Structural and biochemical studies of Cysteine-Rich Domains in morphogen signalling

Samuel C. Griffiths

A thesis submitted in partial fulfilment of the requirements for the degree of *Doctor of Philosophy*



Magdalen College
Division of Structural Biology
University of Oxford

Trinity 2016

Structural and biochemical studies of Cysteine-Rich Domains in morphogen signalling

Samuel C. Griffiths

A thesis submitted in partial fulfilment of the requirements for the degree of *Doctor of Philosophy*
Magdalen College, Division of Structural Biology, University of Oxford
Trinity Term 2016

Abstract

Cell-to-cell communication is vital to the morphogenesis and homeostasis of multicellular organisms and is linked to a variety of diseases. Extracellular Cysteine-Rich Domains (CRDs) are involved in the control of fundamental signalling pathways, mediating crucial interactions with protein and small molecule ligands. The localisation of extracellular ligands and receptors is regulated by the cell surface glycosaminoglycan (GAG) layer. In this thesis, structural and biochemical methods have been used to characterise two multi-domain proteins containing CRDs – HHIP and ROR2.

In Chapter 3, the process of structure solution of the HHIP N-terminal CRD (HHIP-N) by Sulphur-SAD phasing techniques is described. HHIP-N is a novel CRD, with implications for small molecule binding present. Chapter 4 concerns the study of the regulation of HHIP secretion and its interaction with GAGs. Discrete GAG-binding motifs are found on the surfaces of HHIP-N and the C-terminal region of HHIP (HHIP-C). Structural studies with GAGs enabled snapshots of cell surface-bound states. Interactions were dissected using cellular assays and GAG-binding experiments. In addition, the dynamics of HHIP oligomerisation were analysed using solution scattering and ultracentrifugation techniques. Work in this chapter represents a complex mode of HHIP assembly regulated by GAGs.

In chapter 5, structural and biophysical studies of ROR2 were performed. A structure of the ROR2 CRD with its C-terminal Kringle domain was determined. The structure forms a potentially stable structural unit with a marked inter-domain interface. The CRD structure is distinct from related CRDs and suggests that it cannot directly bind Wnt proteins. This was complemented by biophysical and cellular assays.

In summary, the CRD is a fold found in two unrelated signalling pathways tied to a number of diseases and its role in multi-domain protein architectures has been characterised within this thesis. CRDs are involved in a wide range of interactions and these are essential for the precise regulation of signalling pathway activation.

Acknowledgements

Firstly, I must acknowledge my supervisor Christian Siebold for his perpetual insight, enthusiasm, encouragement and support over the past four years. The opportunity to work in STRUBI has been a wonderful one, and has enabled me to interact with a range of individuals, attempt a variety of experiments and gain invaluable experience as a structural biology. It is particularly necessary to acknowledge Benjamin Bishop. His aid in experimental investigation has been invaluable, providing advice and assistance in mammalian cell culture, protein purification, microscopy and other laboratory techniques. This has been supplemented by a number of discussions and comments concerning both Hedgehog signalling and also this thesis. I would also like to thank Ross Robinson and Jonathan Elegheert for laboratory supervision and advice, alongside Tom Walter, Karl Harlos and Kamel El Omari for aid in crystallography, data collection and phasing. I must also acknowledge Robert Gilbert for his aid in performing ultracentrifugation experiments. Further thanks are extended to Armin Wagner and Ramona Duman and the other staff on beamline I23. Additional acknowledgements concern Tomas Malinauskas, who was often on hand to give useful feedback regarding figures. Our laboratory manager, Margaret Jones, has been present at all times to source any and all reagents I have required, for which I am most grateful. Additionally, I would like to thank the members of the Siebold group, whose willingness to listen to me

complain and abide my relentless cynicism has made the difficult times more bearable.

My STRUBI friends and colleagues Michael Grange, Eleanor Healey, Max Renner and Rhys Pryce have always been on hand to provide a sympathetic ear or witty retort in the pursuit of academic clarity, for which I am indebted to them. My time in Oxford has been made particularly enjoyable due to a number of friends I have made in the Magdalen MCR, with too many to list individually. Special thanks to Timothy Flight, Alexander Hess and James Eaton for all the good times. Alongside James, Kirk, Cliff, Lars, Jason and Rob, who have helped me retain my sanity.

I would also like to thank my loving counterpart Shauna for being there for me for the last 8 years, no matter how disagreeable I have become. Additionally, I would like to thank my poor mother for reminding me on a near-daily basis about how happy I should be that she gave birth to me.

This thesis is for the two Eddies in my life, Jones and Howe, who have brought me more joy in the last four years than I had thought possible.

Table of contents

Chapter 1: Introduction 2
1.1 Cell-to-cell communication and multicellular development 2
1.2 Morphogens: The French Flag model 3
1.3 The extracellular matrix 5
1.3.1 The proteoglycan layer and glycosaminoglycan biosynthesis 6
1.3.2 GAGs in gradient formation and growth factor retention 13
1.4 The Hedgehog signal transduction pathway 15
1.4.2 Hedgehog secretion 16
1.4.3 Signal transduction and downstream events 20
1.4.4 Extracellular Hedgehog signalling regulation 23
1.4.5 Small molecules and their function in Hedgehog signal transduction 27
1.5 The Wnt signal transduction pathway 29
1.5.1 Introduction 29
1.5.2 Wnt recognition and signalling output 30
1.5.3 Receptors and modulators 33
1.5.4 Palmitate versus palmitoleate 39
1.6 Cysteine-Rich Domains 41
1.6.1 The CRD family 41
1.6.2 Pockets and grooves 43
1.6.3 Potential roles in the evolution of morphogen signalling: Links to lipid transport and metabolism 49
1.6.4 CRDs and therapeutic potential 51
1.7 Global perspectives and aims 52
Chapter 2: Materials and methods 57
2.1 Molecular Biology 57
2.1.1 cDNA and expression vectors 57
2.1.2 Construct design and PCR 60
2.1.3 Restriction digestion, ligation and transformation 61
2.1.4 Site-directed mutagenesis 62

2.2 Cell culture 63
2.2.1 For small scale expression testing 63
2.2.2 For large scale protein expression 63
2.3 Protein overexpression and purification 64
2.3.1 Western blot analysis 64
2.3.2 Large-scale transfection of HEK293T cells 65
2.3.3 Purification of hexahistidine-tagged proteins 66
2.4 Protein crystallisation 67
2.5 Crystal structure solution 67
2.5.1 Data collection and processing 68
2.5.2 Solution of the phase problem: Experimental phasing 69
2.5.3 HHIP-N Sulphur Single wavelength Anomalous Dispersion (SSAD) 70
2.5.4 ROR2 CRD-Kr Platinum-SAD 71
2.5.5 HHIP-C 71
2.6 Structure solution and refinement 72
2.6.1 HHIP-N:SOS 72
2.6.2 ROR2 CRD-Kr 72
2.6.3 Molecular replacement for HHIP-N-apo and HHIP-C 73
2.6.4 Structure refinement, validation and analysis 73
2.7 Surface Plasmon Resonance 75
2.7.1 Protein-GAG interactions 75
2.7.2 Protein-protein interactions 77
2.8 SEC-MALS 77
2.9 Luciferase reporter assay 78
2.10 Immunofluorescence microscopy 79
2.11 SAXS 80
2.11.1 SAXS background 80
2.11.2 Data collection and processing 80
2.12 Analytical Ultracentrifugation 81
2.13 Heparin-agarose chromatography 83
Chapter 3: The structure of the HHIP N-terminal CRD 85

3.1 Summary 85
3.2 Introduction 86
3.3 Results 88
3.3.1 Purification and crystallisation of HHIP-N 88
3.3.2 HHIP-N data collection 90
3.3.3 Structure solution 94
3.3.4 Structure of HHIP-N 97
3.3.5 Structure analysis 100
3.3.6 Modelling 105
3.3.7 HHIP-like proteins and zHHIP-N 108
3.4 Discussion 114
3.4.1 The HHIP-N CRD as a small molecule-binding motif 115
3.4.2 HHIP-N as a protein-protein interaction motif 119
3.4.3 Outlook for native SAD phasing techniques 121
3.4.4 Further work 127
3.4.5 Concluding remarks 129
Chapter 4: HHIP-GAG interactions 134
4.1 Summary 134
4.2 Introduction 135
4.3 Results 137
4.3.1 HHIP is a secreted molecule regulated by cell surface GAG-interactions with the N-terminal CRD 137
4.3.2 Purification and crystallisation of HHIP constructs 141
4.3.3 Characterisation of HHIP:GAG interactions 144
4.3.4 HHIP-C:SOS complex crystal structure 147
4.3.5 HHIP-C:heparin complex crystal structure 150
4.3.6 GAG-induced changes to HHIP-C 154
4.3.7 Crystalline oligomeric states of HHIP-C 157
4.3.8 Biophysical analysis of HHIP oligomerisation 159
4.3.9 Modelling of the HHIP C-terminal helix (Hx) 164
4.4 Discussion 168

4.4.1 Modes of HHIP oligomerisation 169
4.4.2 Consequences of GAG binding 173
4.4.3 Further work 180
4.4.4 Concluding remarks 182
Chapter 5: Structural insights into the ROR2 extracellular domains 185
5.1 Summary 185
5.2 Introduction 186
5.3 Results 190
5.3.1 Purification and crystallisation of ROR2 190
5.3.2 ROR2 CRD-Kr data collection 192
5.3.3 Structure solution and refinement of ROR2 CRD-Kr 194
5.3.4 Structure of the ROR2 CRD and Kr domains 199
5.3.5 Structural analysis 205
5.3.5 Biophysical characterisation of ROR2 constructs 210
5.3.6 Binding experiments with ROR2 213
5.4 Discussion 216
5.4.1 Functional insights from the ROR2 CRD structure 217
5.4.2 The roles of the ROR2 Ig and Kringle domains 223
5.4.3 Further work 226
5.4.4 Concluding remarks 228
Chapter 6: Summary and concluding remarks 232
6.1 HHIP and GAG binding 233
6.2 ROR2 234
6.3 The CRD in ligand binding 237
6.4 Concluding remarks 240
Bibliography 243
Supplementary Information 258
Abbreviations (not expanded upon the text) 258
Supplemetary Materials 259
Supplementary Bibliography (for the construction of Fig. 3.12B) 264

Chapter 1

Introduction

Chapter 1: Introduction

1.1 Cell-to-cell communication and multicellular development

At no point, at least during non-disease states of their existence, are cells within a multicellular organism totally isolated. In order to coordinate appropriate developmental and homeostatic responses, the cell must precisely integrate signals from the surrounding environment. This is communicated into downstream effects upon intracellular compartments and is achieved by the presence of a multitude of membrane-embedded receptors (Tyson and Novak, 2014). Examples of these are far-reaching and encompass a large number of families; of 6718 identified membrane proteins in the human genome, receptors constitute the largest single functional group (1352 members in 63 groups), followed by transporters (817 members, 89 groups) and enzymes (533 members, 7 groups) (Almén et al., 2009). Beyond the complexity granted via the vast number of receptors, these proteins must be temporally and spatially regulated such that they are expressed appropriately in order to respond to an external stimulus, spanning from a protein ligand (for instance, a growth factor binding to a receptor kinase) to a small molecule (such as adrenaline binding to its cognate G protein-coupled receptor) (Lemmon and Schlessinger, 2010; Ring et al., 2013). This is vital from the view of developmental biology, regulating proliferation, differentiation, adhesion, cell fate and cell motility. A major example of this is the establishment of a nervous system within an organism.

At the point of neurulation, a handful of embryonic cells are committed to neural fates, and these ultimately form the entire nervous system (Copp and Greene, 2009). Individual axons must differentiate and innervate every reach of the body plan, a process reliant on responding to an array of growth cues and undergo distinct morphological cellular changes (Yam and Charron, 2013). An adult nervous system requires constant remodelling and renewal, facilitated by differentiation from a pool of neural stem cells (Gage and Temple, 2013). Such processes are commonplace throughout the body, and their dysregulation can lead to tumourigenesis.

1.2 Morphogens: The French Flag model

In order for development to proceed, each cell within an organism must be imparted with positional identity by signals from the surrounding environment in both a spatially and temporally controlled manner. This task is fulfilled by secreted proteins called 'morphogens,' three well-known members of which include the Hedgehog (Hh), Bone Morphogenetic Protein (BMP) and Wnt families (Perrimon et al., 2012). A morphogen is defined as a gradient-forming molecule capable of organising cells into patterns, produced from a localised source. This is best explained via the 'French Flag' model (Fig. 1.1A), which illustrates that defined thresholds of morphogen concentration are able to instruct cells to adopt particular fates. This concentration is communicated as transcriptional output, enabling cells to

respond to these cues and subsequently influence others within their vicinity (Ashe, 2006; Wolpert, 1996).

Figure 1.1B shows an example of concentration-dependent cell-fate specification, as well as the interplay between different morphogens to appropriately pattern a developing organism. This diagram displays the development of the vertebrate neural tube, and cellular populations within this tissue form the framework for the central nervous system. The ventral notochord (NC) secretes Sonic Hedgehog (Shh) which triggers the differentiation of the floor plate (FP), a structure which consequently also produces Shh (Echelard et al., 1993). Diffusion of Shh forms a morphogen gradient, creating (from high concentration to low concentration): V3 interneurons, motor neurons (MN), and V2, V1 and V0 interneurons (Briscoe and Small, 2015). A high concentration of BMP and Wnt originates from the dorsal end of the embryo to enable the differentiation of the dorsal interneurons (di1-6) (Zechner et al., 2007). This highlights the concept of downstream pathway interaction in the specification of distinct cell types as the result of the interplay between morphogens, alongside amplitude and duration of signal (Briscoe and Small, 2015). Beyond providing positional identification, it is also worth noting that morphogens are capable of behaving as cell survival cues; anti-Shh antibodies trigger widespread apoptosis within the developing neural crest, a process facilitated by its receptor, Patched (Ptc). Via this method, cells which are temporally required to aid in the patterning of a particular tissue but later become redundant can be removed (Mehlen

et al., 2005). These examples display the tightly regulated and highly integrated processes through which different signalling systems control embryonic development, and highlights their implications in disease states.

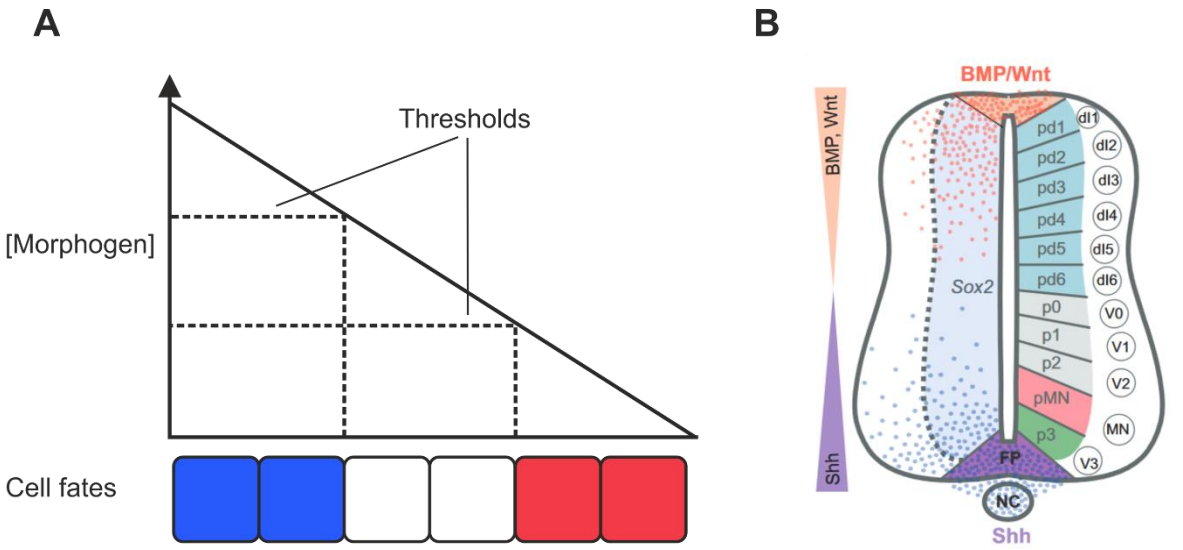


Figure 1.1 – Morphogen signalling and the French Flag model. A) The French Flag model of morphogen signalling. As the concentration of secreted morphogen decreases further from the source, differential signalling thresholds are achieved. This leads to the patterning of cellular populations via cell fate being imparted as a function of signal amplitude (adapted from Wolpert, 1996); B) Schematic of a developing neurula stage embryo. Defined morphogen threshold concentrations elicit differentiation of defined cellular phenotypes, as detailed in the text (From Briscoe and Small, 2015).

1.3 The extracellular matrix

The extracellular matrix (ECM) is a three-dimensional extracellular scaffold containing a host of fibrous proteins (collagens, elastins, fibronectins, laminins), membrane-bound receptors (integrins, DDRs), proteoglycans/glycosaminoglycans and soluble glycoproteins (proteases, glycosidases, vitronectin, growth factors) vital

for the maintenance of surrounding cellular networks (Theocharis et al., 2016). On top of structural and physical support, these dynamic structures also possess roles in cell growth, migration, differentiation, survival and general homeostasis, exhibiting diverse structures throughout tissues and organs. The ECM is organised into two general layers - the 'pericellular' and 'interstitial' matrices, the first of which is in close contact with cells and the second of which forms layers surrounding the cells (Frantz et al., 2010; Theocharis et al., 2016).

1.3.1 The proteoglycan layer and glycosaminoglycan biosynthesis

Proteoglycans (PGs), typically modified by addition of heparan sulphate (HSPGs) or chondroitin sulphate (CSPGs), form layers within the ECM both at the cell surface and are also secreted into the extracellular milieu (Sarrazin et al., 2011). Their makeup comprises diverse protein cores attached to sugar chains, and they possess a general role in the retention of secreted proteins and also function as co-receptors and modulators of signalling pathways (Kim et al., 2011; Lin and Perrimon, 1999). The three loosely overlapping functional regions occupied by PGs are the cell surface, basement membrane and 'extracellular' regions, which appear to possess a 'PG code' element in relation to their resident PGs (Fig. 1.2) (Iozzo and Schaefer, 2015). The major cell surface PGs in vertebrates include GPI-anchored Glypicans and transmembrane-anchored Syndecans (Gould et al., 1992; Watanabe et al., 1995), but other, more specific members are found in the form of Betaglycan (high-affinity TGFβ receptor) (Wang et al., 1991) and Phosphacan (interacts with N-CAM) (Maurel

et al., 1994). The pericellular region is in close contact with the cells and components of the basement membranes (Perlecan, Agin, Collagens) and also to the 'extracellular' PGs (Hyalectans and Versicans), which are associated with non-sulphated and non-protein-coupled hyaluronan sugar chains (Hedman et al., 1979; Iozzo and Schaefer, 2015; Wilusz et al., 2014). Whilst HSPGs can be found at the cell surface and in the pericellular regions, the 'extracellular' PGs favour modification with CS and dermatan/keratin sulphate chains over HS chains (Filmus et al., 1995; Jones et al., 2003; Watanabe et al., 1998).

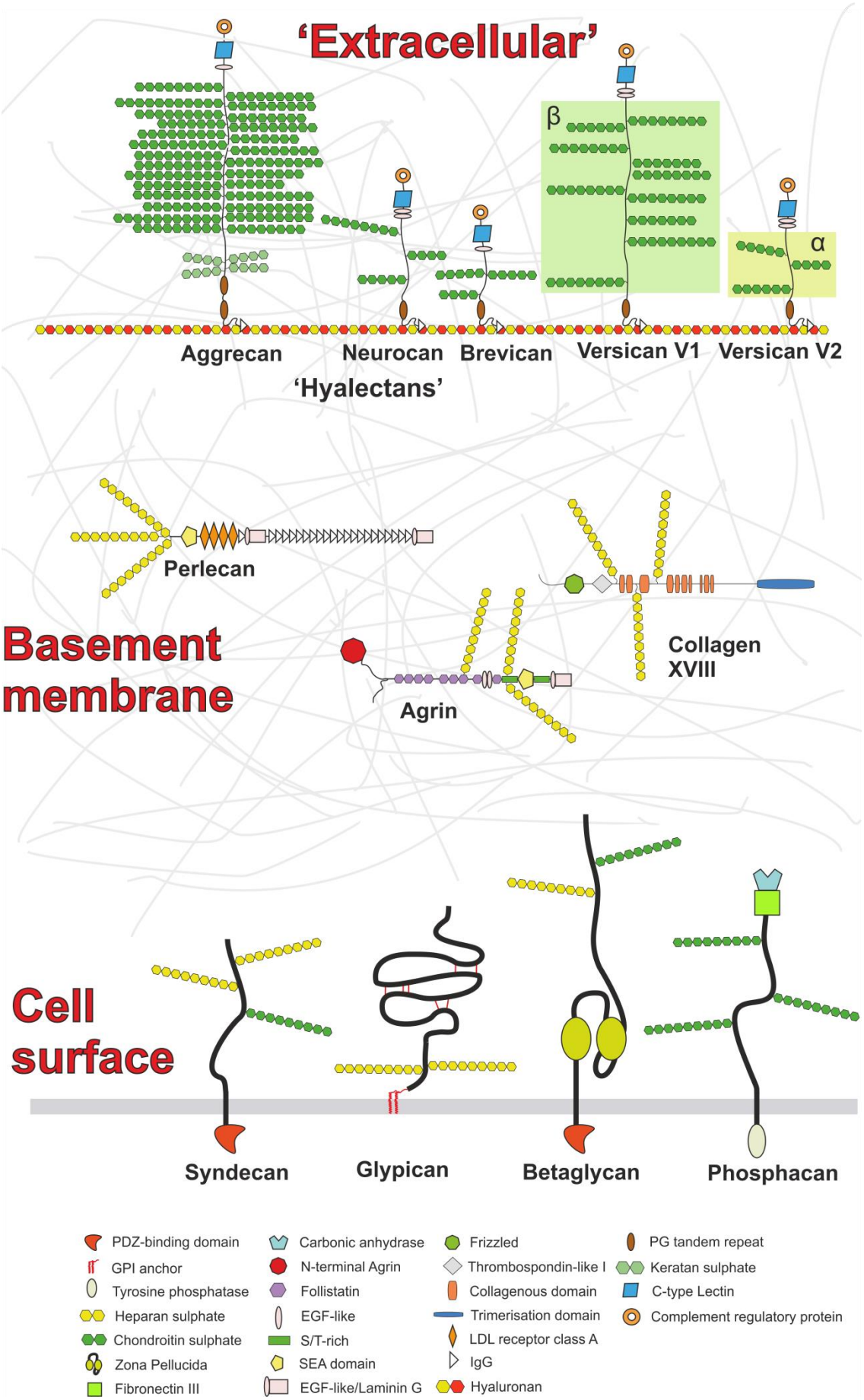

'Extracellular'
β
α
Aggrecan
Neurocan
Brevican
Versican V1
Versican V2
'Hyalectans'
Perlecan
Basement membrane
Agrin
Collagen XVIII
Cell surface
Syndecan
Glypican
Betaglycan
Phosphacan
PDZ-binding domain
GPI anchor
Tyrosine phosphatase
Heparan sulphate
Chondroitin sulphate
Zona Pellucida
Fibronectin III
Carbonic anhydrase
N-terminal Agrin
Follistatin
EGF-like
S/T-rich
SEA domain
EGF-like/Laminin G
Frizzled
Thrombospondin-like I
Collagenous domain
Trimerisation domain
LDL receptor class A
IgG
Hyaluronan
PG tandem repeat
Keratan sulphate
C-type Lectin
Complement regulatory protein

Figure 1.2 – General schematic of the extracellular matrix (above). The ECM is organised into three overlapping extracellular domains, each with specific proteoglycans and functions. Cell surface PGs are membrane-bound by a TM helix or GPI anchor, but enzymatic cleavage can trigger their secretion in order to elicit non-cell autonomous functions. Extracellular matrices comprise PGs within a fibrous network of collagens, elastin and laminins (amongst others; shown as a grey mesh), alongside additional other soluble components. Basement membranes comprise distinct HSPGs, including the collagens, the most abundant of all human proteins. Further from the cell, anchored to hyaluronan sugar chains, are the 'truly extracellular' PGs. These act to give mechanical strength to tissues (e.g. cartilage) as well as binding various proteins. A key of the various domains shown is displayed (adapted from Iozzo and Schaefer, 2015).

The motif for the initiation of HS or CS biosynthesis is a serine flanked by a glycine residue, found within a region rich in both acidic and hydrophobic amino acids. Within the late-ER/*cis*-Golgi, a 'linkage' region is sequentially added to this serine, comprised of the sequence xylose-galactose-galactose-glucoronic acid, to label the protein for glycosaminoglycan (GAG) attachment. HS chain initiation is then performed by the EXTL family of glysosyltransferases, which add an N-acetyl-glucosamine (GlcNAc) moiety followed by a glucoronic acid (GlcA) residue to form the backbone of the polymeric chain. This polymerisation event is performed almost entirely by the enzymes EXT1 and EXT2 in vertebrates. Following this, the HS chain is modified further to include sulphated ('NS') and acetylated ('NA') domains. To form NS domains, GlcNAc residues must first be N-deacetylated and N-sulphated by NDST enzymes at certain regions along the chain to form GlcNS residues. What follows is an epimerisation event; a C5-epimerase converts GlcA residues exclusively next to GlcNS residues iduronic acid (IdoA) via isomerisation of the carboxyl group

on the C5 position of the GlcA. This is met with concomitant 2-O-sulphation of the resultant IdoA moiety by a single 2-O-sulphotranferase to form 2S-IdoA. NS domain sulphation is facilitated further by the modification of GlcNS by a 3-O-sulphotransferase and a 6-O-sulphotransferase (Fig. 1.3) (Carlsson and Kjellén, 2012; Kreuger and Kjellen, 2012).

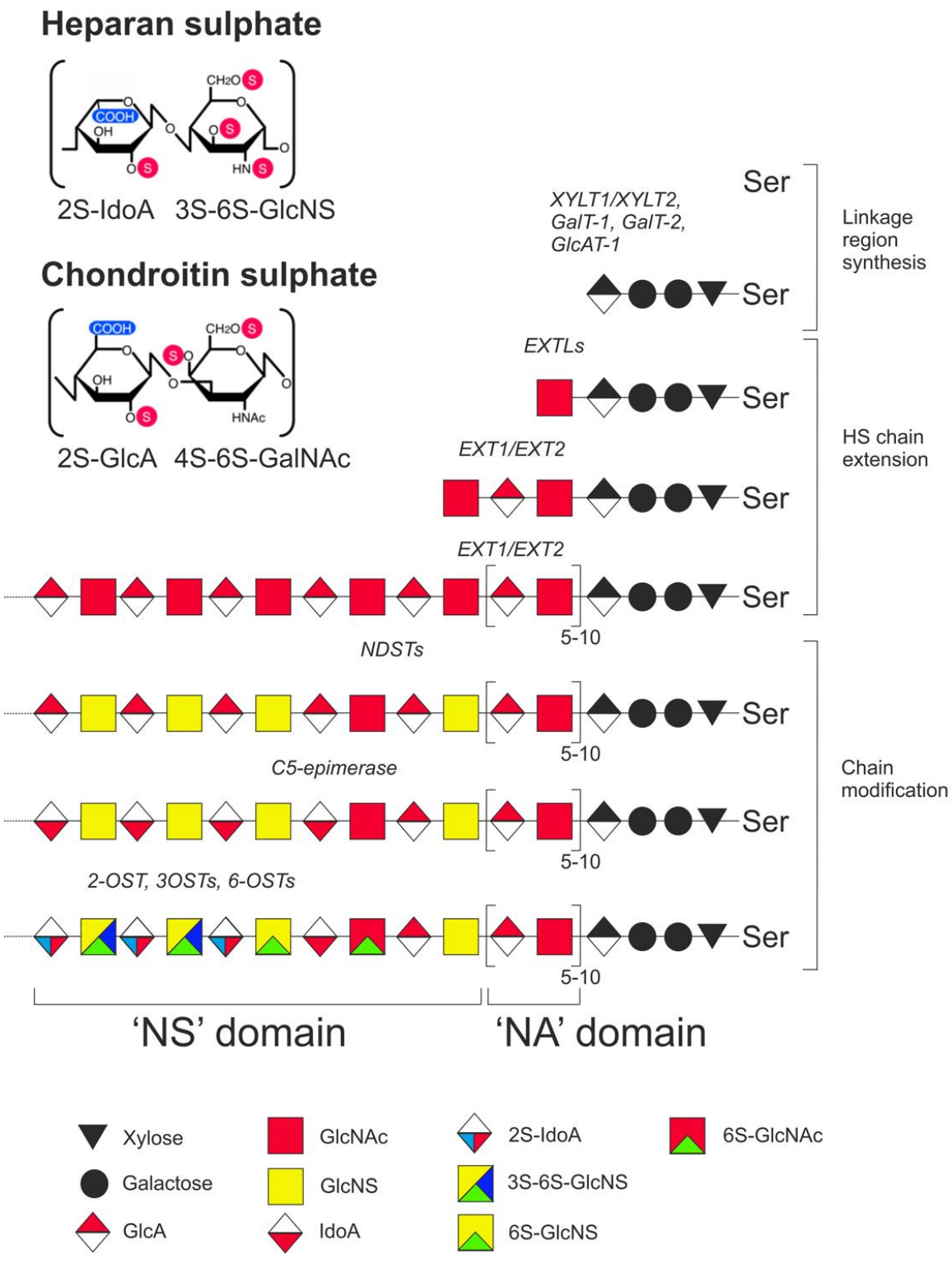


Figure 1.3 – GAG synthesis. A general schematic of the steps undertaken in the modification of a PG core with an HS chain is shown (from top to bottom), with the enzymes in question italicised. A key of the sugars added or created is indicated (beneath). The general structures of the sugars employed in either HS or CS chains is displayed in the top left-hand corner (adapted from Kreuger and Kjellen, 2012; Mikami and Kitagawa, 2013).

CS chains differ to HS chains in that GlcNAc is replaced by N-acetyl-galactosamine (GalNAc) (Fig. 1.3). It is unclear how a particular GAG type is selected during HSPG/CSPG synthesis, but it does appear that 4-O-sulphation of the second Gal residue in the linkage region causes addition of GalNAc instead of GlcNAc and subsequent CS synthesis (Ueno et al., 2001). It is worth noting that once polymerisation has begun, the chain selected is the one polymerised, and mixed HS/CS chains are not observed. Like HS, CS can contain regions of GalNAc-GlcA repeats or undergo C5-epimerisation of GlcA to IdoA; this isomerisation and inclusion of GalNAc-IdoA regions results in the formation of related sugar polymer DS. IdoA imparts greater conformational flexibility upon GAG chains, enhancing the range of possible interactions (Thelin et al., 2013). Therefore, mixed CS/DS chains exist in nature, and do not contain the differential NA/NS domains as seen in HS (Mikami and Kitagawa, 2013).

The GAG polymers found in HSPGs and CSPGs represent highly variable structures, the formation of which is subject to several tightly regulated processing steps. As well as the above, this is expanded upon by the presence of extracellular enzymes such as SULF1 and SULF2 which are capable of removing 6-O-sulphate groups from 3S-6S-GlcNS residues (Kalus et al., 2015). Additionally, the secreted enzyme Heparanase can cleave HS chains from PG cores to release diffusible and biologically active GAG chains (Vlodavsky et al., 2013). It is also of note that the first xylose in the linkage region is phosphorylated and this functions as a switch to

enable GAG elongation; inhibition of phosphorylation prevents chain formation (Wen et al., 2014). Further modes of regulation are poorly understood, but can occur at the levels of both transcription and post-translational modification. Overexpression of EXT1 increases expression and N-linked glycosylation of NDST-1, resulting in increased HS synthesis. Overexpression of EXT2, however, reverses this effect (Presto et al., 2008). This supports a mode of regulation within which HS biosynthesis enzymes form complexes with one another to further regulate the modification process. Additionally, proteins such as matrix metalloproteases are capable of shedding whole PG molecules from the cell membrane to enable regulation over a larger distance (Manon-Jensen et al., 2013).

1.3.2 GAGs in gradient formation and growth factor retention

Once synthesised and secreted, PGs display a number of roles in development, homeostasis and disease. Their most simplistic function is as a scaffold for the retention of secreted growth factors and the formation of gradients in morphogen signalling, and can induce different signalling outputs depending on the sugar composition. RPTPσ, a receptor involved in axon guidance and synaptogenesis, is oligomerised by HS. This organises receptors into tight puncta at the cell surface, enabling concentrated domains of signalling and neurite extension in regions depleted of RPTPσ. Conversely, CS competes with HS for RPTPσ binding but does not facilitate oligomerisation, causing dispersal of RPTPσ across the growth cone surface and widespread inhibition of outgrowth (Coles et al., 2011).

Crystal structures of Shh in complex with both HS and CS were solved, which regulate the oligomeric state of the HhN ligand (Whalen et al., 2013). The binding site in question overlaps with that of several Hh receptors, and functions as a hotspot for interactions; mutations in this site have been shown to affect hedgehog signalling and are associated with several developmental diseases. Both HS and CS chains are thus important for the retention of morphogens in various oligomeric states and the resultant signalling output (Whalen et al., 2013). Genetic knockout of *Drosophila* EXT homologue *tout-velu* (*ttv*), a HS polymerase, leads to a loss of Hh distribution within the developing embryo and supports this data from vertebrate Shh (The et al., 1999). The roles of GAGs within hedgehog signalling are far-reaching, but another example concerns the influence of the Hh co-receptor Ihog in *Drosophila*. Ihog synergises with Hh receptor Ptc in order to facilitate signalling, and the formation of this complex is dependent on a HS chain which acts as a 'staple' to bridge Hh and Ihog (McLellan et al., 2006). On top of this, the FGF ligand requires a GAG chain in order to bind to its receptor and activate signalling (Pellegrini et al., 2000; Schlessinger et al., 2000).

Beyond this ability to influence ligand:receptor interactions in Hh and FGF signalling, GAGs have also been implicated in the modulation of BMP signalling. The *Drosophila* HSPG Dally is able to enhance BMP signalling, presumably via its retention and accompanying increase in local concentration (Dejima et al., 2011). The addition of heparin blocks the interaction between BMP2 and PRDC, a negative regulator of BMP signalling (Nolan et al., 2013). These studies all highlight the

influence of protein:GAG interactions in terms of control of different levels of morphogen signalling and subsequent implications for cell differentiation, growth, homeostasis and cancer.

1.4 The Hedgehog signal transduction pathway

1.4.1 Introduction

The Hh family of morphogens represents an evolutionarily conserved family of proteins vital for proper metazoan development, controlling cell fate, growth and survival. Whilst *Drosophila* possesses a single Hh gene, mammalian Hh orthologues have undergone gene duplication to include three members - Sonic hedgehog (Shh), Indian hedgehog (Ihh) and Desert hedgehog (Dhh) (Varjosalo and Taipale, 2008). The zebrafish Hh family comprises five members, including additional Shh and Ihh genes (Avaron et al., 2007). Besides their roles listed in section 1.2 regarding patterning of the nervous system, Hh ligands have widespread and crucial functions in the development of a number of organs and tissues (Robbins et al., 2012; Varjosalo and Taipale, 2008). Development is mostly dependent on the activities of Shh, the most broadly expressed Hh ligand, whilst Ihh and Dhh appear to be restricted to more tissue-specific developmental patterning events. To add to this, Hh signalling is necessary for the maintenance of various adult stem cell niches and thus regeneration of adult tissue, as well as being involved in the progression of multiple cancers (Petrova and Joyner, 2014). Therefore, the Hh signal is capable of acting as

morphogen, mitogen, and survival factor given the cellular context, and its signalling pathways represent attractive therapeutic targets for translational research.

1.4.2 Hedgehog secretion

The Hh morphogen is synthesised as a 45 kDa pre-pro-protein, containing a signal peptide, N-terminal metalloprotease-like signalling domain (HhN) and C-terminal autoprocessing intein domain. The signal peptide is removed upon entry to the ER, where the C-terminal autoprocessing event also occurs (Varjosalo and Taipale, 2008). During this process, the C-terminal domain undergoes autoproteolysis with concomitant esterification of a conserved catalytic cysteine with the 3β-hydroxyl group of a cholesterol molecule. The free autoprocessing domain is subsequently subjected to proteasome-mediated degradation (Chen et al., 2011). The 17 kDa HhN-cholesterol fragment is then transferred to the Golgi apparatus where Skinny Hedgehog, a member of the membrane-bound O-acyltransferase (MBOAT) family, catalyses the addition of an N-terminal palmitoyl group (Chamoun et al., 2001; Pepinsky et al., 1998). These two modifications are unique; Hh is, to date, the only known protein that undergoes covalent modification by cholesterol. Additionally, whilst other palmitoylated proteins have been characterised (Wnt, Spitz), these exist in the form of an O-linked ester (serine) or S-linked thioester (cysteine) (Nile and Hannoush, 2016). The palmitoyl adduct, however, is coupled through an amide linkage to the N-terminus of HhN (Resh, 2013). Cholesterol modification is not a prerequisite for palmitoyl modification, and forms of HhN can

be produced with either individual modification, both modifications or in an unmodified form.

Processed HhN subsequently remains self- or membrane-associated once modified. There is evidence for a variety of oligomeric forms existing within the context of Hh trafficking (Eaton, 2008). The secretion of HhN depends on the activities of two proteins - SCUBE2 and Dispatched (Disp). Disp, a member of the RND family of transporter proteins (to which HhN receptor Ptc also belongs), recognises distinct structural elements of the cholesterol modification to SCUBE2, and both proteins are key to HhN secretion and extracellular transport (Tukachinsky et al., 2012). HhN can also be assembled into much larger oligomers, both alone in HhN multimers, in lipoprotein particles alongside shed HSPGs or as 'exovesicles' with the lipid components embedded within shed membranes (Sanders et al., 2013; Tanaka et al., 2005). This is thought to be necessary to shield the highly hydrophobic lipid adducts within micellar-type structures, as well as increase local HhN concentrations to modulate signalling potency through avidity effects. A schematic for HhN processing and secretion is displayed in Fig. 1.4 (Eaton, 2008).

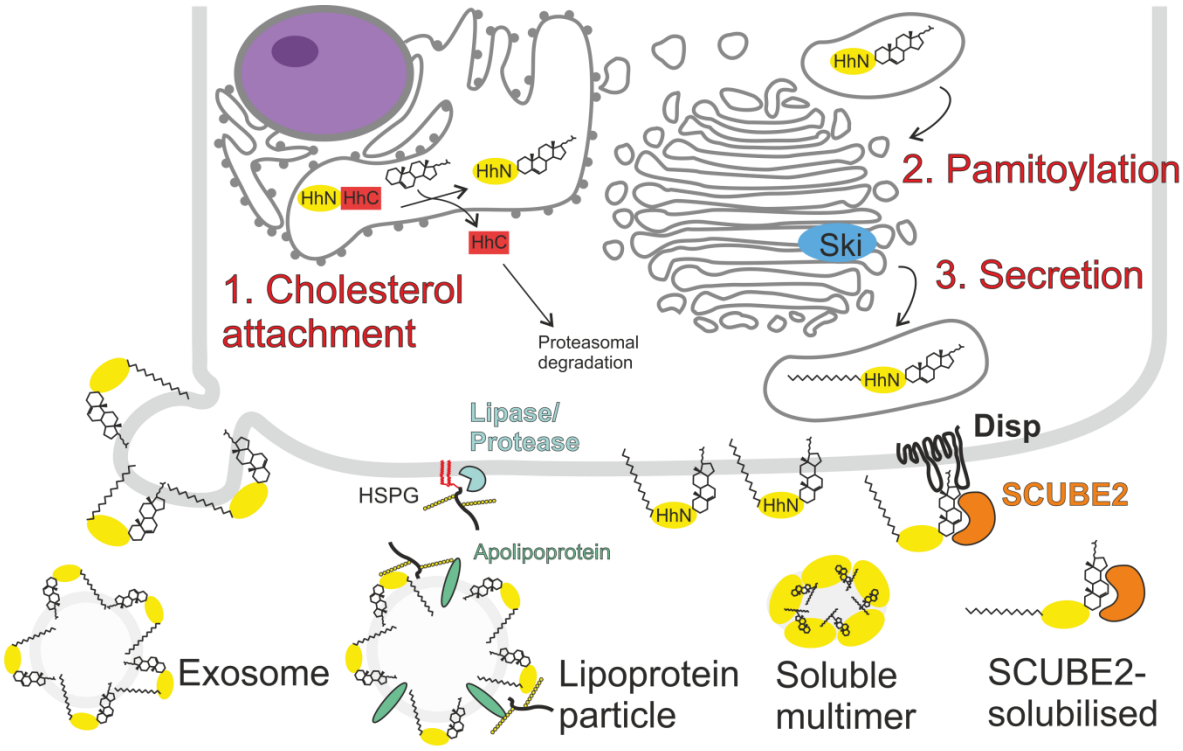


Figure 1.4 – Hh processing and secretion. Full-length Hh undergoes autoproteolysis with concomitant modification of the C-terminal cysteine residue with cholesterol (1). This yields cholesterol-modified HhN, while the proteolysed fragment (HhC) is degraded in the proteasome. Skinny Hedgehog (Ski) then catalyses the attachment of a palmitoyl moiety to the N-terminus of HhN (2). Lipid-modified HhN is then secreted in a Dispatched (Disp)-mediated process (3). This entails a hand-off mechanism involving soluble extracellular SCUBE2. HhN has also been found associated with cellular membranes and in other secreted assemblies, comprising HhN multimers, exosomes shed from the cell surface and complexes with lipoproteins and solubilised HSPGs.

What are the individual functions of these lipid adducts? A study of HhN modified with palmitate only concluded that the hydrophobicity of the N-terminus correlates with signal activation. In this study, chain lengths up to that of C16 (palmitoyl) gave the greatest signalling output, whilst mutation of the N-terminal cysteine to two isoleucine residues also resulted in an 8-fold increase in potency when compared with unmodified HhN (Taylor et al., 2001). A palmitoylated peptide

corresponding to the N-terminal 22 residues of Shh is sufficient to bind to Ptc and activate signalling (Tukachinsky et al., 2016). This links the N-terminal palmitoyl modification to a role in Ptc receptor binding. The role of the cholesterol modification is more complex and less well understood; dually lipidated HhN was seen to accumulate at the cell surface as well as form higher-order molecular weight complexes, whilst palmitoyl-modified HhN was not found in either of these populations (Gallet et al., 2003). This role of the cholesterol moiety in long-range Hh signalling is reinforced by the fact that in *ttv* knockouts in *Drosophila*, HhN can only be detected associated with cells adjacent to its source. Palmitoyl-HhN, however, is detected throughout *ttv* mutant clones. It is therefore likely that the cholesterol adduct is involved in multimerisation and subsequent extracellular transport via HSPGs (Fig. 1.4) (Callejo et al., 2006). It is also worth noting that various saturated and unsaturated fatty acids can be detected in the place of palmitate at the N-terminus, and the abundance of different populations of these can be altered via external serum concentrations and the level of Hh expression (Long et al., 2015). In addition, a reduction in the hydrophobicity of the N-terminal adduct reduces the association of HhN with lipid rafts by modulating both membrane association and multimer formation (Long et al., 2015). Whether multimers form spontaneously or from delivery by SCUBE2, as well as how this truly modulates Hh signalling activation, requires further study. HhN modifications are therefore linked to roles in multimerisation, localisation and signalling potency.

1.4.3 Signal transduction and downstream events

Hh signal transduction is unique in the sense that it does not follow the typical trend of a ligand causing a conformational change in an inactive receptor to activate downstream signalling (Lemmon and Schlessinger, 2010). In the absence of HhN, Ptc represses the Hh signal transducer Smoothened (Smo) via a catalytic mechanism; this is thought to most likely be due to the pumping of some small molecule Smo inhibitor into the ECM (Taipale et al., 2002). Like Disp, Ptc also exhibits homology to bacterial RND transporters, which strengthens the hypothesis that Ptc effects the efflux or influx of some modulatory molecule acting upon Smo (Kuwabara and Labouesse, 2002). This is further supported by the discovery of cyclopamine, a plant-derived oxysterol shown to bind to Smo and cause holoprocensephaly (HPE), a cephalic disorder attributed to an absence of Hh signalling (Chen et al., 2002).

