE	H	7.422
174	ILE	HA	3.796
174	ILE	HB	1.853
174	ILE	HG12	1.412
174	ILE	HG13	1.15
174	ILE	QG2	0.846
174	ILE	QD1	0.78
174	ILE	N	113.379
175	THR	C	175.798
175	THR	CA	62.856
175	THR	CB	70.168
175	THR	CG2	22.263
175	THR	H	7.574
175	THR	HA	3.889
175	THR	HB	4.129
175	THR	QG2	1.111
175	THR	N	109.403
176	MET	C	177.174
176	MET	CA	60.235
176	MET	CB	32.338
176	MET	CG	32.327
176	MET	CE	16.098
176	MET	H	7.152
176	MET	HA	3.541
176	MET	QB	2.07
176	MET	QG	2.455
176	MET	QE	1.862
176	MET	N	119.638
177	GLU	CA	60.422
177	GLU	CB	26.103
177	GLU	CG	36.005

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
177	GLU	H	8.58
177	GLU	HA	4.114
177	GLU	HB2	1.952
177	GLU	HB3	1.872
177	GLU	QG	2.105
177	GLU	N	117.224
178	PRO	CA	65.128
178	PRO	CB	31.141
178	PRO	CG	27.893
178	PRO	CD	50.204
178	PRO	HA	4.304
178	PRO	HB2	1.981
178	PRO	HB3	1.437
178	PRO	HG2	1.978
178	PRO	HG3	1.907
178	PRO	HD2	3.597
178	PRO	HD3	3.423
179	CYS	C	174.808
179	CYS	CA	54.06
179	CYS	CB	42.879
179	CYS	H	7.208
179	CYS	HA	4.679
179	CYS	HB2	3.187
179	CYS	HB3	3.071
179	CYS	N	111.879
180	ARG	C	178.21
180	ARG	CA	59.743
180	ARG	CB	29.486
180	ARG	CG	26.057
180	ARG	CD	43.243
180	ARG	H	7.436
180	ARG	HA	3.161
180	ARG	HB2	1.906
180	ARG	HB3	1.642
180	ARG	QG	1.368
180	ARG	QD	3.049
180	ARG	HE	7.016

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
180	ARG	N	123.75
180	ARG	NE	115.983
181	ILE	C	176.577
181	ILE	CA	63.166
181	ILE	CB	37.781
181	ILE	CG1	27.844
181	ILE	CG2	17.306
181	ILE	CD1	13.593
181	ILE	H	8.13
181	ILE	HA	4
181	ILE	HB	1.844
181	ILE	HG12	1.363
181	ILE	HG13	1.214
181	ILE	QG2	0.831
181	ILE	QD1	0.788
181	ILE	N	116.384
182	LEU	C	178.346
182	LEU	CA	54.697
182	LEU	CB	42.872
182	LEU	CG	27.304
182	LEU	CD1	26.603
182	LEU	CD2	22.845
182	LEU	H	7.771
182	LEU	HA	4.54
182	LEU	HB2	1.805
182	LEU	HB3	1.685
182	LEU	HG	1.504
182	LEU	QD1	0.91
182	LEU	QD2	0.826
182	LEU	N	119.738
183	TYR	C	176.169
183	TYR	CA	59.682
183	TYR	CB	37.18
183	TYR	CD1	131.938
183	TYR	CE1	120.703
183	TYR	H	7.886
183	TYR	HA	4.125