Repression of Smo by Ptc leads to phosphorylation of the transcription factors GLI2 and GLI3 by PKA, CKI and GSK3β, which results in their proteasomal degradation into transcriptional repressor forms GLI2R and GLI3R (Fig. 1.5A) (Jia et al., 2005). Activation of the Hh signalling pathway is initiated by HhN binding to Ptc. Consequently, repression of Smo is released and full-length GLI proteins translocate to the nucleus and activate expression of specific genes (Fig. 1.5B). It is the balance of GLI in repressor and activator forms that dictates the differential expression of Hh target genes to pattern cell populations (Humke et al., 2010).

Hh signalling events in vertebrates occur at the primary cilium, a cellular membrane projection supported by an internal microtubule structure. It is of note that mutations in ciliary components lead to disease characteristics related to defective Hh signalling (Dorn et al., 2012). This highlights the primary cilium as a Hh 'signalling antenna,' the reason for which is currently unknown. In unstimulated cells, Ptc is concentrated within the cilium and around its base (Rohatgi et al., 2007). PKA, CK1 and GSK3β are all associated with the basal body, preventing the entry of GLI proteins into the cilium and facilitating their partial degradation (Fig. 1.5A) (Dorn et al., 2012). Upon stimulation by HhN, Smo is able to enter the primary cilium and is phosphorylated upon its cytoplasmic tail, leading to its translocation to the ciliary tip (Kim et al., 2009). This permits entry of SUFU:GLI2/3 complexes into the primary cilium via microtubule transport by Kinesin homologue Kif7 (Li et al., 2012; Liem et al., 2009). Ciliary entry triggers release of Gli proteins by Sufu, and Kif7 transports Gli back down the cilium and into the nucleus (Fig. 1.5B) (Humke et al., 2010; Tukachinsky et al., 2010). The molecular mechanisms governing this process remain poorly understood in both vertebrates and invertebrates. It is interesting to note that while the mechanism of Ptc-Smo interplay discussed above is conserved in *Drosophila* Hh signalling, the primary cilium is absent in flies. Despite this, Hh signalling still requires microtubule transport via kinesin-like protein COS2 (Endoh-Yamagami et al., 2009).

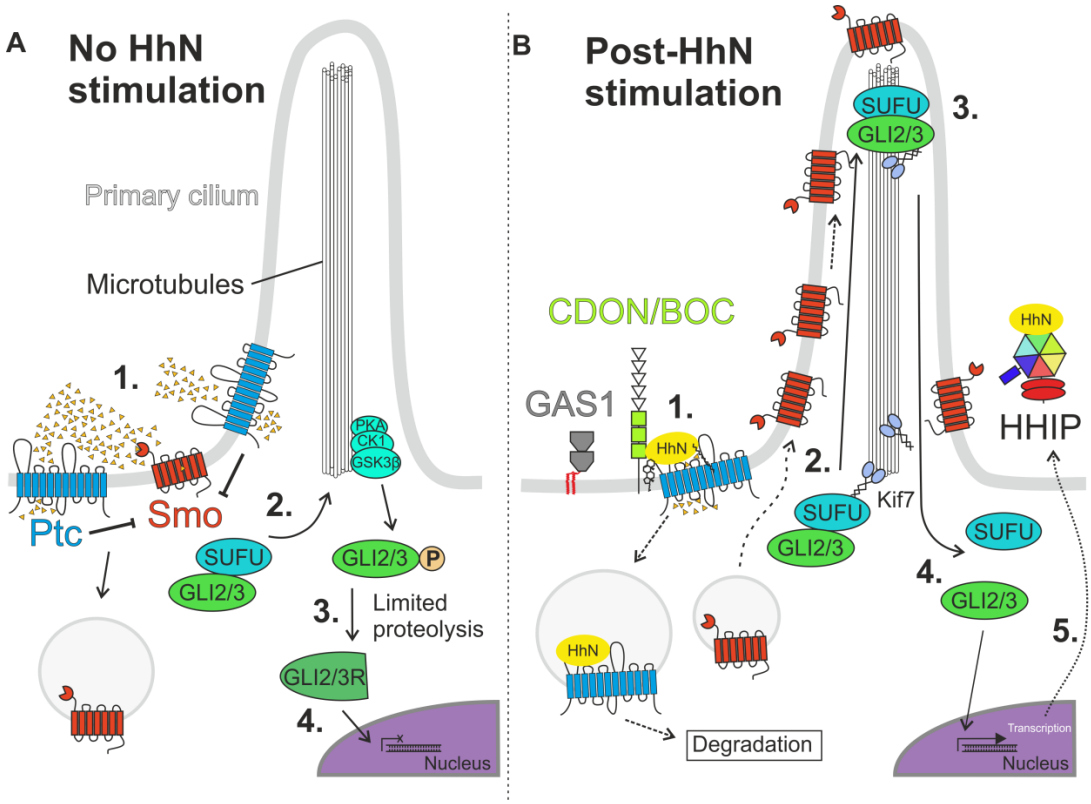


Figure 1.5 – Vertebrate Hh signal transduction. A) Unstimulated conditions. Ptc, via altering the localisation of an unknown small molecule (orange triangles), is able to repress the activity of Smo. Ptc is enriched in the base of the primary cilium to prevent ciliary entry of Smo. Additionally, Ptc retains Smo in intracellular vesicles (1). A complex of Sufu and Gli is subsequently phosphorylated by kinases in the ciliary base (2). Following phosphorylation and ubiquitination, Gli is degraded into a repressor form (GliR) (3). GliR is transported to the nucleus and blocks expression of Hh target genes (4). B) HhN-stimulated state. HhN binds to Ptc to release Smo repression. This is dependent on the presence of additional co-receptors CDON, BOC or GAS1 (1). Activated Smo is transported into and up the primary cilium, which enables Kif7-mediated transport of a Sufu/Gli complex (2). Ciliary transport of the Sufu/Gli complex causes its dissociation, and Kif7 transports Gli back into the cell (3). Full-length Gli enters the nucleus and activates the expression of Hh target genes (4). Hh-responsive genes include proteins involved in negative feedback of signalling, including pseudoreceptor molecule HHIP (5). Figure adapted from (Briscoe and Thérond, 2013).

1.4.4 Extracellular Hedgehog signalling regulation

Hh signalling is regulated at levels from the extracellular matrix to the nucleus, and these processes are of vital importance for the regulation of HhN distribution and function (Fig. 1.5). Three co-receptors are found in CDON, BOC and GAS1. CDON and BOC are homologues of one another, possessing extracellular Ig and Fn domains, whilst GAS1 has a GDNFR-type ectodomain. Knockout studies have shown that removal of one of these proteins has modest effects on Hh signalling, whilst loss of all three co-receptors leads to a complete loss of ventral neural progenitors which mirrors a Hh knockout phenotype (Allen et al., 2011). The involvement of the GAG chains of HSPGs in Hh signalling regulation have been discussed above; beyond this, HSPG glypican-3 (GPC3) acts as a negative Shh receptor in a GAG-independent fashion, and GPC3 knockouts lead to defects associated with overactive Hh signalling (fused ribs, polydactyly and cleft palate) (Capurro et al., 2008).

One downstream target of Hh signalling is the vertebrate-specific Hedgehog Interacting Protein (HHIP), and structural studies have been undertaken within this thesis to further characterise its function in relation to Hh signalling (see Chapters 3 and 4). HHIP is secreted as a negative feedback inhibitor and is able to sequester HhN, facilitating its degradation and a return to unstimulated, Smo-repressed conditions (Chuang and McMahon, 1999). A Zn^{2+} in the HHIP:HhN interface is essential for this high affinity interaction. Two Ca^{2+} ions are also bound at the surface

of HhN in complex with HHIP and, although not within the interface, these increase the affinity of the HHIP:HhN interaction from a K_D of 100 nM to a K_D of 6 nM. Ca^{2+}-binding causes ordering of acidic residues near the interface, and is thought to be necessary to prevent electrostatic repulsion between the two proteins (Bishop et al., 2009). Interestingly, a structural study on the CDON:HhN complex showed that a Ca^{2+} ion is essential for this interaction and the Ca^{2+}-binding site of ShhN is in the same conformation in the CDON:HhN interface as in the HHIP:HhN complex (McLellan et al., 2008). Hence extracellular metal ion concentration represents a further potential mechanism for the modulation of Hh signalling.

HhN complex structures have pointed to the existence of a 'hotspot' region on the surface for the interaction with its binding partners (Fig. 1.6). Despite the fact that HHIP and Ptc-blocking anti-HhN antibody 5E1 occupies more of the site surrounding the Zn^{2+} ion bound by HhN, there is still significant overlap with residues surrounding the Ca^{2+} site (Maun et al., 2010). Negative receptor GPC3 competes with Ptc for HhN binding (Capurro et al., 2008). CDON is capable of competing with Ptc for monomeric HhN, implying that Ptc and CDON bind to the same site on the HhN ligand despite cooperating to activate signal transduction (Allen et al., 2011). This could be explained *in vivo* via multimeric lipid-modified HhN bringing several morphogen molecules from a multivalent cluster in close contact to both the Ptc receptor and CDON co-receptor, suggesting an avidity-based mechanism for signal transduction (Eaton, 2008). Mutational data shows that

residues on the opposite face of HhN to the Zn^{2+} and Ca^{2+} sites are involved in GAS1 binding, implying that GAS1 has evolved to utilise an alternative face of the HhN ligand (Martinelli and Fan, 2009). However, HhN D88V, a mutation which causes HPE and prevents coordination of the Ca^{2+} ion, displays a loss of GAS1 binding (McLellan et al., 2006). Thus, despite the lack of a complex crystal structure, this hotspot region of HhN is still required for interaction with GAS1. CDON, BOC and GAS1 expression is downregulated as a consequence of Hh signalling, whilst that of HHIP is upregulated. This negative feedback loop system has been shown to be essential for cell patterning, and it is important to note that Hh signalling systems are dynamic and ligand input is modulated based on the receptors present (Allen et al., 2007). *Drosophila* CDON/BOC homologue IHOG also occupies a different face upon HhN (McLellan et al., 2006). Whether this truly represents a mode of regulation applicable to vertebrates or a facet of evolution requires further structural characterisation of HhN interaction partners.

The difference in HHIP as a negative regulator likely arises from the fact that it has an affinity greater than 100-fold higher for monomeric HhN than CDON (6 nM versus 1.3 μM), and would ultimately occupy many more binding sites upon a multivalent HhN particle. Ultimately, experiments have yet to test the consequence of multiple valence of HhN on ligand:receptor dynamics (Beachy et al., 2010; Bishop et al., 2009).

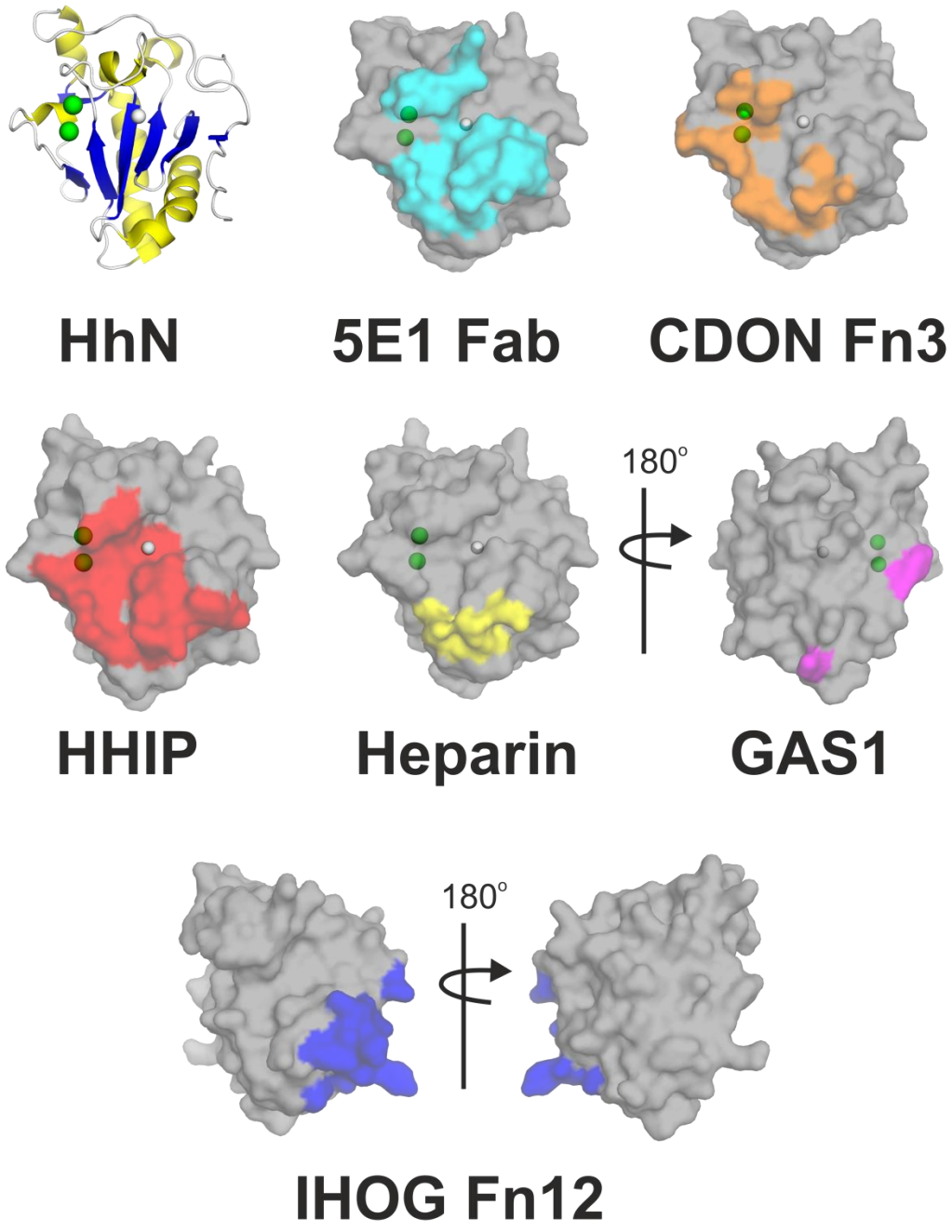


Figure 1.6 – A hotspot for HhN:receptor interactions. HhN shown in cartoon representation (top left-hand corner) and the surface residues utilised in interfaces with the indicated binding partners are coloured. The two Zn^{2+} ions are shown as green spheres, and the Ca^{2+} ion as a white sphere. The mutations probing the interaction of HhN with GAS1 were mapped onto the HhN surface from (Martinelli and Fan, 2009). Complex structures were analysed using PDBsum (Laskowski, 2001) (PDB accession codes: 3N1R, 3MXW, 3D1M, 2WFX, 4C4N, 2IBG; adapted from Beachy et al., 2010). (See figs. 1.15 and 4.1).

1.4.5 Small molecules and their function in Hedgehog signal transduction

Ever since the discovery of Ptc as a hypothesised pump with the ability to catalytically repress Smo, the discovery of the true cognate small molecule binding partner which controls Hh signalling has been of paramount importance. Dysregulation of the Hh pathway is responsible for 30% of human cancers, including basal cell carcinoma (BCC), medulloblastoma, melanoma and prostate, breast, ovarian, pancreatic, lung and cervical cancers (Gonnissen et al., 2015). Ergo, the responsiveness of Smo to such a molecule is of great interest for the treatment of such diseases.

Smo is a member of the Frizzled class of GPCR-containing proteins, which also includes the Frizzled proteins involved in Wnt recognition. Smo contains both an extracellular cysteine-rich domain (CRD) and 7 transmembrane helices (7TM), the two of which constitute distinct small molecule binding sites. Modulators of Hh signalling such as SANT-1 and the anti-tumour agent Vismodegib bind within the 7TM region, as do agonists of Smo activity such as SAG (Byrne et al., 2016; Wang et al., 2013, 2014). Oxysterols such as 20(S)-hydroxycholesterol (20(S)-OHC) are capable of activating Hh signalling, but do so through direct binding to the extracellular CRD of Smo (Byrne et al., 2016; Huang et al., 2016; Myers et al., 2013; Nachtergaele et al., 2013). Full pathway activation in response to Shh or SAG depends on the presence of the CRD, and this implies that there is some role of the CRD being capable of allosterically modulating the 7TM region in response to oxysterol-like molecules

(Nachtergaele et al., 2013). This finding was corroborated by the structure of the CRD and 7TM regions of Smo, which was observed to contain a natively co-purified cholesterol molecule that brings the CRD in contact with the 7TM domain (see section 1.6.2 and Fig 1.12). Site-directed mutagenesis of residues involved in this interaction leads to a loss of signal activation by ShhN, and cholesterol is able to directly activate Hh signalling (Byrne et al., 2016; Huang et al., 2016; Luchetti et al., 2016).

Cholesterol has been shown to play a role in the facilitation of Hh signalling, and blocking of cholesterol synthesis has been linked to a reduction of Hh signalling *in vivo*. This is unlikely to be due to a loss of the cholesterol modification upon HhN, as Hh pathway inhibition induced via cholesterol depletion cannot be rescued by the addition of oxysterols (Blassberg et al., 2016; Cooper et al., 2003). To add to this, the homology between Ptc and cholesterol transporter NPC1 suggests an endogenous sterol-like molecule exerting control over the pathway, as suggested by the above discussion of the roles of cholesterol, oxysterols and cyclopamine. However, NPC1 acts to transport cholesterol from an oxidising compartment into the cell membrane, whilst the current hypothesis of Ptc1 repression of Smo requires the movement of an antagonist molecule in the opposite direction (Gong et al., 2016). Therefore, as Hh signalling is activated by cholesterol, the true role of Ptc may be the removal of cholesterol from cellular membranes to block Hh signalling. A mechanism for this would require extensive further study.

In conclusion, small molecules exert stringent control over the Hh signalling pathway, and the Smo antagonist Vismodegib has been approved as an anti-cancer drug (Sekulic et al., 2012). Given its roles in developmental diseases and cancer progression alongside the emerging problems of drug resistance, Hh signalling represents an ever-present opportunity for therapeutic targeting.

1.5 The Wnt signal transduction pathway

1.5.1 Introduction

The Wnt family of secreted glycolipoproteins contains 19 members and elicits a diverse range of roles ranging from development to adult cell homeostasis and disease (MacDonald et al., 2009). As morphogens, Wnt ligands are involved in neural tube patterning, CNS regionalisation and skeletal, muscular and vascular development, and their dysregulation leads to a range of developmental disorders. Wnts have roles in adult homeostasis mechanisms via influencing stem cell niches in bone, muscle, blood and intestinal tissues, as well as in axon guidance, synaptogenesis and CNS injury repair. As a consequence, a number of proliferative cancers are associated with aberrant Wnt signalling, as well as cardiovascular disease, Alzheimer's disease, dementia and osteoporosis (Ciani and Salinas, 2005; Inestrosa and Varela-Nallar, 2014; Regard et al., 2012).

Wnts are conserved across all metazoans, comprising a cysteine-rich core domain of some 350-400 residues (MacDonald et al., 2009). Similar to HhN, Wnts undergo palmitoleate modification upon secretion, catalysed by ER-resident MBOAT Porcupine (Gao and Hannoush, 2014). The major difference is that whilst HhN modification favours a saturated C16:0 palmitate, the preferred modification for Wnt is an unsaturated C16:1 palmitoleate moiety. Unlike HhN, however, this modification is a prerequisite for secretion (alongside glycosylation), and involves the esterification of a serine side chain (Janda and Garcia, 2015). An additional parallel to HhN secretion is the existence of Golgi-localised transmembrane protein Wntless which, like Disp in Hh signalling, is necessary for secretion of lipidated Wnt (Bänziger et al., 2006). Wnts are solubilised in the ECM by forming higher-order oligomers and being incorporated into lipoprotein particles, as well through interactions with HSPGs. HSPGs are vital for creating Wnt gradients and facilitating proper signalling (Dhoot et al., 2001; Fuerer et al., 2009).

1.5.2 Wnt recognition and signalling output

The 19 mammalian Wnts engage a variety of different receptors to tune specific cellular responses, the largest class of which is that of the Frizzled (FZD) family. The 10 mammalian FZDs display a GPCR architecture similar to that of Smo, comprising an extracellular CRD and a 7TM region (Dijksterhuis et al., 2014). Three types of downstream Wnt signalling are known and are categorised into two subtypes - 'canonical' (β-catenin-dependent) signalling and 'non-canonical' (β-

catenin-independent) signalling. β-catenin-dependent signalling involves activation of TCF/LEF target genes. β-catenin-independent signalling involves the Planar Cell Polarity (PCP) and Ca^{2+}-dependent pathways, more associated with cell polarity and migration (MacDonald et al., 2009).

Wnt signalling is believed to predominantly involve the formation of an asymmetric ternary signalling complex at the cell surface. Canonical signalling is initiated by the bridging of a FZD receptor and a co-receptor molecule from the Lipoprotein-Related Protein family (LRP5/6) (Fig. 1.7A) (Bourhis et al., 2010). In the absence of a Wnt ligand, intracellular β-catenin forms a complex with Axin, APC, GSK3 and CK1 (Amit, 2002; Willert et al., 1999). This leads to the phosphorylation of β-catenin by GSK3 and CK1 and subsequent proteasomal degradation (Angers et al., 2006). In the absence of β-catenin, TCF/LEF transcription factors in the nucleus act to repress Wnt-responsive genes (Cavallo et al., 1998). Upon stimulation via a Wnt ligand, the FZD cytoplasmic tail recruits the large scaffold protein Dishevelled (DVL) (Rothbächer et al., 2000). This enables recruitment of Axin, GSK3 and CK1, and the two kinases are thought to phosphorylate the intracellular portion of LRP5/6 (Zeng et al., 2007). This prevents the degradation of β-catenin and enables its transport into the nucleus in order to associate with TCF/LEF and regulate transcription (MacDonald et al., 2009).

β-catenin-independent signalling comprises two non-canonical pathways (Fig. 1.7B-C). FZD seems to be necessary in almost all instances of Wnt signalling,

although FZD-independent mechanisms have been speculated upon as a mode of signal activation (Billiard et al., 2005; Liu et al., 2008; Niehrs, 2012). PCP signalling has been observed with all Wnt receptors besides LRP5/6, and involves the activation of small GTPases Rac1 and RhoA alongside JNK. This plays a role in the regulation of actin cytoskeletal dynamics and microtubule stability (Bryja et al., 2008; Fanto et al., 2000). The second β-catenin-independent signalling mechanism, the Wnt-Ca^{2+} pathway, involves the activation of PLC and production of Ca^{2+} as a second messenger (Kuhl, 2000; Luna-Ulloa et al., 2011). This, in turn, regulates actin cytoskeletal dynamics and cell motility (Wang et al., 2010). It is worth noting that how specific Wnt signalling pathways are selected for remains poorly understood at present.

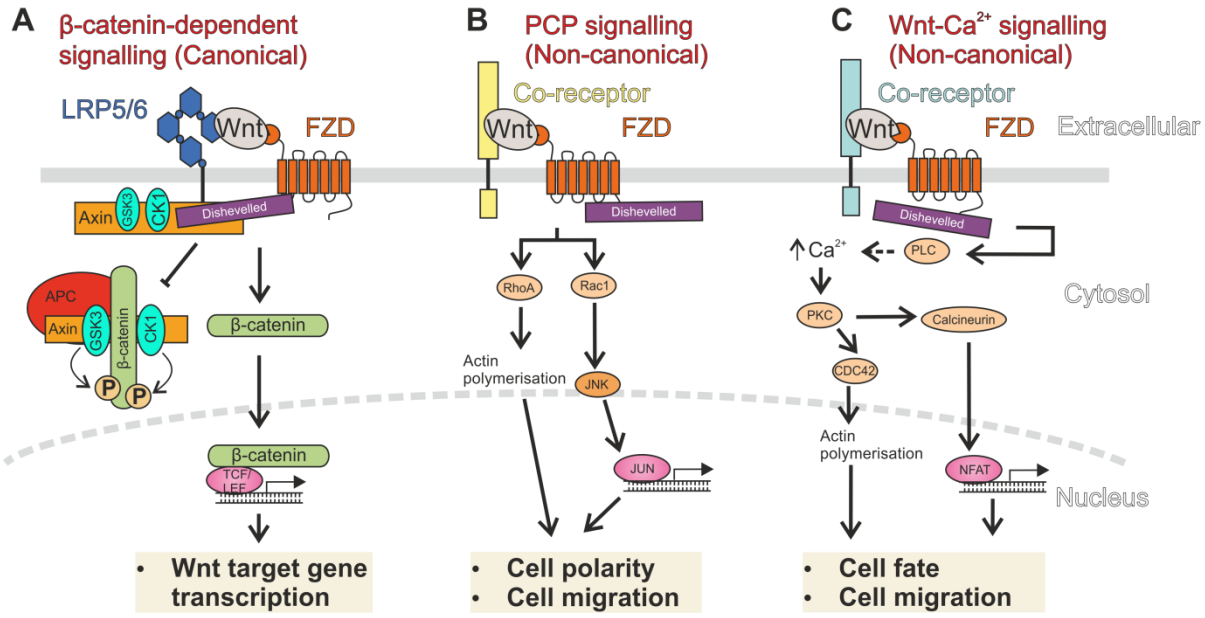


Figure 1.7 – Wnt signalling pathways. A) Canonical Wnt signalling pathway. Formation of an asymmetric FZD:Wnt:LRP5/6 complex leads to stabilisation of β-catenin and enables transcription of Wnt target genes. Wnt ligand binding blocks the formation of the APC/Axin/GSK3/CK1 destruction complex which leads to phosphorylation and degradation of β-catenin; B) Non-canonical PCP signalling pathway. Wnt ligand binding employs the use of a co-receptor such as ROR1/2 in the place of LRP5/6 to modulate cytoskeletal dynamics through small GTPases such as Rac/RhoA and JNK-mediated transcriptional activation; C) Non-canonical Ca^{2+} pathway. A Wnt ligand binds a co-receptor such as Ryk or PTK7 in place of LRP5/6 to activate alternative pathways involved in the regulation of the actin cytoskeleton.

1.5.3 Receptors and modulators

How is a specific downstream Wnt signalling pathway selected for and regulated? The ubiquitous nature of FZDs was made more transparent by the structural characterisation of the Wnt8:FZD8 complex. Wnt8 binds to the FZD8 CRD as a monomer (Fig. 1.8A), making contacts at two distinct sites by gripping the CRD like a 'hand.' Most notably, the amino-terminal palmitoleate moiety of Wnt8 projects into a deep groove upon the FZD8 CRD surface, forming a 'lipid thumb' that interacts

with residues conserved across FZD family members. Additionally, the carboxyl-terminal region of Wnt8 forms an elongated 'index finger' of hydrophobic contacts with the opposing face of FZD8, utilising yet more highly conserved residues (Janda et al., 2012). As Wnts and FZDs appear to be well conserved at both interaction sites, the variability in signalling output is likely to be a consequence of different Wnt ligands being able to recruit other co-receptors into a complex with themselves and FZD (MacDonald and He, 2012). The definition of all 19 Wnt ligands as being canonical, non-canonical or 'intermediate' is still poorly defined, and it appears that there is a great deal of receptor overlap and promiscuity consolidated by defined expression of specific Wnt-binding partners in particular cells at precise times (Grumolato et al., 2010; Janda et al., 2012; Mikels and Nusse, 2006). Wnt signalling is controlled yet further by an array of receptors and modulators beyond FZD and LRP5/6 (Fig. 1.8B). Of note is the number of domains playing a role in these process, and also the notion that the same combination of receptors can activate more than one type of Wnt signalling (possibly as a function of a particular Wnt ligand) (MacDonald and He, 2012; Niehrs, 2012).

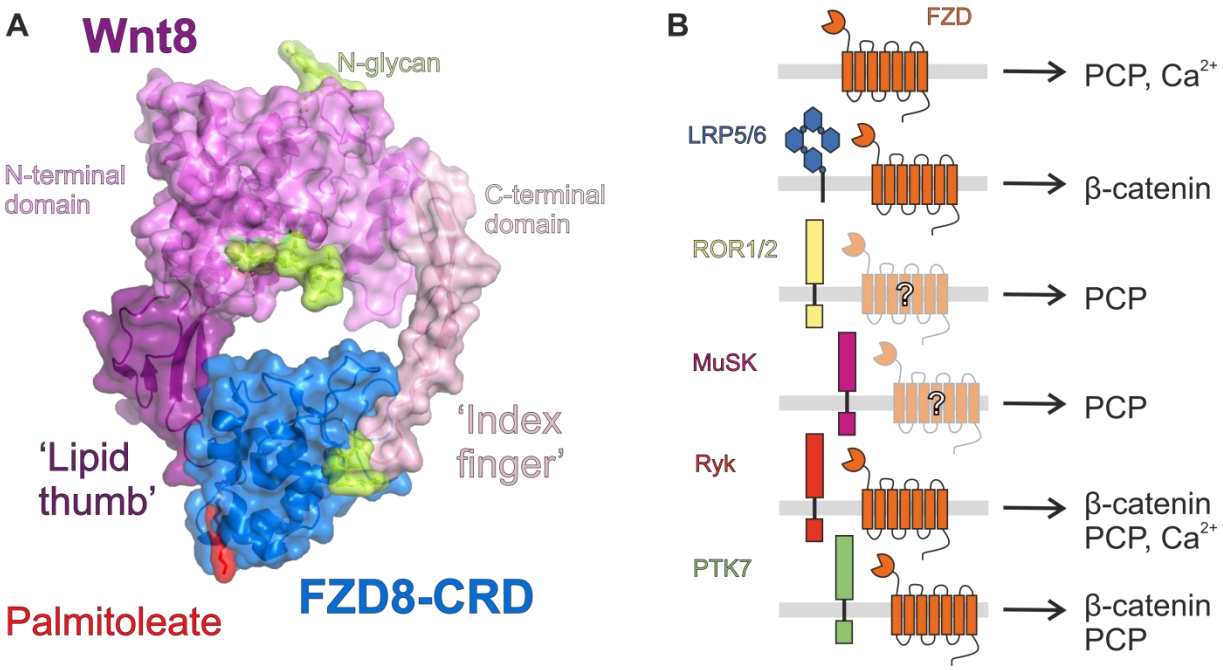


Figure 1.8 – Wnt:receptor interactions. A) The structure of Wnt8 in complex with the CRD of FZD8, with separate domains highlighted in different colours. The Wnt ligand makes contacts with FZD8 at two specific surface hotspot regions (PDB accession code 4F0A); B) A number of co-receptors function alongside FZD receptors. The combination of Wnt, FZD and co-receptor is linked to the activation of specific canonical or non-canonical downstream signalling pathways. ROR1/2 and MuSK have been suggested to be capable of activating signalling in the absence of a FZD receptor (Banerjee et al., 2011; Liu et al., 2008).

LRP5/6 exclusively activates canonical Wnt signalling, engaging ligands in an asymmetric complex with FZD (Figs. 1.7A and 1.8B). The extracellular domains of LRP5/6 contain approximately 1600 amino acids, and are composed of four tandem β-propeller/EGF repeats followed by three LDLR repeats. Distal elements of these domains mediate binding to different Wnts; Wnt9b has been shown to bind to β-propeller/EGF repeats 1 and 2 (P1E1-P2E2), whilst Wnt3a targets repeats 3 and 4 (P3E3-P4E4) (Bourhis et al., 2010). Structural studies of LRP6 P1E1-P4E4 have

indicated that this region forms a kinked horseshoe-like structure, giving a platform that has been shown to be able to accommodate the simultaneous binding of different Wnt ligands (Chen et al., 2011a). Owing to the poor solubility of Wnt ligands, the extent to which particular family members engage and activate signalling through particular FZD receptors and which represent specific canonical or non-canonical pairs requires further study. Co-receptor expression patterns in particular cell types clearly play a large role in determining ultimate signalling output (Dijksterhuis et al., 2014).

A common domain type present in Wnt signalling is that of the FZD-type CRD (Fz-CRD; see section 1.6). These are found within receptors FZD, ROR1/2, MuSK and the secreted FZD-related proteins (sFRP1-5) (Niehrs, 2012). The role of these domains was made clear from the Wnt8:FZD8 complex structure, functioning as a module to interact with the palmitoleate moiety upon the Wnt ligand (Fig. 1.8A) (Janda et al., 2012). The use of a structurally comparable ligand to grant different signalling outputs from 19 different Wnts may rely on subtle deviations between residues in the 'index finger' region as well as precise expression of particular receptors. The 'lipid thumb' region is well conserved between FZD receptor CRDs, suggesting a common lipid recongition mechanism for all members (Janda et al., 2012).

MuSK and ROR1/2 are both receptor tyrosine kinases (RTKs), containing multiple extracellular domains alongside their Fz-CRDs (Mendrola et al., 2013).

ROR1/2 contain extracellular Ig, Fz-CRD and Kringle domains (the subject of Chapter 5 of this thesis; see Fig. 1.15) alongside an intracellular kinase domain. Both have been heavily implicated in the binding of Wnt5a and activation of non-canonical signalling pathways, but whether this proceeds through a ROR2:Wnt or a tripartite ROR2:Wnt:FZD complex remains unknown (Fig. 1.8B) (Billiard et al., 2005; Oishi et al., 2003). A mode of Wnt:ROR2 interaction is difficult to hypothesise without an understanding of how kinase activity influences signalling; current evidence does not support dimerisation of Wnt, and how this would activate RTK signalling via ligand-induced dimerisation of the receptor is not known (Janda and Garcia, 2015; Lemmon and Schlessinger, 2010). MuSK uses Ig1-3 to engage cell-surface ligand LRP4 at the neuromuscular junction (NMJ) and cluster acetylcholine receptors (Kim et al., 2008; Zong et al., 2012). Evidence suggests that Wnt dimerises MuSK through a simultaneous interaction with its Fz-CRD and also LRP4 (mirroring the asymmetric LRP5/6-Wnt-FZD complex), and that this leads to signalling through elements of the PCP pathway (Zhang et al., 2012). Questions remain to be answered about how different cell-surface components and indeed, different Wnts, influence the activities of MuSK and ROR proteins and subsequent signal activation modes.

A number of extracellular Wnt signalling modulators exist to alter signal transduction (Fig. 1.9A). One mechanism involves targeting the palmitoleate moiety of Wnt ligands. Wnt Inhibitory Factor-1 (WIF-1), antagonises Wnt by sequestration of Wnt ligands in the same manner as the sFRP members, albeit not by the use of a Fz-

CRD (Dann et al., 2001; Malinauskas et al., 2011; Wawrzak et al., 2007). Another mode of regulation is via the carboxyesterase Notum, a Wnt deacylase, which enzymatically removes the lipid adduct from the thumb region of Wnt ligands in order to antagonise signalling (Kakugawa et al., 2015). Dkk and Kremen synergise to promote the internalisation of the LRP5/6 co-receptors (Mao et al., 2002; Zebisch et al., 2016). Additionally, FZD receptors at the cell surface can be ubiquitinated by transmembrane E3 ubiquitin ligases, leading to their internalisation and degradation. Secreted R-Spondins form a complex with ubiquitin ligase ectodomains to trigger their internalisation and increase the lifetime of FZD receptors at the cell surface (Zebisch and Jones, 2015). In conclusion, an array of receptors and modulators exist to regulate Wnt signal transduction at a number of steps, contributing to the wide range of pathway outcomes.

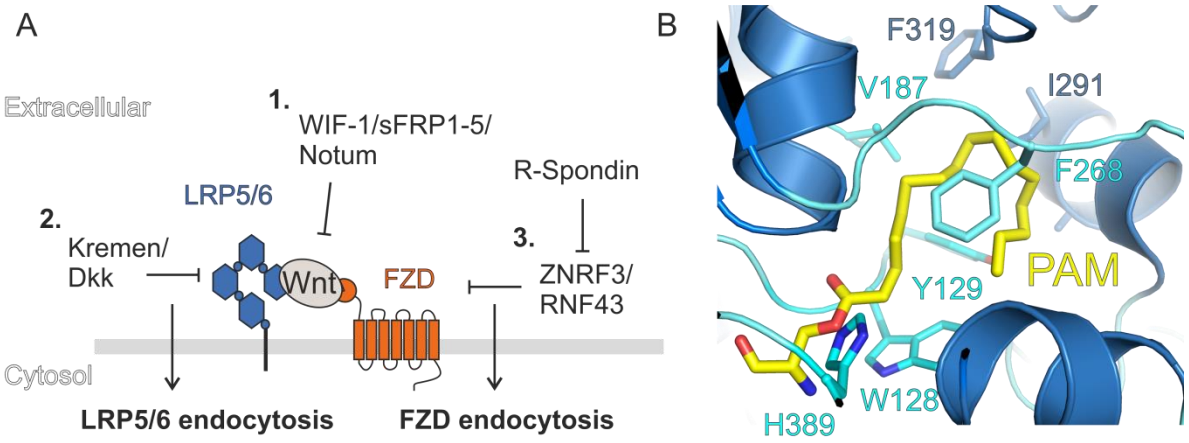


Figure 1.9 - Modulators of Wnt signalling. A) Signal activation can be modulated through any one of the three proteins constituting the asymmetric ternary Wnt:receptor complex. Soluble proteins may bind to the lipid moiety important for Wnt:FZD interactions (WIF-1, sFRP) or remove it altogether (Notum) (1). Alternatively, soluble protein Dkk and membrane-bound Kremen can form a complex with LRP5/6 and trigger its internalisation and removal from the cell surface (2). FZD may also be targeted for endocytosis through ubiquitination by ZNRF3 or RNF43 E3 ubiquitin ligases. The R-Spondins form complexes with ZNRF3/RNF43 and block this process (3); B) Notum active site, shown in complex with a palmitoleate-modified Wnt peptide (PAM). The lipid is kinked and makes numerous van der Waals contacts with Notum residues (PDB ID: 4UZQ). Adapted from (Kakugawa et al., 2015).

1.5.4 Palmitate versus palmitoleate

Despite the differences in constituents and signal transduction between the Hh and Wnt pathways, there are similarities in the forms of ligand acylation and the use of a FZD-class GPCR protein. Whilst it has been shown that the favoured acyl modification on both HhN and Wnt comes in the form of C16:0 and C16:1 chains respectively, these are not exclusive. HhN has been shown to be secreted with saturated and unsaturated C14 to C18 fatty acids, but HhN-C16:0 activates signalling approximately 2.5 fold more than HhN-C16:1 (Long et al., 2015). Porcupine is able to

catalyse transfer of saturated and unsaturated chains of C13-C18 fatty acids, but displays a preference for monounsaturated palmitoleate (C16:1) (Gao and Hannoush, 2014; Takada et al., 2006). What reason is there for this preference for a monounsaturated C16 fatty acid in Wnt signalling? From published work, palmitoleate was assigned as the modification in the Wnt8:FZD8 structure and the lipid adopts a linear conformation which would very likely be the same as a fully saturated C16:0 palmitate moiety (Janda et al., 2012). Mass spectrometry analysis could not unambiguously identify the lipid within this context, and it remains possible that a C16:0 fatty acid is in fact attached to the Wnt ligand. A possible reason for the selectivity of Porcupine for a C16:1 lipid may be suggested by the structure of Notum bound to a Wnt-palmitoleate peptide. The Notum active site preferentially accommodates C14 and C16 lipids with a C9-C10 *cis*-double bond (Fig. 1.9B) (Kakugawa et al., 2015). This suggests that susceptibility to signal modulation by Notum plays a part in the choice of lipid attachment on Wnt ligands via Porcupine, as well as triggering different changes within FZD receptors to modulate signalling outputs. It is also conceivable that *cis*-palmitoleate favours association with lipid rafts and in multivalent clusters in a manner more easily facilitated than by a saturated palmitate molecule (Nile and Hannoush, 2016). Finally, it is worth noting that the limited studies on this matter have utilised Wnt3a only. Further study may point to various modifications (*cis*/*trans*, saturated and unsaturated) on any of the 19 Wnts, broadening the complexity of signal output.

1.6 Cysteine-Rich Domains

1.6.1 The CRD family

The relevance of the Cysteine-Rich Domain (CRD) in relation to the Hh and Wnt signal transduction pathways has been discussed in detail above. CRDs are found in a range of secreted or membrane-attached proteins involved in the trafficking of small molecules to the formation of protein-protein interfaces, and can be found in the form of individual soluble domains up to components of multi-domain proteins (Byrne et al., 2016; Gong et al., 2016; Ohto et al., 2016; Wibowo et al., 2013).

The limited characterisation of CRDs points to a relatively mobile evolutionary unit, exhibiting notable structural variability across members. As the name suggests, CRDs contain cysteines involved in between 5 to 9 intramolecular disulphide bonds. Typically, there are 5 core α-helices in a splayed-open bundle fold; however, additional helices and β-sheets have been observed to contribute to the overall structure in various members (Fig. 1.10A). Structural comparison of members highlights a helical fold relationship between members, as well as designating the occurrence of 3 'loops' that typically constitute a ligand-binding site (numbered 1-3 and coloured red in Fig. 1.10A) (Bazan and de Sauvage, 2009).

Bioinformatics approaches have unearthed CRDs in a variety of proteins, relating them to roles in several biological contexts (Fig. 1.10B). These can be found as single domains (FRα/β, JUNO) or as part of multi-domain architectures. Within

multi-domain frameworks, the N-terminal CRDs have been more thoroughly structurally characterised (HHIP, Smo, FZD, NPC1, sFRP), but less well-understood proteins have been identified with CRDs in intermediate positions (ROR1/2, MuSK, RECK, Collagen-α1) (Bishop et al., 2009; Byrne et al., 2016; Dann et al., 2001; Green et al., 2014; Kwon et al., 2009; Pei and Grishin, 2012; Stiegler et al., 2009). It worth noting that these CRDs are present within the genomes of metazoans and fungi; however, CRDs have also been identified within plants and protists based upon sequence conservation (Pei and Grishin, 2012).

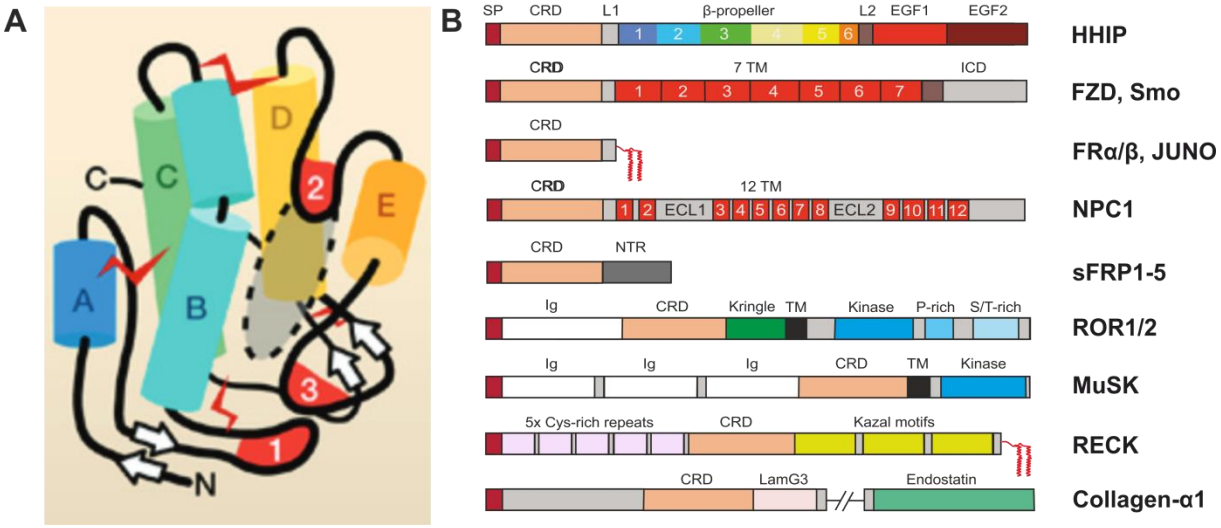


Figure 1.10 – Cysteine-Rich Domains. A) Helical architecture of a typical CRD. Up to 3 ligand-binding loops are observed in all CRDs, numbered 1-3. A pocket or groove enclosed by these elements is shown as a dashed oval. Average disulphide bonds are displayed as red lightning bolts, and are found in varying numbers between different CRD family members (from Bazan and de Sauvage, 2009); B) CRDs and their presence within many proteins of various functions (adapted from Pei and Grishin, 2012).

1.6.2 Pockets and grooves

A sequence-based bioinformatics approach was utilised by Pei and Grishin to classify CRDs throughout metazoan species. This study resulted in the conclusion that there were two classes of CRD based on the pattern of cysteines and conserved disulphides within their sequences; Frizzled-CRDs (Fz-CRDs) and HHIP-FRBP-NPC1-CRDs (HFN-CRDs). Fz-CRDs possess 5 disulphide bonds whilst HFN-CRDs are more varied, possessing between 5 and 9 disulphide bonds (Pei and Grishin, 2012).

Further insight into the evolution of CRDs can be obtained from structural phylogenetic analysis (Fig. 1.11A). This supports the classification on sequence presented above, placing the HFN-CRDs in one defined branch versus the Fz-CRDs, and further insight can be gained from physical inspection from the structures. The ligand-bound structures of FRα, FRβ, RFBP and NPC1 CRDs reveal deep pockets occupied by bound small molecules (Fig. 1.11B) (Chen et al., 2013; Kwon et al., 2009; Monaco, 1997; Wibowo et al., 2013). The structures of CRDs from Smo and Fz8, however, exhibit more surface-exposed 'grooves' which are occupied by their cognate ligands (Figs. 1.11C and 1.8A) (Byrne et al., 2016; Janda et al., 2012). Hence, structural characterisation of CRDs enables the assignment of general ligand recognition mechanisms to designate two groups of CRDs - 'pockets' and 'grooves' (Nachtergaele et al., 2013).

Conformational change of pocket-CRDs is important to their function. The residues in regions around the binding pocket undergo structural rearrangement upon ligand binding to establish an enclosed pocket within which the ligand is bound (Fig. 1.11D) (Wibowo et al., 2013). One variation in this context is highlighted by sperm-egg fusion protein JUNO. The sequence and structure of mJUNO indicate the presence of three loops surrounding a pocket (Fig, 1.10A), but two of these are disordered. Despite having an overall fold similar to FRα and FRβ (Cα RMSD of 1.4 Å and 1.6 Å, respectively), JUNO displays no folate binding capacity. JUNO's cognate function entails binding to sperm surface receptor IZUMO1 to mediate sperm-egg fusion, a species-specific interaction. Humanisation of the loop 1 in mJUNO resulted in the loss of binding to mIZUMO1, whilst humanisation of loop 3 converted mJUNO to also having hIZUMO1 binding capacity (Han et al., 2016). The N-terminal CRD of HHIP (Fig. 1.10B; the subject of Chapter 3 of this thesis) is proposed to belong to the pocket-type group of CRDs (Nachtergaele et al., 2013; Pei and Grishin, 2012). The groove- or 'Fz-type' CRD class are typically more associated with proteins involved in morphogen signalling regulation, and therefore the cognate function of a 'pocket' within Hh signalling remains to be understood.

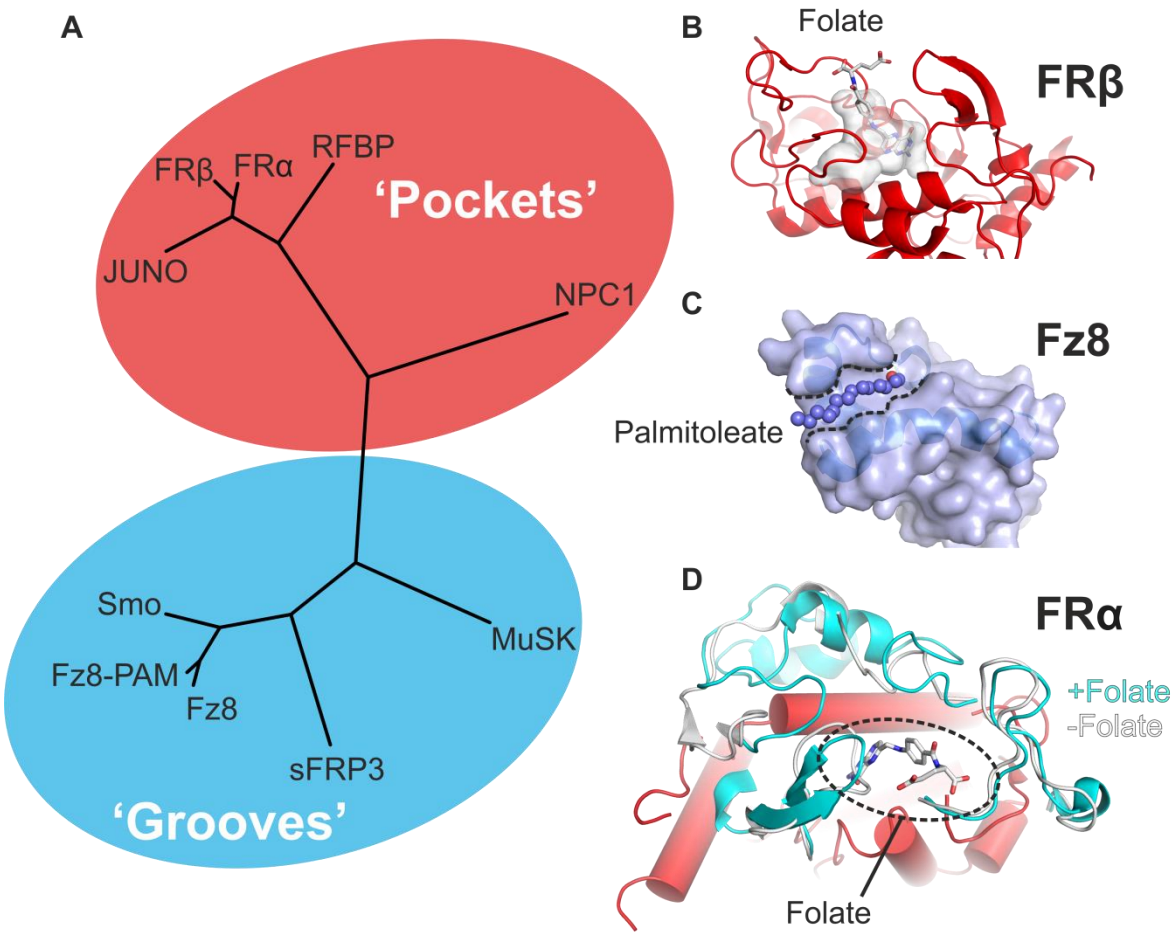


Figure 1.11 – Pockets and grooves as structural motifs as ligand recognition by CRDs. A) Structural phylogeny of currently published CRDs. Analysis of the two distinct branches displays structures with 'pockets' on one branch, and structures with 'grooves' on the other; B) A 'pocket' (represented as a white surface) is observed in FRβ, containing a bound folate molecule (PDB ID: 4KMZ); C) A 'groove' is visible in the Fz8 CRD (outlined using black dashed lines), and is occupied by a palmitoleate molecule (PDB ID: 4F0A); D) Changes in the loops surrounding the folate-binding pocket of FRα upon folate recognition. The folate molecule is shown in stick representation and helices unchanged between the apo- and folate-bound forms of FRα are displayed as red cylinders. The apo-structure loops are displayed in white whilst those from the folate-bound structure are displayed in cyan (PDB IDs: 4KMX, 4LRH).

The role of CRDs does not exclusively involve the recognition of small molecule ligands. Despite Fz8 making intricate contacts via its groove with the palmitoleate moiety of Wnt8, a second interface between both proteins exclusively involves protein:protein contacts (Fig 1.8A) (Janda et al., 2012). The same binding mechanisms can, without structural data, be assumed for the CRDs in ROR1/2, MuSK and the sFRP family. NPC1, the protein responsible for cholesterol incorporation into cell membranes, has a CRD with a deep pocket lined with hydrophobic residues for the retention of sterols (Kwon et al., 2009). Beyond this function, however, there must be a recognition event between the NPC1 CRD and the soluble lysosomal protein NPC2 which scavenges free cholesterol and delivers it to NPC1 (Xu et al., 2007). This notion of CRDs as playing a large role in protein:protein interactions is contributed to by the fact that the NPC1 CRD displays an extensive interface with the extracellular domain C of NPC1, to which it must also deliver cholesterol in order to enable membrane delivery (Gong et al., 2016). This is similar to the structure of full-length Smo, in which the CRD makes contacts with a long extracellular loop extending from the TMD (ECL3) and a linker domain (LD) (Fig 1.12A). A cholesterol molecule bound in a groove of the CRD makes additional contacts with ECL3, and mutations to residues involved in ligand coordination impair Smo-activated signal transduction (Fig. 1.12B) (Byrne et al., 2016).

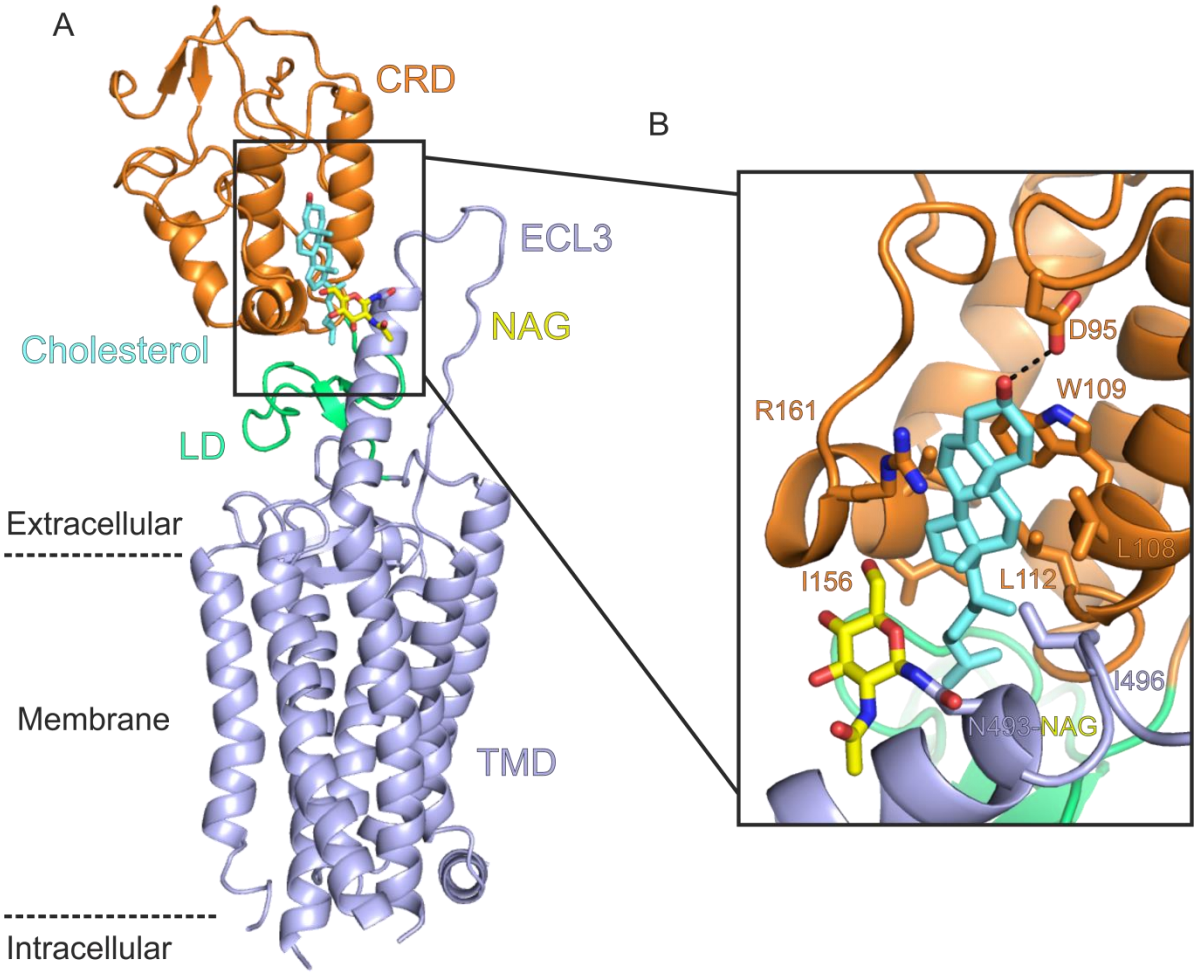


Figure 1.12 – The structure of the Smoothened CRD and transmembrane domain. A) The extracellular region and transmembrane domain (TMD) of human Smoothened. The CRD contains a cholesterol molecule (cyan) and makes contacts with an extended loop from the TMD (ECL3) which runs between TMD helices 6 and 7, as well as with the linker domain (LD); B) Close-up of cholesterol interactions. Cholesterol forms a single hydrogen bond with the side chain of D95, which utilises the 3β-hydroxyl group of the sterol. A number of additional van der Waals interactions stabilise Smo:ligand interactions. Several CRD residue side chains shield the α-face of cholesterol; additionally, R161 shields the β-face alongside ECL3 residue I496 and an N-linked sugar (NAG – N-acetyl-glucosamine). (PDB ID: 5L7D).

CRD-like motifs have also been identified in unexpected places. Structures of human Glypican-1 and *Drosophila* Glypican homologue Dally-like (Dlp) display the presence of an HFN-CRD region to which long helical insertions have been added (Fig. 1.13), and HhN and Wnt ligands have been shown to bind to these core proteins in the absence of any GAG modifications (Capurro et al., 2008; Kim et al., 2011; Svensson et al., 2012). No pocket or groove is apparent, however, and studies have shown that only mutations at a site at the opposite end of the Dlp core protein to this pseudo-CRD abolish the interaction with HhN (Kim et al., 2011). However, the location of the Wnt-glypican interface has yet to be determined (Capurro et al., 2005). A divergent Fz-CRD has also been identified in human FAM155B. Interestingly, the fungal homologue Mid1 functions as a stretch-activated calcium channel, and further studies must be carried out to validate the function of this within a non-fungal species (Pei and Grishin, 2012).

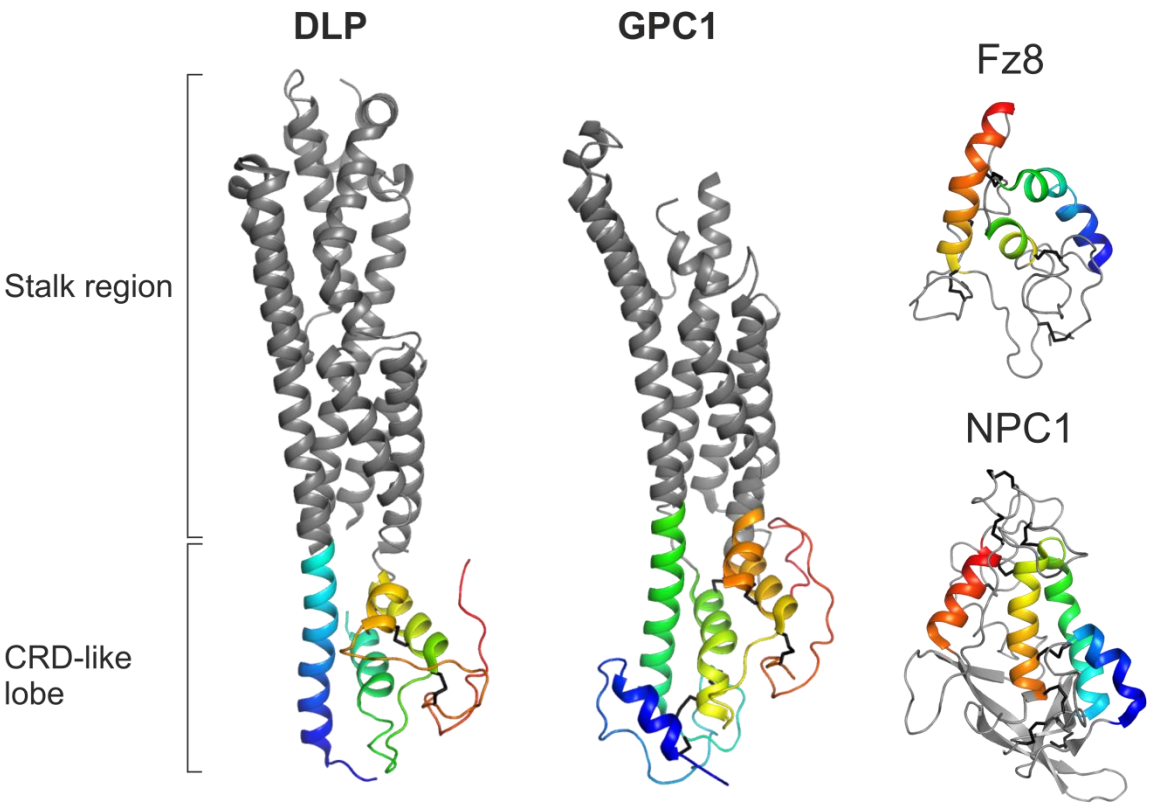


Figure 1.13 – A divergent CRD found in Glypicans. An N-terminal CRD-like lobe is observed in GPC1 and DLP. Four helices, comparable to the core helices seen in the Fz8 and NPC1 structures, form a tight CRD-type helical bundle. Fz8 and NPC1 are displayed in the right-hand panel for comparison. CRDs and the Glypican CRD-like region are coloured in rainbow representation. The stalk regions of GPC1 and DLP do not contribute to this CRD-like structure and are coloured in grey (PDB IDs: 3ODN, 4ACR, 1IJY and 3GKH).

1.6.3 Potential roles in the evolution of morphogen signalling: Links to lipid transport and metabolism

The study of CRDs may give insight into the evolution of morphogen signalling systems due to the roles of lipids in the Hh and Wnt signalling systems. As there are multiple Fz-CRDs within Wnt signalling, Smo is of note as it represents a FZD receptor gene duplication that activates Hh signalling and does not bind to a

Wnt ligand (Pei and Grishin, 2012). Whilst Smo does not exhibit direct binding to a lipid bound covalently to a protein, the CRD can directly bind to cholesterol and related sterols (Byrne et al., 2016; Huang et al., 2016; Luchetti et al., 2016). Additionally, a cavity within the 7TM region is a target for the binding of hydrophobic molecules (Sharpe et al., 2015). The HhN receptor Ptc contains a sterol-sensing domain (SSD), mutations within which eliminate Smo repression whilst retaining HhN-binding ability (Strutt et al., 2001). A SSD is also found with Disp, and mutations in this region prevent secretion of cholesterol-modified HhN (Tukachinsky et al., 2012).

An indication as to the prevalence of CRDs within the regulation and signalling capacity of lipid-modified morphogens may lie with small insect pheromone and odourant-binding proteins (PBP/OBPs). While there is insufficient sequence similarity to directly assume an evolutionary relationship, the topology of these proteins resembles that of the CRD helical bundle, supported by 3 or more disulphide bonds (Figs. 1.10A and 1.14). The *Drosophila* sensible, a sensory projection rich in olfactory neurons, contains LUSH, a secreted PBP. LUSH binds 11-*cis*-vaccenyl acetate, triggering a conformational change and binding of a GPCR-like receptor to activate signalling pathways associated with mate recognition (Laughlin et al., 2008; Tegoni et al., 2004). The preservation of FZD GPCR-like domains in Hh and Wnt signal transduction, as well as the responsiveness of Smo to hydrophobic molecules and the presence of small molecule binding functions within CRDs, points to the

evolution of such morphogen signalling networks from some ancient chemosensory system still functional in insects (Bazan and de Sauvage, 2009; Byrne et al., 2016; Nachtergaele et al., 2013). Ptc is homologous to bacterial RND transporters with roles in the efflux of toxic molecules, which may also have evolved from this same signalling system (Hausmann et al., 2009).

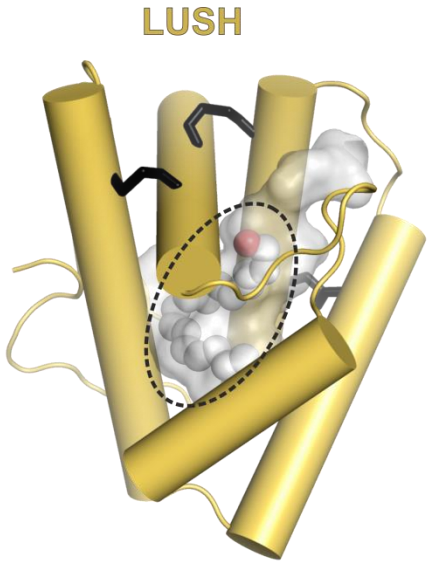


Figure 1.14 – Structure of LUSH, a fly pheromone-binding protein. LUSH displays a helical topology enclosing a central ligand-binding pocket (shown as a white surface). A hydrophobic ligand is observed to be bound within this cavity (dashed oval), displayed in sphere representation (PDB ID: 1DQE).

1.6.4 CRDs and therapeutic potential

The presence of small molecule-binding motifs important for signalling within CRDs highlights them as strong candidates for drug discovery and one such is example is Smo (see section 1.4.5). The responsiveness of the Smo CRD to sterols alongside the capacity for antagonism by sterol-alkyne derivatives represents an opportunity for cancer treatment, as well as expanding these into combinatorial approaches concerning the 7TM GPCR-like region (Byrne et al., 2016; Nachtergaele et

al., 2013; Wang et al., 2014). Dysregulation of Wnt signalling through related FZD proteins is also implicated in cancer, and preliminary design of small molecule Wnt-FZD antagonists has granted success in blocking signal transduction (Lee et al., 2015).

Antifolates have been prescribed in cancer treatments for some 50 years, and the structural characterisation of FRα/β with these should enable new methods for targeting diseased cells with similar molecules (Jackman, 2004; Wibowo et al., 2013). The same can be said of RFBP, the overexpression of which serves as a biomarker for prostate and breast cancer (Johnson et al., 2009). As discussed, the function of the HHIP CRD has yet to be discovered but, given the role of small molecules within CRD function and also their potency regarding Hh signalling, this may yet prove to be another target for the treatment of Hh signalling-related disorders (Bazan and de Sauvage, 2009; Bishop et al., 2009). A large amount of structural information has been accumulated thus far and further therapeutic prospects may be realised from structural studies of uncharacterised CRDs such as ROR1/2, RECK and full length FZDs.

1.7 Global perspectives and aims

Cellular communication is vital for organismal development and homeostasis, and a number of mechanisms are used to regulate signalling through various ligands. One such mechanism includes the use of common protein domains, such as CRDs, to interact with molecules of a variety of shapes and sizes. Another includes the

regulation of both ligand and receptor distribution via polymeric chains of heavily sulphated sugars at the cell surface and throughout the ECM. Contacts between CRDs and other intrinsic domains have been shown to be involved in the functions of NPC1 and Smo (Byrne et al., 2016; Gong et al., 2016). The study of other CRD-containing proteins within a multi-domain architecture is important to gain a more complete functional understanding. Two examples of multi-domain CRD-containing proteins studied in this thesis include HHIP (see section 1.4.5) and ROR2 (see section 1.5.2) (Fig. 1.15A). HHIP contains an uncharacterised N-terminal CRD and structural studies of this and the associated functional implications are described in Chapter 3. The remainder of HHIP comprises a β-propeller, 2 EGF repeats and a predicted C-terminal helix. The C-terminal 6-bladed β-propeller and EGF repeats of HHIP have been structurally characterised (Bishop et al., 2009), and constitute an elongated molecule (Fig. 1.15B). The role of these C-terminal domains is the sequestration of HhN ligands and downregulation of Hh signalling, and this is dependent primarily on contacts made with blade 3 of the HHIP β-propeller (Bishop et al., 2009) (Fig. 1.15C and D). HHIP has been shown to be secreted, but the molecular mechanisms underlying this process are unknown (Coulombe et al., 2004). To this effect, the roles of the N- and C-terminal regions of HHIP in the context of GAG interaction and protein complex assembly are the focus of Chapter 4. A second CRD-containing protein studied in this work is ROR2, which comprises a CRD in between a N-terminal Ig domain and a C-terminal Kringle domain (Fig. 1.15A) (Green et al., 2014).

The role of ROR2 is poorly understood, but has been linked to both canonical and non-canonical Wnt signalling via both FZD-dependent and FZD-independent pathways (Liu et al., 2008; Mikels and Nusse, 2006). Structural characterisation of the ROR2 extracellular domains is detailed in Chapter 5, alongside a discussion of the functional implications.

Both the presence of CRD-containing receptors and the regulation of protein distribution by GAG chains in the ECM are involved in the regulation of signalling and are fundamentally linked to development and disease. In this study, a combination of structural, biophysical and *in vivo* approaches were used to investigate the mechanisms involved in molecular interactions and signal transduction of CRDs within multi-domain architectures.

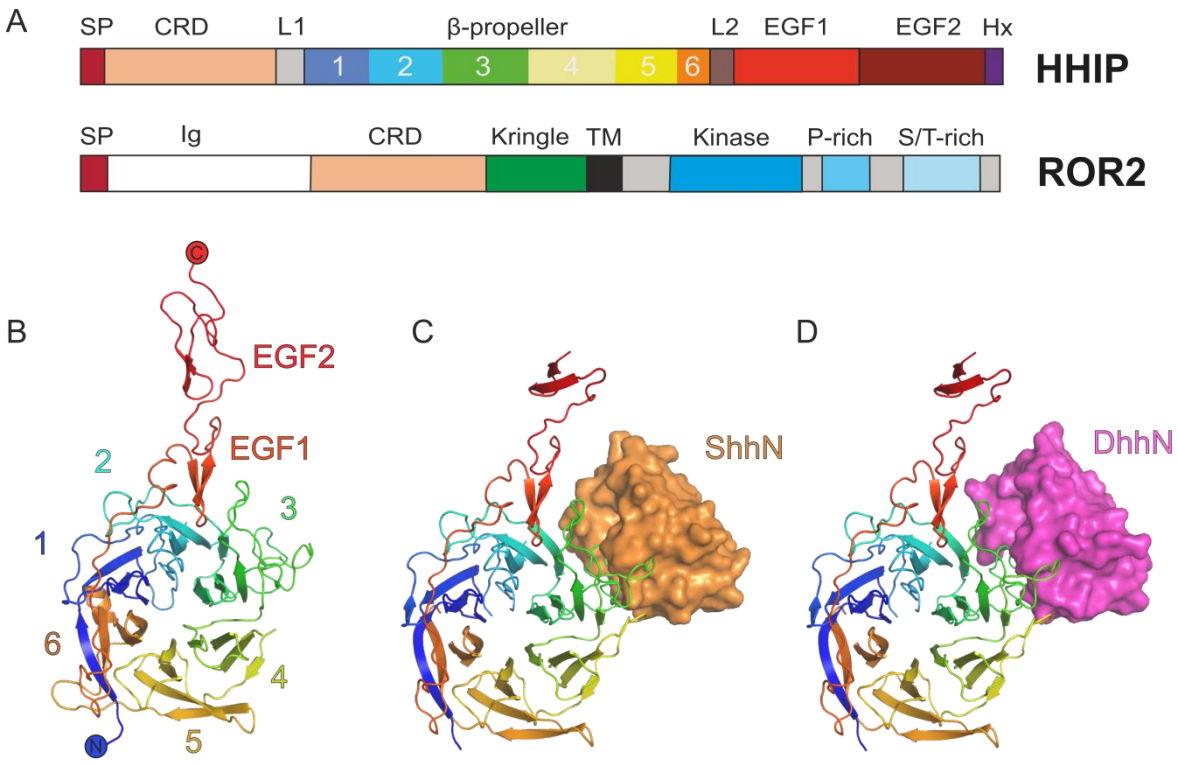


Figure 1.15 – Proteins studied in this thesis. A) Domain architectures for HHIP and ROR2. Both contain a signal peptide (SP) which is removed by signal peptidase in the endoplasmic reticulum during protein expression. CRDs are found in both proteins, alongside a host of other domains. (L1 and L2 – Linkers 1 and 2, Hx – C-terminal helix; TM – transmembrane helix); B) Structure of HHIP-C (β-propeller and 2 EGF repeats). The protein is coloured as a spectrum of blue to red from the N- to the C-terminus. All 6 blades of the β-propeller are numbered (PDB ID: 2WFT); C) HHIP-C:ShhN complex structure (PDB ID: 2WFX); D) HHIP-C:DhhN complex structure (PDB ID: 2WG3). Both HhN ligands bind to a region encompassing blade 3 and part of blade 2 of the HHIP-C β-propeller.

Chapter 2

Materials and methods

Chapter 2: Materials and methods

2.1 Molecular Biology

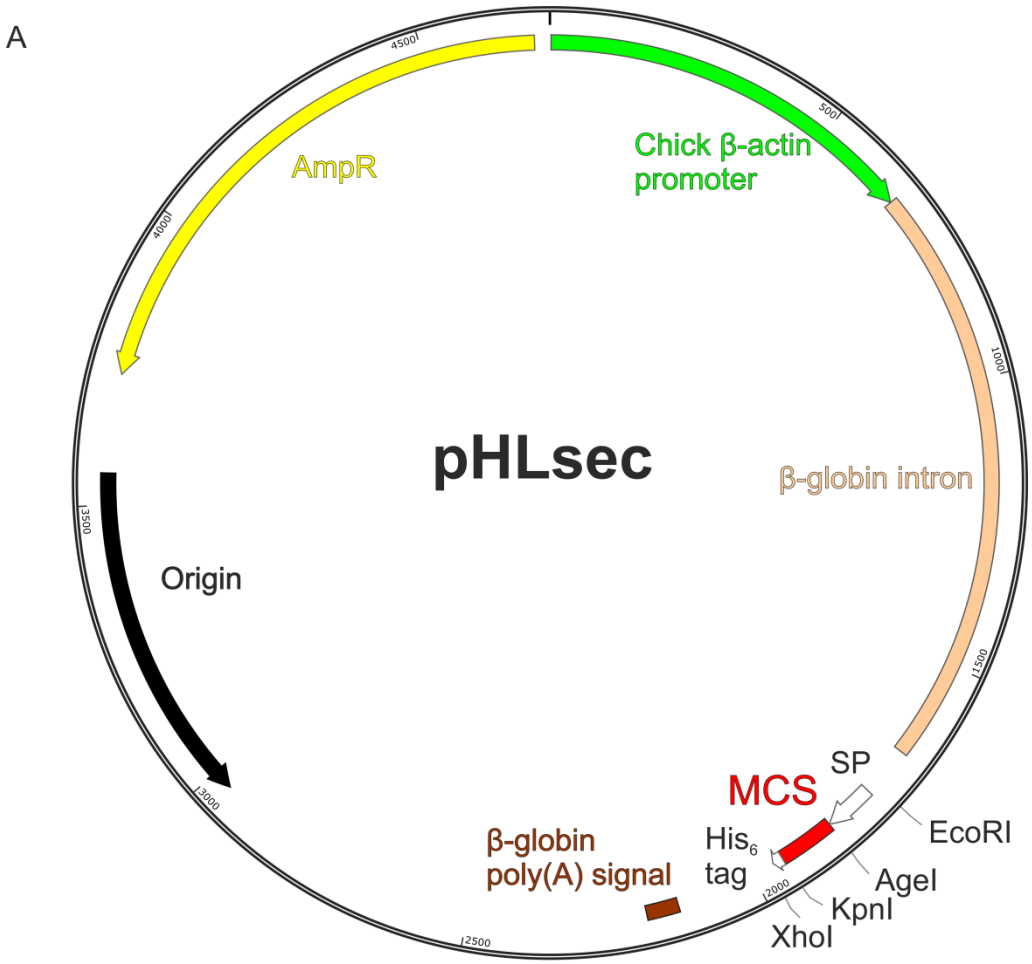
2.1.1 cDNA and expression vectors

Constructs of human HHIP (Genbank ID: 20143973) comprising the N-terminal domain (HHIP-N: residues 39-209), C-terminal domain (HHIP-C: residues 213-670) and a full-length construct only lacking the C-terminal helix (HHIP FL ΔHx: residues 39-670) were cloned into the pHLsec vector for mammalian protein expression, resulting in a fusion comprising an N-terminal signal sequence and a C-terminal His_6-tag (Aricescu et al., 2006) (Fig 2.1). HHIP FL ΔHx contained 7 point mutations (R185A, K186A, R189A, K204A, R210A, K211A and K213A) in order to prevent proteolysis (Bishop et al., 2009). For immunofluorescence microscopy experiments, construct boundaries were maintained but primers were altered to place an N-terminal HA tag after the pHLsec signal sequence and a stop codon at the C-terminus of each construct. Additionally, a variant of HHIP-C was cloned (213-700) containing an N-terminal His_6-tag after the signal peptide, as well as the C-terminal 30 residues of HHIP (Hx) followed by a stop codon. The chimeric human/zebrafish HHIP (Genbank ID: 751998145) construct comprised residues 39-79 of human HHIP and residues 75-183 of zebrafish HHIP (hzHHIP-N). A variety of constructs were cloned for both human HHIPL1 (Genbank ID: 187936955) and human HHIPL2 (Genbank ID: 124248546) into pHLsec. For HHIPL1, constructs comprising the full-

length version (19-782) as well as the CRD and β-propeller (19-605) and the CRD alone (19-180) were cloned into pHLsec. A further 12 constructs for the CRD and a construct comprising the C-terminal SRCR domain (668-782) were cloned into both pHLsec and pHL1D4 backgrounds. For HHIPL2, constructs comprising the full-length (42-724), CRD and β-propeller (42-621) and the CRD alone (a total of 12 constructs) were cloned into pHLsec (see Supplementary Table 3.2).

Constructs comprising the human ROR2 (Genbank ID: 19743898) ECD (34-403), and Ig-CRD (60-307), CRD-Kr (171-396) and CRD (171-307) constructs were cloned into pHLsec and the pHLAviTag3 vectors. Additionally, human BMPR1b (Genbank ID: 219521908) ECD constructs (14-126, 14-112, 20-126, 20-112) were cloned into the pHLsec and pHLAviTag3 vectors.

A


AmpR
Chick β-actin promoter
pHLsec
β-globin intron
Origin
MCS
SP
EcoRI
His6 tag
AgeI
KpnI
XhoI
β-globin poly(A) signal
4500
5000
1000
1500
2000
2500
3000
3500
4000

B

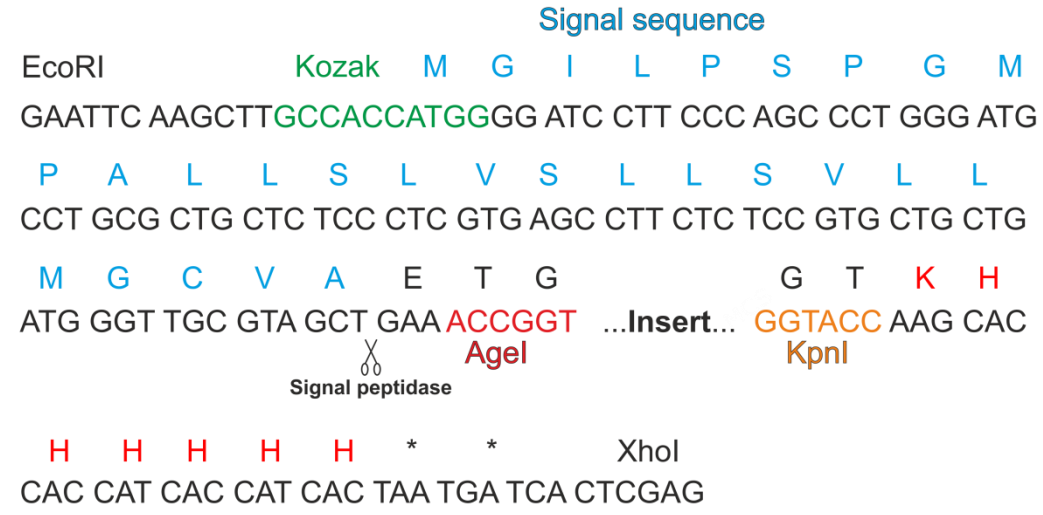

Signal sequence
EcoRI Kozak M G I L P S P G M
GAATTC AAGCTTGCCACCATGGGG ATC CTT CCC AGC CCT GGG ATG
P A L L S L V S L L S V L L
CCT GCG CTG CTC TCC CTC GTG AGC CTT CTC TCC GTG CTG CTG
M G C V A E T G G T K H
ATG GGT TGC GTA GCT GAA ACCGGT ...Insert... GGTACC AAG CAC
AgeI
KpnI
Signal peptidase
H H H H H * * XhoI
CAC CAT CAC CAT CAC TAA TGA TCA CTCGAG

Figure 2.1 – Cloning into pHLsec (above). A) Vector map. pHLsec contains an ampicillin resistance marker (AmpR) and bacterial replication origin for cloning using *E.coli*. The chick β-actin promoter and β-globin intron are used to enhance expression of the target gene. Cloning into the multiple cloning site (MCS) using AgeI/KpnI gives the insert of choice an N-terminal signal peptide (SP) and a C-terminal hexa-histidine (His_6) tag. The β-globin poly(A) signal is attached to the 3' end of the transcribed mRNA and is important for nuclear export; B) The multiple cloning site of pHLsec. Inserts cloned in using AgeI/KpnI are expressed with an N-terminal signal sequence (blue) and a C-terminal His_6-tag (red). The Kozak sequence (green) plays a role in translation initiation. This signal sequence is enzymatically removed by signal peptidase in the endoplasmic reticulum during the secretory process (site displayed). The signal sequence can also be altered via cloning using EcoRI instead of AgeI. Additionally, alternative tags can be introduced via cloning using XhoI (* = Stop codon). Adapted from (Aricescu et al., 2006).

2.1.2 Construct design and PCR

Construct design was facilitated by ternary structure prediction using HHpred (https://toolkit.tuebingen.mpg.de/hhpred), alignment with related sequences using MultAlin (Corpet, 1988), domain annotation from UniProt (http://www.uniprot.org/) and disorder prediction using RONN (Yang et al., 2005). PCRs were performed in 50 µL volumes containing: 50-100 ng template DNA, 0.2 pmol/µL of both forward and reverse primers, 0.25 mM dNTPs, 1x Pyrobest reaction buffer II (containing 1 mM Mg^{2+}) and 0.2 µL Pyrobest™ DNA Polymerase (Takara Clontech). Each PCR was performed in a T-100™ Thermal Cycler (Bio-Rad), following the scheme -

- An initial 30 second melting step at 98°C;
- 35 repeats of the following 3 cycles -
    - 98°C (melting), 10 seconds;
    - 58°C (annealing), 30 seconds;

- 72°C (extension), varied in length of time according to the size of the insert in question;

- A final extension step at 72°C for 10 minutes

PCRs were kept at 4°C prior to restriction digest. PCR products were analysed based on molecular weight through agarose gel electrophoresis and were subsequently purified using a QIAquick PCR purification kit (Qiagen).

2.1.3 Restriction digestion, ligation and transformation

Restriction digests were performed in 50 µl volumes containing 30 µl purified PCR product, 5 µL Cutsmart™ Buffer and 1 µL each of AgeI and KpnI restriction enzymes (New England Biolabs) for cloning into pHLsec and its variants. Digests were incubated at 37°C for 2 hours, separated by agarose gel electrophoresis and gel-purified using a QIAquick gel extraction kit (Qiagen). The pHLsec vector backbone was subject to the same treatment. The concentration of both insert and vector were estimated using the A_{260} (Nanodrop) and both were mixed at a 3:1 molar ratio. 10 µL of quick ligation reaction buffer and 1 µL of Quick T4 DNA Ligase (New England Biolabs) were added to this mixture and the ligation reaction allowed to proceed at 25°C for 5 minutes. 1 µL of this ligated product was then mixed with 30 µL of DH5α library efficiency competent *E. coli* cells (Invitrogen) on ice, and incubated for 30 minutes. Competent cells were transformed by heat shock at 42°C for 45 seconds and incubation on ice for a further 2 minutes. Transformed cells were recovered via the addition of 200 µL of S.O.C. medium (Thermo Fisher) and shaking at 37°C for 30

minutes. 200 μL of this transformation mixture was plated on LB-agar plates containing 100 μg mL^{-1} Ampicillin and incubated overnight at 37°C. 2 individual colonies were selected from each plate, inoculated into LB medium containing 100 μg mL^{-1} Ampicillin and grown at 37°C with shaking. Cells were pelleted via centrifugation at 2425 xg for 10 minutes and plasmid DNA was purified using a QIAprep spin miniprep kit (Qiagen). 1 μL Plasmid DNA was subjected to a restriction digest for 1 hour at 37°C in a 10 μL volume containing 1 μLCutsmart™ Buffer and 0.1 μL each of AgeI and KpnI restriction enzymes (New England Biolabs). The presence of correct inserts was verified via agarose gel electrophoresis, and Sanger sequencing was performed by either Source BioScience or GATC Biotech.

2.1.4 Site-directed mutagenesis

Site-directed mutagenesis was performed on double-stranded plasmid DNA using the two-step overlap extension PCR method (Heckman and Pease, 2007). Four primers were utilised over two separate initial reactions, comprising forward and reverse primers corresponding to the start and end points of the construct of interest, and two complimentary primers of approximately 21 bp corresponding to the intragenic target site for the point mutation(s). DNA from these PCRs was purified via gel purification using a QIAquick gel extraction kit (Qiagen), and these fragments were mixed in a 1:1 molar ratio to a final weight of DNA of 50-100 ng to be used as a template for a second PCR. Forward and reverse primers from above were used to span the entire gene containing the mutation(s) of interest, and subsequently full-

length inserts containing the mutation(s) of interest were gel purified. Inserts were then subjected to restriction digest, ligation and transformation into *E. coli* DH5α competent cells as described above.

2.2 Cell culture

2.2.1 For small scale expression testing

A 175 cm^2 T175 flask (Greiner Bio-One) of 90-100% confluent HEK293T cells grown in Dulbecco's Modified Eagle's Medium (DMEM high glucose) (Sigma) containing 10% Fetal Bovine Serum (FBS), 2 mM L-Glutamine and a 1:100 dilution of MEM non-essential amino acids ('10% DMEM') (all Life Technologies) was decanted and cells were washed with 5 mL sterile PBS pH 7.4 (Life Technologies). The HEK293T cell monolayer was then incubated for 5 minutes with 5 mL 0.05% Trypsin-EDTA (Life Technologies) to digest the extracellular matrix and facilitate resuspension of cells. Cells were dislodged, 40 mL 10% DMEM was added to the cell suspension, and 2 mL aliquots were added per well of a 6-well dish (Greiner Bio-One). Cells were incubated at 37°C, 5% CO_2 for 20 hours in order to enable them to reach confluency prior to transfection.

2.2.2 For large scale protein expression

As in section 2.2.1, a T175 flask of HEK293T cells was split using trypsin, and the suspension was supplemented with 20 mL of 10% DMEM. This cell suspension was added to 200 mL of 10% DMEM in an expanded surface roller bottle (2125 cm^2,

Greiner Bio-One). Cells in roller bottles were then incubated for 3 days at 37°C, 5% CO_2 prior to transfection.

2.3 Protein overexpression and purification

2.3.1 Western blot analysis

To test the suitability of expression constructs for protein purification, expression vectors were purified using an endotoxin-free Miniprep kit (Qiagen) and transfected into HEK293T cells at 90-100% confluency in 6-well dishes. 4 μg of DNA was transfected per well with the addition of 8 μg of lipofectamine (Thermo Fisher Scientific), and cells were left at 37°C, 5% CO_2 for 3 days. Supernatant was harvested via centrifugation at 3380 xg for 10 minutes. The remaining cell pellet was resuspended in 2 mL of PBS and re-pelleted via centrifugation at 3380 xg for a further 10 minutes in order to remove residual cell culture medium. The cleared supernatant from this step was discarded and the washed cell pellet was resuspended in 2 mL of fresh PBS for further analysis.

Proteins were separated using Sodium Dodecyl Sulphate Polyacylamide Gel Electrophoresis (SDS-PAGE) on a 4 – 12% Bis-Tris PAGE gel (Life Technologies) run at a voltage of 180 V for 40 minutes. Proteins were then blotted onto a nitrocellulose membrane at a voltage of 15 V for a period of 60 minutes using NuPAGE buffer (Life Technologies) supplemented with 10% ethanol in a Bolt™ Mini Blot module (Life Technologies). Blotted membranes were then blocked in 10% milk in water at 25°C

for 10 minutes, before incubation with monoclonal anti-pentaHis antibody (Qiagen) or Rho-1D4 antibody (University of British Columbia) diluted 1:1000 in 5% milk/PBS for 1 hour. Membranes were then washed 5 times for 2 minutes each in a buffer of PBS, 0.1% Tween-20 before being probed a second time with goat anti-mouse antibody coupled with horseradish peroxidase (Sigma) diluted 1:2000 in 5% milk for 1 hour at 25°C. Following a further 5 washes for 2 minutes each in PBS, 0.1% Tween-20, blots were developed using Clarity™ Western ECL substrate (Bio-Rad) via incubation for 5 minutes at 25°C. Signal was detected with a ChemiDoc™ MP imaging system (Bio-Rad).