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
183	TYR	HB2	2.965
183	TYR	HB3	2.903
183	TYR	QD	6.459
183	TYR	QE	7.004
183	TYR	N	121.945
184	ASN	C	175.612
184	ASN	CA	52.67
184	ASN	CB	38.058
184	ASN	H	8.228
184	ASN	HA	4.885
184	ASN	HB2	2.886
184	ASN	HB3	2.805
184	ASN	HD21	7.568
184	ASN	HD22	6.795
184	ASN	N	115.992
184	ASN	ND2	111.975
185	THR	C	175.844
185	THR	CA	61.815
185	THR	CB	72.484
185	THR	CG2	21.264
185	THR	H	7.649
185	THR	HA	4.682
185	THR	HB	4.469
185	THR	QG2	1.288
185	THR	N	109.094
186	THR	C	175.033
186	THR	CA	63.174
186	THR	CB	68.529
186	THR	CG2	21.201
186	THR	H	8.261
186	THR	HA	4.026
186	THR	HB	4.134
186	THR	QG2	0.981
186	THR	N	112.067
187	PHE	C	175.398
187	PHE	CA	59.89
187	PHE	CB	39.668

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
187	PHE	CD1	131.517
187	PHE	CE1	131.784
187	PHE	H	7.764
187	PHE	HA	4.153
187	PHE	QB	2.884
187	PHE	QD	6.844
187	PHE	QE	7.025
187	PHE	N	118.139
188	PHE	CA	56.245
188	PHE	CB	38.81
188	PHE	CD1	132.999
188	PHE	CE1	130.252
188	PHE	H	7.67
188	PHE	HA	4.662
188	PHE	HB2	2.789
188	PHE	HB3	2.574
188	PHE	QD	6.754
188	PHE	QE	7.029
188	PHE	N	118.447
189	PRO	CA	62.432
189	PRO	CB	32.67
189	PRO	CG	26.937
189	PRO	CD	50.293
189	PRO	HA	4.398
189	PRO	HB2	2.14
189	PRO	HB3	1.387
189	PRO	QG	1.651
189	PRO	HD2	3.327
189	PRO	HD3	3.128
190	LYS	C	177.977
190	LYS	CA	60.279
190	LYS	CB	31.599
190	LYS	CG	24.915
190	LYS	CD	28.946
190	LYS	CE	42.607
190	LYS	H	8.565
190	LYS	HA	3.58

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
190	LYS	QB	1.753
190	LYS	HG2	1.401
190	LYS	HG3	1.314
190	LYS	QD	1.587
190	LYS	QE	2.836
190	LYS	N	118.753
191	PHE	C	173.946
191	PHE	CA	57.841
191	PHE	CB	36.807
191	PHE	CD1	133.281
191	PHE	CE1	132.398
191	PHE	CZ	131.403
191	PHE	H	6.795
191	PHE	HA	4.481
191	PHE	HB2	3.487
191	PHE	HB3	2.661
191	PHE	QD	7.036
191	PHE	QE	6.9
191	PHE	HZ	7.343
191	PHE	N	110.638
192	LEU	C	175.398
192	LEU	CA	53.169
192	LEU	CB	44.901
192	LEU	CG	24.522
192	LEU	CD1	25.563
192	LEU	CD2	25.567
192	LEU	H	7.204
192	LEU	HA	4.71
192	LEU	HB2	1.792
192	LEU	HB3	1.024
192	LEU	HG	1.826
192	LEU	QD1	0.736
192	LEU	QD2	0.691
192	LEU	N	119.653
193	ARG	C	177.099
193	ARG	CA	55.25
193	ARG	CB	30.355

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
193	ARG	CG	27.16
193	ARG	CD	42.817
193	ARG	H	7.09
193	ARG	HA	4.328
193	ARG	QB	1.742
193	ARG	QG	1.587
193	ARG	QD	2.791
193	ARG	N	116.807
194	CYS	C	175.022
194	CYS	CA	54.449
194	CYS	CB	34.958
194	CYS	H	8.21
194	CYS	HA	4.588
194	CYS	HB2	2.881
194	CYS	HB3	2.733
194	CYS	N	117.271
195	ASN	C	175.08
195	ASN	CA	53.188
195	ASN	CB	40.003
195	ASN	H	7.801
195	ASN	HA	4.811
195	ASN	HB2	2.839
195	ASN	HB3	2.67
195	ASN	HD21	7.657
195	ASN	HD22	6.952
195	ASN	N	120.232
195	ASN	ND2	114.087
196	GLU	C	176.447
196	GLU	CA	57.928
196	GLU	CB	29.822
196	GLU	CG	36.338
196	GLU	H	9.086
196	GLU	HA	4.225
196	GLU	QB	1.911
196	GLU	QG	1.924
196	GLU	N	123.435
197	THR	C	174.853