2.3.2 Large-scale transfection of HEK293T cells

Expression plasmids were purified in milligram quantities from DH5α *E. coli* cells via Endotoxin-free Giga- or Mega-prep kits (Qiagen) (Aricescu et al., 2006). DNA was chloroform extracted and transfected into roller bottles containing approximately 90% confluent HEK293T at a ratio of 2 mg of DNA to 4 mg of transfection reagent PEI per litre of media. Due to the presence of glycosylation sites in all constructs, media was supplemented with 5 μM kifunensine, an N-linked glycosylation inhibitor (Toronto Research Chemicals). This results in the inhibition of α-Mannosidase and the addition of N-linked glycans no further than $Man_9GlcNAc_2$ (Chang et al., 2007). Cells were grown at 37°C, 5% CO_2 for 4-7 days. In the case of hzHHIP-N, 50 mM L-Arginine (Sigma) was added to each roller bottle in order to

compete with Golgi-localised furin endoproteases and block proteolytic degradation (Thomas, 2002).

2.3.3 Purification of hexahistidine-tagged proteins

Media containing overexpressed recombinant protein was harvested by centrifugation at 1935 xg for 30 minutes, and cleared supernatant was subsequently filtered using a 0.22 μM filter (Amicon). Filtered supernatant was buffer exchanged into PBS and concentrated to a volume of 500 mL via diafiltration using a 5 kDa or a 10 kDa concentration device (GE Healthcare). 25 mM imidazole was subsequently added and supernatant buffered to pH 8. Supernatant was then bound to a 5 mL HisTrap HP™ column (GE Healthcare), washed in steps of 25 mM and 50 mM imidazole until a stable baseline was achieved and His_6-tagged proteins were eluted using 300 mM imidazole. Eluted sample was then injected onto an S200 16/60 column (GE Healthcare) in a running buffer of 10 mM HEPES pH 7.5, 150 mM NaCl. HHIP-N and HHIP FL ΔHx were further purified using a 1 mL HiTrap™ Heparin column over a gradient of 10 mM to 1M NaCl in 10 mM HEPES pH 7.5. Purity was analysed by SDS-PAGE and concentration estimated spectroscopically at 280 nm (Nanodrop).

2.4 Protein crystallisation

Prior to crystallisation trials, all purified proteins were subjected to a PCT™ Pre-Crystallisation Test (Hampton) to estimate an appropriate concentration for crystal growth. Nanolitre crystallisation trials were set up using a Cartesian Technologies robot (100 nL precipitant solution plus 100 nL protein) in 96-well Greiner plates and placed within a Formulatrix storage vault at 21°C or 4°C with visualisation system (Walter et al., 2005)

HHIP-N was crystallised at a concentration of 6.8 mg mL^{-1} with the addition of 10 mM Sucrose Octasulphate (SOS) (Toronto Research Chemicals). HHIP-C crystallisation trials were performed at 7 mg mL^{-1} with either 10 mM SOS, 3 mM heparin dp10 or 1.5 mM heparin dp30 (both Iduron). Trials of the human/zebrafish HHIP-N chimeric construct were carried out at 7 mg mL^{-1} with the addition of 10 mM SOS. Crystallisation of the ROR2 CRD-Kr construct was performed at 25 mg mL^{-1}. All samples were deglycosylated via the addition of 0.2 μL EndoF1 per 50 μL final volume prior to dispensing.

2.5 Crystal structure solution

All crystals were harvested using litholoops™ (Hampton) matching crystal size, cryoprotected using glycerol or ethylene glycol and flash cooled in liquid nitrogen. For each crystal, the mother liquor of the drop was supplemented with the amount of cryoprotectant indicated and 0.5 μL of this mixture pipetted on top of the

drop prior to harvesting. Data were collected on beamlines I02, I03, I04, I04-1, I23 and I24 at Diamond Light Source, Harwell, UK.

2.5.1 Data collection and processing

Diffracted X-rays from protein crystal produce a regularly spaced pattern related to the cell parameters and space group symmetry. The intensities of detected X-rays are related to the electron density. Data processing of collected diffraction patterns is required to produce a series of intensities, which was performed in this thesis using XIA2 (Winter, 2010)

Certain data quality indicators can be utilised to assess the data consistency and select appropriate resolution cutoffs. One such indicator is R_{merge}, which measures the spread in data values for the merging of **N** reflections I(*hkl)i* compared to their average intensity $\bar{I}(hkl)$ -

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{N} |I(hkl)i - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{N} |I(hkl)i|}$$

R_{merge} is a poor indictor of data quality as it rises with the multiplicity of the data. Judgements based on this would thereby lead to rejection of higher precision and thus higher accuracy data (Karplus and Diederichs, 2015). An improved data quality assessment is the precision-indicating merging R factor $R_{p.i.m}$;

$$R_{p.i.m} = \frac{\sqrt{\frac{1}{N-1}} \sum_{hkl} \sum_{i=1}^{N} |I(hkl)i - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{N} |I(hkl)i|}$$

$R_{p.i.m}$ has been shown to correlate well with the improved statistics in model refinement, and is a better choice of data quality indicator than R_{merge} (Weiss, 2001).

The above statistics have previously been used to choose resolution cutoffs, typically based on the shell where R_{merge} does not exceed ~0.8 and the signal-to-noise ratio, $I/\sigma I$, is 2.0 (Karplus and Diederichs, 2012). This leads to the discarding of high resolution information content, which produces lower-quality models than are possible. The $CC_{1/2}$ is a more appropriate metric to select a resolution limit. Higher resolution information leads to greater model quality. This entails splitting data in two random half-datasets and calculating a correlation coefficient. A $CC_{1/2}$ of 0.3 in the highest resolution shell is used as a guide to select resolution cutoffs in this thesis. This has been observed to lead to improvement in both R factors in refinement and electron density map quality (Karplus and Diederichs, 2012).

2.5.2 Solution of the phase problem: Experimental phasing

Experimental phasing entails the identification of the location of a marker atom from which initial phases for reconstruction of the protein electron density can be determined. Anomalous dispersion takes advantage of the absorption of X-rays by an element at a particular wavelength known as the 'absorption edge'. This causes a change in the atomic scattering factor $f_{(s,\lambda)}$ –

$$f_{(s,\lambda)} = f^{0}_{(s)} + f'_{(\lambda)} + if''_{(\lambda)}$$

Anomalous scattering factor contributions $f'_{(\lambda)}$ and $f''_{(\lambda)}$ are independent of the scattering angle and cause perturbations in the atomic scattering (Hendrickson,

2014). Scattering from the protein alone is given by $f_{(s)}^{0}$. Friedel's law states that reflections separated by 180° possess the same amplitude but have the opposite sign. Ergo, structure factors $\mathbf{F}_{hkl}$ and $\mathbf{F}_{-(hkl)}$ have equal intensities and conjugate phase angles ($\alpha_{hkl} = -\alpha_{-h-k-l}$). Anomalous scattering breaks Friedel's law and the intensities are no longer equal. Phases for anomalously scattering atoms can be approximated and used in calculation in an initial electron density map for the protein (Rupp, 2009). These can be density-modified using programs such as SHELXE, DM and RESOLVE (Cowtan, 2010; Sheldrick, 2010). Once a map has been calculated, automated model building can be performed (Cowtan, 2006; Terwilliger et al., 2008).

2.5.3 HHIP-N Sulphur Single wavelength Anomalous Dispersion (SSAD)

Crystals of HHIP-N:SOS were grown in 0.1 M Imidazole/MES pH 6.5, 0.1 M carboxylic acids (0.02 M Sodium formate, 0.02 M Ammonium acetate, 0.02 M Sodium citrate, 0.02 M Sodium oxamate and 0.02 M Potassium sodium tartrate), 10% w/v PEG 20 000 and 20% v/v PEG MME 550. Crystals were cryoprotected in 15% glycerol and diffraction data were collected at a wavelength of 1.7712 Å. Inverse beam and mini-kappa goniometry-based approaches were utilised on Diamond Light Source beamlines I03 and I04 respectively to minimize radiation damage between Bijvoet pairs (El Omari et al., 2014; Hendrickson, 2014). 33 data sets were collected with oscillations of 0.1° and exposure times of 0.01 s over a range of 360-720° each. Datasets were processed individually and non-isomorphous crystals and various

anisotropic regions discarded. A native dataset was also collected on beamline I03 at a wavelength of 0.97949 Å.

The structure of HHIP-N-apo was solved from crystals grown in 0.1 M HEPES pH 7.0, 0.2 M NH_4Cl and 20% w/v PEG 6000 and cryoprotected in 30% ethylene glycol. Data was collected on beamline I24 at a wavelength of 0.96862 Å.

2.5.4 ROR2 CRD-Kr Platinum-SAD

Crystals of ROR2 CRD-Kr were grown in 0.1 M HEPES pH 7.5, 1.5 M $LiSO_4$. These were soaked with $KPtCl_6$ for approximately 60 minutes prior to cryoprotection in 20% glycerol. Data were collected on beamline I03 at a wavelength of 1.05 Å, and data from two crystals were merged together. For native data collection, crystals were cryoprotected with 20% glycerol prior to flash cooling and data were collected on beamline I03 at a wavelength of 0.97625 Å.

2.5.5 HHIP-C

Crystals of HHIP-C in complex with SOS were grown in 0.1 M Tris/Bicine pH 8.5, 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 10% w/v PEG 20 000 and 20% v/v PEG MME 550 and cryoprotected in 20% glycerol and 10 mM SOS. Data were collected on beamline I04 at a wavelength of 0.97949 Å. HHIP-C:heparin (30-mer) crystals were grown in 0.1 M bis-Tris pH 6.5, 0.2 M sodium citrate and 20% w/v PEG 3350 and cryoprotected using 30% ethylene glycol. Data was collected on beamline I03 at a wavelength of 0.97631 Å. HHIP-C:Heparin data were processed using the program DIALS (Gildea et al., 2014).

2.6 Structure solution and refinement

2.6.1 HHIP-N:SOS

A Sulphur substructure was initially solved for the HHIP-N:SOS complex using the HKL2MAP interface for the SHELX suite of programs (Pape and Schneider, 2004; Sheldrick, 2010). An anomalous signal to 4 Å resolution was observed using SHELXC, and searching for 6 Sulphur sites as disulphides in SHELXD gave a solution with a reasonable CC_{weak} vs CC_{all}. This was used as an input file for PHENIX AutoSol (Terwilliger et al., 2009), the phases from which were subjected to automated model building using PHENIX AutoBuild (Terwilliger et al., 2008). This process was repeated with the inclusion of the model produced from this first AutoBuild job as a starting model for MR-SAD and phase improvement (see chapter 3 for a more in-depth discussion of HHIP-N:SOS structural solution).

2.6.2 ROR2 CRD-Kr

Initial phases for ROR2 CRD-Kr construct were obtained using PHENIX AutoSol with Pt-SAD (Terwilliger et al., 2009). Four strong sites were visible in the resultant electron density map, and further density modification and automated model building was performed using the program *Buccaneer* from the CCP4 suite (Cowtan, 2006). This produced an interpretable model for most of the CRD (residues 174-304) but phases were not of high enough quality to properly trace and refine the Kr domain (308-396). As a consequence of this, the CRD was used as a molecular replacement search model in *Phaser* (McCoy et al., 2007) against higher resolution

native data (2.7 Å resolution). This solution was fixed, and a second search was performed using an ensemble of 3 homology models for the Kr domain. This resulted in appreciable molecular replacement scores and the structure was manually built using COOT (Emsley and Cowtan, 2004) and refined initially in *Refmac5* (Murshudov et al., 2011).

2.6.3 Molecular replacement for HHIP-N-apo and HHIP-C

For both HHIP-N-apo and the HHIP-C:SOS and HHIP-C:Heparin structures, the molecular replacement method (MR) was used for phasing. MR relies on the knowledge that similar sequences exhibit similar structures. The program *Phaser* (McCoy et al., 2007) attempts to place a model in the unit cell, from which a structure factor summation is calculated (F_{calc}). *Phaser* then calculates the probability that experimental F values will be observed, given F_{calc}. This is output as the Log-Likelihood Gain (LLG) and the Translation Function Z-score (TFZ) (Evans and McCoy, 2008; McCoy et al., 2007). For the solution of HHIP-N-apo, the HHIP-N:SOS structure solved via S-SAD was used as an MR search model. For both ligand-bound HHIP-C structures, HHIP-C-apo (PDB ID: 2WFT) was used as a search model (McCoy et al., 2007).

2.6.4 Structure refinement, validation and analysis

During structure refinement, the R_{work} measures the agreement between structure factors calculated from the protein model and from the experimental diffraction data. This alone is a poor metric for the judgement of crystallographic

structure quality, as data can be overfit (Kleywegt, 2000). The R_{free} was introduced to improve model quality, calculated using 5% of the experimental data not used during refinement (Brünger, 1992). The R_{work}/R_{free} gap can be used to assess if a refinement strategy is appropriate. These values are used alongside r.m.s.d bond angles and lengths, and peptide bond geometry (Ramachandran statistics) to assess the model quality (Kleywegt, 2000).

The final models for HHIP-N:SOS, ROR2 CRD-Kr and HHIP-C:Heparin were refined using AutoBUSTER (https://www.globalphasing.com/). The HHIP-N-apo and HHIP-C:SOS structures were refined using PHENIX (Adams et al., 2010). Between refinement cycles, models were manually manipulated using COOT (Emsley and Cowtan, 2004). The stereochemical properties of all refined structures were assessed using MolProbity (Davis et al., 2007). Structural alignment of CRDs and Cα RMSD calculation was performed with the program SHP (Riffel et al., 2002; Stuart et al., 1979) and the resulting phylogenetic tree plotted using the program PHYLIP (Felsenstein, 1989). Electrostatic potentials were calculated using APBS (Baker et al., 2001) and all structures were prepared for figures using PyMOL (The PyMol Molecular Graphics System, Version 1.7, Schrödinger, LLC). Sequence alignments were performed using the MultAlin server (Corpet, 1988) and the program ALINE (Bond and Schüttelkopf, 2009).

2.7 Surface Plasmon Resonance

2.7.1 Protein-GAG interactions

Surface Plasmon Resonance (SPR) is an optical technique used for the detection of protein-protein interactions. The carboxymethylated dextran surface of a sensor chip is immersed in aqueous buffer, and 'ligand' molecules are immobilised upon this. When a solution of 'analyte' molecules capable of binding to the ligand is flown over the sensor chip, retention of the analyte causes a change in the refractive index at the surface of the sensor chip. This causes a change in θ_{spr}, which can be measured using a photo-detector array to monitor changes in the angle of laser light shone upon the chip. These changes are measured in the form of Response Units (RUs), with 1 RU corresponding to a change in θ_{spr} of 10^{-4} degrees (Van Der Merwe).

CM5 Sensor chips for use in SPR experiments comprise 4 flow cells within which ligand:analyte interactions are monitored. The process of ligand immobilisation in SPR involves primary amine coupling of streptavidin to the dextran surface (O'callaghan et al., 1999). Biotinylated ligands are then deposited upon and retained by the surface through the high affinity streptavidin:biotin interaction (~10^{-15} M) (Driegen et al., 2005). In each case, streptavidin is coupled to each flow cell but one cell is kept ligand-free as a means of subtracting any non-specific interaction between the analyte and streptavidin.

SPR experiments were performed using a Biacore T200 machine (GE Healthcare) in 10 mM HEPES pH 7.5, 120 mM NaCl, 0.05% (v/v) Tween-20, at 25°C.

Proteins were buffer exchanged into running buffer and concentrations calculated from the absorbance at 280 nm using molar extinction coefficient values. Heparin (Iduron; average molecular weight >9000 Da), heparan sulphate (HS) from porcine mucosa (Iduron) and chondroitin sulphate (CS) sodium salt from shark cartilage (Sigma) from were biotinylated using EZ-link Biotin-LC-Hydrazide (Thermo Fisher Scientific) as described previously (Malinauskas et al., 2011). Biotinylated sugars were immobilised upon a CM5 sensor chip to which 3000 RU of streptavidin were covalently coupled (O'callaghan et al., 1999). Post-experiment, the chip was regenerated using running buffer with an increase of ionic strength to 1.5 M NaCl at 30 μL min^{-1} for 120s. HHIP constructs were injected at a flow rate of 5 μL min^{-1} for binding studies. In all experiments, the trace returned to baseline following regeneration. All data were analysed using SCRUBBER2 (Biologic) and GraphPad Prism Version 6.04 (GraphPad Software, La Jolla California USA). Best-fit binding curves were calculated using non-linear curve fitting of a one-site - total binding model ($Y = [R_{max}*X / (Kd + X)] + NS*X + Background$, where X is analyte concentration and the level of non-specific binding is assumed to be proportional to the analyte concentration; hence, NS equals the slope of non-specific binding). The background value was set to zero as the data had been previously referenced. R_{max} and K_D values quoted are determined for the specific binding portion only.

2.7.2 Protein-protein interactions

Biotinylated protein ligands for SPR were expressed in 6-well dishes. 3 μg of an Avi-tagged construct in the pHLAviT3 background was co-transfected with 1 μg of the BirA-pDisplay plasmid (Howarth and Ting, 2008), and cells were grown with the addition of 2 mM D-Biotin to cell culture medium. Supernatant was harvested after 3 days' expression at 37°C, 5% CO_2 via centrifugation at 3380 xg for 10 minutes, and buffer exchanged via dialysis into running buffer (10 mM HEPES pH 7.5, 120 mM NaCl, 0.05% (v/v) Tween-20). Ligands were immobilised upon a CM5 sensor chip at densities of 200-1000 RU, to which 3000 RU of streptavidin had been coupled.

For measurements between the hBMPR1b ECD and ROR2 constructs (ECD, CRD-Kr and CRD), purified hBMPR1b was buffer exchanged into running buffer and injected at a concentration of 200 μM at 5 μL min^{-1}.

2.8 SEC-MALS

Size Exclusion Chromatography Multi-Angle Light Scattering (SEC-MALS) can be utilised to accurately determine the molecular weight of a particle and determine solution oligomerisation state. This operates on the angular dependence of light scattered by macromolecules being related to the radius of gyration and molecular weight (Wyatt, 1993). 100 μL samples were injected onto an S200 10/30 column (GE Healthcare) equilibrated in a running buffer of 10 mM HEPES pH 7.5, 150 mM NaCl at the concentrations specified in the Results chapters. A Wyatt Dawn

HELEOS-II multi-angle light scattering (MALS) detector and Wyatt Optilab rEX refractive index monitor recorded both the refractive index and light scattering once separated via SEC. ASTRA software (Wyatt Technology) was utilised in data analysis.

2.9 Luciferase reporter assay

Wnt reporter assays were performed using Super TOP-FLASH (STF) HEK293 cells stably transfected with a luciferase reporter controlled by 7 Wnt-responsive LEF/TCF binding sites (Kakugawa et al., 2015). This enables signalling output to be measured as function of luciferase intensity following stimulation of cells with Wnt ligands alone and in the presence of binding partners. 100 μL of STF cells at a concentration of 20,000 cells mL^{-1} were seeded per well in a 96-well nunc plate (Thermo Fisher) and incubated at 37°C, 5% CO_2 for 24 hours. Media was removed and 50 μL of Wnt3a conditioned medium from L cells (Willert et al., 2003) was added, with and without ROR2 CRD-Kr or positive control protein sFRP4 up to a final volume of 100 μL. Cells were incubated at 37°C, 5% CO_2 for 48 hours and luciferase activity measured using the Dual-Glo luciferase reporter assay system (Promega).

2.10 Immunofluorescence microscopy

Immunofluorescence (IF) microscopy enables the subcellular localisation of target proteins. Antibodies to specific proteins tagged with a fluorescent dye stain the region in which a target protein is located. The confocal volume is visualised, excited using a laser and fluorescence is detected according to the emission spectrum of the dye in question. HEK 293T cells were seeded at a concentration of 10,000 cells mL^{-1} in poly-D-lysine coated 35 mm dishes (MatTek) and incubated at 37°C, 5% CO_2 for 18 hours. Media was exchanged to 2% DMEM and cells were transfected with 2 μg DNA of HA-tagged constructs using lipofectamine. 6 hours post-transfection, media was removed and replaced with 0.5% DMEM. Cells were then incubated for 2 days at 37°C, 5% CO_2. Media was removed and cells were washed with PBS before fixation for 15 minutes with 3% para-formaldehyde and quenching in 0.3 M glycine for 3 minutes. Fixed cells were stored at 4°C in PBS.

For the staining process, fixed cells were blocked for 10 minutes using 1% BSA/PBS, before HA tag probing using a Mouse HA Epitope Tag Antibody (Thermo fisher) at 1 μg mL^{-1} in 1% BSA/PBS at 25°C for 1 hour. Cells were washed in PBS 3 times for 10 minutes, before incubation with an Alexa Fluor® 633-conjugated Goat anti-Mouse IgG antibody (λ_{ex}: 633 nm; λ_{ex}: 647 nm) (Thermo Fisher) at a concentration of 2 μg mL^{-1} in 1% BSA/PBS for 1 hour at 25°C. Unbound antibody was removed via washing 3 times in PBS for 10 minutes, with one final wash containing 0.5 μg mL^{-1} Hoechst 33342 nuclear stain (λ_{ex}: 353 nm; λ_{ex}: 483 nm) (Chazotte, 2011).

Stained cells were kept light safe and stored in PBS at 4°C overnight. Immunofluorescence was detected using a Leica TCS SP8 WLL Confocal SMD Microscope. Images were processed in Fiji (Schindelin et al., 2012).

2.11 SAXS

2.11.1 SAXS background

Small-Angle X-ray Scattering (SAXS) is a technique complimentary to crystallography. It exploits X-ray scattering through a solution of macromolecules and gives information about flexibility, shape and size in the resolution range 50 – 10 Å. This can be used to verify particular assemblies seen in crystal structures or model absent domains (Rambo and Tainer, 2013). The χ^2 statistic is used to measure the discrepancy between experimental SAXS curves and SAXS curves calculated for models. A value of $\chi^2 = 1$ indicates a perfect fit (Pelikan et al., 2009).

2.11.2 Data collection and processing

SAXS data were collected on beamline B21 at Diamond Light Source. Measurements were taken over a range of $0.0044\ Å^{-1} < q < 0.04\ Å^{-1}$ at a wavelength of 1 Å. Readings were taken from sample in a capillary tube at a concentration of 1.3 mg mL^{-1} for the ROR2 ECD. Data were collected from glycosylated proteins. Buffer subtraction followed by data reduction and analysis was performed using the program SCATTER (Rambo and Tainer, 2013). A homology model was generated for the N-terminal Ig domain of ROR2 using the HHpred server

(https://toolkit.tuebingen.mpg.de/hhpred), corresponding to residues 59-151. This was manually placed alongside the crystal structure model for the CRD-Kr domains using COOT (Emsley and Cowtan, 2004) and missing residues added using MODELLER (Šali and Blundell, 1993). All-atom ensemble modelling (including branched N-glycans) of the ROR2 ECD, ROR2 CRD-Kr and the modelled Ig domain was performed using Allosmod (Weinkam et al., 2012) Automated selection of the most ideal ensemble of models in agreement with experimental SAXS data was performed using MES (Pelikan et al., 2009). Theoretical scattering profiles were calculated and fit to experimental data using FoXS (Schneidman-Duhovny et al., 2013), and the Allosmod-FoXS server was used to perform this process automatically (Guttman et al., 2013).

2.12 Analytical Ultracentrifugation

Sedimentation Velocity Analytical Ultracentrifugation (SV-AUC) provides information about the size of molecules in solution and enables the identification of low-affinity interactions not detectable by other techniques. This operates based on the principle that mass is redistributed in a gravitational field until gravitational potential and chemical energies equate. The distribution of protein molecules over time can be monitored by the optical properties of this solution (A_{280} and interference optics), and the change in concentration used to calculate sedimentation and

diffusion coefficients (s and D). The ratio s/D is related to the molecular mass of the species undergoing sedimentation (Cole et al., 2008).

HHIP-N and HHIP-C were dialysed into 10 mM HEPES pH 7.5, 120 mM NaCl prior to performing AUC experiments. SV-AUC experiments were performed at 20°C using a Beckman Optima XL-1 analytical ultracentrifuge (Beckman Instruments). For this, a scanning absorbance of 280 nm and interference optics were utilised. HHIP-N samples were spun at a concentration of 0.7 mg mL^{-1}, alone and with the addition of 0.09 mM heparin dp30 (Iduron). HHIP-C samples were spun at a concentration of 1 mg mL^{-1}, alone and with the addition of 0.16 mM heparin dp30 (Iduron). In both cases, the sugar was spun alone in a separate chamber at the concentration specified and analysed using interference optics. Samples were contained within 12 mm Epon sector-shaped two-channel centerpieces and spun at 400,000 rpm (An60Ti rotor, Beckman Coulter Inc., CA), with 80 sample distribution scans taken in 6 minute intervals, alongside interference optics. Data were analysed via the use of the program SedFit for the distribution of size and shape (Brown and Schuck, 2006). This enabled plots of c(s,ff0), where ff0 corresponds to the frictional coefficient ratio (1 for a perfect sphere; >1 for alternative molecular shapes). This enabled contour plots of c(s,M), where M corresponds to the molecular weight of the protein. A molecular weight distribution was calculated over a range of frictional coefficient ratios, which was ultimately selected based on a uniform distribution of molecular weights within this range. For the proteins in isolation, ff0 was set to a value of 1.0 – 2.0, whilst values spanning 1.0 – 2.5 were utilised for protein:heparin

complexes and heparin in isolation. As it was not possible to know the size or shape of individual protein:heparin species in solution, optimization of the ff0 ratio was required and is ultimately non-ideal due to the result of both self- and heparin-induced oligomerisation to extremely large sizes. In all cases, a partial specific volume value of 0.73 mL g^{-1} was used.

2.13 Heparin-agarose chromatography

COS-7 cells transfected with HA-tagged HHIP constructs were lysed in heparin column buffer (50 mM Tris-HCl pH 7.2, 150 mM NaCl, and 5 mM EDTA) with 1% NP-40 containing complete mini protease inhibitor cocktail (Roche). The soluble fraction was separated by centrifugation at 16000 xg for 30 min at 4°C. Heparin-agarose (2 mL) (Sigma-Aldrich) was equilibrated with 10 ml of column buffer. Cleared lysates were diluted 1:25 in heparin column buffer and bound to the column. Heparin-agarose columns were washed with 10 ml of heparin column buffer, and proteins eluted in a step gradient (4 ml per elution) of column buffer supplemented with increasing NaCl concentrations. Dot blot analysis was used to analyse eluates for the presence of HA-tagged HHIP proteins. Dot intensities were quantified using FIJI (Schindelin et al., 2012), and plotted as the relative signal level compared with the total intensity measured across all eluted fractions.

Chapter 3

The structure of the HHIP N-terminal CRD

Chapter 3: The structure of the HHIP N-terminal CRD

3.1 Summary

The CRD is a domain found in a wide range of proteins, possessing a host of different functions. The function of the HHIP N-terminal CRD (HHIP-N) remains elusive, with the C-terminal region (HHIP-C) functioning to bind and sequester HhN ligands. Here, the structure of HHIP-N was solved via the use of multi-crystal averaging and Sulphur-SAD methods. The structure of HHIP-N displays a novel bipartite fold with distinct N-terminal and C-terminal sub-domains held together by 6 disulphide bonds. The N-terminal sub-domain is observed to bind to GAG mimic SOS, whilst the C-terminal region possesses a 'classical' CRD fold composed of 5 core α-helices. Structural phylogenetic analysis identifies HHIP-N as a pocket-type CRD, with similarities to FRα/β, JUNO, RFBP and NPC1. Additionally, two disordered loops line the central pocket of HHIP-N, and structural comparisons to related CRDs suggest a role of these loops in forming a ligand-binding pocket. Small molecules have a role in much of the cognate functions of CRDs and the notion of sterols influencing Hh signalling is well-known. The structure of HHIP-N therefore suggests a role in small molecule recognition, as well as having evolved a specific cell surface binding sub-domain previously unseen in CRDs.

3.2 Introduction

The Hedgehog-Interacting Protein (HHIP) is a negative feedback inhibitor of Hh signalling with roles in the development of the lung (Chuang, 2003), cartilage (Chuang and McMahon, 1999) and brain (Wilson and Stoeckli, 2013), as well as being a biomarker in the prognosis of several cancers (Evangelista et al., 2006). HHIP has been shown to bind to HhN ligands to antagonise Hh signalling (Bishop et al., 2009). Knockouts of HHIP display defects in the endochondral skeleton, lungs, pancreas and stomach, and mutant mice die shortly following birth due to respiratory defects (Chuang, 2003). Lung defects are linked to hyper-activated Hh signalling, whilst defects in the pancreas, skeleton and stomach are due to impaired growth (Chuang, 2003; Kawahira, 2003). Additionally, HHIP expression in commissural axons crossing the spinal chord converts ShhN from a chemoattractant to a chemorepellent via competing with Ptc for HhN interaction (Bourikas et al., 2005; Wilson and Stoeckli, 2013). Thus, HHIP is involved in the regulation of cellular proliferation and differentiation induced by Hh signalling and this activity has been mapped to an interaction between HHIP-C and HhN (see Fig. 1.15B-D and section 4.2) (Bishop et al., 2009; Bosanac et al., 2009).

The function of HHIP in terms of Hh signal antagonism has thus far only been considered within the context of HHIP-C and there is a general lack of information about the function of HHIP-N. This is most likely due to the difficulty in purification of a full-length version of HHIP, as the linker region between HHIP-N and HHIP-C

is prone to degradation (Bishop et al., 2009). As discussed in sections 1.4.5 and 1.6, CRDs have ties to a number of small molecule-binding functions and Hh signalling is subject to stringent control by sterols and cholesterol (Byrne et al., 2016; Nachtergaele et al., 2013). To add to this, lipid modifications upon the HhN ligand may undergo physical interaction with a CRD, an interaction seen to occur between palmitoleated Wnt and the CRD of receptor FZD (Janda et al., 2012). Therefore, a potential role of the HHIP-N CRD may be to add an additional HhN binding site, or to regulate the availability of molecules involved in Hh signalling activation. To gain insight into the functional role of HHIP-N, the protein was purified, crystallised and structurally characterised.

3.3 Results

3.3.1 Purification and crystallisation of HHIP-N

A construct comprising HHIP residues 39-209 was identified as the best candidate for large-scale expression in HEK293T cells, and could be readily purified (Fig. 3.1A-C) (B. Bishop, personal communication). Crystals were grown from SEC-purified protein, but these diffracted poorly. Sucrose Octasulphate (SOS) is a mimic of sulphated cell surface GAG molecules and can interact with solvent-exposed basic residues to facilitate crystallisation (Chang et al., 2015; Kakugawa et al., 2015). Improved crystals were grown upon the addition of SOS, but were difficult to mount and also to reproduce (Fig 3.1D-E).

It was reasoned, based on the change in crystal morphology seen upon the addition of SOS, that HHIP-N binds to GAGs. It was also reasoned that the issues with crystal growth were due to degradation of a basic residue-rich region at the N-terminus, and an additional heparin column purification step was added (Fig. 3.1F). Purification via this method gave two separate peaks with the removal of lower molecular weight contaminant in the higher-affinity peak (Fig. 3.1G; arrow). Crystallisation of this peak with and without SOS gave crystals which grew and diffracted reproducibly, and could be more easily harvested (Fig. 3.1H-I).

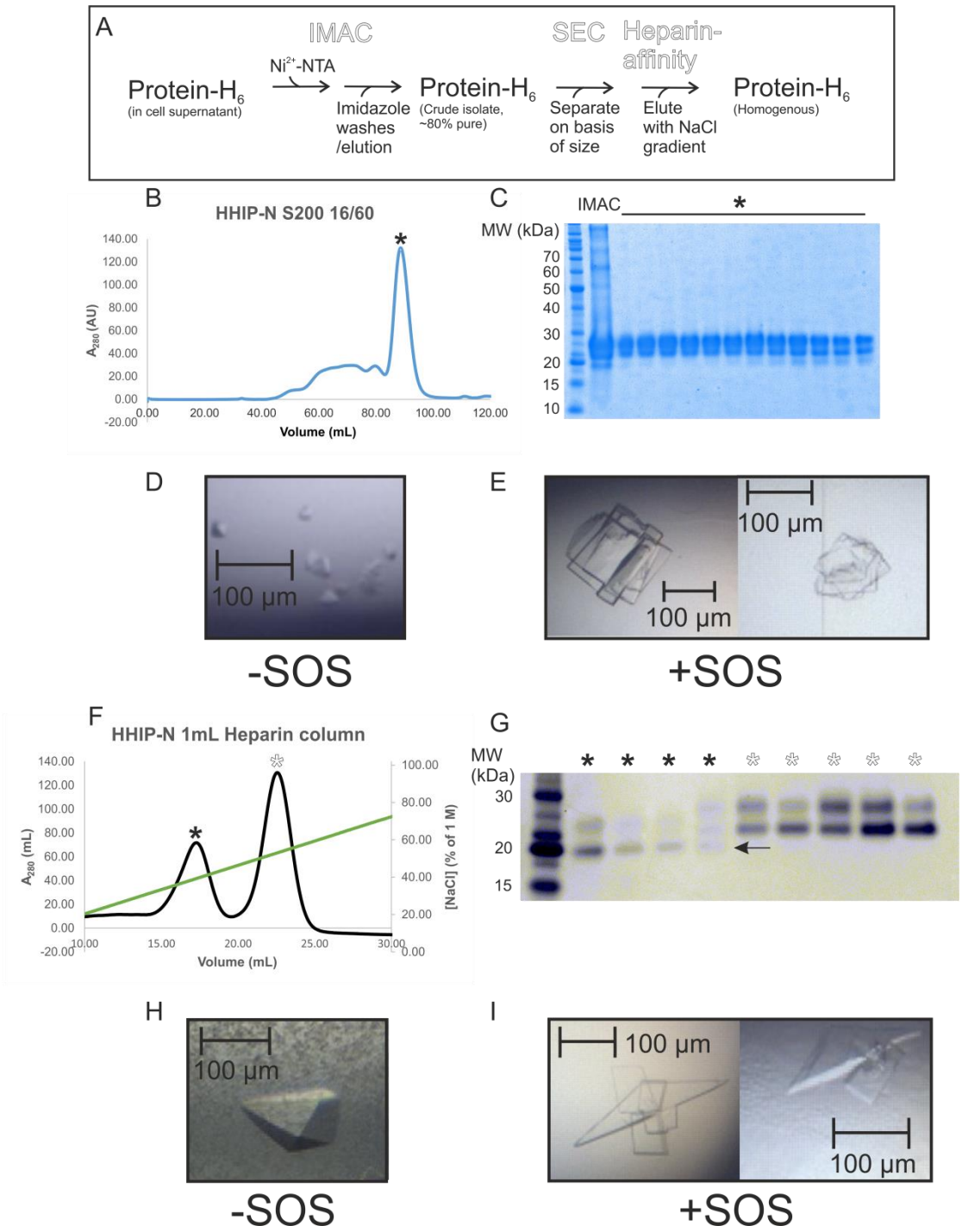


Figure 3.1 – Purification and crystallisation of HHIP-N. A) Purification strategy (see section 2.3.3 for acronyms and methods); B-E) Purification via SEC and crystallisation trials; F-I) Addition of a heparin column purification step following SEC and subsequent crystallisation trials; B) SEC UV trace; C) SDS-PAGE analysis of major SEC peak; D) Crystallisation trials without SOS; E) Crystallisation trials with SOS; F) Post-SEC heparin column UV trace; G) SDS-PAGE analysis purification process from IMAC, SEC and heparin column. Black and white asterisks refer to the indicated heparin column peaks in F). An arrow is placed next to a contaminant which is removed from the second peak (white asterisk); H) Crystal grown from the second heparin column peak in the absence of SOS; I) Crystal grown from the second heparin column peak in the presence of SOS.

3.3.2 HHIP-N data collection

Native data were collected from crystals of HHIP-N:SOS and HHIP-N-apo, extending to resolutions of 2.75 Å and 2.6 Å, respectively (Table 3.1). However, it was not possible to determine either structure by molecular replacement methods. This involved the use of homology models computed from the structures of other CRDs (FRα and NPC1) produced using the HHpred server (https://toolkit.tuebingen.mpg.de/hhpred) and *ab initio* modelling with Rosetta (Raman et al., 2009) and AMPLE (Bibby et al., 2012). Normal mode analysis of computed models was also attempted using elNEMO (Suhre and Sanejouand, 2004). Heavy metal soaks resulted in a decline in diffraction quality, and attempts to crystallise the full-length protein (minus the C-terminal 30 residues) with a stabilised domain linker resulted in crystals of HHIP-C only. Consequently, a native Sulphur-SAD (S-SAD) approach was utilised. The presence of 12 cysteines in 6 disulphide bonds in a 21.1 kDa protein gives a predicted Bijvoet ratio of 1.7% when using the wavelength selected. The Bijvoet ratio is a measure of the expected anomalous contribution to the total scattering, and a theoretical value can be calculated using the equation $(|\Delta F^{\pm}|)/(|F|) = (2N_A/N_P)^{1/2} (f''/Z_{eff})$, where N_A = number of anomalous scattering atoms, N_P = number of non-hydrogen atoms in the ASU, f'' = imaginary scattering contribution of anomalous atom and Z_{eff} = effective number of electrons per average protein atom (6.7). S-SAD was therefore a viable option for an experimental phasing strategy (Weinert et al., 2014).

S-SAD data was collected using both mini-kappa multi-axis goniometry (I04) and inverse beam (I03) techniques. The collection of highly multiplicity data is paramount when utilising a wavelength of 1.7712 Å (f'' = 0.7 elections) (Liu et al., 2011), and to this effect 33 data sets were collected from a total of 12 crystals. Data were collected from different regions of the same crystal (i.e. via collecting via the rotation method, translating in the X-direction to a new region of the crystal and repeating) and also from multiple sweeps of the same region of a crystal until radiation damage was witnessed via a drop in the number of Bragg peaks detected per image (Fig. 3.2A).

Inclusion of data with a weaker anomalous signal can 'poison' the resulting merged data and complicate structure solution (Akey et al., 2014). Individual datasets were each processed individually to decide which images to merge from all datasets in order to maximise the anomalous signal. Plots of resolution and diffraction maxima detected for each dataset versus image number were analysed to identify images to remove. This is illustrated by Fig. 3.2A, where the dashed oval corresponds to a region that would not be included in processing of this particular dataset.

The CC_{anom} and SigAno statistics were used to judge which images to include in the final merging process (Fig. 3.2B and C; both statistics are defined in the figure legend). Based on both criteria, for the 'accepted' dataset, an anomalous signal is detected to ~4.6 Å based on the CC_{anom}, whilst a signal is detected to ~4.0 Å using

SigAno as a data quality indicator. For the 'rejected' dataset, neither criterion could detect anomalous signal to a resolution higher than 5 Å. Also, the smoother drop of CC_{anom} and SigAno with resolution down to the anomalous signal limit for the 'accepted' dataset is not witnessed for the 'rejected' dataset, which is indicative of numerous poor resolution shells for the 'rejected' dataset. Data merged from a single dataset which did not give an anomalous signal according to these criteria above a resolution of 5 Å was removed from the final merging process.

For the final merging process, XIA2 was used (Winter, 2010). Non-isomorphism was not an issue, and BLEND was not required (Foadi et al., 2013). A total of 24 data sets from 8 crystals were used, giving a final multiplicity of 222.4 and an anomalous slope of 1.03 (Table 3.1).

Table 3.1

	HHIP-N:SOS S-SAD	HHIP-N:SOS Native	HHIP-N-apo Native
Data collection			
Beamline	I03/I04, DLS	I03, DLS	I24, DLS
Space group	$P4_22_12$	$P4_22_12$	$P4_32_12$
Unit-cell parameters			
a, b, c (Å)	59.8, 59.8,108.7	60.4, 60.4, 108.5	59.1, 59.1, 98.7
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. of crystals / data sets	8 / 24	1 / 1	1 / 1
Wavelength (Å)	1.7712	0.9763	0.9686
Resolution (Å)	108.7-2.9 (3.0-2.9)	54.2-2.75 (2.85-2.75)	50.7-2.6 (2.7-2.6)
No. of unique reflections	4801 (345)	5609 (534)	5640 (420)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Multiplicity	222.4 (20.9)	11.9 (10.9)	12.3 (12.3)
$I/\sigma(I)$	41.7 (2.0)	12.0 (1.1)	22.3 (1.4)
R_{merge} (%)	19.8 (137.3)	11.3 (182.7)	5.8 (180.7)
R_{pim} (%)	1.2 (30.9)	3.5 (57.4)	2.4 (75.2)
$CC_{1/2}$	1.0 (0.8)	1.0 (0.4)	1.0 (0.5)
Anomalous slope	1.03 (0.0)		

Table 3.1 – Data collection statistics for HHIP-N:SOS (SSAD and native) and HHIP-N-apo. Values in parenthesis correspond to the highest resolution shell.

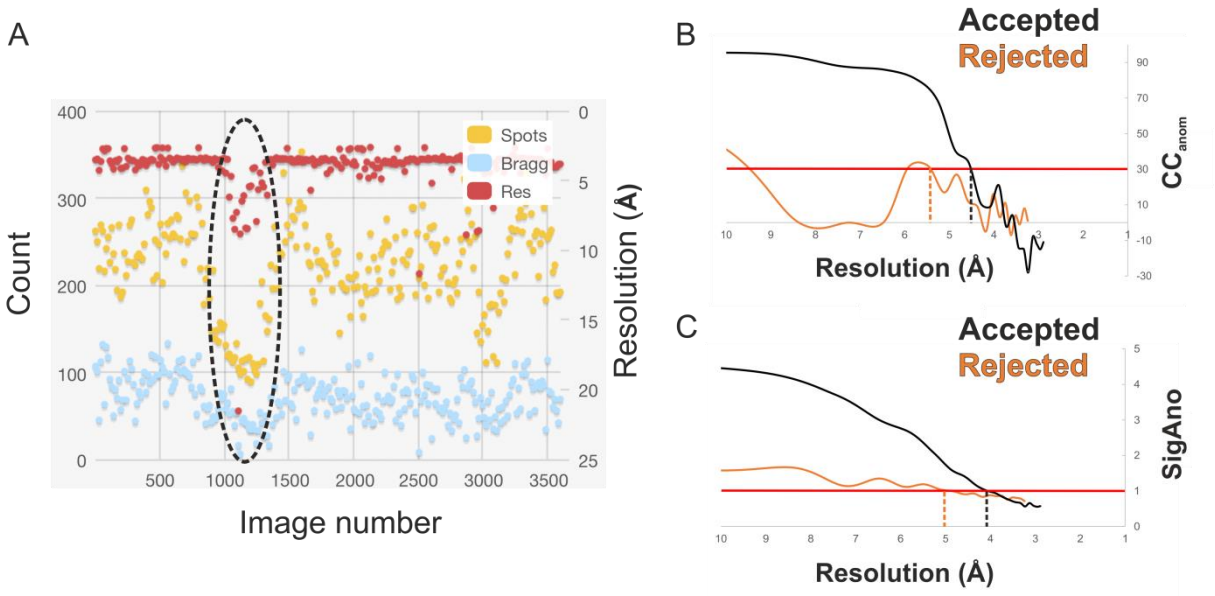


Figure 3.2 – HHIP-N S-SAD data collection and processing. A) A graph of image number (X-axis) versus the number of spots detected (yellow) and Bragg peaks (blue) on the primary Y-axis, with resolution (red spots) on the secondary Y-axis. The dashed oval indicates a region that will be removed from the final data processing strategy; B) Graph of resolution vs. CC_{anom} from the XSCALE log file; C) Graph of resolution vs. SigAno from the XSCALE log file. In both B) and C) a red line indicates the limit for detected anomalous signal (30 for the CC_{anom}, 1.0 for the SigAno). Colour-coded dashed lines corresponding to either data set indicate the limit to which such a signal is detected for either dataset. CC_{anom} – % correlation between random half-sets of anomalous intensity differences. SigAno - $(|F_{(+)}-F_{(-)}|/\sigma$; mean anomalous difference versus standard deviation of anomalous intensity.