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
197	THR	CA	63.694
197	THR	CB	68.841
197	THR	CG2	21.608
197	THR	H	8.067
197	THR	HA	3.964
197	THR	HB	3.951
197	THR	QG2	1.012
197	THR	N	113.403
198	LEU	C	176.215
198	LEU	CA	55.656
198	LEU	CB	42.291
198	LEU	CG	26.85
198	LEU	CD1	24.96
198	LEU	CD2	23.206
198	LEU	H	7.448
198	LEU	HA	3.964
198	LEU	HB2	1.079
198	LEU	HB3	0.643
198	LEU	HG	1.204
198	LEU	QD1	0.761
198	LEU	QD2	0.617
198	LEU	N	121.377
199	PHE	C	179.412
199	PHE	CA	53.495
199	PHE	CB	39.248
199	PHE	CD1	132.342
199	PHE	CE1	130.432
199	PHE	H	7.542
199	PHE	HA	5.202
199	PHE	HB2	3.008
199	PHE	HB3	2.814
199	PHE	QD	7.259
199	PHE	QE	7.161
199	PHE	N	117.704
200	PRO	CA	62.949
200	PRO	CB	32.242
200	PRO	CG	27.435

Appendix A. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
200	PRO	CD	50.414
200	PRO	HA	4.528
200	PRO	HB2	2.214
200	PRO	HB3	1.891
200	PRO	HG2	1.852
200	PRO	HG3	1.792
200	PRO	HD2	3.754
200	PRO	HD3	3.492
201	THR	C	174.175
201	THR	CA	61.863
201	THR	CB	69.742
201	THR	CG2	21.498
201	THR	H	8.452
201	THR	HA	4.324
201	THR	HB	4.123
201	THR	QG2	1.076
201	THR	N	114.025
202	LYS	CA	57.661
202	LYS	CB	33.467
202	LYS	CG	24.484
202	LYS	CD	28.887
202	LYS	CE	42.239
202	LYS	H	7.831
202	LYS	HA	4.046
202	LYS	HB2	1.692
202	LYS	HB3	1.605
202	LYS	QG	1.248
202	LYS	QD	1.417
202	LYS	QE	3.185
202	LYS	N	127.089

APPENDIX B. HUMAN SMO CRD ATOM ASSIGNMENTS

Residue Number	Residue Name	Atom	Chemical Shift
65	GLY	CA	44.866
65	GLY	H	8.568
65	GLY	N	110.631
66	ARG	CA	54.092
66	ARG	CB	33.568
66	ARG	H	8.399
66	ARG	N	118.431
67	ALA	CA	52.56
67	ALA	CB	18.342
67	ALA	H	8.508
67	ALA	N	126.813
68	ALA	CA	50.575
68	ALA	CB	20.85
68	ALA	H	7.964
68	ALA	N	125.589
70	CYS	CA	56.115
70	CYS	CB	42.206
70	CYS	H	8.373
70	CYS	N	126.153
71	GLU	CA	56.12
71	GLU	CB	30.757
71	GLU	H	9.52
71	GLU	N	123.866
75	TYR	CA	56.313
75	TYR	CB	41.653
75	TYR	H	6.985
75	TYR	N	115.809
76	ASN	CA	52.918
76	ASN	CB	38.726
76	ASN	H	8.345
76	ASN	N	116.528
77	VAL	CA	59.527
77	VAL	CB	34.128
77	VAL	H	7.128
77	VAL	N	115.617
78	CYS	CA	54.99
78	CYS	CB	47.639
78	CYS	H	9.088