3.3.3 Structure solution

An initial substructure for HHIP-N was solved using the SHELX program suite (Sheldrick, 2010). SHELXC indicated the presence of an anomalous signal to 4 Å and a signal of CC_{weak}/CC_{all} of 23.8/42.4 was given via searching for 6 S atoms in 6 disulphides in SHELXD (Fig. 3.3). Chain tracing and density modification was not possible using the associated program SHELXE and subsequently the site locations file and experimental data were used as input files for the PHENIX AutoSol wizard (Terwilliger et al., 2009). This gave a FOM (figure of merit, the mean phase error

(Rupp, 2009)) of 0.34 and the resulting map was automatically traced (Fig. 3.4A). This process was repeated using the traced model, resulting in an improvement of the FOM to 0.63 (Fig. 3.4B). Following manual rebuilding and further rigid-body refinement, a model with R_{work}/R_{free} of 31.1%/35.1% was obtained. At this point in refinement unambiguous extra electron density for a SOS molecule became visible (Fig. 3.4C). Cycles of refinement resulted in reduction of R_{work}/R_{free} to 25.8%/27.6% in the final model (Fig. 3.4D and Table 3.2).

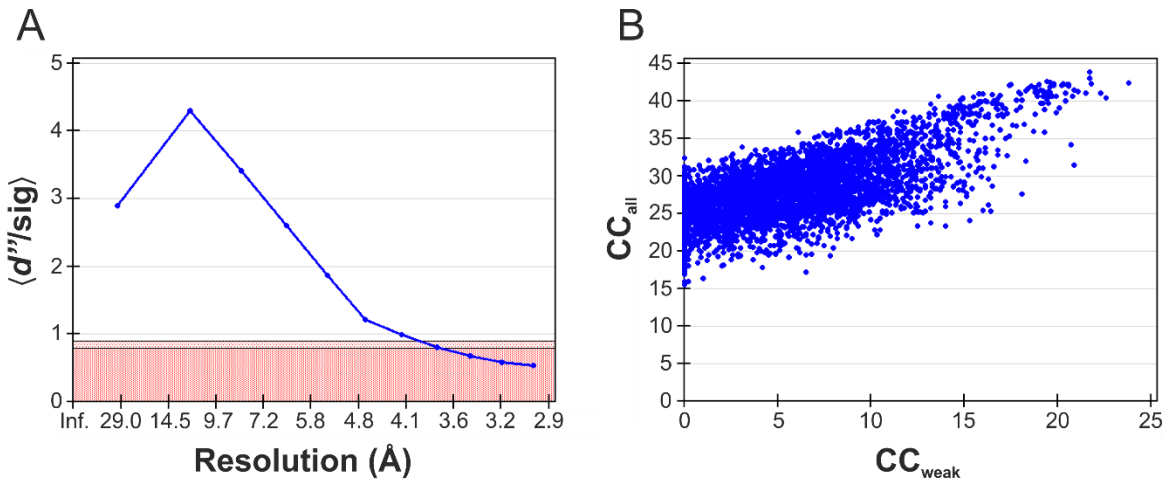


Figure 3.3 – SHELX processing of HHIP-N:SOS S-SAD data. A) Output from SHELXC, with ⟨*d''*/sig⟩ plotted as a function of resolution, indicating the signal to noise of the anomalous differences. The red portion of the graph displays the level at which the anomalous signal is considered non-existent; B) Output from SHELXD, displaying the contrast between observed and calculated Bijvoet pair differences. CC_{all} = Correlation coefficient between anomalous differences calculated from the anomalous substructure and the experimental anomalous differences; CC_{weak} = As for CC_{all}, but for 30% of reflections which were not used in this process.

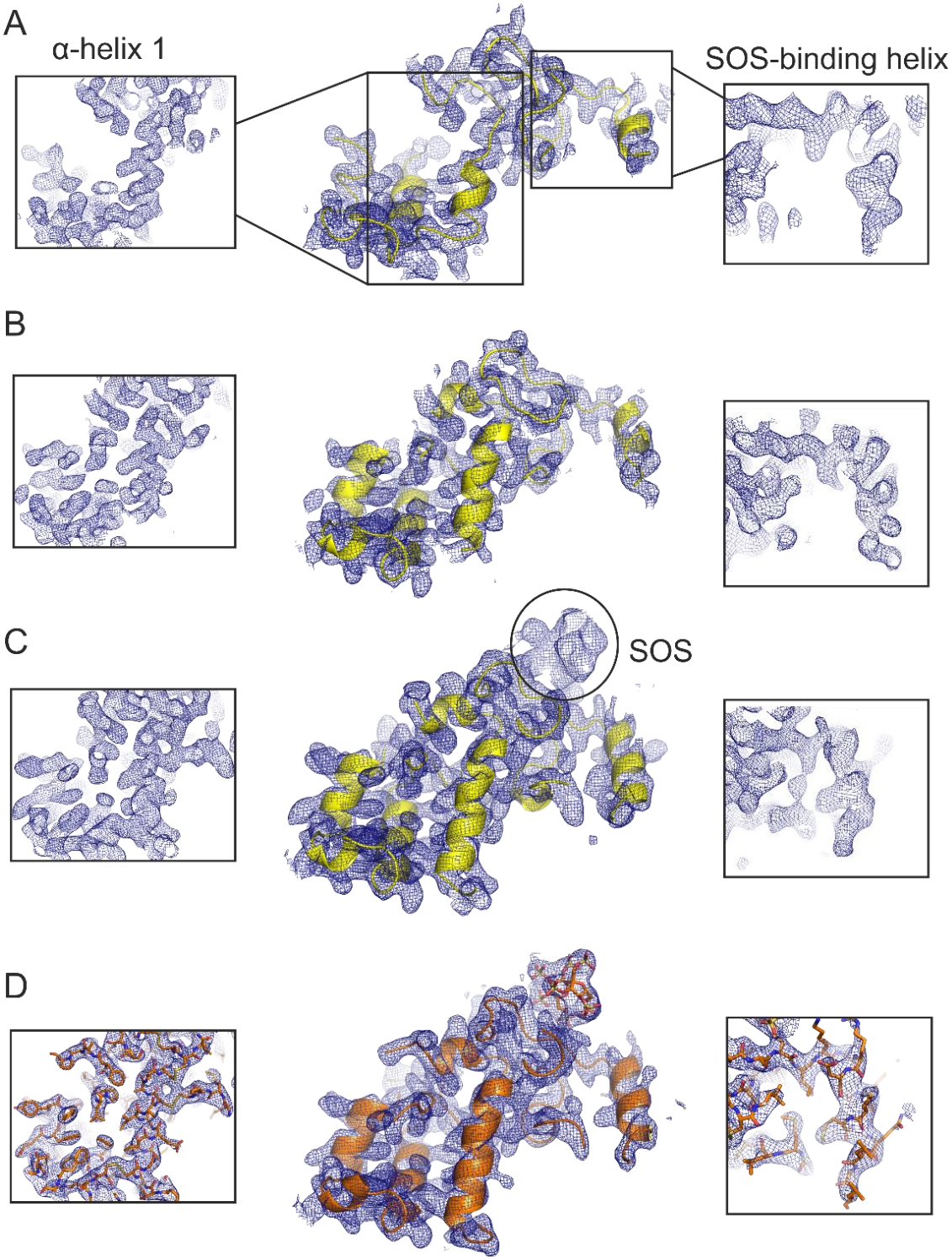

A
α-helix 1
SOS-binding helix
B
C
SOS
D

Figure 3.4 – Improvement of HHIP-N:SOS electron density maps (above). Blue meshes show $2F_o$-Fc electron density contoured at 1σ, and models for each step are displayed**.** Two well-resolved features of HHIP-N, the first α-helix of the CRD (α-helix 1) (left) and the SOS-binding helix (right), are displayed either side of the model from structure solution process at each point (centre); A) Output from the first cycle of phasing/autobuilding cycle; B) Second cycle of phasing/autobuilding. Both A) and B) were extended to 2.9 Å during model building; C) Second cycle of refinement in *Refmac5*, with electron density for a SOS molecule visible for the first time; D) Final refined model and $2F_o$-Fc map. Both C) and D) were extended to 2.75 Å resolution and the final refined model is coloured orange versus yellow for partially built models.

3.3.4 Structure of HHIP-N

The structure of HHIP-N:SOS displays an elongated globular fold, comprising an α-helical motif common to all CRDs that is stabilised by 6 disulphide bonds (Fig. 3.5A). The overall fold of HHIP-N is bipartite, with two distinct sub-domains observed in the form of a C-terminal CRD region and an N-terminal GAG-binding domain. The N-terminal region contains a single helix and is stabilised by an intra-sub-domain disulphide bond (C39-78) and 2 inter-sub-domain disulphide bonds with α1 of the CRD region (C69-C112 and C79-115). The CRD region is composed of 5 helices (α1-α5) stabilised by 3 intra-CRD region disulphide bonds (C103-C152, C141-170 and C144-168). Two disordered loops (17 and 6 residues, respectively) called DL1 and DL2 (Fig. 3.5A) join the N- and C-terminal regions and connect α2 and α3 in the CRD, respectively. Surface electrostatics display a contiguous region of positive charge on the top of the GAG-binding domain, to which three SOS molecules are coordinated (Fig. 3.5B). HHIP-N crystallises as a monomer with one SOS molecule bound in the asymmetric unit (ASU), while the other two shown are contributed by

crystal contacts from symmetry-related molecules upon the 4-fold axis. Superposition of HHIP-N:SOS and HHIP-N-apo crystal structures according to Cα backbone RMSD highlights the GAG-binding sub-domain as being conformationally divergent between the two states, whilst the CRD region is largely conformationally conserved (Fig. 3.5C) The refinement statistics for both HHIP-N:SOS and HHIP-N-apo are displayed in Table 3.2.

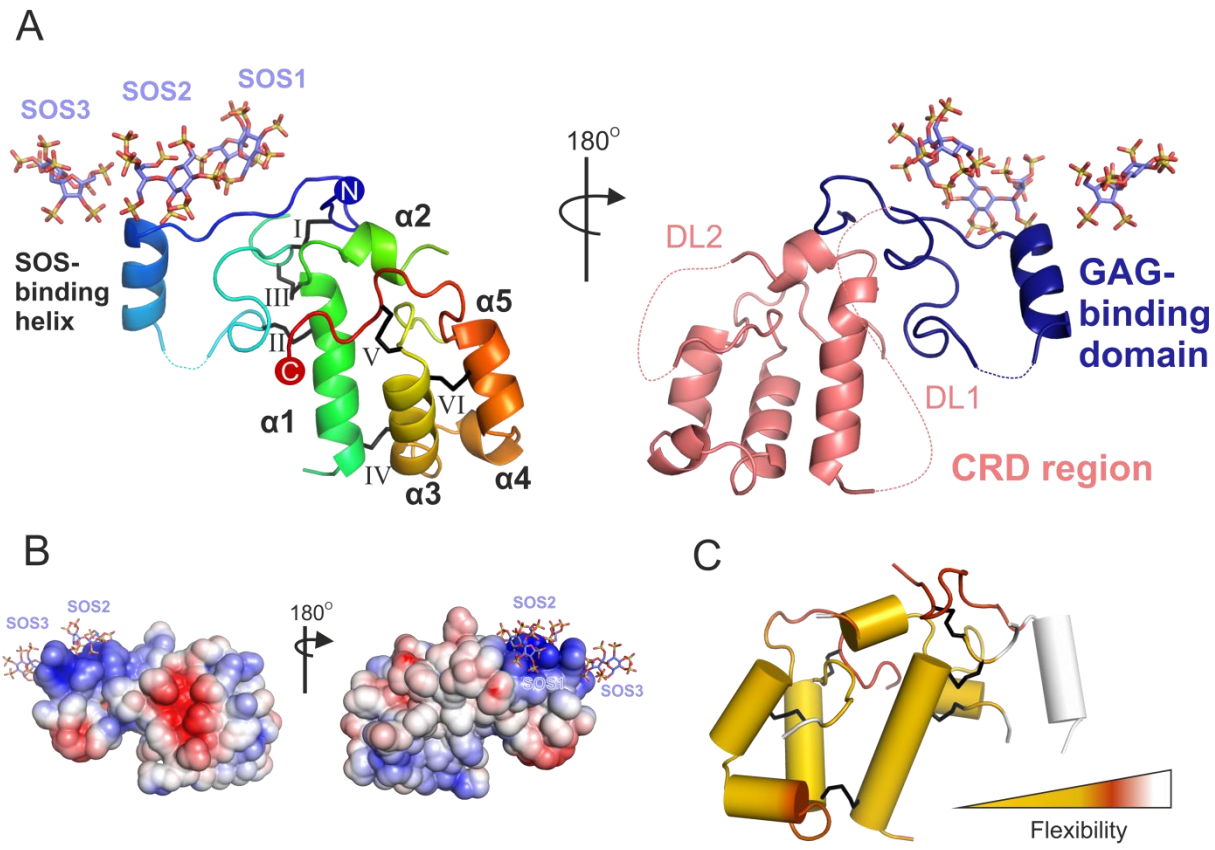


Table 3.2

	HHIP-N:SOS	HHIP-N-apo
Refinement		
Resolution	54.0-2.75	50.7-2.6
No. reflections (test set)	5609 (288)	5640 (281)
R_{work}/R_{free} (%)	26.7/28.6	24.7/26.1
No. atoms		
Protein	964	807
SOS	55	
Mean B factor (Å^2)		
Protein	97.0	105.4
SOS	195.2	
RMSD bond lengths (Å)	0.007	0.002
RMSD bond angles (°)	0.90	0.55
Ramachandran plot (%)		
Favoured	97.3	97.9
Allowed	2.7	2.1
Outliers	0	0

Figure 3.5 and Table 3.2 – Structure of HHIP-N and associated refinement statistics (above). A) The structure of the HHIP-N:SOS complex shown in two orientations. Helices in the CRD region are numbered α1-5 and disulphide bonds are numbered with Roman numerals. 3 SOS molecules are numbered 1-3, alongside a helix which becomes ordered upon SOS binding (left). The two sub-domains are coloured blue (N-terminal) and salmon (C-terminal). Disordered loops are drawn as dotted lines and labelled DL1 and DL2 (right); B) Electrostatic surface potential of the HHIP-N:SOS complex, calculated using APBS and shown from -5 KT/e to +5 KT/e (Blue: positive charge; red: negative charge). C) Structural changes to HHIP-N upon SOS binding, analysed using ProSMART (Nicholls et al., 2014). The HHIP-N:SOS complex structure is coloured yellow to red according to the Cα RMSD of the best-matching aligned fragment pair between HHIP-N:SOS and HHIP-N-apo, with white corresponding to regions that cannot be aligned and adopt different conformations.

3.3.5 Structure analysis

Structural characterisation of HHIP-N reveals a distinct folded sub-domain functionally linked to interactions with the cell surface GAG layer and HHIP secretion which has been demonstrated using biochemical techniques (see Chapter 4). A classical CRD fold is present and structure-based evolutionary analysis identifies the closest homologues as NPC1, RFBP, FRα/β and JUNO. This analysis places the HHIP-N CRD on the same branch of the tree as the 'pocket-type' CRDs in Fig. 3.6A (Nachtergaele et al., 2013). Structural superposition of HHIP-N with other CRDs displays the presence of various secondary structural elements involved in ligand recognition which occupy the same location as disordered loops DL1 and DL2 of HHIP-N is predicted to occupy (Figs. 3.6B-G). In each case, a central helix aligned with α1 of HHIP-N is contiguous in sequence to a ligand-binding element as also seen for DL1 of HHIP-N, alongside four other helices in the case of the 'pocket-type'

CRDs (Figs. 3.6B-E). The N-terminal GAG-binding region of HHIP-N does not superpose with any element of other CRD structures. HHIP-N is well aligned with FRβ (RMSD = 1.55 Å for 66 Cα positions), with two of the three loops involved in folate ligand binding occupying similar regions in space as the general locations of DL1 and DL2 (Fig. 3.6B). The arrangement of five helices is also similar in the NPC1 CRD (RMSD = 1.99 for 61 Cα positions), along with the inclusion of a long β-pleated sheet. This sheet insertion aligns in sequence with DL2 of HHIP-N and is occupies a similar region in space, but does not contact the cholesterol ligand. A loop which does not align in sequence with HHIP-N DL2 also occupies a similar position as DL2 would be expected to, and this does make contacts with cholesterol. A long helix projecting toward α1 (labelled) makes several hydrophobic contacts with the cholesterol ligand, and can be thought of comparable to DL1 (Fig. 3.6C). Structural superposition of HHIP-N and RFBP (RMSD = 1.53 for 71 Cα positions) exhibits a very similar situation as for the comparison between HHIP-N and FRβ, with comparable ligand-binding elements of RFBP occupying the same regions as DL1 and DL2 of HHIP-N (Fig. 3.6D). Sperm-egg fusion protein JUNO, also known as FRδ (Han et al., 2016), exhibits a very similar fold to that of other folate receptors (RMSD = 0.42 and 0.46 for 161 and 157 Cα positions for FRα and FRβ respectively). The loops that FR α/β utilise in folate binding are partially conserved within JUNO, which has evolved the use of these in protein-protein recognition at the expense of small molecule binding (Han et al., 2016). JUNO is relatively well superposed with HHIP-N (RMSD = 1.54 for 63 Cα positions) with five core CRD helices. Loops involved in IZUMO

ligand binding occupy structural similar positions to DL1 and DL2 (Fig. 3.6E). Sequence alignment of HHIP-N to other pocket-like CRDs exhibits alignment between DL1 and DL2 and the ligand-binding elements of these family members (Fig. 3.7A).

Two CRDs observed to bind small molecules, Frizzled 8 (Fz8) and Smoothened (Smo), occupy the 'groove-type' arm of the CRD structural phylogeny in Fig. 3.6A (Byrne et al., 2016; Janda et al., 2012). Both CRDs do not have the same arrangement of core helices as the pocket-type family members analysed, but do exhibit a long helix structurally analogous to HHIP-N α1. This is linked to a long helix involved in the recognition of cognate ligands, which occupies a region comparable with the potential position of DL1 in HHIP-N. No ligand-binding sequence region comparable to DL2 is visible in these groove-type CRDs (Figs. 3.6F and G). Ligand-binding elements of both groove-type CRDs align well in sequence with DL1 of HHIP-N in either case (Fig. 3.7B). PDB accession codes and RMSD values between equivalent Cα positions are given in Supplementary Table 3.1. Taken together, evolutionary analysis suggests that the CRD region of HHIP-N may be involved in the binding of some currently unknown small molecule.

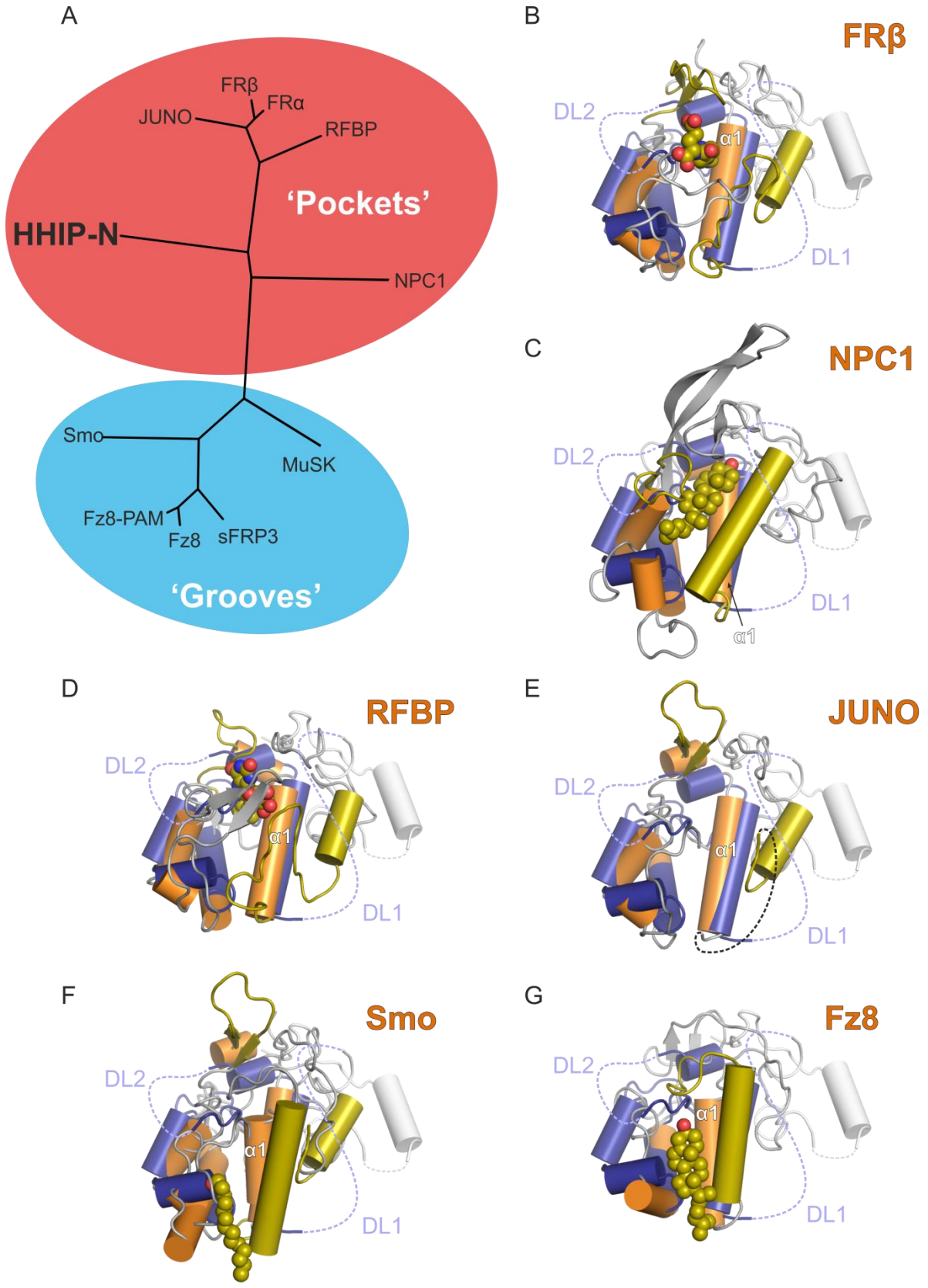

A
FRβ
FRα
JUNO
RFBP
'Pockets'
HHIP-N
NPC1
Smo
MuSK
Fz8-PAM
Fz8
sFRP3
'Grooves'
B
FRβ
DL2
α1
DL1
C
NPC1
DL2
DL1
α1
D
RFBP
DL2
α1
DL1
E
JUNO
DL2
α1
DL1
F
Smo
DL2
α1
DL1
G
Fz8
DL2
α1
DL1

Figure 3.6 – Structural analysis of HHIP-N (above). A) Structural phylogenetic tree of all published CRD structures including HHIP-N; B-G) Structural superposition of HHIP-N with indicated CRD family members. The aligned CRD region of HHIP-N is displayed in dark blue, with the non-superposed N-terminal GAG-binding region in white. The disordered loops DL1 and DL2 of HHIP-N are displayed as dashed purple lines. A grey dashed line in the GAG-binding region corresponds to disordered residues 61-65. Helices from aligned CRDs are displayed in orange and loops in grey, with ligand-binding elements which align with the expected location of DL1 and DL2 of HHIP-N coloured in mustard yellow. Other structural elements of superposed CRDs not aligned with HHIP-N are displayed in dark grey. Bound ligands are shown as spheres. CRD helices superposed with the long helix α1 of HHIP-N are indicated. In the case of JUNO (E), disordered residues 71-83 are displayed as a dashed black line. Only the CRD regions superposed with HHIP-N are displayed (see supplementary figure 3.1).

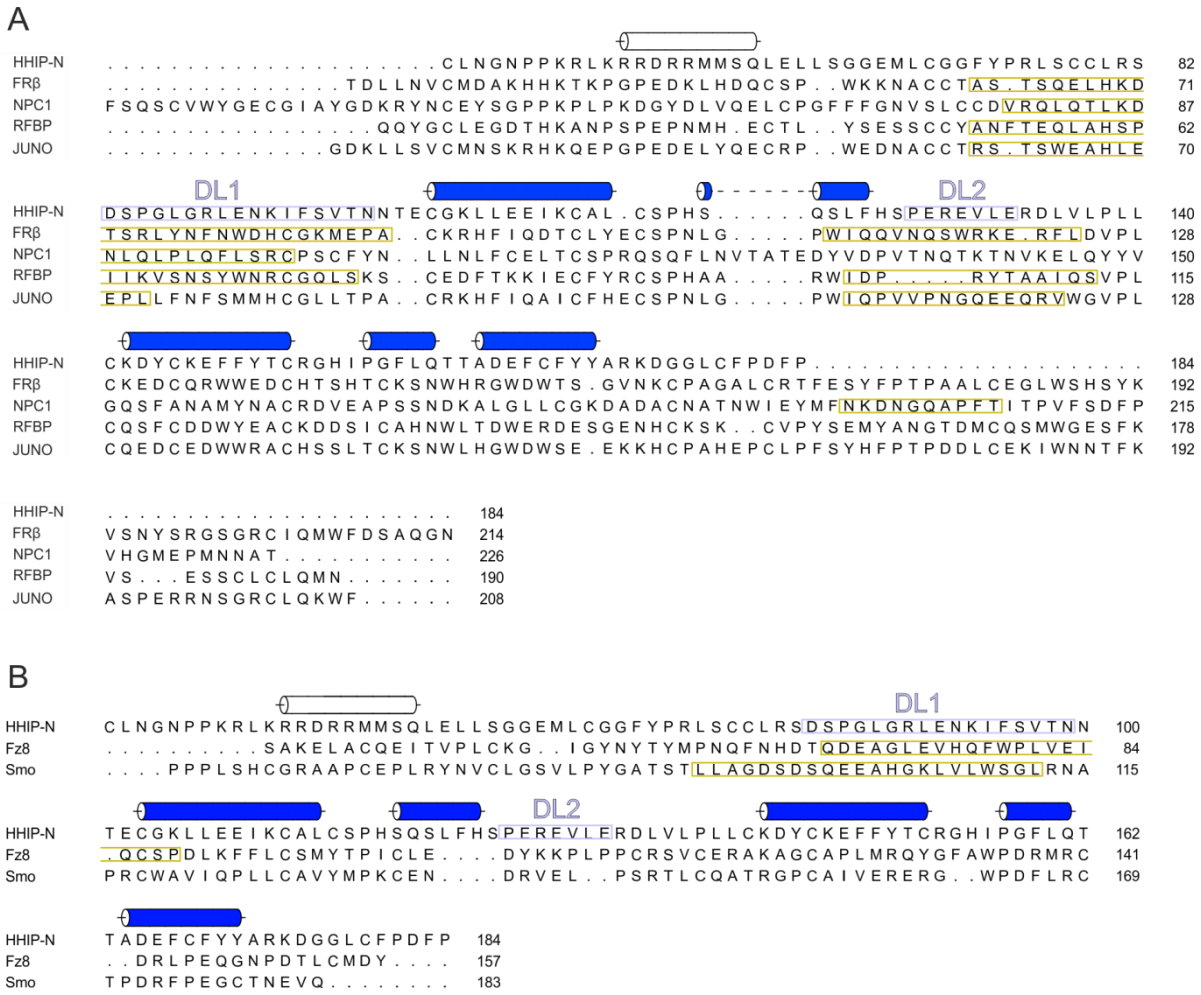


Figure 3.7 – Sequence alignment of CRDs analysed in Figure 3.6. A) Pocket-type CRDs; B) Groove-type CRDs. HHIP-N helices are coloured as in Fig. 3B-G. DL1 and DL2 are outlined using light purple boxes. Loops from related CRDs involved in ligand recognition that are structurally superposed with DL1 and DL2 are outlined in yellow boxes.

3.3.6 Modelling

From the structures of related CRDs (Fig. 3.6) it was suspected that DL1 and DL2 may play a role in the binding of a small molecule binding partner. Mass spectrographic analysis of purified HHIP-N did not identify any candidate small molecules. This meant that there were no rational choices to analyse via binding experiments (see section 3.4). Similar to the cholesterol-binding groove of Smo, a hydrophobic surface region is observed in HHIP-N in the structures solved in this study which would potentially be enclosed by DL1 and DL2 (3.8A-B). DL1 and DL2 were modelled using comparative loop modelling via the program *Modeller* (Šali and Blundell, 1993). *Modeller* uses a database of known protein structures as a template for restraints to produce three-dimensional structural models for residues missing from structures due to disorder. The lowest energy structure which proposes a ligand-binding pocket shows that modelling of DL1 and DL2 partially encloses this hydrophobic patch, as is likely to be seen in certain solution states due to loop flexibility (Fig. 3.8C-D).

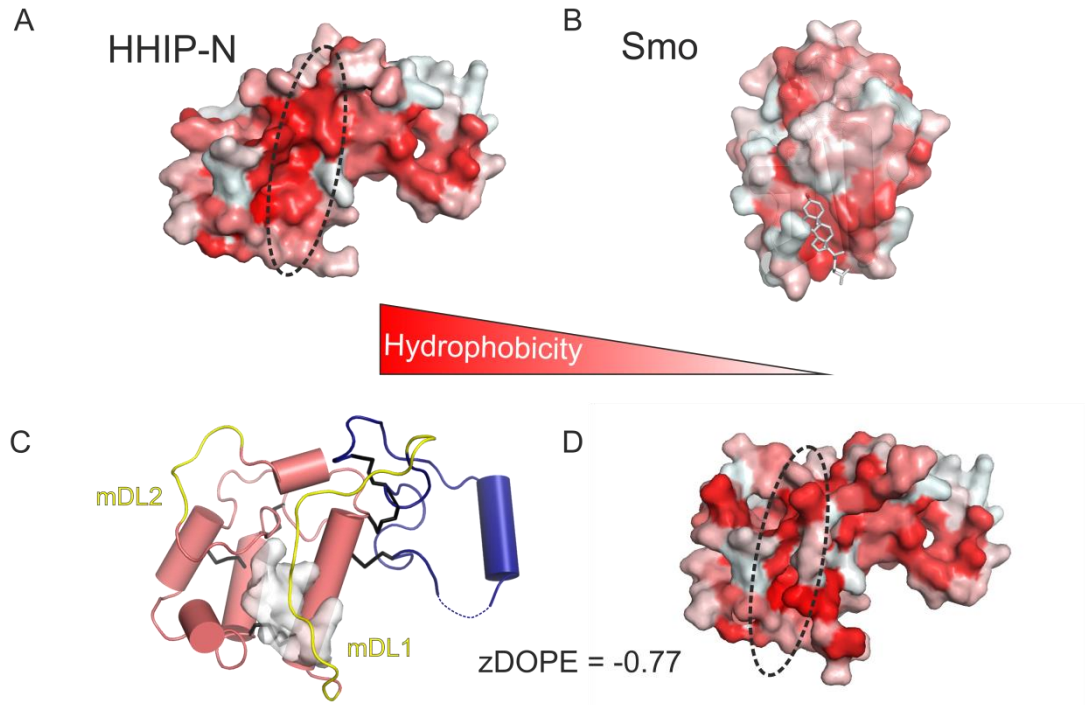


Figure 3.8 – Modelling the potential small molecule-binding pocket of HHIP-N. A) Surface analysis of HHIP-N reveals a large hydrophobic region; B) Hydrophobic surface representation of Smo, showing a comparable region of higher hydrophobicity comprising the cholesterol-binding groove; C) Modelling of HHIP-N DL1 and DL2 using *Modeller* (mDL1 and mDL2) produces a potential binding pocket enclosed by DL1 and the hydrophobic region from A). From the five models produced by *Modeller*, this model had the lowest zDOPE (normalised Discrete Optimised Protein Energy) score and a visible pocket. The zDOPE score is a probability score and represents an atomic distance-dependent statistical potential. A negative value indicates a favourable conformation (Shen and Sali, 2006); D) Surface representation of HHIP-N with mDL1 and mDL2 showing folding of these loops over the hydrophobic region from A) and a reduction in the exposed hydrophobic surface area.

Several ligands were docked into this proposed model of HHIP-N using *AutoDock Vina*, as well as into FRβ and the hSMO CRD (Trott and Olson, 2009). This predicts the docking of cholesterol and cyclopamine into the modelled HHIP-N pocket with a favourable binding energy (Fig. 3.9A and Table 3.3). Cognate ligands were also docked into FRβ and Smo as a control experiment, which predicts an almost completely correct docking mode (Fig. 3.9B and C). However, as Table 3.3

shows, other bulky ligands including haem and methotrexate could also be docked into all models (including HHIP-N without modelling of DL1 and DL2) with favourable binding energies. Therefore, these results should be interpreted no further than suggesting the likelihood of small molecule binding by HHIP-N, alongside the discussions of structural and sequence alignment with structurally characterised CRDs in section 3.3.5.

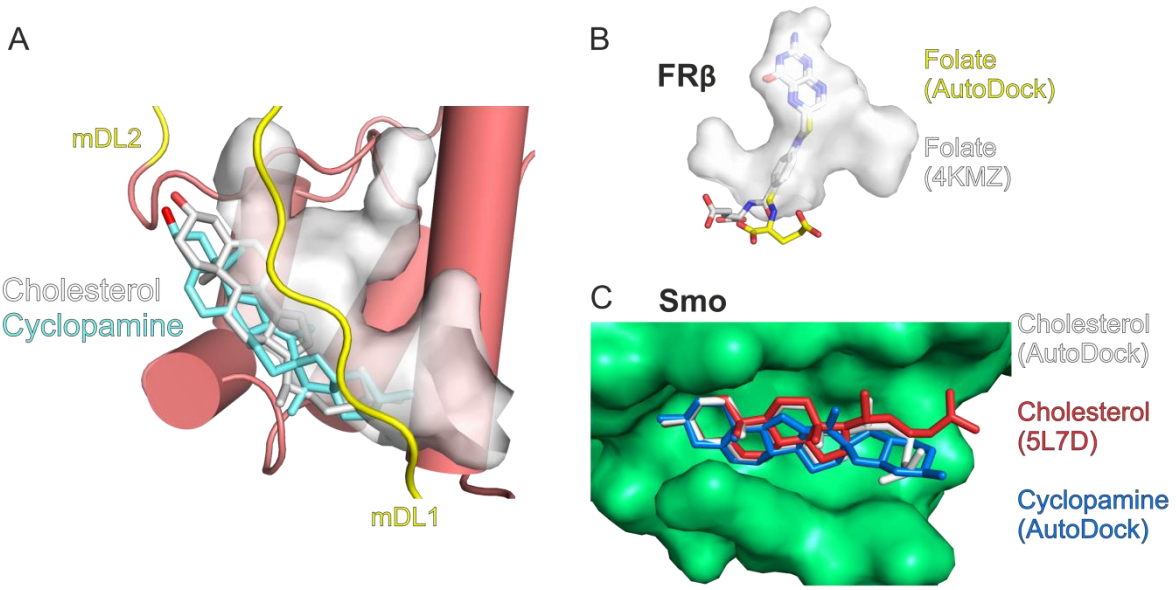


Table 3.3

Ligand	HHIP-N (modeller)	HHIP-N	FRβ	SMO
Cholesterol	-7.8	-7.4	-10.7	**-9.3**
Palmitate	-4.8	-4.4	-8.0	-5.6
Cyclopamine	-9.0	-8.5	-8.6	-10.7
Folate	-8.3	-7.7	**-10.7**	-8.0
Ethanol	-2.6	-2.6	-3.1	-2.6
Haem	-8.1	-6.9	-8.3	-7.3
Methotrexate	-7.9	-7.5	-10.5	-7.9

Figure 3.9 and Table 3.3 – Docking of small molecules into the modelled HHIP-N pocket. A) Analysis of the binding modes of two molecules known to influence Hh signal transduction to the proposed binding pocket of HHIP-N by *AutoDock Vina*; B) Validation of docking predictions by comparison of docked structures of FRβ to the experimental FRβ:folate complex structure; C) Validation of docking by comparison of a docked cholesterol in the Smo CRD to the experimental Smo:Cholesterol structure. The docking of cyclopamine into Smo is shown as a comparison to that of structurally characterised ligand cholesterol. Predicted binding energies for each ligand to the structures indicated are shown in kcal mol^{-1} in Table 3.3, with the docking energies for FRβ and the Smo CRD with structurally characterised ligands shown in bold.

3.3.7 HHIP-like proteins and zHHIP-N

Vertebrate genomes contain 2 genes with sequence similarity to HHIP in the forms of HHIP-like 1 and HHIP-like 2 (HHIPL1 and HHIPL2). Whilst both HHPL1 and HHIPL2 have not been functional characterised, both contain similar domain architectures to HHIP with N-terminal CRDs and structural characterisation of these might aid in the functional characterisation of the HHIP-N CRD region (Fig. 3.10A and B). A set of HHIPL1 and HHIPL2 constructs were cloned that comprised full-length as well as individual domain constructs with a focus on the CRD (Yang et al., 2005) (see Supplementary Table 3.2 for boundaries). These constructs were subjected to expression testing (Fig. 3.10C-F). Analysis of HHIP-N is displayed as a positive control, and shows the expected shift to a higher molecular weight in the secreted fraction due to N-linked glycosylation (Fig. 3.10C). His_6-tagged constructs of HHIPL1 and HHIPL2 with domains other than the CRD (Fig. 3.10A) were not secreted (Fig. 3.10D). A more thorough screen of CRD constructs with a number of start and end points and both His_6 and 1D4 tags did not identify a candidate for large-scale protein production (Fig 3.10E-F). A single construct was secreted, comprising the C-terminal Scavenger Receptor Cysteine-Rich (SRCR domain) of HHIPL1 (lanes annotated with '13' in Fig. 3. 10E).

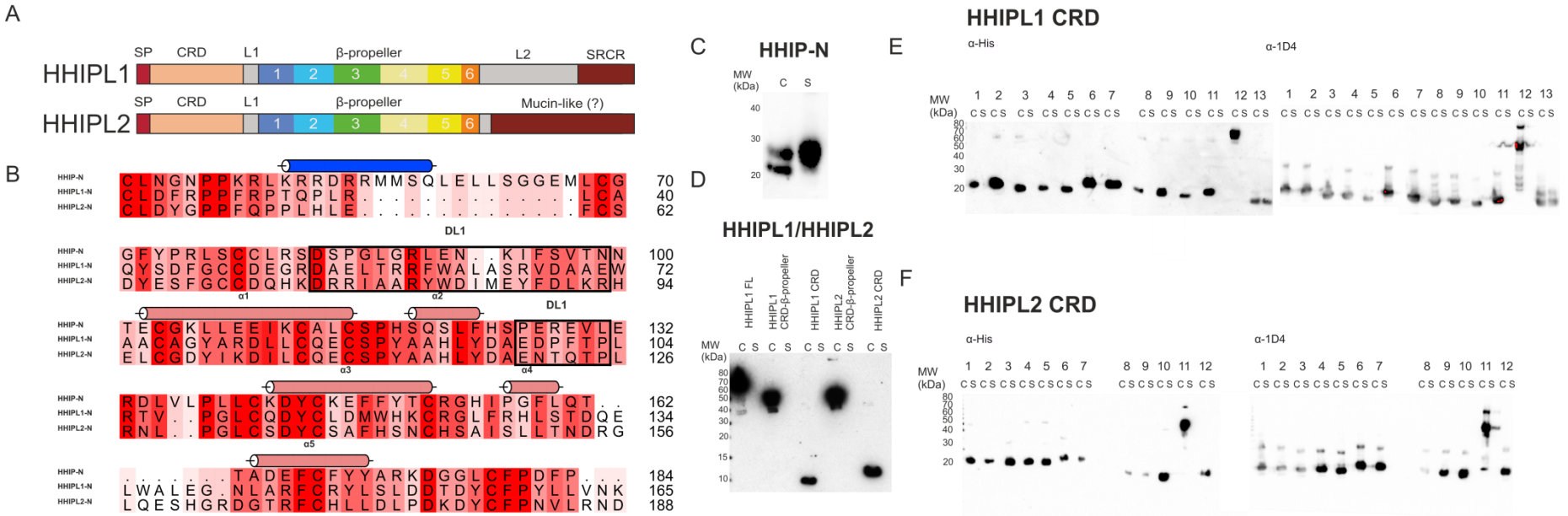


Figure 3.10 – HHIP-N homologue study as a means of elucidating the function of the CRD; A) HHIP-like protein domain architecture; B) HHIP-N and HHIP-like CRD sequence alignment, with conservation coloured as a red gradient. HHIP-N secondary structure is numbered α1-5 and DL1 and DL2 of HHIP-N are indicated with black boxes; C) HHIP-N secretion from HEK293T cells by western blot; D) Expression and secretion tests of full-length and domain-truncated versions of HHIPL1 and HHIPL-2, detected using anti-His_6 antibody; E) Expression and secretion analysis of CRD constructs of HHIPL1 cloned into His_6-tagged and 1D4-tagged backgrounds and analysed via western blots using antibodies against the tags indicated; F) As in E, but with HHIPL2. In C-F, C = cell fraction, S = supernatant. (See Supplementary Table 3.2 for construct boundaries in relation to the lanes numbered)

An additional insight came from the alignment of HHIP-N sequences from other species (Fig 3.11A). The disordered loops which form the structures of other pocket-type CRDs are well conserved between humans (HHIP-N), mice (mHHIP-N; 94% identity in DL-1, 83% in DL2) and pigs (pHHIP-N; 88% identity in DL1, 83% in DL2). More variation is displayed by frog HHIP-N (xHHIP-N; 43% identity in both DL1 and DL2), and zebrafish HHIP-N (zHHIP-N; 31% identity in DL1, 14% in DL2) displays the largest amount of variation in these regions. Residues constituting the secondary structural elements of the CRD (α1-5) are well conserved between all species. Due to the availability of an expression construct of zebrafish HHIP-N (zHHIP-N), it was reasoned that structural characterisation of this HHIP-N homologue may aid in a greater functional understanding of the CRD region. As basic residues are conserved in the N-terminal regions of HHIP-N orthologues (see Fig. 3.11A and figure legend), it is likely that zHHIP-N also binds to GAGs and this knowledge was utilised in the design of a purification strategy (see Fig. 3.1A).

Attempts to purify zHHIP-N using a similar protocol as established for HHIP-N gave a variety of proteolysed fragments, presumably due to N-terminal degradation. To overcome this issue, a chimeric construct comprising the human N-terminal GAG-binding domain and the zebrafish CRD was cloned (hzHHIP-N). This was performed in an attempt to mimic the purification strategy of HHIP-N. It is the basic residue-rich N-terminal region which binds to GAGs, and zHHIP-N is likely

cleaved in this region by Furin proteases (Thomas, 2002). However, proteolysis was still an issue; the major species in the purification of HHIP-N elutes at a concentration of 550 mM NaCl from a heparin column, which was seen to be shifted to ~350 mM with hzHHIP-N (Figs 3.1E and 3.11B). Subsequently, 50 mM L-arginine was added to the cell culture medium, which reduced the occurrence of this proteolysed species and gave a major peak that eluted at ~ 450 mM NaCl (Figs. 3.11B and D). The rationale for the addition of L-arginine was that a basic amino acid may act as a competitive inhibitor to enzymes which cleave regions rich in basic-rich sequence motifs. It was possible to grow crystals of only this major peak of hzHHIP-N (Fig. 3.11E), but these did not appear for 25 days and only diffracted to 12 Å resolution (Fig. 3.11F).