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
78	CYS	N	125.011
79	LEU	CA	54.803
79	LEU	CB	40.469
79	LEU	H	9.576
79	LEU	N	127.547
80	GLY	CA	44.911
80	GLY	H	7.752
80	GLY	N	103.062
81	SER	CA	57.403
81	SER	CB	63.607
81	SER	H	7.977
81	SER	N	117.765
82	VAL	CA	62.918
82	VAL	CB	31.214
82	VAL	H	8.767
82	VAL	N	126.97
83	LEU	CA	51.393
83	LEU	CB	40.838
83	LEU	H	8.209
83	LEU	N	130.061
85	TYR	CA	54.328
85	TYR	CB	38.472
85	TYR	H	6.591
85	TYR	N	116.586
86	GLY	CA	45.745
86	GLY	H	8.171
86	GLY	N	102.141
87	ALA	CA	51.542
87	ALA	CB	22.069
87	ALA	H	8.627
87	ALA	N	125.651
88	THR	CA	58.538
88	THR	CB	71.209
88	THR	H	9.046
88	THR	N	112.456
89	SER	CA	56.615
89	SER	CB	65.857
89	SER	H	8.316
89	SER	N	112.546
90	THR	CA	61.864

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
90	THR	CB	67.864
90	THR	H	8.374
90	THR	N	116.183
91	LEU	CA	57.392
91	LEU	CB	43.533
91	LEU	H	7.857
91	LEU	N	120.104
92	LEU	CA	54.971
92	LEU	CB	39.288
92	LEU	H	8.457
92	LEU	N	113.989
93	ALA	CA	49.524
93	ALA	CB	17.887
93	ALA	H	6.923
93	ALA	N	121.028
94	GLY	CA	46.066
94	GLY	H	6.749
94	GLY	N	118.69
95	ASP	CA	52.367
95	ASP	CB	41.471
95	ASP	H	7.187
95	ASP	N	118.08
96	SER	CA	57.935
96	SER	CB	66.627
96	SER	H	7.356
96	SER	N	114.015
97	ASP	CA	53.474
97	ASP	CB	42.061
97	ASP	H	9.119
97	ASP	N	122.753
98	SER	CA	56.16
98	SER	CB	66.723
98	SER	H	7.797
98	SER	N	110.808
99	GLN	CA	60.106
99	GLN	CB	29.619
99	GLN	H	9.56
99	GLN	N	121.247
100	GLU	CA	60.593
100	GLU	CB	28.653

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
100	GLU	H	9.075
100	GLU	N	121.483
101	GLU	CA	58.99
101	GLU	CB	29.949
101	GLU	H	8.079
101	GLU	N	122.149
102	ALA	CA	55.045
102	ALA	CB	18.189
102	ALA	H	8.423
102	ALA	N	123.013
103	HIS	CA	60.525
103	HIS	CB	29.259
103	HIS	H	8.35
103	HIS	N	116.352
104	GLY	CA	46.704
104	GLY	H	7.749
104	GLY	N	104.621
105	LYS	CA	57.303
105	LYS	CB	30.476
105	LYS	H	7.835
105	LYS	N	121.153
106	LEU	CA	58.239
106	LEU	CB	42.051
106	LEU	H	8.105
106	LEU	N	121.045
107	VAL	CA	66.258
107	VAL	CB	31.432
107	VAL	H	7.633
107	VAL	N	116.905
108	LEU	CA	57.476
108	LEU	CB	40.605
108	LEU	H	7.284
108	LEU	N	122.085
109	TRP	CA	59.865
109	TRP	CB	30.311
109	TRP	H	8.289
109	TRP	N	120.219
110	SER	CA	61.491
110	SER	CB	62.728
110	SER	H	7.84