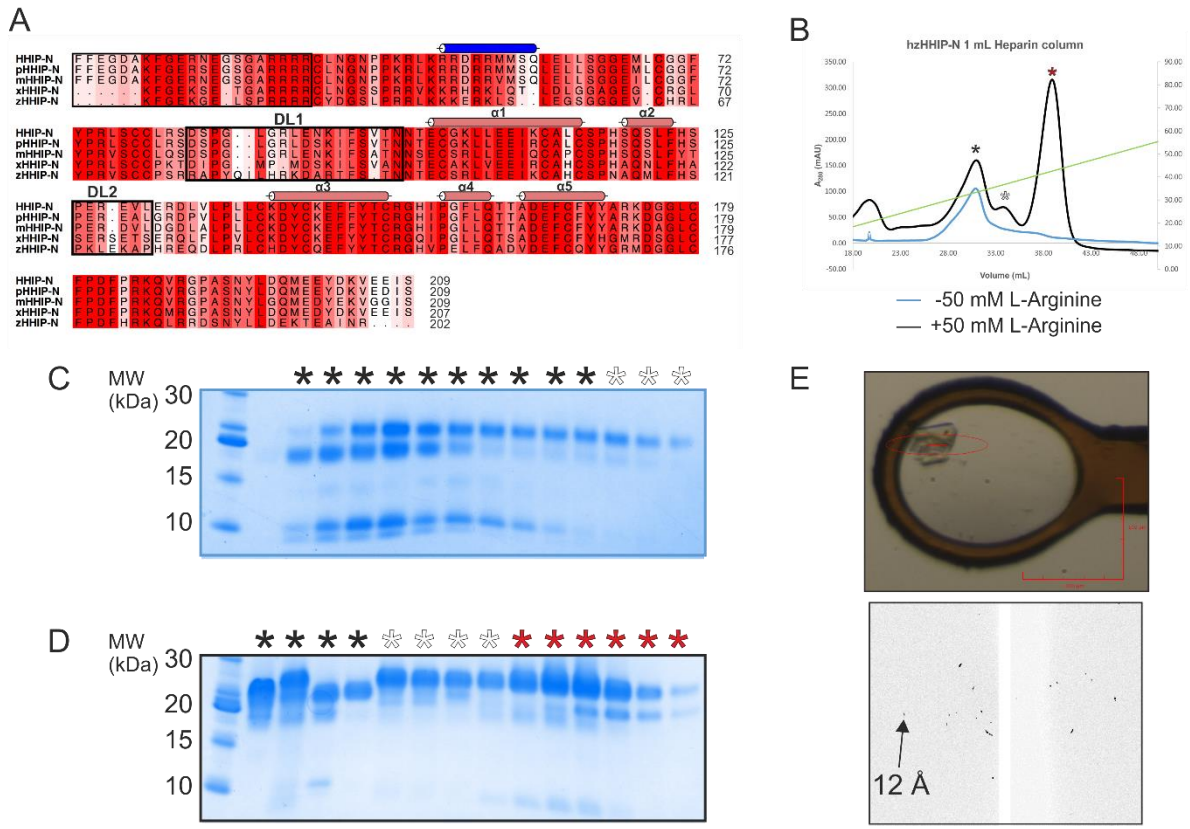


Figure 3.11 – Initial characterisation of zebrafish HHIP-N. A) Alignment of HHIP-N sequences from five species, with conservation coloured as a red gradient. Helices constituting the CRD are numbered α1-5 and DL1 and DL2 of HHIP-N are indicated with black boxes. A box at the N-terminus identifies a region of HHIP-N not included in the expression construct used in this study but may also contain a GAG-binding site; B-D) Purification of the hzHHIP-N chimera via heparin affinity chromatography with and without 50 mM L-arginine in the cell culture medium, and subsequent analysis of purity via SDS-PAGE. B) UV trace for heparin column purification; C) (-)L-Arginine; and D) (+)L-Arginine; E) Crystal of hzHHIP-N from the major peak in B) (red asterisk) mounted for data collection; F) detected diffraction peak following an X-ray diffraction experiment.

3.4 Discussion

The physiological roles of CRDs are far-ranging and diverse, relying on measures of structural plasticity in order to fulfil their functions. Here, the three-dimensional structure of HHIP-N has been elucidated and this raises several questions concerning its native function. Our research has demonstrated that HHIP is a novel secreted cell-surface binding Hh signalling antagonist, involving regulation through a residue cluster at the very N-terminus of HHIP-N (see Chapter 4). The structure of HHIP-N:SOS presents a structural model for this proteoglycan recognition, displaying a peripheral N-terminal GAG-binding subdomain (Fig. 3.5). Structures determined in the presence and the absence of GAG mimic SOS identify a disorder-to-order transition as being the outcome of this interaction (Fig. 3.5B and C; see Chapter 4 for a more in-depth discussion of HHIP:GAG interactions). Despite an observed conformational change upon GAG binding, the fold of the C-terminal α-helical CRD region remains unaltered (Fig 3.5C). Two disordered loops of HHIP-N are predicted to map to the location of small molecule recognition elements of several structurally characterised CRDs, identifying these sequence motifs as potentially being linked to the binding of some small molecule (Figs. 3.6 and 3.7).

3.4.1 The HHIP-N CRD as a small molecule-binding motif

Structural alignment of HHIP-N with published pocket-type CRDs suggest a role in the recognition of a small molecule ligand. The α-helices of the HHIP-N CRD region align well with related CRDs, representing a 'core' fold. Helix α1, adjacent to ligand-binding loop 1 of related CRDs, is especially well superposed (Fig 3.6). A basic comparison can be drawn with FRβ (RMSD of 1.55 Å for 66 Cα positions), in which structural elements utilised in folate and antifolate binding align with the positions of HHIP-N DL1 and DL2 (Fig. 3.6B) (Wibowo et al., 2013). The same scenario is observed for the CRD of NPC1 (RMSD of 1.99 Å for 61 Cα positions), albeit to a more complex degree; both loops previously identified via sequence similarity between CRDs are positioned near to the same location of HHIP-N as DL2 is thought to occupy. One such region, an extended β-hairpin, is not involved in contacts with the ligand and it is this region that aligns with HHIP-N DL2 in sequence (Fig. 3.6C). An additional helix aligned to the predicted location HHIP-N DL1 is also utilised in cholesterol recognition (Kwon et al., 2009). Comparatively, mJUNO is the only pocket-type CRD analysed (RMSD of 1.54 Å for 63 Cα positions) which does not display a folded structural element overlapping with a position comparable to that of DL1 of HHIP-N (Fig. 3.6E). This may be explained by the hJUNO:IZUMO complex structure, in which disordered loop 1 from the mJUNO structure is used in contacts with the receptor in the human complex structure (Aydin et al., 2016) (Figs. 3.6E and 3.13B). Additionally, all three loops in one of the two mJUNO molecules from the

ASU are disordered, and therefore this ordered structure of loop 2 may be a result of crystal packing (Han et al., 2016). Therefore, sequences aligned to DL1 are conserved as ligand recognition elements across pocket-type CRDs.

Structural comparison between HHIP-N and Fz-type CRDs with characterised ligand binding sites also supports the hypothesis of DL1 and DL2 being able to recognise a small molecule binding candidate. Both Fz8 and Smo (RMSD of 2.20 Å for 42 Cα positions and 2.35 for 41 Cα positions respectively) possess long helices which align with the predicted location and the sequence of DL1 in HHIP-N, and conserved residues within this helix form several major contacts with their respective groove-bound aliphatic ligands (Figs. 3.6F-G and 3.7B) (Byrne et al., 2016; Janda et al., 2012). Modelling of DL1 enables the docking of small molecules implicated in Hh signalling into a pocket formed by DL1 and a selection of hydrophobic residues, but this analysis is speculative without supporting experimental data (Figs. 3.8 and 3.9). In conclusion, the functions of all CRDs are dependent on residues which align both in sequence and in the expected location of HHIP-N DL1. The specificity imparted by residues in this region has been critical in the evolution of specific ligand binding functions. Taken together, these facts suggest that DL1 and possibly DL2 of HHIP-N are involved in the recognition of some binding partner.

A thorough understanding of the function of the HHIP-N CRD is not possible in the absence of physiological ligand candidates; mass spectrometry analysis failed to provide any hits. PDB-wide searches using the DALI server did not give further

insight, identifying the expected CRD candidates and also glypicans as structurally-related hits (see section 1.6.2 and Fig. 1.13) (Holm and Rosenstrom, 2010). The ProFunc server also identified these same published CRDs through fold recognition methods, and therefore does not aid in the suggestion of a function for HHIP-N (Laskowski et al., 2005).

Modelling of DL1 and DL2 suggests the formation of a hydrophobic pocket, into which cholesterol could be docked with a favourable binding energy (Fig. 3.9A). Attempts to soak HHIP-N crystals with cholesterol dissolved in various solvents at high concentrations (10 - 20 mM) did not result in any bound ligands after structure solution. Attempts to study cholesterol-binding in solution were also problematic, as HHIP-N is highly stable and cannot be melted by either Thermofluor analysis or Circular Dichroism (Fig. 3.12) (Walter et al., 2012). Functional insight could not be gained from the biophysical study of HHIP-like proteins from the human genome or the zebrafish HHIP-N protein, as neither could be produced in crystallisation-quality forms (Figs. 3.10 and 3.11). Therefore, more work is required before homologue studies lend support for any hypothesis regarding any HHIP-N function which is not related to the binding of GAGs (see chapter 4).

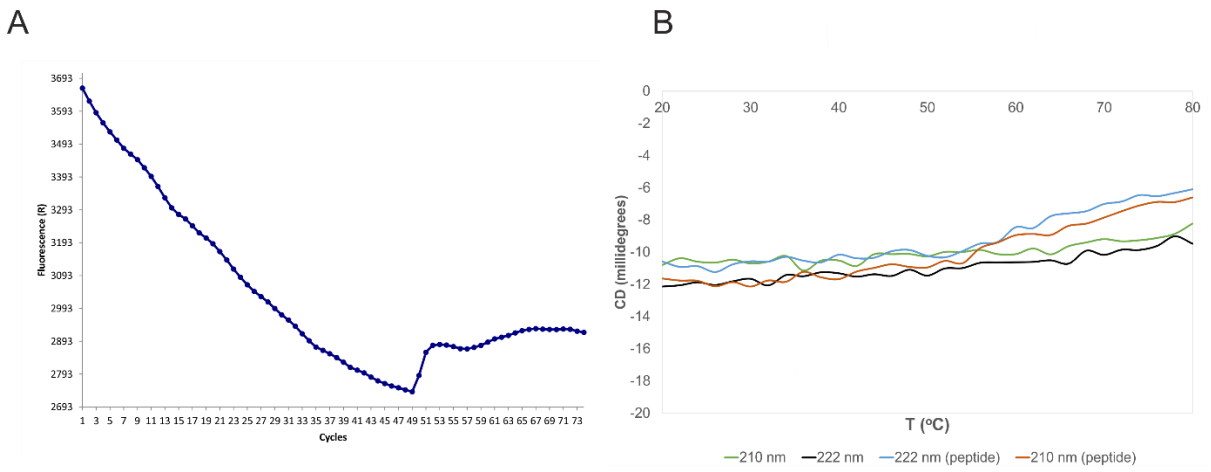


Figure 3.12 – **Attempts to probe HHIP-N:small molecule interactions (above)**. A) Melting curve measured for HHIP-N via Thermofluor using SYPRO orange over the range 25 – 100^{o}C; B) Circular Dichroism spectra were measured from 20 – 80^{o}C for HHIP-N in the absence and the presence of a HhN-palmitate peptide.

A second hypothesis regarding the physiological function of the HHIP-N CRD concerns interaction with the lipid modified HhN ligand itself. The Wnt8:Fz8 interaction and ensuing signal transduction involves recognition of the palmitoleate modification of Wnt by a groove on the Fz8 surface (Janda et al., 2012). The sFRPs are presumed to prevent receptor activation via highjacking this mode of interaction. Carboxyesterase Notum negatively regulates Wnt signalling by the removal of this lipid adduct (Kakugawa et al., 2015). Wnt signalling has therefore evolved regulatory modes around the palmitoleate modification. HhN contains an N-terminal palmitoyl moiety, the presence of which is required for full pathway activation (Taylor et al., 2001). No negative regulator of Hh signalling has yet been discovered which targets the lipid adduct and surrounding region of HhN. It therefore remains possible that HHIP-N may recognise the palmitate modification, and a palmitoylated HhN

peptide was used to test this hypothesis. Co-crystallisation of HHIP-N with this peptide did not yield crystals of a complex, and analysis via SPR showed no binding (data not shown). Whether this requires full-length palmitoyl-HhN remains to be tested. It has been reported that HHIP-N does not influence Hh signalling in a cellular assay with full-length HhN, but was not validated with purified proteins (Bosanac et al., 2009). Another hypothesis is that HHIP-N binds the cholesterol modification upon the C-terminus of HhN, which is known to be important in multimeric particle clustering and long range Hh signalling (Gallet et al., 2003; Peters et al., 2004). In summary, evolutionary comparisons suggest that the HHIP-N CRD binds a ligand involved in Hh signalling, but more work is required to identify and characterise this biological function.

3.4.2 HHIP-N as a protein-protein interaction motif

The structure of mouse JUNO (mJUNO) grants insight into the potential role of the HHIP-N CRD. JUNO, also known as FRδ, has evolved the capacity for protein-protein interactions at the expense of folate binding capacity (Bianchi et al., 2014). Only loop 2 of mJUNO is ordered in this structure and this can be structurally aligned with the expected location of DL2 of HHIP-N (Figs. 3.6E and 3.13A). Loops 1 and 3 are visible in the human JUNO (hJUNO) apo-structure, whilst loop 2 is disordered (Fig. 3.13A) (Aydin et al., 2016; Han et al., 2016). Both loops 1 and 3 of hJUNO contact receptor IZUMO and play a role in receptor specificity (Fig. 3.13B) (Aydin et al., 2016; Han et al., 2016; Ohto et al., 2016). Given the role of sequence

elements related to DL1 of HHIP-N in both pocket- and groove-containing CRDs (Figs. 3.6 and 3.7), a potential function of HHIP-N may entail binding to a protein counterpart and not a small molecule ligand as hypothesised.

NPC1 also represents an example of a pocket-type CRD in such a process. Under resting conditions, the NPC1 CRD remains stably associated with lumenal domains C and I (Fig. 3.13C) (Gong et al., 2016). NPC2 must disrupt this arrangement and binds to both the CRD and domain C to facilitate cholesterol transfer to the CRD (Zhao et al., 2016). The reorganisation of the cholesterol-binding loops of NPC1 thought to be related to DL1 and DL2 of HHIP-N (Figs. 3.6C and 3.7A) in the process of cholesterol transfer has not been investigated (Gong et al., 2016; Kwon et al., 2009; Zhao et al., 2016). Therefore, the NPC1 CRD performs functions in both the binding of small molecule ligands and in protein:protein recognition through sequence elements related to DL1 and DL2 of HHIP-N, which may be related to the physiological function of HHIP-N.

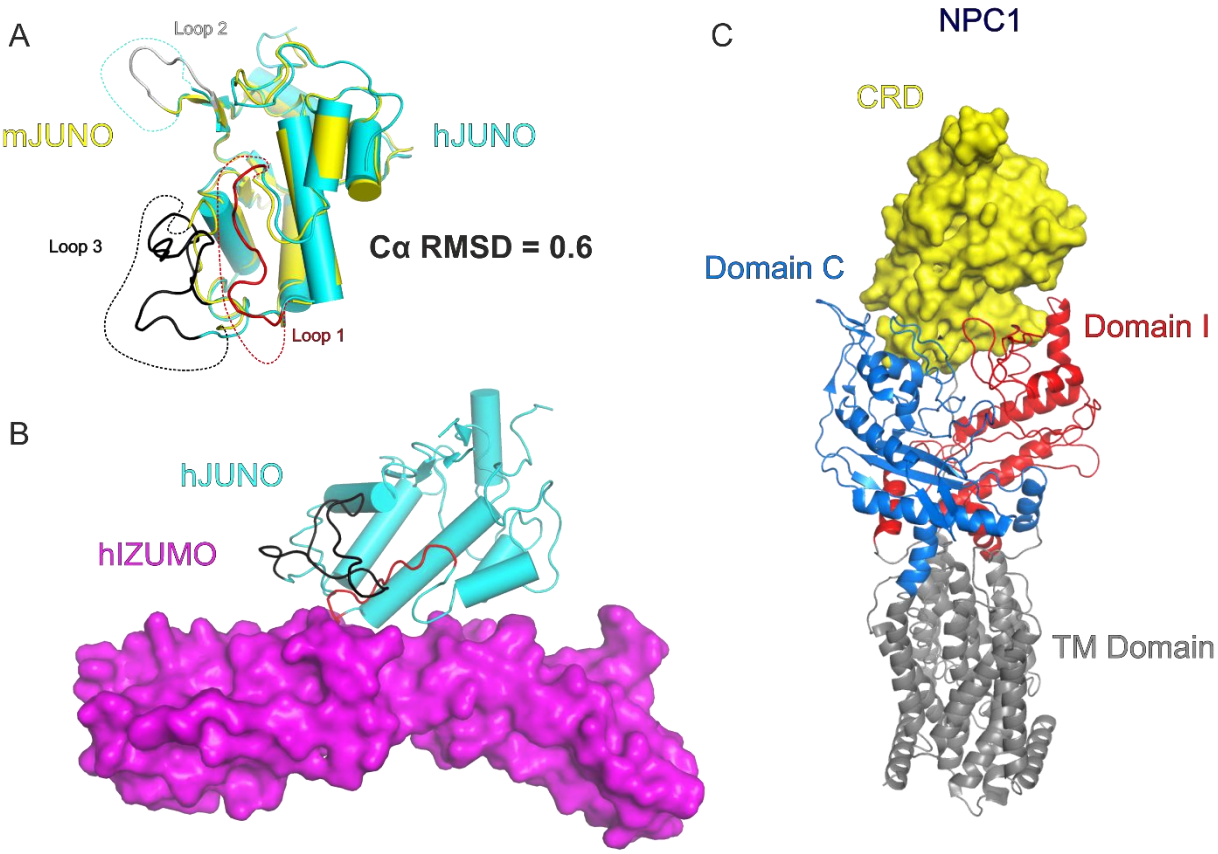


Figure 3.13 – CRDs and protein:protein interactions (above). A) Superposition of mJUNO (yellow, PDB ID: 5EJN) and hJUNO (cyan, PDB ID: 5F4E). Loops 1 and 3 from hJUNO are coloured red and black respectively, and the corresponding disordered loops from mJUNO are displayed as dashed lines of the same colour. Loop 2 from mJUNO is shown in white, and the corresponding disordered loop from hJUNO is shown as a cyan dashed line; B) Complex between hJUNO and receptor IZUMO (magenta) (PDB ID: 5JKE). hJUNO is coloured as in A; C) Full-length structure of NPC1 with component domains coloured as indicated (PDB ID: 3JD8).

3.4.3 Outlook for native SAD phasing techniques

This study joins a growing number of crystallographic studies that have utilised S-SAD as a phasing methodology. The opportunity for rapid structure solution via native SAD experimental phasing is particularly exciting given that it does not necessitate the derivatisation of crystals by heavy elements or

selenomethionine-labelling. However, the major challenge is that at the wavelength at which the f'' of S atoms is highest (4.1 electrons; 5.0 Å), ~90% of X-rays are absorbed by the atmosphere (Liu et al., 2013). Studies have attempted to negate the effects of this and maximise anomalous signal in a vacuum or an atmosphere of helium (Wagner et al., 2016; Weinert et al., 2014). In the absence of either, a wavelength of 1.7712 Å was selected to obtain the maximum possible anomalous signal (f'' = 0.7 electrons) whilst still permitting X-ray transmission (85%). In order to achieve the necessary levels of anomalous signal, a high multiplicity gained from data from several statistically equivalent crystals is required to forego radiation damage (Liu et al., 2012).

A recent innovation in this area is long wavelength data collection in-vacuum beamline I23 at DLS. This can deliver wavelengths of up to 5.9 Å and is equipped with a curved Pilatus 12M detector. This covers an angular range of $2\theta = 100^{o}$, which is necessary for the detection of reflections at a resolution of above 3 Å at a wavelength of 4 Å. Whilst radiation damage will remain an issue in all X-ray data collection, in-vacuum data enables data collection at wavelengths that rapidly give appreciable anomalous signal from intrinsic protein atoms. This will reduce the need for such high multiplicity and thus the time spent collecting data (Wagner et al., 2016).

A second consideration in S-SAD data collection is the precise measurement of diffraction intensities. Anomalous phasing is dependent on accurate measurement of

Bijvoet difference amplitudes, due to the breakdown of Friedel's law and the associated use in phase estimation (Fig. 3.14A; see materials and methods). Due to the susceptibility to radiation damage, it is essential that the delay between measurements of Bijvoet differences is as short as possible. This is achieved via inverse beam (take diffraction image, rotate crystal 180º, take second image; repeat in 0.1º increments over the entire crystal) and mini-kappa multi-axis goniometry (measure Bijvoet pairs on the same image) data collection strategies (El Omari et al., 2014; Hendrickson, 2014).

S-SAD data collection of HHIP-N produced the highest recorded multiplicity for any study published thus far (Table 3.1 and Fig. 3.14B). The multiplicity required for S-SAD structure solution versus the Bijvoet ratio does not show an obvious trend, and other assessments on data quality (i.e. CC_{anom}) must be utilised. This highlights that the true limiting factor in S-SAD structure solution by this method is the way in which Bijvoet pair intensities are measured, and hence experimental design is key to successful phasing.

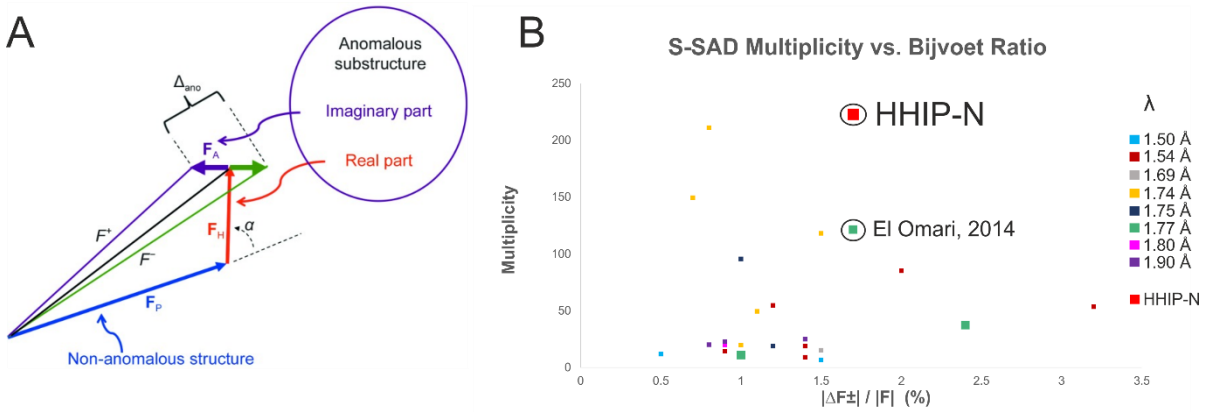


Figure 3.14 – S-SAD phasing considerations. A) Diagram relating structure factors to associated anomalous differences. F^+ and F^- correspond to structure factors for Bijvoet pairs for reflections (h, k, l) and (-h, -k. –l), related by 180° about the real axis. F_p corresponds to the non-anomalous scattering contributions. F_H corresponds to the scattering contribution from the real component of the anomalous substructure (f_o + f'), whilst those from the imaginary part (if'') is represented as F_A. Taken from (Terwilliger et al., 2016); B) A plot of multiplicity versus Bijvoet ratio for studies utilising only S-SAD in order to solve structures. Data points were plotted using the colours indicated for wavelengths over the range 1.5 – 1.9 Å. The studies in question are referenced in the supplementary bibliography.

The next highest multiplicity for a study with the same experimental set up, Bijvoet ratio (1.7%) and the same order of crystallographic symmetry as HHIP-N ($P4_22_12$ versus $P4_12_12$) was a multiplicity of 122 from the structure solution of the HCV E1 ectodomain. Sulphur substructure solution from SHELXD was also of higher quality in the HCV E1 study, with a CC_{weak}/CC_{all} of ~23/55 (Fig. 3.3). Whether a lower multiplicity would have sufficed for HHIP-N was not analysed, and the 'ideal' amount of data to collect is not known (Fig. 3.14B) (El Omari et al., 2014).

Site occupancy is also a consideration for S-SAD. Considering all potential S sites in HHIP-N (Cys, Met, SOS and disulphides; 30 S sites) gives a theoretical Bijvoet ratio of 2.1%. However, SHELXD plots of identified peak number showed the largest

drop and thus the likelihood of a correct solution after 6 S sites, which is confirmed via alignment of the final HHIP-N structure with an anomalous difference map calculated from the experimental data (Fig. 3.15A-B). Hence both the treatment of 12 sulphurs as 6 'super-sulphurs' in disulphide bonds and the trimming of data to 4 Å resolution was vital for correct substructure solution (Debreczeni et al., 2003). This process was likely made more difficult by the lack of an especially high solvent content (53%) or any NCS (Kleywegt and Read, 1997). This discussion of occupancy raises the issue of radiation damage within S-SAD phasing. Disulphide bonds are the first to be broken by radiation damage and this may present an issue for location of a S marker atom substructure (Carugo and Carugo, 2005). This further highlights the necessity for extremely accurate data collection strategies as there is no way of recovering poorly measured Bijvoet pairs, which are imperative for experimental phasing. As with the HCV E1 study, the use of several different software packages was also necessary for the phasing and refinement of HHIP-N (Fig. 3.3 and 3.4) (El Omari et al., 2014). This study also highlights the necessity to remove proteolysed fragments from crystallisation trials in order to produce reproducible diffraction-quality crystals (Fig. 3.1D-E versus H-I).

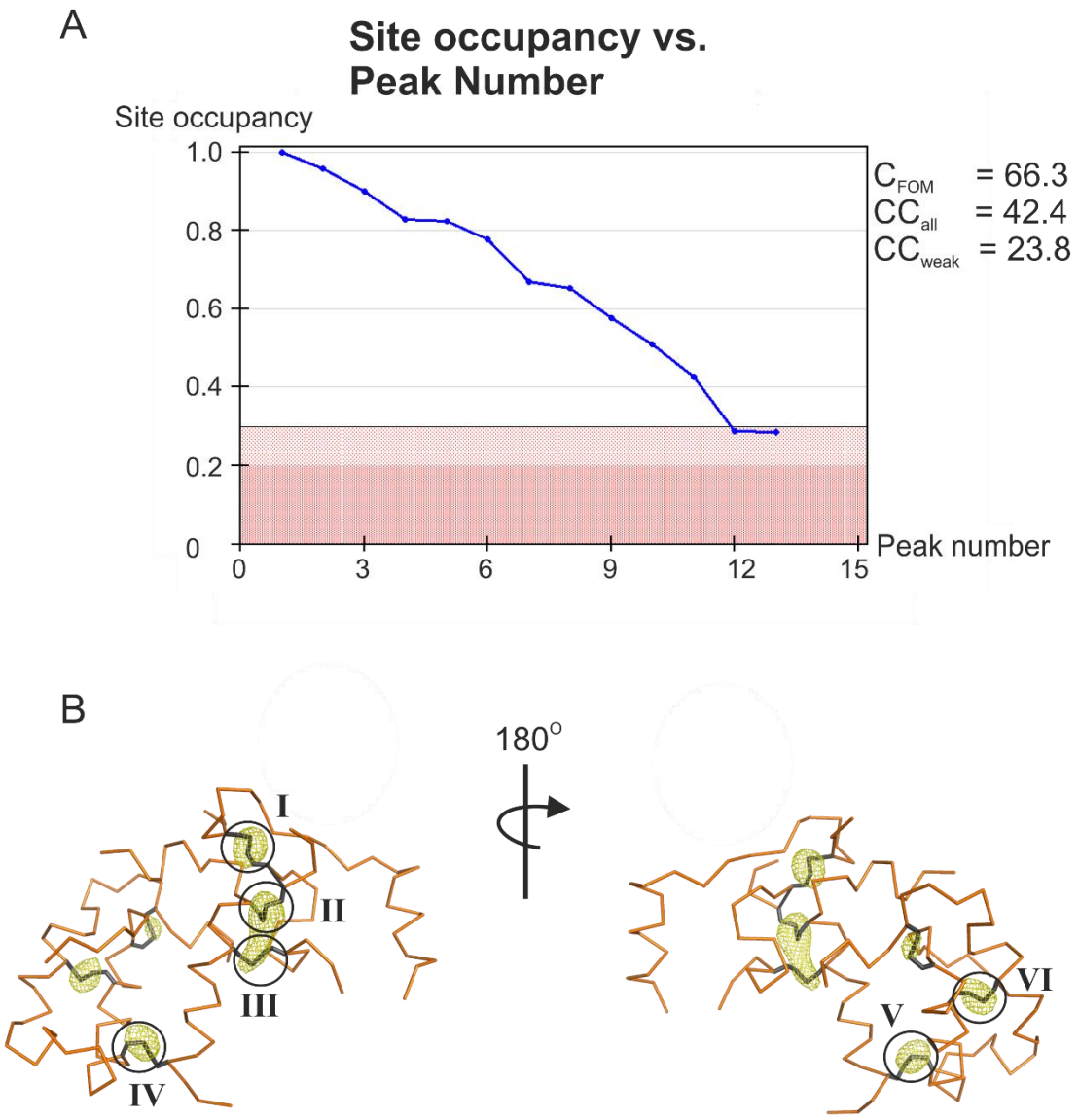


Figure 3.15 – HHIP-N S substructure. A) SHELXD site occupancy graph (top solution) and B) anomalous difference map shown as a yellow mesh for HHIP-N calculated using the PHENIX Maps utility (Adams et al., 2010). This is contoured at 4σ. Disulphides in the respective sites are circled and numbered I – VI. C_{FOM} = CC_{all} + CC_{weak}.

3.4.4 Further work

The logical next step in the study of HHIP-N is identification of the function of loops DL1 and DL2 (Fig. 3.6). Study of the role of HHIP-N *in vivo* represents a challenge because the protein possesses an assumed dual role in both cell surface-binding and ligand recognition. Thus, without candidate residues to mutate within the C-terminal CRD region, one cannot easily attribute resultant HHIP-N knockout phenotypes to a role involving this portion of the protein. Insight may be gained from observation of mutational phenotypes in a model organism such as mouse in which the residues in DL1 and DL2 have been changed to those of frog of zebrafish. The role of the GAG-binding sub-domain will be discussed in Chapter 4.

Screening of fragment libraries via SPR or crystal soaks may aid in the discovery of potential ligand candidates (Patel et al., 2014). As discussed, cholesterol displays an important role in Hh signalling and a cholesterol binding assay has been developed which would give a rapid answer as to whether HHIP-N also binds to cholesterol or related sterols with influence of small molecules over Smo (Byrne et al., 2016; Nachtergaele et al., 2013). An initial bead-based binding assay (Fig. 3.16) suggests that HHIP-N does bind to both cholesterol and 20-(S)-hydroxycholesterol, and free sterols compete with this process (compare lane 5 to 4, lane 7 to 6 and lane 9 to 8). This experiment requires repetition with appropriate controls. More extensive computational modelling and molecular dynamics is an additional technique available; solution study of DL1 and DL2 via NMR would be particularly difficult

due to the expense of producing spin-labelled protein from HEK293T cells as well as the issue of low yield.

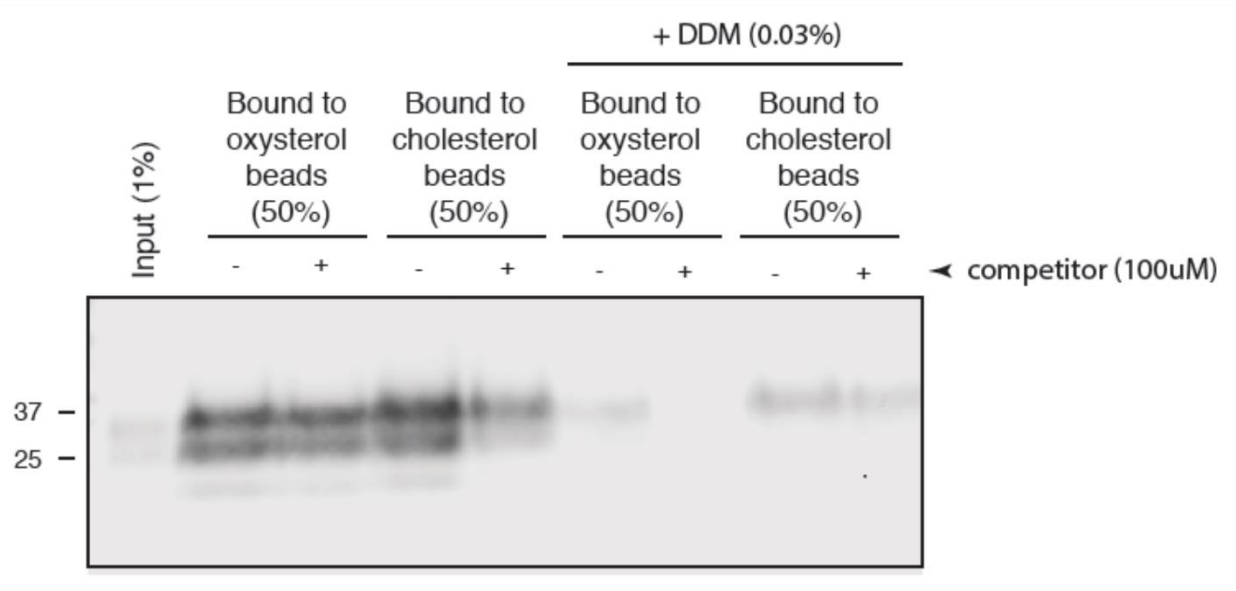


Figure 3.16 – HHIP-N:sterol bead binding assay. 2.5 μg of HHIP-N was bound to oxysterol (1:1000)- or cholesterol (1:50)-coupled beads by incubation for 16 hours at 4°C followed by 1 hour at 25°C, both with and without 0.03% DDM. This was performed in the absence (-) and the presence (+) of 100 μM of a competitor molecule (20-(S)-hydroxycholesterol for oxysterol beads and cholesterol for cholesterol beads). Beads were then analysed using SDS-PAGE to examine binding and competition (Rohatgi R, Stanford University School of Medicine).

High-quality crystals of HHIP-N homologues could not be grown, and therefore further insight into the function of the HHIP-N CRD region from these was not possible (Figs. 3.10 and 3.11). However, the use of 50 mM L-arginine as an inhibitor of secretory proteases was a success, enabling isolation of a species with an increased affinity for heparin, as well as increasing yield (Fig. 3.11B-D). This method

may serve as an inexpensive protein purification rescue strategy on top of the proven use of histone deacetylase inhibitors such as valproic acid (Chang et al., 2015).

3.4.5 Concluding remarks

As more structures of Hh receptors become available, the more important the study of relevant physiological Hh signalling states becomes. As it stands, there are no published structures of lipid-modified Hh either in the multimeric state or in complex with modulatory proteins or receptors. A selection of speculative functions of HHIP-N are displayed in Fig. 3.17. The mode of HhN:Ptc interaction has recently been discovered to be more similar to that of Fz:Wnt than previously thought, with two separate portions of the HhN structure making contacts with either extracellular loop of Ptc (see Fig. 1.5B) (Janda et al., 2012; Tukachinsky et al., 2016). The role of HHIP is the antagonism of Hh signalling and any probable function of HHIP-N would complement this (Chuang and McMahon, 1999). Whilst signalling inhibition by HHIP-N would be enhanced by recognition of the palmitoyl moiety, the N-terminal sequence extension of HhN to which the palmitate is attached (contacts loop 1 of Ptc in Fig. 3.17A) would likely render HHIP-N unable to bind to the same HhN molecule as HHIP-C (Fig. 3.17A). As HhN forms multimers, this does not necessarily make such an interaction impossible. Small molecules capable of Smo activation have been developed and, in the absence of any knowledge of the cognate ligands for Ptc and Smo, it remains possible that Ptc may pump an activator of Smo back into the cell as means of signal transduction antagonism (Huang et al., 2016; McCabe and

Leahy, 2015). Therefore, HHIP-N may act to 'mop up' any such molecules, thereby preventing active signalling (Fig. 3.17B). Additionally, HHIP-N could destabilise Smo via removal of cholesterol or related oxysterols which activate Hh signalling (Figs. 3.16 and 3.17C) (Byrne et al., 2016). HHIP-N may cause disassembly of HhN multimers via recognition of the C-terminal cholesterol modification (Fig. 3.17D) (Eaton, 2008). It also remains possible, given the evolution of JUNO, that a currently unknown protein binds to HHIP-N via DL1 and DL2 or that some as-yet uncharacterised small molecule with influence over the Hh pathway binds to the predicted pocket (Fig. 3.17E) (Han et al., 2016).

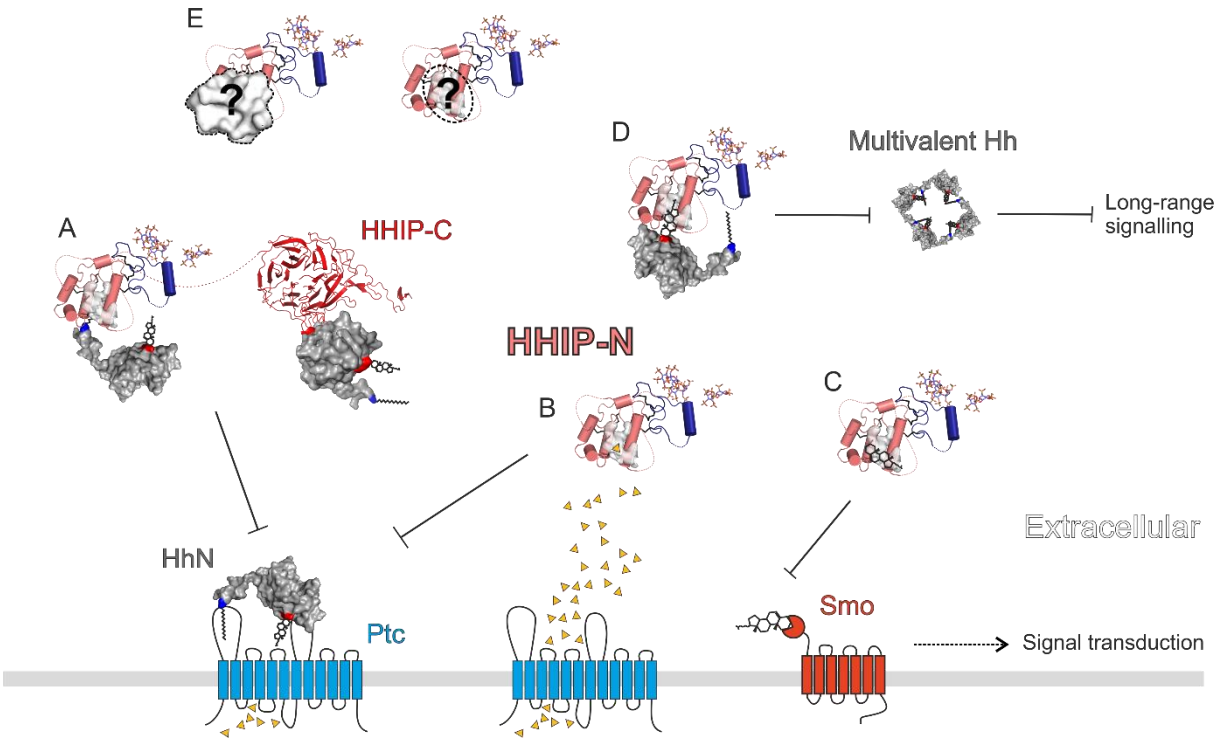


Figure 3.17 – Possible roles for the HHIP-N C-terminal CRD region. A) Competition with Ptc for the palmitate modification on the N-terminus of HhN; B) Recognition of some activator of Hh signalling, given the appropriate cellular conditions; C) Removal of cholesterol from Smo, thereby destabilising the interaction between the Smo CRD and TM region and dampening signal activation; D) Recognition of the HhN C-terminal cholesterol adduct, thereby preventing multimer formation and long-range Hh transport; E) Use of DL1 and DL2 to bind an unknown protein or ligand.

Every aspect of Hh signalling is tightly regulated and, given the consequences of Hh signalling dysregulation, it seems unlikely that the cognate Smo ligand would be left to freely diffuse within the ECM. Cholesterol trafficking in the distantly related NPC1/NPC2 system is tightly regulated, and evolution suggests a common ancestor between this process and Hh signalling (Bazan and de Sauvage, 2009). At any rate, evolution is not meaninglessly expansive and comparison with the related Wnt pathway suggests a number of potential functions to be fulfilled by this small

molecule-binding motif. The primary cilium represents a Hh 'signalling antenna' and the lipid composition of this area of the cell is likely to be important in the control of Hh signalling. HHIP function in the primary cilium has not been investigated, and binding of HHIP-N to specific lipids found in this organelle may be important to the fine control of Hh signalling.

Chapter 4

HHIP-GAG interactions

Chapter 4: HHIP-GAG interactions

4.1 Summary

HHIP is a multi-domain protein comprising an N-terminal CRD (HHIP-N), a C-terminal β-propeller and 2 EGF repeats (HHIP-C), and a 30-residue C-terminal stretch, predicted to form a helix. HHIP demonstrates cell surface-binding properties in a GAG-dependent manner, and binding sites were identified in HHIP-N and HHIP-C using biophysical and cellular studies. Crystal structures of HHIP-N and HHIP-C were solved in complex with GAG mimic SOS in order to investigate the molecular basis of these interactions. A HHIP-C:heparin crystal structure was also solved and, taken together, these results suggest the formation of large multimeric HHIP:GAG assemblies at the cell surface. This model was validated in solution by analytical ultracentrifugation. In addition, modelling of the 30 C-terminal amino acids of HHIP proposes the formation of an amphipathic helix that may either mediate further ECM contacts or regulate HHIP avidity at the cell surface.

4.2 Introduction

The multi-domain architecture of vertebrate Hh antagonist HHIP suggests the potential for functions beyond steric interference of the HhN:receptor complexes. As discussed in Chapter 3, the N-terminal CRD (HHIP-N) comprises a fold commonly associated with small molecule binding, and cholesterol and oxysterols constitute known effectors of Hh signalling (Byrne et al., 2016; Huang et al., 2016; Luchetti et al., 2016). HHIP also contains a β-propeller and 2 EGF repeats (HHIP-C) and a predicted C-terminal helix (Hx) (Fig. 4.1A). The most well-characterised role of HHIP is Hh antagonism, with HHIP being produced by cells in response to Hh signal activation as a means of negative feedback (Fig. 4.1B) (Coulombe et al., 2004; Ochi et al., 2006). Structural studies revealed that HHIP-C binds to HhN through the interaction of blades 2 and 3 of the HHIP β-propeller with a hotspot region on the HhN surface (Bishop et al., 2009; Bosanac et al., 2009)(Fig. 4.1C). Hh signalling inhibition by HHIP is known to be both cell autonomous and non-cell autonomous in nature due to the capacity of HHIP to be secreted (Coulombe et al., 2004; Kwong et al., 2014). A number of secreted proteins interact with GAG chains attached to the cell surface PG layer and this is critical to the regulation of spatial and temporal dynamics, oligomerisation and interactions with various receptors (Nolan et al., 2013; Schlessinger et al., 2000; Whalen et al., 2013). Cellular and biophysical experiments with HHIP constructs were combined with crystallography in order to investigate and characterise HHIP:GAG interactions.

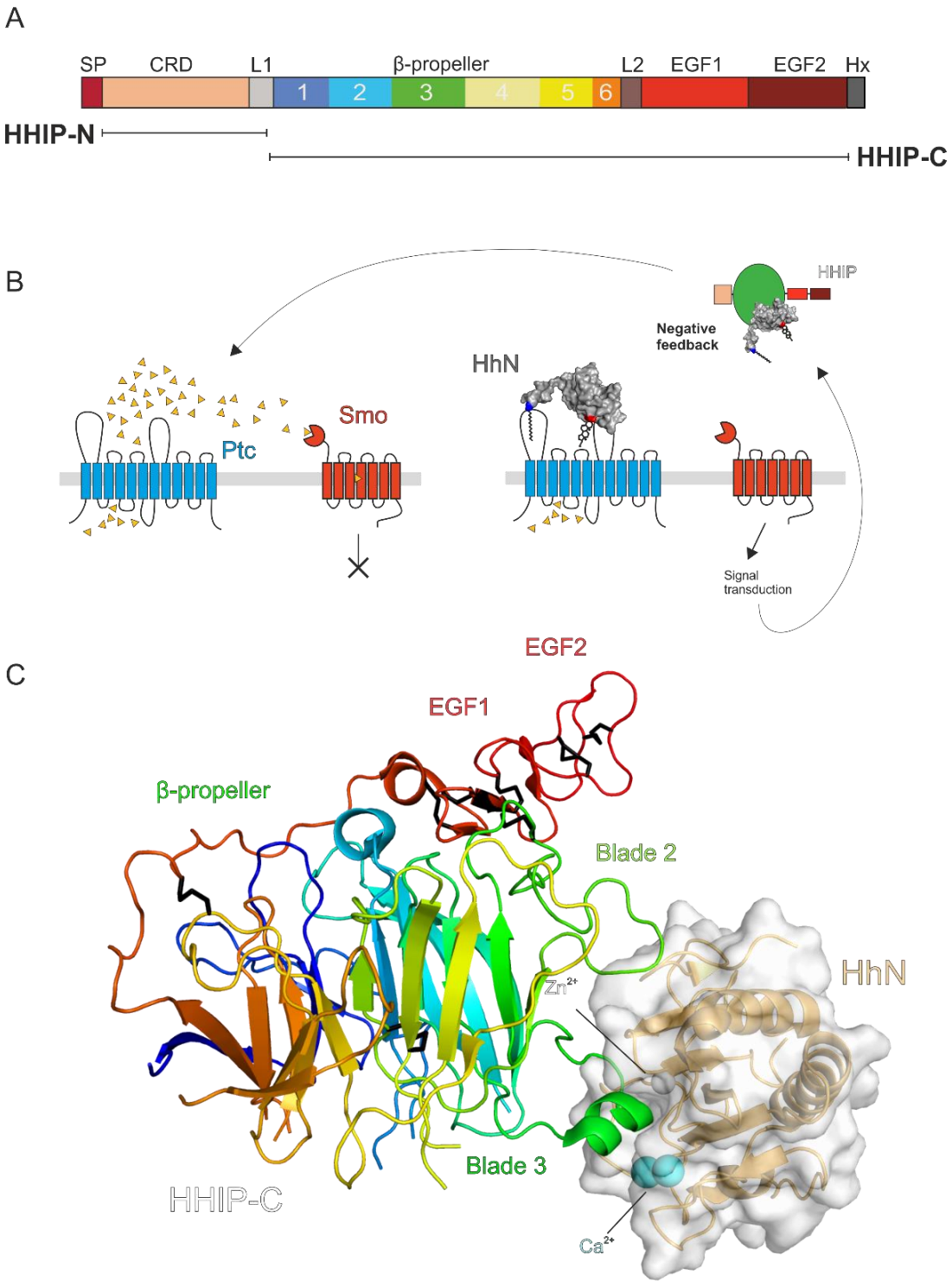


Figure 4.1 – The Hedgehog-Interacting Protein (HHIP). A) Domain organisation of HHIP, with HHIP-N and HHIP-C labelled; B) Schematic of physiological HHIP function. Under resting conditions, Ptc inhibits signal transduction via the repression of Smo (left). HhN:Ptc interaction removes Ptc-induced repression of Smo, activating signalling. One gene target of activated Gli transcription factors is HHIP, which sequesters the Hh ligand and returns the cell to the resting state (right); C) HHIP:HhN crystal structure (PDB: 2WFX). Blade 3 contacts a receptor hotspot site on the HhN surface and interacts with the HhN Zn^{2+}-binding site.

4.3 Results

4.3.1 HHIP is a secreted molecule regulated by cell surface GAG-interactions with the N-terminal CRD

HHIP has been previously reported to be both membrane-associated and secreted, the mechanisms surrounding which are poorly understood (Chuang and McMahon, 1999; Coulombe et al., 2004). Constructs of full-length HHIP (including Hx) were generated with different tags at the N- and C-termini and these were tested for secretion in NIH3T3 cells (Fig. 3.2A). Full-length HHIP is detected in both cell lysates and cell supernatants at a molecular weight of ~75 kDa, corresponding to the full-length protein. As detection of the protein using antibodies to tags at the N- and C-termini gave equal molecular weights, HHIP secretion does not depend on proteolysis (Figs. 3.2B-D). There are two regions of HHIP that contain stretches of basic residues predicted to be involved in binding to the cell surface GAG layer, found within the N-terminal CRD and the N-terminal portion of Hx (Fig. 3.2E). These were analysed using a heparin chromatography assay. HHIP FL elutes from heparin beads at 550 mM NaCl, and removal of Hx reduces this value to 450 mM (Fig. 3.2F). Alanine mutation of the arginine residues in Hx does not reproduce this reduction in heparin binding, and Hx therefore mediates cell surface binding via some other mechanism (see section 4.3.9) (Fig. 3.2G). Removal of the whole CRD leads to a loss of heparin binding affinity (Fig. 3.2H). The HHIP-N CRD was subsequently purified

and crystallised with GAG mimic SOS in order to structurally characterise this interaction (as discussed in Chapter 3).

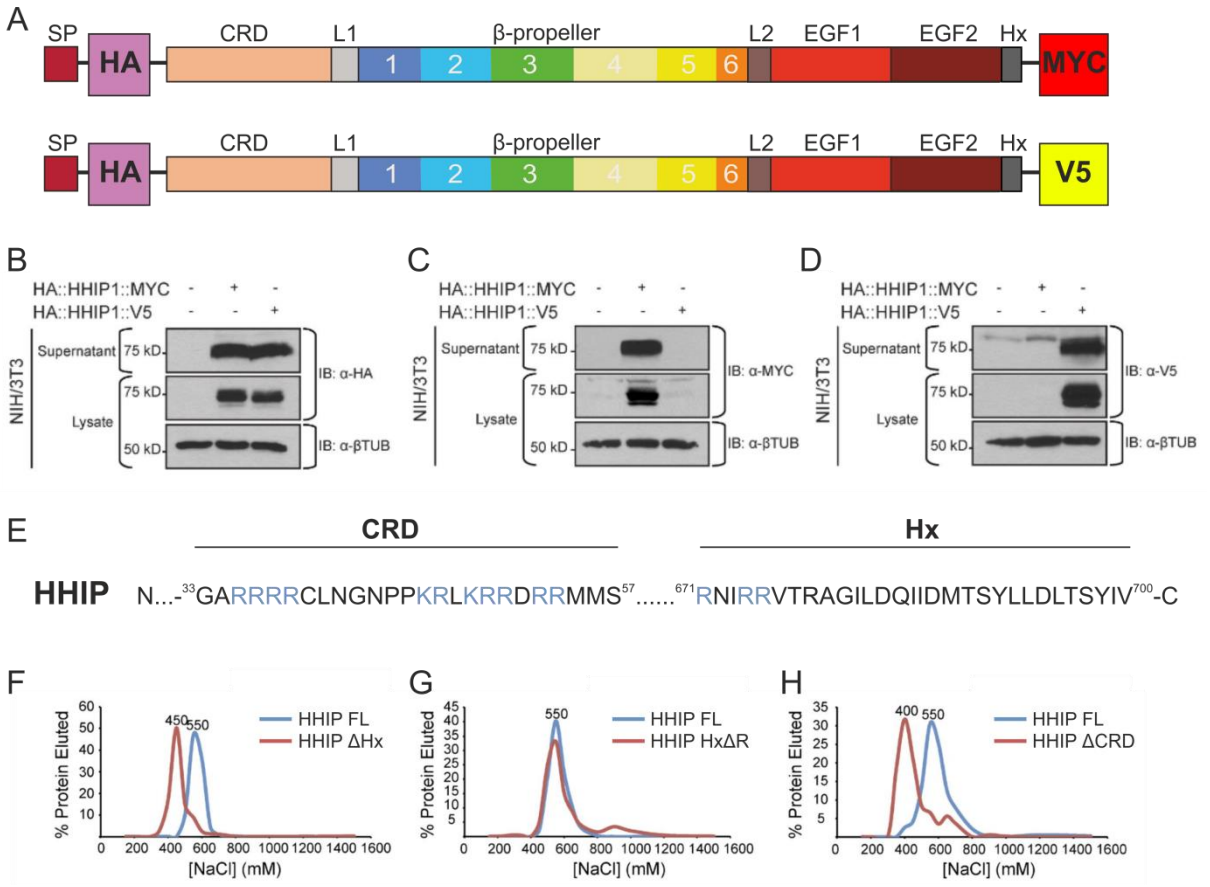


Figure 4.2 – Secretion of HHIP. A) HHIP FL (full-length) was cloned with an N-terminal HA tag and a C-terminal Myc tag (HA-HHIP-Myc) or V5 tag (HA-HHIP-V5). (SP - Signal peptide, removed during secretion to yield an N-terminal HA tag); B-D) Immunoblotting of NIH3T3 cell lysates and supernatants expressing HA-HHIP-Myc and HA-HHIP-V5 tags, comprising detection with B) anti-HA antibody, C) anti-Myc antibody; D) anti-V5 antibody. An anti-β-Tubulin antibody was used as a loading control for cell lysates; E) HHIP sequence motifs proposed to interact with GAGs of the PG layer; F-H) Heparin-agarose chromatography of HA-HHIP constructs. In each case, HHIP FL (blue) was used as a positive control; F) HHIP ΔHx (deletion of the C-terminal 30 residues), G) HHIP HxΔR (alanine mutations of the arginine residues at the N-terminus of Hx), H) HHIP ΔCRD (deletion of the HHIP-N CRD). Mutant traces are shown in red. (Experiments performed by AM Holtz, University of Michigan).

The structure of HHIP-N in both an apo- and a SOS-bound form were determined via X-ray crystallography (as outlined in Chapter 3). The CRD region of the apo-structure of HHIP-N is folded, with 5 α-helices and 6 disulphide bonds observed. Certain portions of the N-terminal GAG-binding sub-domain are ordered in this structure due to the disulphide bonds in these regions, but it is for the most part unstructured (Fig. 4.3A; dashed line). Addition of GAG molecules causes a conformational change in the N-terminal sub-domain, leading to ordering of residues 49-60. Residues 50-58 fold into the 'SOS-binding helix' (Fig. 4.3B), and a number of basic residues within the N-terminal sub-domain are involved in SOS binding. The only residue within the SOS-binding helix that makes contacts with more than one SOS molecule is R47. The only SOS-binding residue within the SOS-binding helix is R51 (Fig 4.3C). The flexibility of this region is reflected by the crystallographic B-factor, with residues in this helix displaying higher values than those in the CRD helices. The SOS molecules possess the largest B-factor values for the entire structure, representative of multiple conformations (Fig. 4.3D). Therefore, a cluster of basic residues within an N-terminal sub-domain of HHIP-N causes folding of this region upon binding to GAGs.

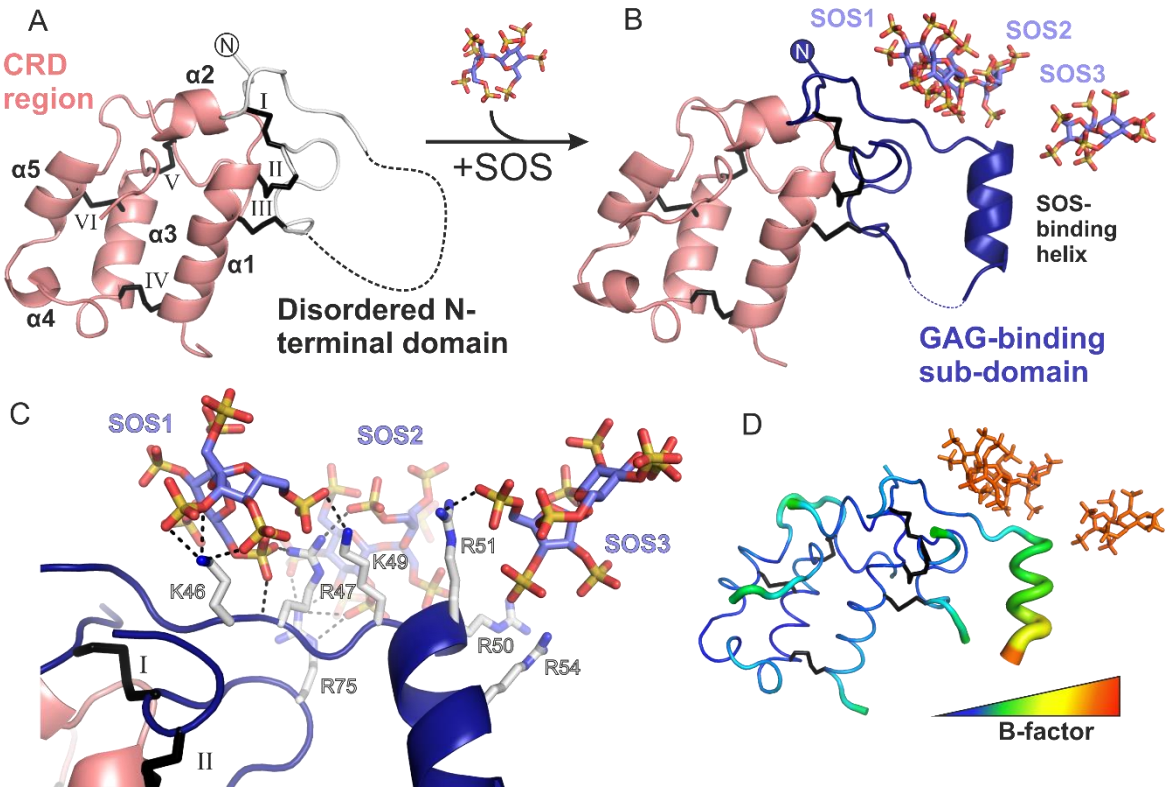


Figure 4.3 – HHIP-N:SOS interactions. A) The HHIP-N-apo structure comprises the CRD region (salmon) with 5 α-helices (α1-5) and 6 disulphide bonds (I-VI). The residues of the N-terminal GAG-binding sub-domain are only partially ordered in this structure (white backbone), with the remainder of the domain being disordered (dashed line); B) The addition of GAGs (such as molecular mimic SOS) to HHIP-N causes folding of the N-terminal GAG-binding sub-domain (blue). The CRD fold remains unchanged (salmon). Three SOS molecules are coordinated to this GAG-binding sub-domain (SOS 1-3); C) Close-up view of HHIP-N:SOS interactions. The side chains of residues coordinating SOS molecules are displayed, with hydrogen bonds shown as dashed black lines. It is important to note that two SOS molecules come from symmetry contacts, and each SOS is bound by all contacts shown; D) HHIP-N:SOS complex structure, coloured according to B-factor using a rainbow spectrum. The thickness of the protein backbone is also used to represent B-factor values. The B-factor ('temperature factor') is a measure of the mean squared deviation of the atomic positions ($B=8\pi(u^2)$, where u is the atomic displacement and can be calculated using diffraction intensity values) (Rupp, 2009).

The interaction between HHIP-N and GAG molecules was validated using a heparin binding assay as used for Fig. 4.2. Basic residues from HHIP-N involved in contacts with SOS were mutated to alanine residues within the context of HHIP FL and heparin binding was analysed (Fig 4.4A). These mutations reduce the affinity of HHIP for heparin and verify that the HHIP-N:SOS structural model reflects physiological HHIP-N:GAG interactions (Fig. 4.4B).

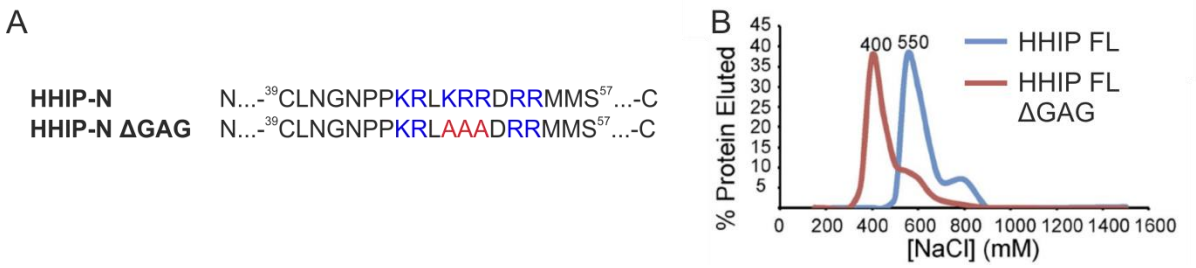


Figure 4.4 – HHIP-N GAG-binding site mutations. A) Residues in the N-terminal region of HHIP-N and those mutated to analyse the GAG-binding site. Basic residues are highlighted in blue. HHIP-N ΔGAG refers to a mutant with a K49A/R50A/R51A mutation (red); B) Heparin-agarose chromatography experiment of HHIP FL (red) and a HHIP FL mutant with the HHIP-N ΔGAG mutation (blue). (Experiment performed by AM Holtz, University of Michigan).

4.3.2 Purification and crystallisation of HHIP constructs

HHIP constructs were cloned and purified to further investigate HHIP:GAG interactions and identify GAG binding motifs in HHIP-C. HHIP-C was purified to homogeneity using SEC (Fig. 4.5A-B). Crystals were grown via co-crystallisation with an excess of either SOS or 30-mer Heparin (Fig. 4.5C-D). HHIP FL ΔHx co-eluted

from an S200 16/60 column alongside contaminants and was subjected to a third purification step using heparin affinity chromatography (Fig. 4.5E-F). This enabled separation of a single species (Fig. 4.5G-H). HHIP FL ΔHx could not be crystallised and all crystals obtained only contained HHIP-C. Whether the observed contaminant (observed at ~60 kDa in Fig. 4.5H in the lanes marked with blue and red asterisks) comprises proteolysis-generated HHIP-C would require MS analysis or N-terminal sequencing. Proteolytic cleavage in the linker between CRD and the β-propeller domain was reported previously (Bishop et al., 2009).

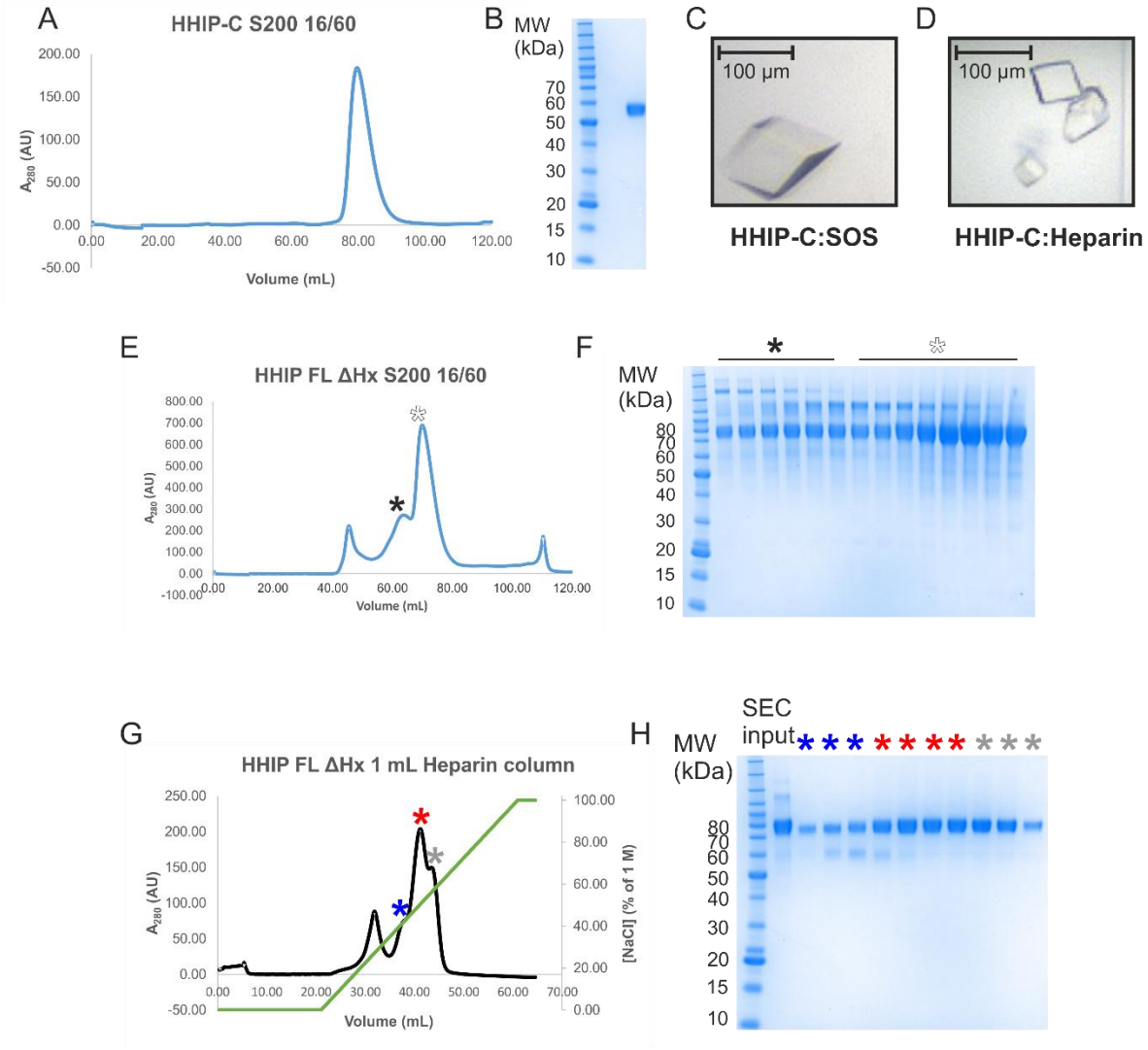
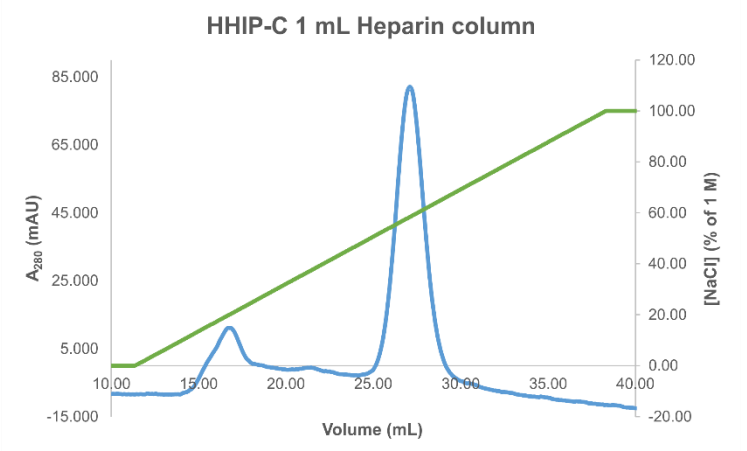


Figure 4.5 – Purification of HHIP constructs. A-D) Purification of HHIP-C. A) UV trace from SEC and B) purity assessment via SDS-PAGE of the pooled peak; C) Crystals of the HHIP:SOS complex; D) Crystals of the HHIP:C heparin complex; E-H) Purification of HHIP FL ΔHx, E) UV trace from SEC and F) analysis of purity by SDS-PAGE, with asterisks used to indicate the peaks in question; G) UV trace for purification of the major SEC peak (white asterisk, E and F) by heparin affinity chromatography. The green line represents an NaCl gradient from 10 mM to 1 M concentration; H) Analysis of heparin affinity-purified material purity by SDS-PAGE. (HHIP-N comprises residues 18-209, HHIP-C residues 210-700 and HHIP FL ΔHx residues 18-670). See Fig. 3.1 and section 2.3.3 for an outline of the purification strategy.

4.3.3 Characterisation of HHIP:GAG interactions

To characterise HHIP-GAG interactions in solution, heparin affinity chromatography was initially performed. As previously shown in this thesis, HHIP-N and HHIP FL ΔHx can bind to heparin columns (Figs. 3.1F and 4.5G). HHIP-C is also capable of binding to heparin, with an affinity comparable to that of HHIP-N (Fig. 4.6). Ergo, there is an additional GAG-binding region within HHIP-C. SPR equilibrium binding experiments with biotinylated GAGs were subsequently performed to further characterise HHIP:GAG interactions (Fig. 4.7 and Table 4.1). HHIP constructs bind to three cell surface GAGS; heparin (Fig. 4.7A), HS (Fig. 4.7B) and CS (Fig. 4.7C). Values reported in Table 4.1 indicate that there are discrete GAG binding sites in both HHIP-N and HHIP-C, verified by the tighter binding with HHIP FL ΔHx. This is in agreement with heparin column results (see Fig. 4.6).



Peak elution

	[NaCl] (mM)
HHIP-N	510
HHIP-C	510
HHIP FL ΔHx	560

Figure 4.6 - HHIP-C Heparin column. UV trace for HHIP-C elution from heparin affinity chromatography (blue line). A gradient of NaCl from 10 mM to 1 M is shown (green line). The NaCl concentrations at which HHIP-N, HHIP-C and HHIP FL ΔHx elute from heparin affinity columns are shown alongside.

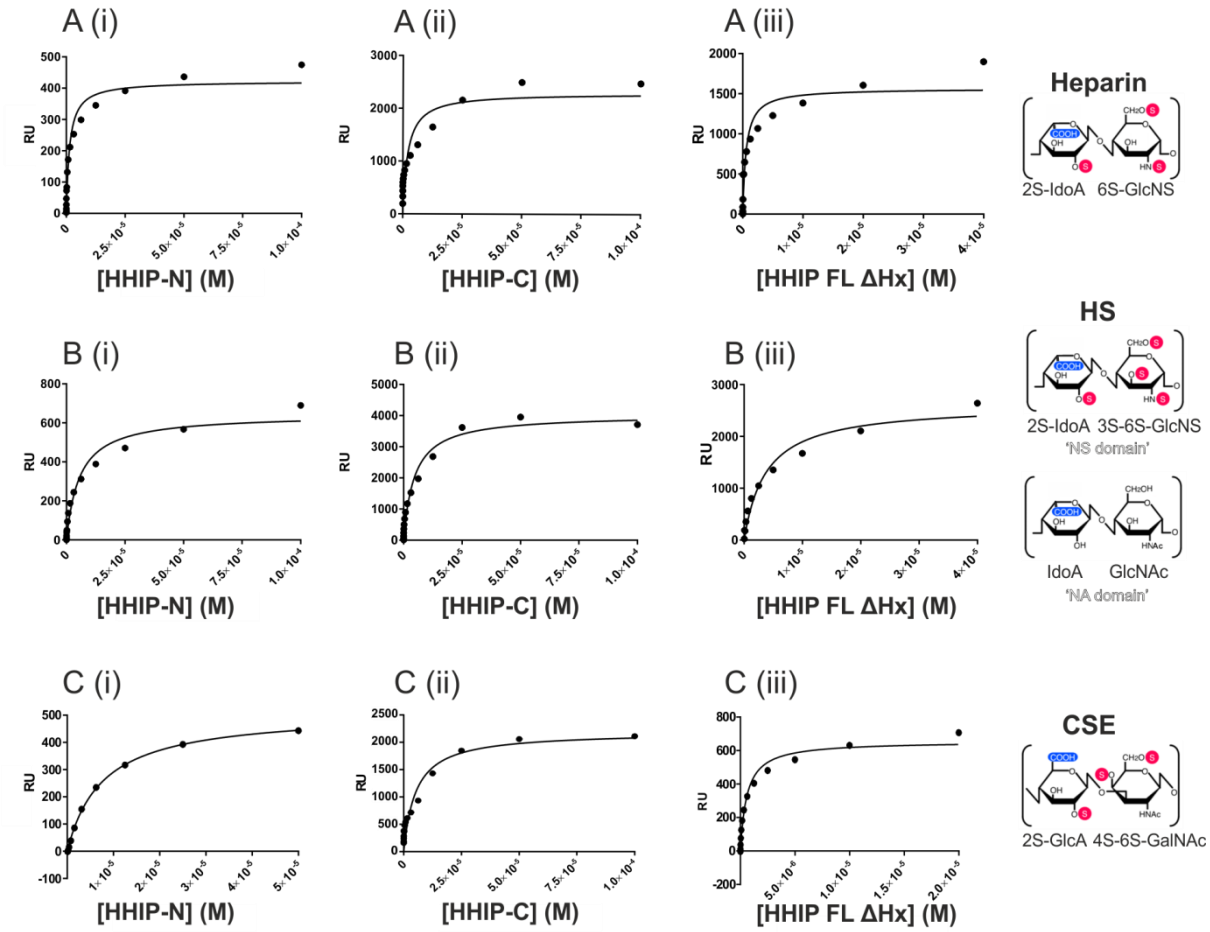


Table 4.1 - Protein:GAG K_D values (nM)

	HHIP-N	HHIP-C	HHIP FL ΔHx
Heparin	1400	2100	570
Heparan Sulphate	5700	4600	3800
Chondroitin Sulphate	7600	5300	590

Figure 4.7 and Table 4.1 – HHIP:GAG equilibrium SPR interaction study. Binding curves are displayed for interactions with A) heparin; B) HS; C) CSE. Sugar chemistries are displayed on the right of each row of curves (see section 1.3.1 for a description of GAG synthesis and NS and NA domains in HS). For each column, data are annotated as i) HHIP-N; ii) HHIP-C; iii) HHIP FL ΔHx. Dissociation constants (K_D) for each interaction are displayed in Table 4.1, in the same layout as the curves displayed. For each ligand, 100 RU were immobilised on the chip.

Regulation of HHIP secretion *in vivo* may involve cleavage events. To assess whether individual HHIP domains can bind to the PG layer under physiological conditions, immunofluorescence microscopy was used. Cells were transfected with HA-tagged HHIP constructs, fixed and subjected to cell-surface antibody staining. HHIP FL ΔHx, HHIP-C and HHIP-N are all retained at the cell surface (Fig. 4.8A-C). This technique was validated using a GPI-anchored RGMB with an HA tag, which shows well-defined cell surface staining (Fig. 4.8D). An unrelated growth factor, NT3, is known to bind to GAGs and was used as a negative control (Kanato et al., 2009). Antibody staining is not observed in response to NT3-1D4, indicating that the anti-HA primary antibody is specific for HA-tagged constructs and does not bind to the cell surface (Fig. 4.8E). Experimental design is shown in Fig. 4.8F.

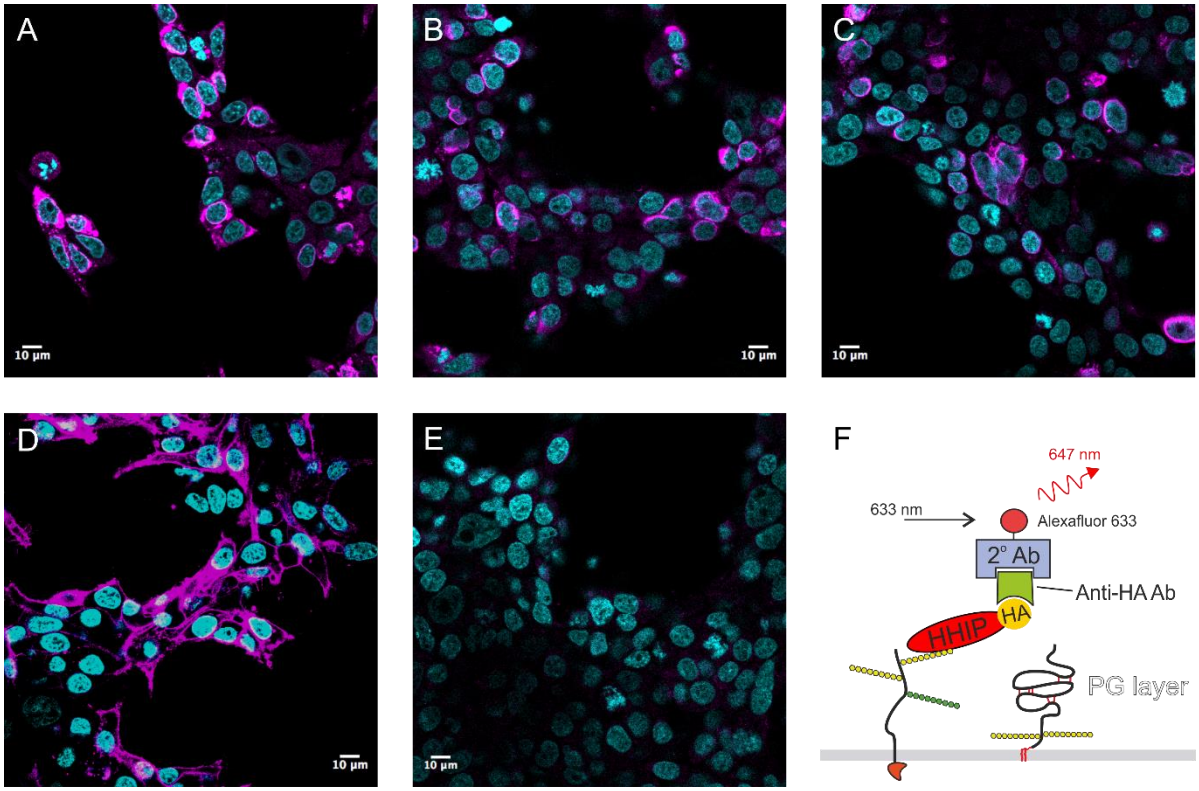


Figure 4.8 – Immunofluorescence microscopy of HHIP constructs. HEK293T cells were transfected with constructs comprising - A) HHIP FL ΔHx N-HA; B) HHIP-C N-HA; C) HHIP-N N-HA; D) RGMB GPI N-HA; E) NT3-1D4 – fixed, stained with an anti-HA antibody and an Alexa Fluor 633-labelled secondary antibody, washed and imaged. Antibody staining is displayed in magenta, nuclear stain Hoechst in cyan; F) Experimental design.

4.3.4 HHIP-C:SOS complex crystal structure

To identify residues responsible for the additional GAG-binding motifs in HHIP-C, a crystal structure of HHIP-C in complex with GAG mimic SOS was solved in space group $P3_121$ to a resolution of 2.40 Å. 2 HHIP molecules and four SOS molecule are observed in the ASU, with approximately equal cell dimensions as seen for the apo-form of HHIP-C (PDB ID: 2WFT) (Bishop et al., 2009). Each HHIP-C chain binds to 2 SOS molecules, which are coordinated between two symmetry-related HHIP-C molecules in a head-to-tail dimer. These form contacts between the β-

propeller of one chain and the 2 EGF repeats of a symmetry-related chain (Fig. 4.6A). Both HHIP-C chains are effectively identical, displaying a Cα RMSD of 0.0030 Å over 439 equivalent positions. This interaction involves hydrogen bonds with residues from the β-propeller of chain A (K277, R350, K514, K569) and from EGF repeats 1 and 2 (R635 and R651) and β-propeller blade 2 (R328, K329) of chain B (Figs. 4.9B and 4.11B). Data collection and refinement statistics for HHIP-C:SOS are shown in Table 4.2.

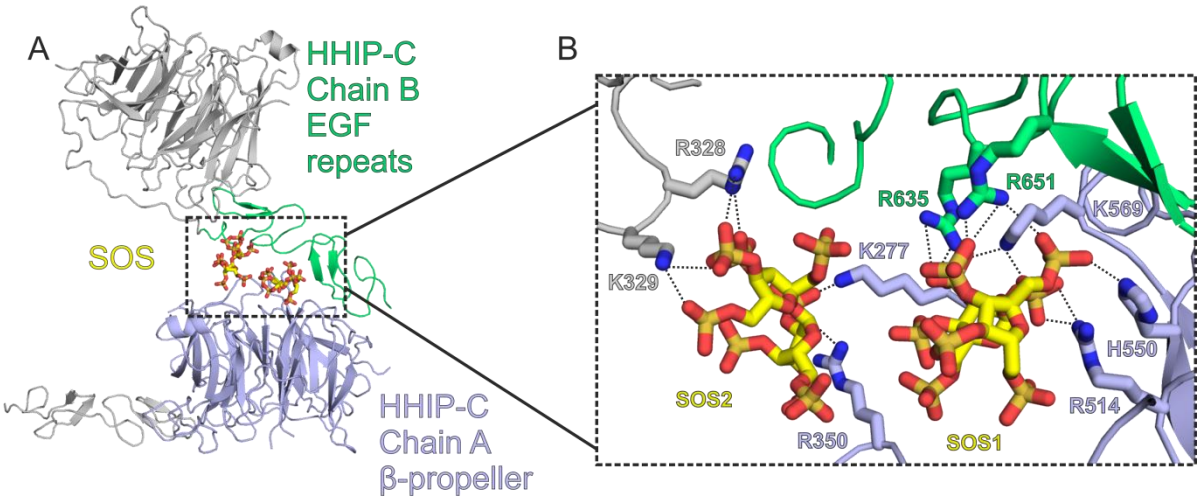


Figure 4.9 – HHIP-C:SOS complex crystal structure. A) The arrangement of symmetry-related HHIP-C chains (cartoon representation) about 2 coordinated SOS molecules (stick representation; boxed); B) Close-up view of the HHIP-C SOS-binding site. Residues from the β-propeller are coloured in slate, whilst those from the EGF repeats are coloured in green. Additionally, two hydrogen bonds are made via residues from blade 2 of HHIP-C chain B, coloured grey. Hydrogen bonds are displayed as dashed lines.

Table 4.2

HHIP-C:SOS	
Data collection	
Beamline	I04, DLS
Space group	$P3_121$
Unit-cell parameters	
a, b, c (Å)	100.6, 100.6, 311.9
α, β, γ (°)	90.0, 90.0, 120.0
No. of crystals / data sets	1 / 1
Wavelength (Å)	0.97949
Resolution (Å)	76.0-2.4 (2.44-2.40)
No. of unique reflections	72635 (3547)
Completeness (%)	100.0 (99.9)
Multiplicity	19.9 (20.0)
$\langle I/\sigma(I)\rangle$	21.4 (0.8)
R_{merge} (%)	9.7 (431.9)
R_{pim} (%)	2.2 (97.7)
$CC_{1/2}$	1.0 (0.3)
Refinement	
Resolution	58.0-2.4
No. reflections (test set)	72499 (3664)
R_{work} / R_{free} (%)	19.0 / 22.7
No. atoms	
Protein	6843
SOS	220
Water	37
Ca^{2+}	2
Glycerol	6
Mean B factor (Å^2)	
Protein	95.3
SOS	239.8
Water	70.8
Ca^{2+}	136.6
Glycerol	108.8
RMSD bond lengths (Å)	0.008
RMSD bond angles (°)	1.0
Ramachandran plot (%)	
Favoured	98.7
Allowed	1.3
Outliers	0

Table 4.2 - HHIP-C:SOS complex data collection and refinement statistics. Data collection values in parenthesis correspond to the highest resolution shell unless otherwise stated.

4.3.5 HHIP-C:heparin complex crystal structure

To further investigate the mode of HHIP-C:GAG interaction, HHIP-C was crystallised with a 30-mer heparin oligosaccharide and the crystal structure was determined in a C2 space group at a resolution of 3.35 Å (Table 4.3). The arrangement of HHIP-C chains in a head-to-tail dimer with a centrally coordinated heparin molecule is the same as in the HHIP-C:SOS complex structure (Fig. 4.10A). Heparin is coordinated by two distinct patches of basic surface residues on one face of the β-propeller and EGF repeat 1 and occupies a similar location to that of SOS (Fig. 4.10B). This GAG-binding site does not overlap with the acidic HhN-binding surface region of HHIP-C. The interface between HHIP-C and heparin is dominated by a number of hydrogen bonds with basic residues (Figs. 4.10B and 4.11A), and the oligosaccharide adopts a right-handed conformation as seen in other heparin-bound structure (Figs. 4.10D) (Waksman and Herr, 1998). All sugars within the oligosaccharide occupy the low-energy chair conformation (Agirre et al., 2015). Heparin residues I – V are regular, showing the expected helical pitch of 4 sugars/turn (Khan et al., 2010). A kink between sugars VI and VII distorts the helical pitch of the remainder of the molecule (Fig. 4.10D). A detailed summary of HHIP-C:GAG interactions is displayed in Figure 4.11.

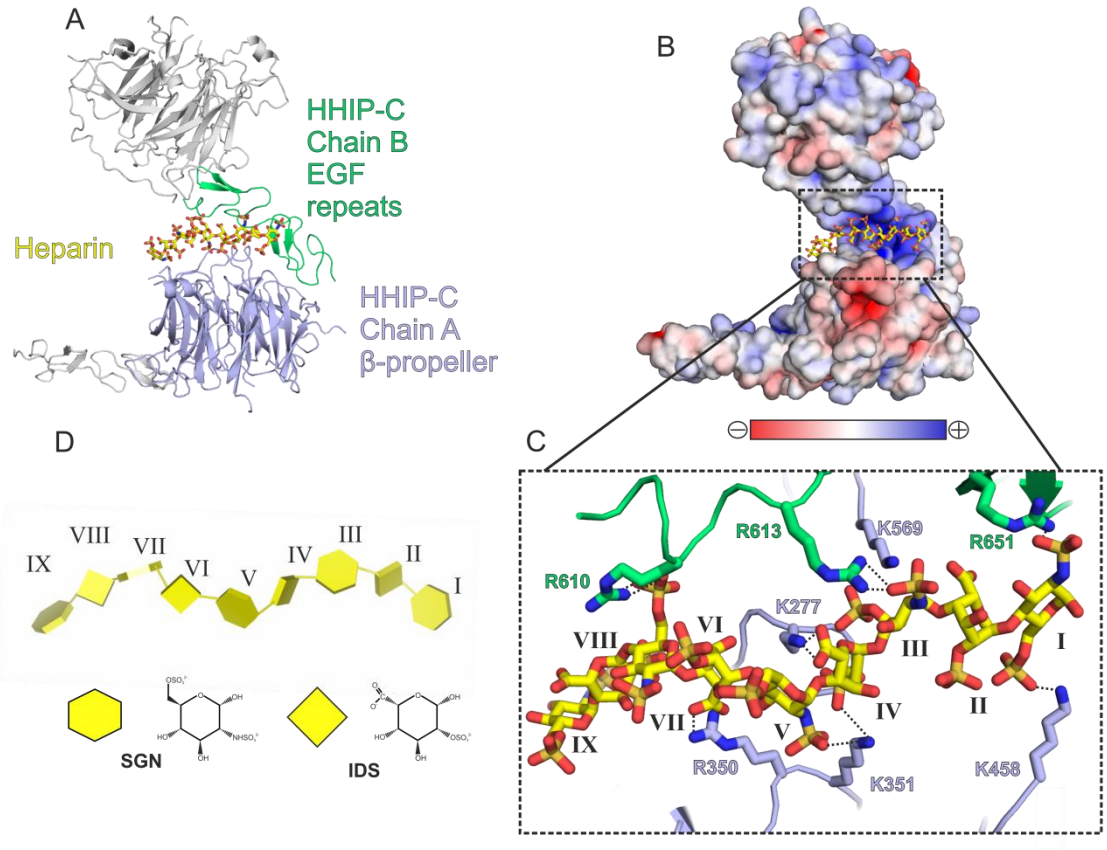


Figure 4.10 – HHIP-C:heparin complex crystal structure. A) Arrangement of two HHIP-C chains (A and B; cartoon representation) around a centrally coordinated heparin molecule (stick representation); B) Electrostatic surface potential shown from red (-9 kT/e) to blue (+9 kT/e) with heparin chain in stick representation (boxed); C) Close-up view of HHIP-C:heparin interactions, with residues coloured as in A (green – EGF repeats of HHIP-C chain B; slate – β-propeller of HHIP-C chain A). Sugar residues are numbered using Roman numerals (even, O2-sulphoiduronic acid; odd, N,O6-disulphoglucosamine) and major hydrogen-bonding interactions are shown using dashed lines; D) Representation of heparin chain assembled using GlycoBlocks (McNicholas and Agirre, 2017). Individual sugars are displayed as shown. SGN – N-O6-disulphoglucosamine; IDS, O2-sulphoiduronic acid.

Table 4.3

HHIP-C:heparin			
Data collection		**Refinement**	
Beamline	I03, DLS	Resolution	85.2–3.35
Space group	C2	No. reflections (test set)	37842 (1893)
Unit-cell parameters		R_{work} / R_{free} (%)	24.3 / 25.3
a, b, c (Å)	171.4, 105.6, 150.1	No. atoms	
α, β, γ (°)	90.0, 96.3, 90.0	Protein	10335
No. of crystals / data sets	1 / 1	Heparin	160
Wavelength (Å)	0.97630	Mean B factor (Å^2)	
Resolution (Å)	85.2-3.35 (3.41-3.35)	Protein	96.3
No. of unique reflections	38203 (1859)	Heparin	189.8
Completeness (%)	99.3 (98.9)	RMSD bond lengths (Å)	0.009
Multiplicity	6.7 (6.8)	RMSD bond angles (°)	1.0
⟨*I*/*σ*(I)⟩	5.2 (0.7)	Ramachandran plot (%)	
R_{merge} (%)	22.8 (232.6)	Favoured	98.8
R_{pim} (%)	9.5 (95.5)	Allowed	1.2
$CC_{1/2}$	1.0 (0.3)	Outliers	0

Table 4.3 - HHIP-C:heparin complex data collection and refinement statistics. For data collection statistics, values in parenthesis correspond to the highest resolution shell unless otherwise stated.