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
110	SER	N	112.283
111	GLY	CA	46.31
111	GLY	H	8.015
111	GLY	N	111.133
112	LEU	CA	55.889
112	LEU	CB	41.383
112	LEU	H	7.922
112	LEU	N	119.636
113	ARG	CA	58.458
113	ARG	CB	29.153
113	ARG	H	7.057
113	ARG	N	121.204
114	ASN	CA	53.244
114	ASN	CB	37.77
114	ASN	H	7.079
114	ASN	N	113.584
115	ALA	CA	49.184
115	ALA	CB	17.697
115	ALA	H	7.853
115	ALA	N	125.41
117	ARG	CA	58.668
117	ARG	CB	29.348
117	ARG	H	8.845
117	ARG	N	116.909
118	CYS	CA	57.851
118	CYS	CB	38.364
118	CYS	H	7.532
118	CYS	N	115.296
119	TRP	CA	59.286
119	TRP	CB	29.708
119	TRP	H	9.272
119	TRP	N	122.445
120	ALA	CA	54.169
120	ALA	CB	18.806
120	ALA	H	7.601
120	ALA	N	113.919
121	VAL	CA	60.897
121	VAL	CB	32.006
121	VAL	H	7.059
121	VAL	N	104.379

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
122	ILE	CA	62.563
122	ILE	CB	36.844
122	ILE	H	9.012
122	ILE	N	123.603
123	GLN	CA	61.609
123	GLN	CB	23.615
123	GLN	H	7.417
123	GLN	N	124.242
125	LEU	CA	57.354
125	LEU	CB	40.939
125	LEU	H	6.297
125	LEU	N	117.729
126	LEU	CA	57.501
126	LEU	CB	40.998
126	LEU	H	8.454
126	LEU	N	116.525
127	CYS	CA	59.057
127	CYS	CB	40.887
127	CYS	H	8.011
127	CYS	N	116.116
128	ALA	CA	54.389
128	ALA	CB	18.63
128	ALA	H	7.43
128	ALA	N	120.222
129	VAL	CA	64.114
129	VAL	CB	32.878
129	VAL	H	7.392
129	VAL	N	114.186
130	TYR	CA	61.808
130	TYR	CB	39.626
130	TYR	H	8.867
130	TYR	N	115.377
131	MET	CA	53.432
131	MET	CB	34.209
131	MET	H	7.561
131	MET	N	114.692
133	LYS	CA	57.887
133	LYS	CB	31.573
133	LYS	H	8.432
133	LYS	N	121.485

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
134	CYS	CA	54.235
134	CYS	CB	36.734
134	CYS	H	7.505
134	CYS	N	128.623
135	GLU	CA	55.98
135	GLU	CB	33.189
135	GLU	H	8.603
135	GLU	N	126.856
136	ASN	CA	54.267
136	ASN	CB	36.849
136	ASN	H	9.374
136	ASN	N	124.032
137	ASP	CA	55.07
137	ASP	CB	38.663
137	ASP	H	7.761
137	ASP	N	111.749
138	ARG	CA	54.987
138	ARG	CB	32.71
138	ARG	H	7.293
138	ARG	N	118.465
139	VAL	CA	59.134
139	VAL	CB	33.74
139	VAL	H	9.109
139	VAL	N	125.421
140	GLU	CA	56.549
140	GLU	CB	30.463
140	GLU	H	8.206
140	GLU	N	126.308
141	LEU	CA	51.69
141	LEU	CB	40.686
141	LEU	H	8.228
141	LEU	N	124.66
143	SER	CA	56.767
143	SER	CB	65.101
143	SER	H	7.938
143	SER	N	113.49
144	ARG	CA	60.111
144	ARG	CB	29.343
144	ARG	H	8.436
144	ARG	N	123.139

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
145	THR	CA	66.274
145	THR	CB	68.46
145	THR	H	8.113
145	THR	N	114.103
146	LEU	CA	57.543
146	LEU	CB	42.032
146	LEU	H	7.4
146	LEU	N	123.411
147	CYS	CA	60.097
147	CYS	CB	37.029
147	CYS	H	6.975
147	CYS	N	114.788
148	GLN	CA	58.404
148	GLN	CB	27.241
148	GLN	H	9.146
148	GLN	N	122.935
149	ALA	CA	53.979
149	ALA	CB	18.176
149	ALA	H	7.834
149	ALA	N	120.735
150	THR	CA	62.771
150	THR	CB	72.025
150	THR	H	7.17
150	THR	N	101.283
151	ARG	CA	58.497
151	ARG	CB	30.728
151	ARG	H	7.362
151	ARG	N	119.293
152	GLY	CA	46.648
152	GLY	H	8.977
152	GLY	N	105.493
154	CYS	CA	52.293
154	CYS	CB	40.849
154	CYS	H	7.359
154	CYS	N	109.833
155	ALA	CA	55.667
155	ALA	CB	17.957
155	ALA	H	7.446
155	ALA	N	126.103
156	ILE	CA	62.983

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
156	ILE	CB	37.631
156	ILE	H	8.319
156	ILE	N	115.894
157	VAL	CA	65.461
157	VAL	CB	31.77
157	VAL	H	7.715
157	VAL	N	121.439
158	GLU	CA	58.554
158	GLU	CB	29.821
158	GLU	H	6.958
158	GLU	N	118.562
159	ARG	CA	57.319
159	ARG	CB	30.724
159	ARG	H	7.484
159	ARG	N	115.69
160	GLU	CA	57.288
160	GLU	CB	30.585
160	GLU	H	7.638
160	GLU	N	117.391
161	ARG	CA	56.929
161	ARG	CB	31.426
161	ARG	H	7.82
161	ARG	N	116.655
162	GLY	CA	43.749
162	GLY	H	7.955
162	GLY	N	108.456
163	TRP	CA	54.034
163	TRP	CB	28.994
163	TRP	H	8.387
163	TRP	N	121.443
165	ASP	CA	57.769
165	ASP	CB	39.321
165	ASP	H	8.931
165	ASP	N	121.675
166	PHE	CA	56.906
166	PHE	CB	36.906
166	PHE	H	7.552
166	PHE	N	112.988
167	LEU	CA	52.917
167	LEU	CB	43.643

Appendix B. Continued.

Residue Number	Residue Name	Atom	Chemical Shift
167	LEU	H	7.532
167	LEU	N	119.315
168	ARG	CA	56.122
168	ARG	CB	30.429
168	ARG	H	7.251
168	ARG	N	118.248
169	CYS	CA	52.541
169	CYS	CB	34.684
169	CYS	H	8.249
169	CYS	N	123.03
170	THR	CA	59.851
170	THR	CB	68.048
170	THR	H	8.411
170	THR	N	115.601
172	ASP	CA	54.686
172	ASP	CB	38.818
172	ASP	H	8.532
172	ASP	N	113.366
173	ARG	CA	56.361
173	ARG	CB	33.227
173	ARG	H	7.159
173	ARG	N	118.602
174	PHE	CA	54.961
174	PHE	CB	39.168
174	PHE	H	7.755
174	PHE	N	114.404
176	GLU	CA	56.3
176	GLU	CB	30.512
176	GLU	H	8.554
176	GLU	N	120.929
177	GLY	CA	46.064
177	GLY	H	8.03
177	GLY	N	117.379

VITA

Rajashree Rana was born in Calcutta, India in 1984. She received her Bachelors of Science specializing in Biotechnology from the Bangalore University in 2005. She completed her Masters in Science in Medical Biochemistry from Kasturba Medical College, Manipal University in 2008. She joined the Integrated Program in Biomedical Sciences at the University of Tennessee Health Science Center for graduate studies in 2009. She carried out her dissertation project in Dr. Jie Zheng's lab in the Department of Structural Biology at St. Jude Children's Research Hospital. She is expected to receive a Ph.D. degree in December 2014.