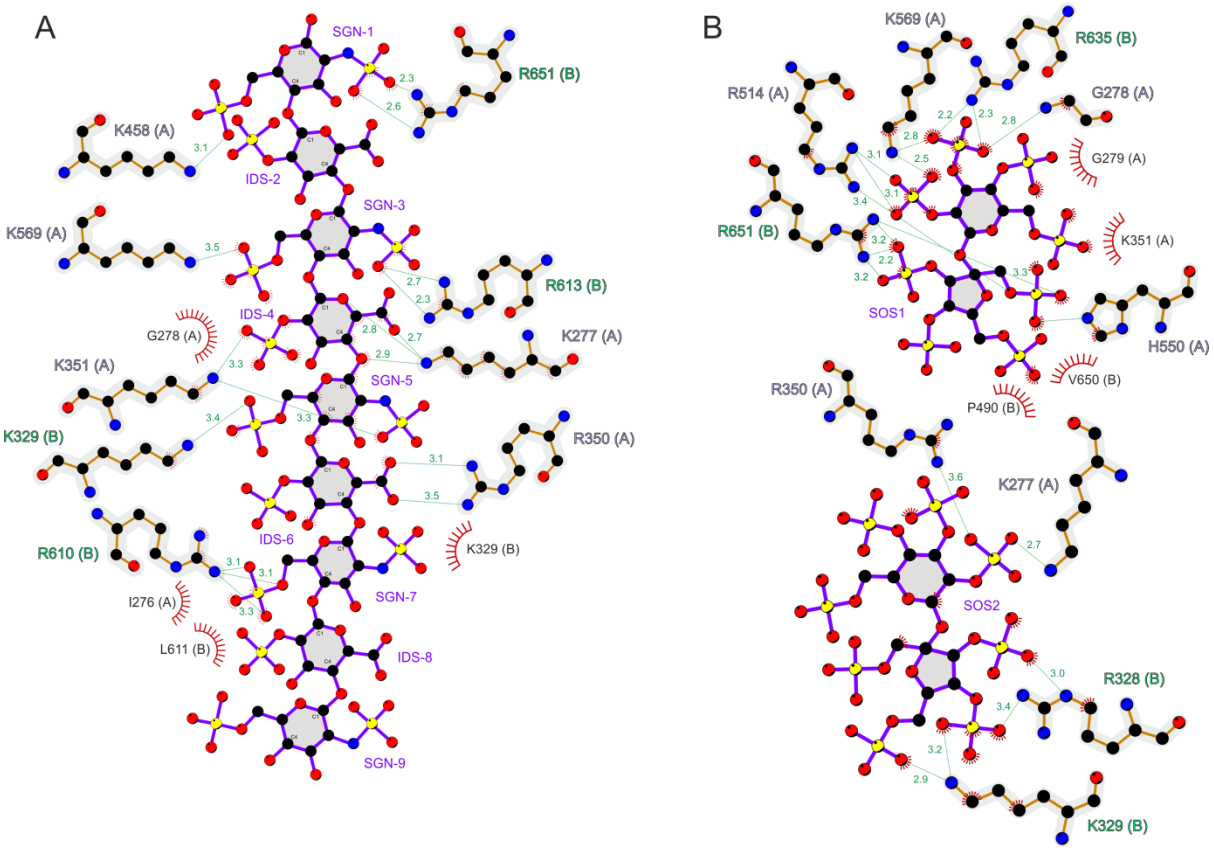


Figure 4.11 – HHIP-C:GAG interaction schematic. Residues contacting A) heparin and B) SOS are displayed. Atoms are coloured as follows: N, blue; O, red; S, yellow; C, black. Covalent bonds are displayed, coloured violet (GAG) and orange (HHIP-C), and brackets after amino acids correspond to HHIP-C chains as designated in Figs. 4.9 and 4.10. Hydrogen bonds are displayed as green dashed lines and labelled with lengths in Å. Red 'eyelashes' correspond to hydrophobic interactions. SGN, N,O6-disulphoglucosamine; IDS, O2-sulphoiduronic acid. Figure calculated using PDBSUM (Laskowski, 2001).

4.3.6 GAG-induced changes to HHIP-C

Both SOS and heparin are coordinated by the same face of the HHIP-C β-propeller. A major difference between both GAG-bound structures and the apo-structure is seen in the conformation of the loop linking blades 4 and 5 (residues 473-490). This is indicated by a curved arrow in Figs. 4.12A-C. Interestingly, no residues from this loop are identified as interacting with GAG ligands. Additionally, a loop in blade 1 (black dashed oval; Fig. 4.12A-C) and a helix in blade 6 (red dashed oval; Fig. 4.12A-C) displays a reduction in B-factor in both GAG-bound structures versus the HHIP-C apo-structure. The loop from blade 1 contains residue K277, used in interactions with a GAG molecule (Fig. 4.11). The helix from blade 6 contains residue K569 which is within hydrogen bonding distance of both the heparin and SOS molecules (Fig 4.11). GAG ligands have the highest B-factors of the HHIP-C:GAG structures (Fig. 4.12A and C). Therefore it is possible that a residue in this helix contacts the ligands in these structures, due to binding in multiple conformations (Bell et al., 2013; Veverka et al., 2009). The fold of the EGF repeats is conserved.

On the opposing face of the β-propeller, a loop of high B-factor involved in ShhN binding occupies the same conformation in both the HHIP-C:SOS and HHIP-C-apo structures (arrow; Fig. 4.12D and E). This loop is partially disordered in the HHIP-C:heparin complex structure and in a different conformation (arrow; Fig. 4.12F). HHIP-C:SOS and HHIP-C-apo crystallise in the same space group with similar unit cell dimensions (space group $P3_121$; a, b, c, α, β, γ = 101 Å, 101 Å, 306 Å,

90°, 90°, 120° for HHIP-C-apo; a, b, c, α, β, γ = 101 Å, 101 Å, 312 Å, 90°, 90°, 120° for the HHIP-C:SOS complex – see Table 4.2). The HHIP-C:heparin complex crystallises with larger unit cell dimensions in space group C2 (Table 4.3). Thus, deviations between HHIP-C:heparin and the HHIP-C:SOS/HHIP-C-apo structures that do not contact GAG ligands are probably due to differences in crystal packing. In summary, GAG ligands cause limited conformational changes in the HHIP-C structure and act to rigidify certain structural elements on the β-propeller surface.

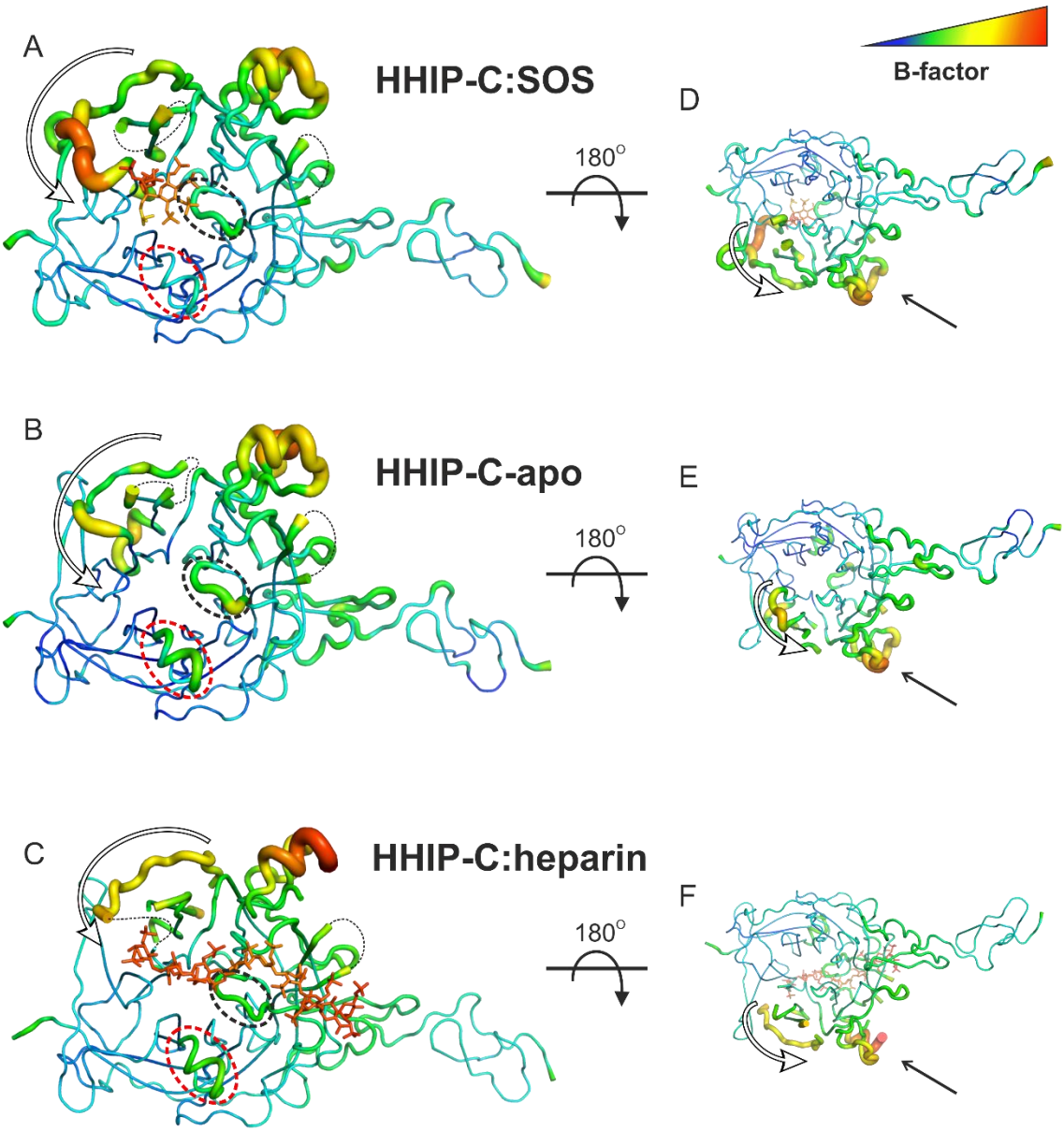


Figure 4.12 – GAG-induced changes to the HHIP-C structure. (Down page) A-C) GAG ligand-binding face of HHIP-C; D-F) view of HHIP-C rotated through 180° in the X-axis versus A-C). A and D – HHIP-C:SOS; B and E – HHIP-C-apo; C and F – HHIP-C:heparin. All structures are rainbow coloured according to a B-factor gradient, and the roles of indicators of structural indicators (arrows and dashed ovals) are discussed in the text (section 4.3.6). Additionally, thickness of the backbone cartoon representation is used as a secondary indicator of B-factor value (HHIP-C-apo PDB ID: 2WFT).

4.3.7 Crystalline oligomeric states of HHIP-C

The HHIP-C:heparin complex crystallises with 3 HHIP-C molecules in the ASU (Fig. 4.13A). This is in contrast to both the HHIP-C:SOS complex (Fig. 4.13B) and the HHIP-C apo-form (Fig. 4.13C), both of which form head-to-head dimers in the ASU. Previous analysis by MALS did not suggest the presence of dimers in solution, and this was assumed to be an artefact of crystallisation (Bishop et al., 2009). The HHIP-C:heparin complex crystal form containing three chains in the ASU more likely represents a 'dimer and a half' arrangement, given how chains are arranged in other HHIP structures. The head-to-tail interaction seen as a crystal contact mediated by the heparin chain (Fig. 4.10A) is also observed within the ASU. It is likely that heparin also mediates this interaction; unfortunately, the electron density map was not of a high enough quality to build a sugar molecule (Supplementary Fig. 4.1). The 'head-to-head' interface between β-propellers observed in all crystal structures comprises a buried surface area of 981 $Å^2$ and this may well correspond to a true oligomeric state in solution (Krissinel and Henrick, 2007). The oligomeric solution states of HHIP constructs were therefore analysed further to investigate their potential existence and gain insight into related functional roles (section 4.3.8).

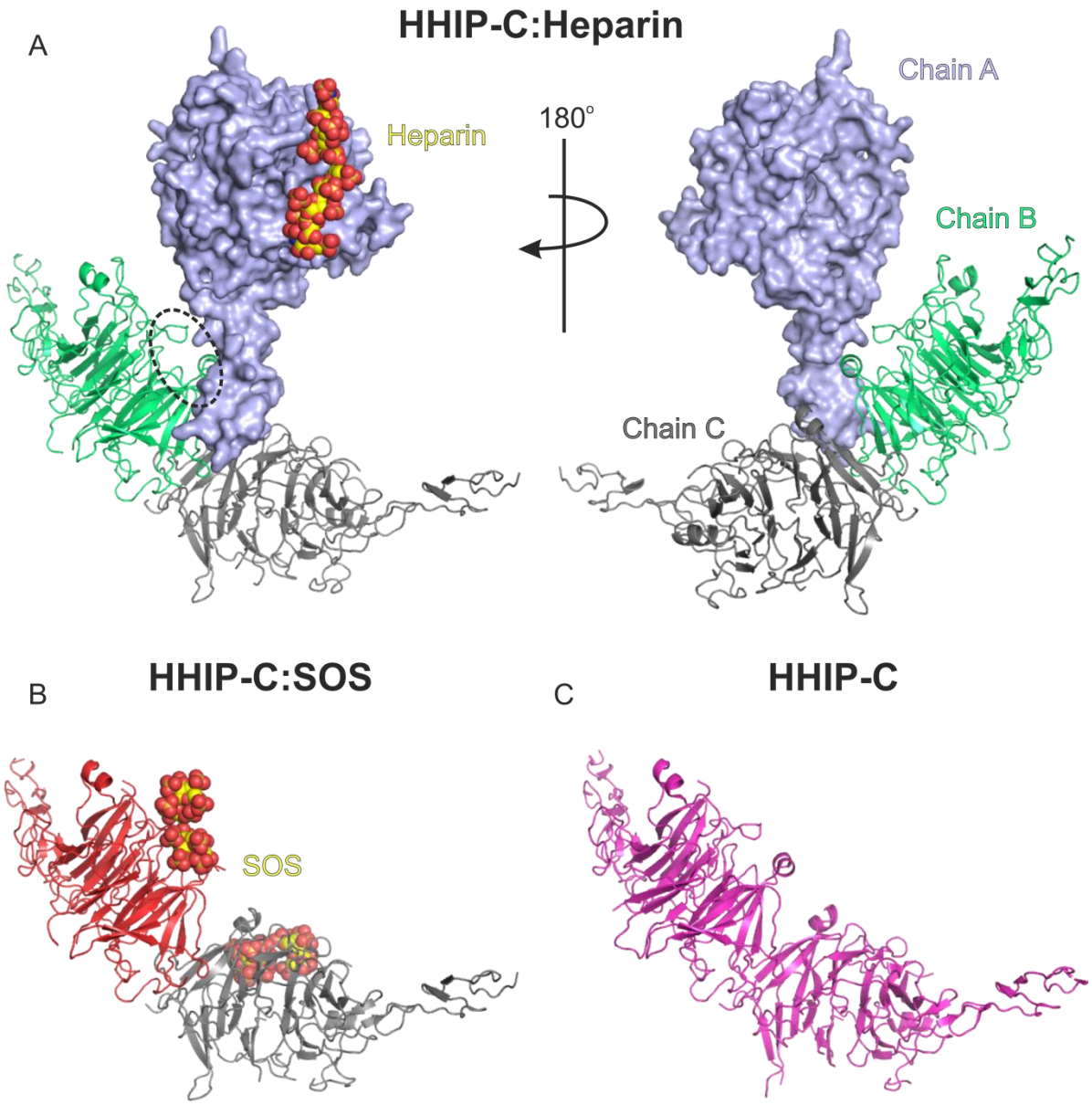


Figure 4.13 – HHIP-C ASU contents. A) HHIP-C:heparin complex structure. Heparin is displayed as spheres. Chains A and B are coloured as in Fig. 4.10. A region at which heparin is likely to be present in the ASU is indicated by a dashed oval (see Supplementary Figure 4.1); B) HHIP-C:SOS complex structure. SOS is shown as spheres; C) HHIP-C apo-form (PDB ID: 2WFT).

4.3.8 Biophysical analysis of HHIP oligomerisation

MALS experiments on SEC-purified HHIP-N resulted in a major peak (~90% of eluted material) corresponding to a monomeric molecular weight of ~22 kDa. A minor peak consistent with a dimer of ~42 kDa is also visible (Fig. 4.14). This ratio of peak heights was conserved over the concentration range 48 – 192 μM. Therefore, a mode of Hh signalling inhibition enhancement by HHIP via both protein-protein and GAG-induced oligomerisation may represent a mode of regulation, also suggested by HHIP-C:GAG crystal structures.

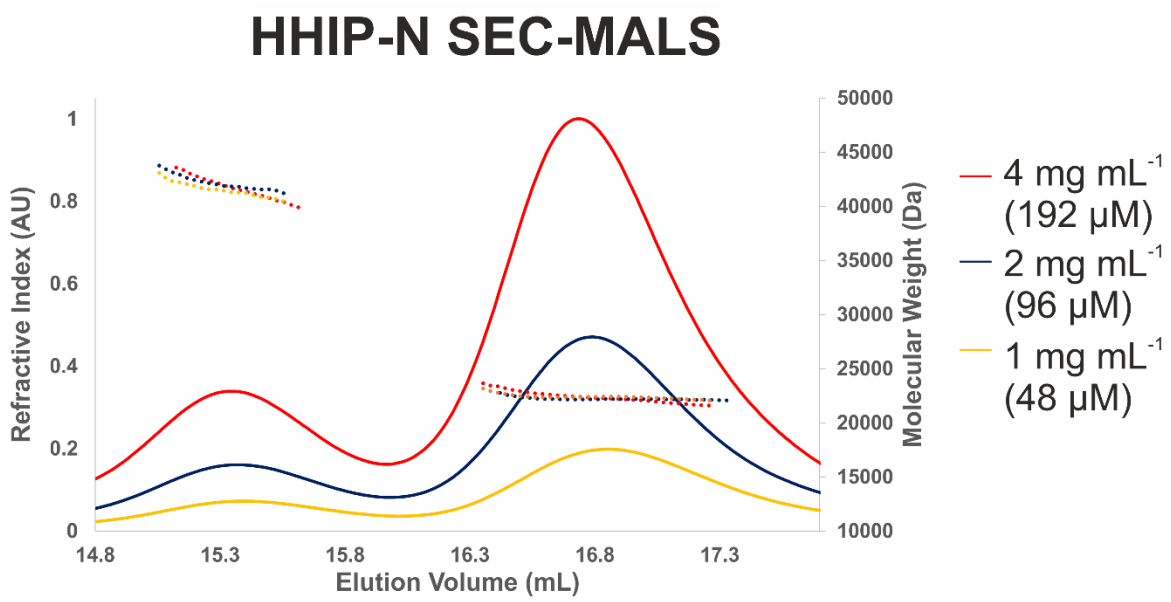


Figure 4.14 – Analysis of the oligomeric state of HHIP-N. SEC coupled with Multi-Angle Light Scattering (SEC-MALS) was used to analyse purified HHIP-N. SEC-purified HHIP-N was injected onto an S200 10/30 column over a series of concentrations. The eluted peaks are shown as solid lines corresponding to refractive index readings, whilst the corresponding calculated molecular weights are shown as dotted lines for each peak.

MALS experiments were confirmed using sedimentation velocity AUC (SV-AUC) experiments with HHIP-N at 34 μM, again resulting primarily in a monomeric species (peak at ~2.5 S; 21 kDa). A dimer is also visible (peak at ~4 S) (Fig. 4.15A), despite AUC being performed at a lower concentration than MALS experiments. Surprisingly, the addition of 30-mer heparin does not assemble large HHIP complexes but instead gives a major peak at ~4 - 5 S, corresponding to a species of 50 kDa (Fig. 4.15B). This suggests a model in which two HHIP-N molecules are bound to one 30-mer heparin molecule. Minor species at higher molecular weights (~100 kDa) are visible, as are monomers, likely representing transient interactions. The 30-mer heparin alone shows a population at ~9 kDa, in agreement with the calculated theoretical molecular weight (Fig. 4.15C). The spread of observed values likely represents difficulties in accurately estimating frictional coefficients for such an elongated and negatively charged molecule (Khan et al., 2010).

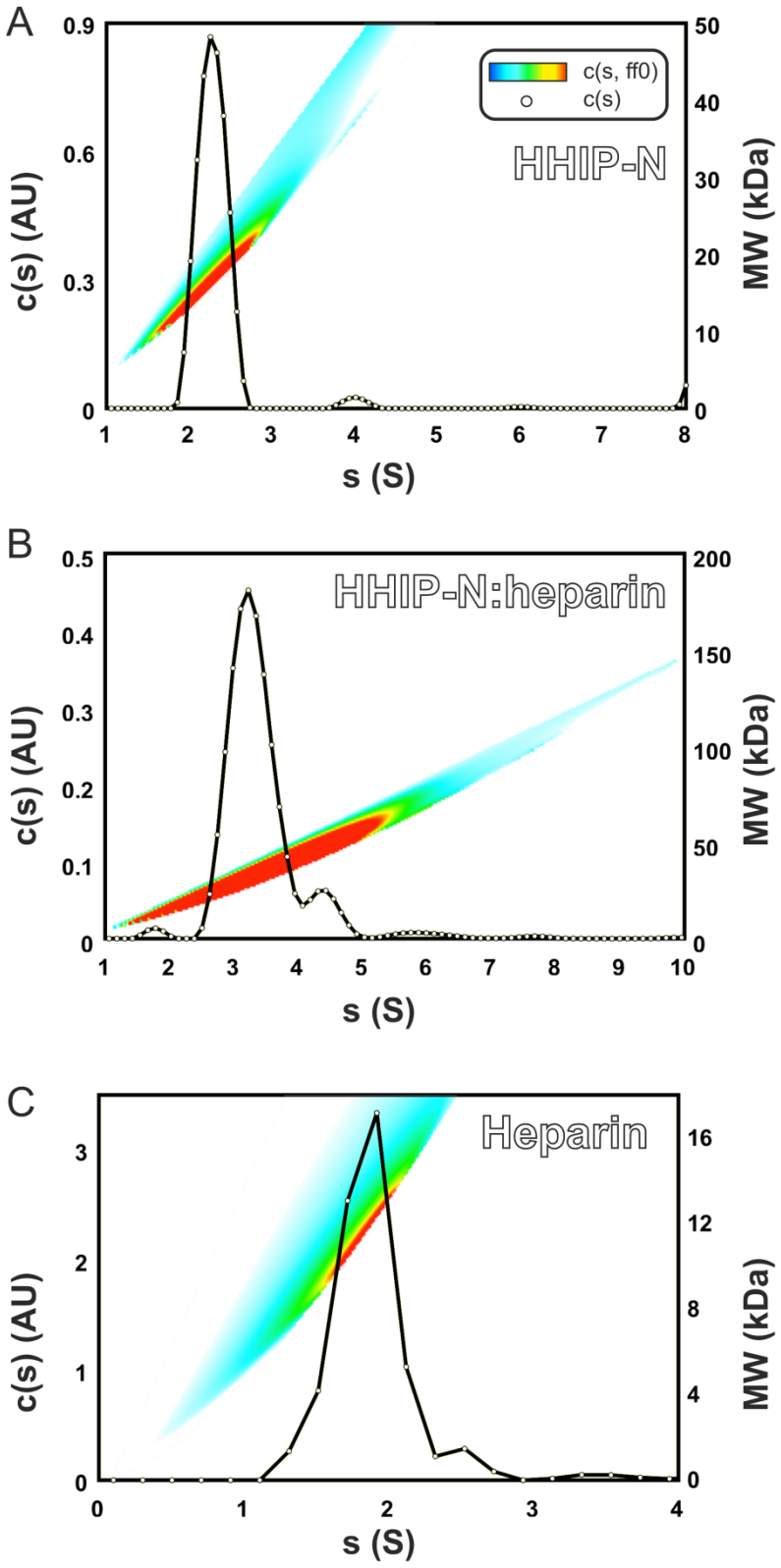


Figure 4.15 – SV-AUC experiments of HHIP-N with heparin. Distributions of c(s) and calculated molecular weight are shown for A) HHIP-N; B) HHIP-N:heparin 30-mer; C) 30-mer heparin. Experiments were performed at a HHIP-N concentration of 0.7 mg mL^{-1} (34 μM). The relative populations of species at calculated molecular weights is displayed as a heat map, coloured light blue to red.

The HHIP-C:GAG assemblies proposed by complex crystal structures (Figs. 4.9, 4.10 and 4.13A) suggests the potential for HHIP to undergo oligomerisation induced via both GAG-binding and self-association. SV-AUC experiments at a concentration of 19 μM show that HHIP-C is predominantly monomeric (~53 kDa; ~4 S) in solution but a shift to dimers of ~100-120 kDa (~5 S) is seen, as well as a small population of higher order oligomers at ~150-200 kDa (~6 S) (Fig. 4.16A). Upon addition of heparin, a range of rapidly-sedimenting larger molecular weight species was observed, indicating the capacity of HHIP-C to form higher-order oligomeric GAG-induced solution assemblies. These vary from more transient populations between 100 kDa and 1 MDa to more stable distributions from 1 to 8 MDa, as judged by the rapid increase in the size of complexes over the latter range (Fig. 4.16B). Heparin alone gave a single monomeric species, sedimenting at ~1.6 S, in agreement with previous published results (Fig. 4.16C) (Khan et al., 2010).

Combining this AUC data with crystal packing it is possible to propose a mode of GAG-induced clustering. Within this model, heparin forms contacts between the β-propeller and EGF repeats of adjacent HHIP-C chains. These are held in solution via GAG-free HHIP dimers contacting the GAG-bound HHIP chains (Fig. 4.16D-F). This could lead to a regularised arrangement of HHIP molecules along GAG chains *in vivo*, modulated by protein:protein and protein:GAG interactions.

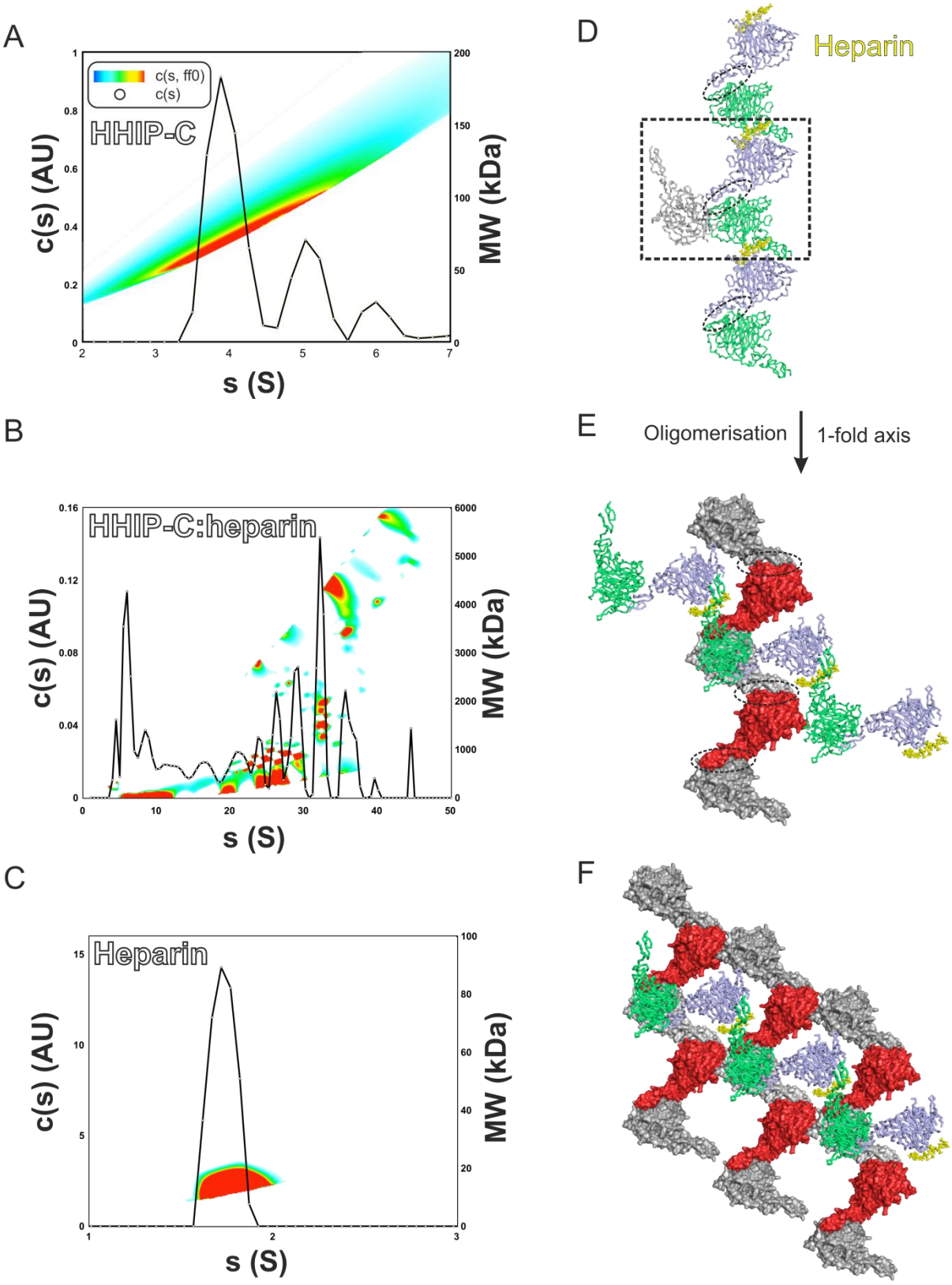

A
c(s, ff0)
c(s)
HHIP-C
c(s) (AU)
MW (kDa)
s (S)
B
HHIP-C:heparin
c(s) (AU)
MW (kDa)
s (S)
C
Heparin
c(s) (AU)
MW (kDa)
s (S)
D
Heparin
E
Oligomerisation
1-fold axis
F

Figure 4.16 – HHIP-C heparin-mediated oligomerisation (above). SV-AUC experimental curves for A) HHIP-C alone; B) HHIP-C:heparin and C) 30-mer heparin alone. Experiments were performed at a concentration of 19 μM, and graphs are displayed as in Fig. 4.15; D-F) Proposed model for HHIP-C oligomerisation *in vivo* (from crystal packing). D) Heparin-linked head-to-tail dimers are assembled along a 2-fold axis. Heparin is displayed as yellow sticks. A dashed oval indicates a second site to which heparin is also likely to be coordinated (see Supplementary Figure 4.1). The contents of the asymmetric unit are indicated via a dashed box; E) 1-fold axis (in surface representation), assembled via heparin-mediated contacts. This is linked to the arrangement of HHIP-C chains in D) by a head-to-head dimer (shown in grey in D); F) Extension to other asymmetric units on the same axis as in E). A combination of protein:protein-mediated and protein:GAG-mediated interactions may regulate HHIP-C assemblies at the cell surface.

4.3.9 Modelling of the HHIP C-terminal helix (Hx)

To gain insight into the function of the Hx, the only motif of HHIP not studied thus far by biophysical methods, a variant of HHIP-C including these 30 C-terminal amino acids was cloned with an N-terminal His_6-tag (HHIP-C/Hx N-H_6). Despite being able to purify, crystallise and solve an X-ray structure of this construct, extra electron density could not be observed for the Hx (data not shown). To analyse if the Hx was present, HHIP-C and HHIP-C/Hx N-H_6 were treated with amidase PNGase F and separated via SDS-PAGE. PNGase F removes entire glycan chains (Tarentino and Plummer, 1994). Comparison of both proteins with and without complete deglycosylation on a low percentage SDS-PAGE indicates the likely absence of the Hx (Fig. 4.17), suggesting that the C-terminal region was proteolytically cleaved during expression.

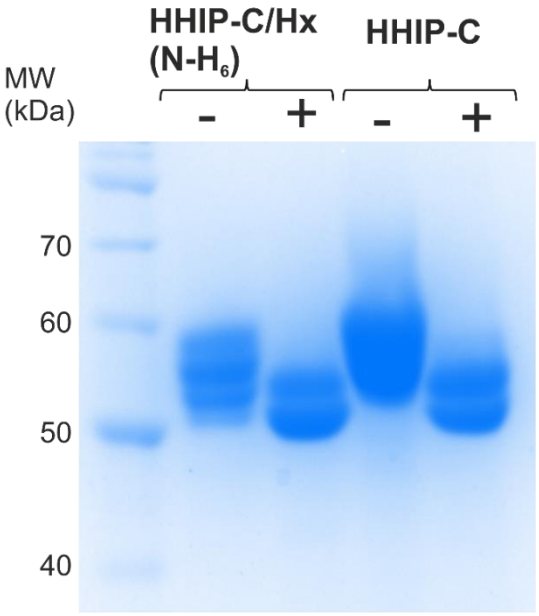


Figure 4.17 – PNGase F treatment of HHIP-C expressed with Hx. HHIP-C expressed with the C-terminal 30 residues and an N-terminal His_6-tag (HHIP-C/Hx N-H_6) and HHIP-C alone were deglycosylated for 48 hours at 25°C with PNGase F and separated according to molecular weight on an 8% SDS-PAGE gel (lanes marked: **+**). Untreated protein was run as a control (lanes marked: **-**).

Studies have revealed that full-length HHIP is secreted from cells, and that the Hx is therefore not a transmembrane domain (Coulombe et al., 2004; Kwong et al., 2014) (Fig. 4.2B-D). Modelling of the Hx predicts a long amphipathic helix (Fig. 4.18A). This is indicated by a basic cluster of amino acids at the N-terminus of the helix (dashed circle, left), with hydrophobic amino acids projecting from the same face. An acidic region is seen on the opposing face of the helix (see dashed oval in the right-hand panel of Fig. 4.18A). Therefore, the Hx is unlikely to be membrane-embedded. Hydrophobicity analysis reveals that the basic/hydrophobic face possesses a hydrophobic stretch of amino acids (Fig. 4.18B) and that flexibility due to a glycine residue (G680) may break the helix between these basic and hydrophobic clusters (Fig. 4.18C). This may play a role in placing the hydrophobic face of Hx perpendicular to the remainder of HHIP and using it to interact with the membrane bilayer. It is also possible that it acts like a leucine zipper, adding another region to regulate HHIP oligomerisation ((Krylov and Vinson, 2001)). The basic residues are conserved across HHIP species, with the exception of zebrafish (Fig. 4.18D), and experimental evidence does not

suggest that these residues contribute to heparin binding (Fig. 4.2F and G). Despite this, the hydrophobic region is conserved, meaning that the functional role of the Hx is likely to be related to these residues.

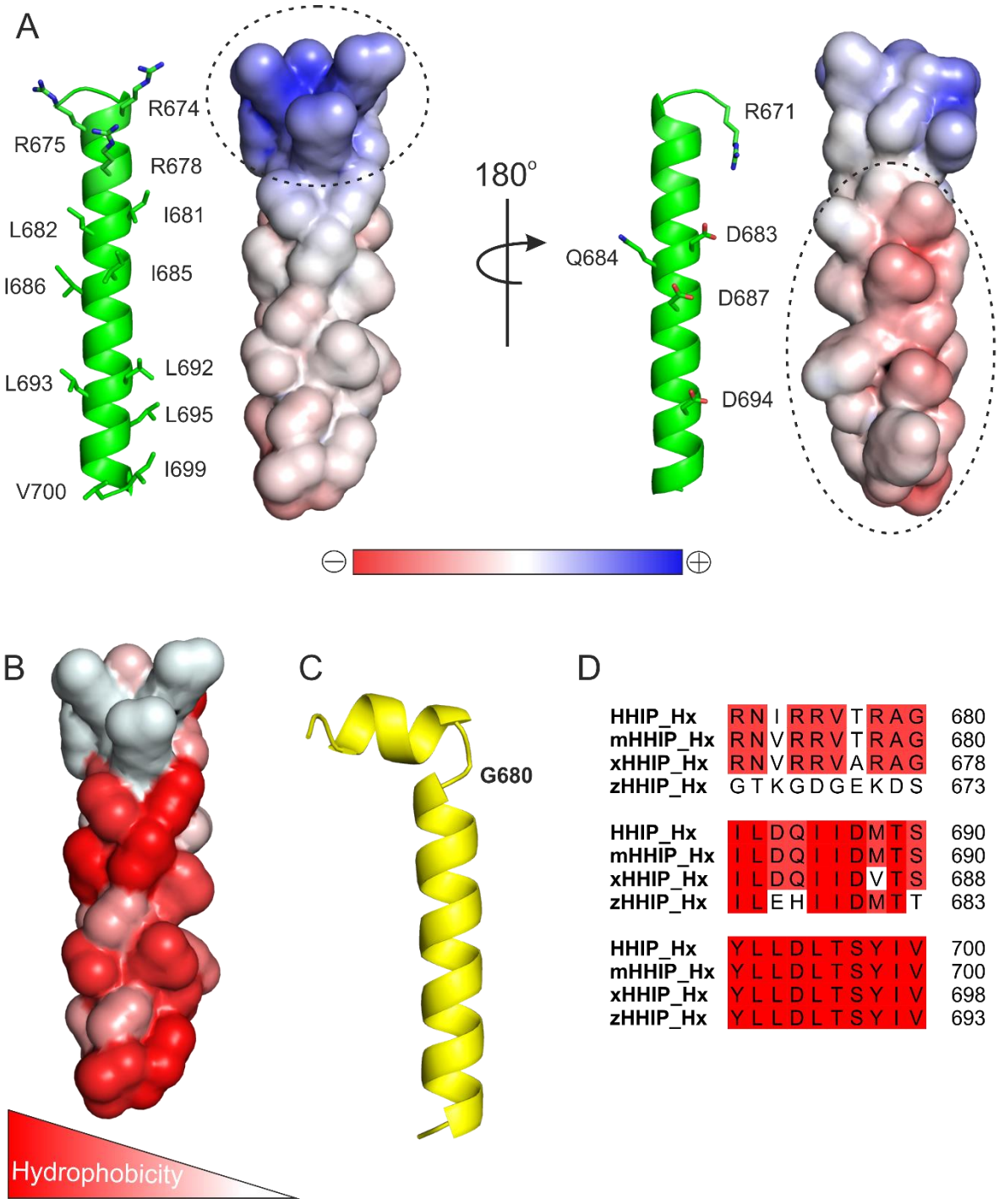


Figure 4.18 – Modelling of the C-terminal helix (Hx) of HHIP. A) Amphipathic nature of Hx. Electrostatic surface potential (plotted from -8 kT/e to 8 kT/e) of two sides of the helix with residue contributions shown alongside in cartoon representation; B) Potential membrane-binding face of Hx, coloured according to hydrophobicity; C) Alternative model of Hx, in which flexibility around residue G680 results in a different conformation; D) Sequence alignments of Hx from different species (human, mouse, *Xenopus*, zebrafish). Hx was modelled using the Robetta full-chain protein structure prediction server with the standard Rosetta comparative modelling) protocol (Raman et al., 2009).

4.4 Discussion

Secreted proteins frequently undergo regulation through components of the proteoglycan layer at the cell surface and within the extracellular matrix, which is largely facilitated by the GAG chains of proteoglycans. HHIP has previously been reported as secreted (Coulombe et al., 2004), but it was not known how this process was regulated. In this chapter, GAG-binding motifs have been identified within both the N- and C-terminal regions of HHIP. These interactions have been dissected biophysically and demonstrate a complex mode of HHIP:GAG complex regulation. SPR and heparin-affinity chromatography displayed approximately equivalent GAG binding affinities for HHIP-N and HHIP-C, enhanced in the context of the full HHIP architecture. Crystal structures of HHIP-N and HHIP-C solved with GAG mimic SOS and heparin identified residues involved in these interactions, and suggest a mode of HHIP oligomerisation regulated through protein:protein and protein:GAG interactions. Assembly in solution through these interactions was verified using SV-AUC, confirming that heparin enables HHIP oligomerisation into a diverse range of high molecular weight complexes. HhN, the binding partner for HHIP, undergoes comparable assembly into large multimeric complexes and this is important for appropriate signalling (Chen et al., 2004; Eaton, 2008). GAG-induced multimerisation may represent a further method via which HHIP fine-tunes regions of Hh signalling through antagonism of HhN:Ptc interactions.

4.4.1 Modes of HHIP oligomerisation

HHIP is a secreted protein that interacts with the PG layer. Regions in the HHIP-N CRD and the Hx were initially identified as containing residues involved in this process (Fig. 4.2). To investigate this further, HHIP constructs were expressed and purified from HEK293T cells for *in vitro* biophysical studies. As production of the full-length version of HHIP was hampered by proteolysis, the protein was split into its N- and C-terminal regions for *in vitro* studies (Fig. 4.1A).

HHIP-N:SOS possesses a fold comprising a distinct N-terminal GAG-binding sub-domain and a C-terminal CRD motif (Fig. 4.3B), with the former domain undergoing a disorder-to-order transition upon GAG binding (Fig. 4.3A). Crystallisation with GAG mimic SOS enabled the identification of a cluster of basic residues in the N-terminal sub-domain (Fig. 4.3C), and alanine mutations in this region lead to a reduction in heparin binding affinity (Fig. 4.4). Two residues in the helix induced by SOS recognition are not within hydrogen-bonding distance of the SOS ligand. The atomic displacement is largest in this region of HHIP-N, which is likely due to multiple conformations of basic side chains. HHIP-N:GAG interactions are promiscuous, with binding to heparin, HS and CS observed (Fig. 4.7 and Table 4.1). Solution of a HHIP-N:heparin complex structure displayed an ordered N-terminal sub-domain, whilst electron density for the ligand was not observed (data not shown). This is likely explained by the coordinated heparin ligand being in a variety of conformations. A similar scenario is observed in the Hh:Ihog structure;

heparin is required to facilitate Ihog:HhN interaction, but electron density for a sugar ligand was not present in the complex crystal structure (McLellan et al., 2006).

Structures of HHIP-C in complex with both SOS and heparin identify residues involved in GAG recognition (Figs. 4.9, 4.10 and 4.11). HHIP-C, like HHIP-N, is capable of binding to heparin, HS and CS (Fig. 4.7 and Table 4.1). A cluster of basic residues that may constitute a GAG-binding motif is suggested by modelling of the Hx and sequence analysis (Figs. 4.18A and 4.2E). Removal of the Hx from full-length HHIP results in a reduction of affinity for heparin (Fig 4.2F), but this cannot be reproduced with alanine mutations of these basic residues (Fig. 4.2G). This means that Hx plays an alternative role in the regulation of HHIP:GAG interactions. The predicted hydrophobic face may act as another HHIP oligomerisation site, comparable to a leucine zipper (Fig. 4.18B) (Krylov and Vinson, 2001). This region may be subject to regulation via proteolysis (Fig. 4.17). Therefore, the Hx may influence heparin binding by increasing HHIP avidity though stabilisation of dimers.

For *in vitro* GAG-binding studies, it was possible to produce a variant of HHIP containing a stabilised linker between HHIP-N and HHIP-C lacking the Hx. This allowed the identification and confirmation of discrete GAG-binding sites by equilibrium SPR experiments. HHIP FL ΔHx binds GAGs with higher affinity than the two component domains in isolation and heparin, the most heavily sulphated of the three GAGs, gives the lowest K_D (570 nM for heparin versus 3800 nM for the less sulphated analogue HS; Table 4.1). Whilst these experiments serve to highlight the

potency of the full-length protein in relation to GAG binding it should be noted that absolute K_D values are unlikely to be correct. Both HHIP-N and HHIP-C oligomerise in solution, which likely contributes to avidity effects under experimental conditions (Figs. 4.14, 4.15A and 4.16A). This can be seen from the data points in Fig. 4.7, in which equilibrium is not fully achieved. Additionally, GAG molecules bridge adjacent HHIP-C chains (Figs. 4.9 and 4.10), suggesting a model within which GAGs lead to the formation of large HHIP oligomers. This was confirmed in solution via AUC (Figs. 4.15B and 4.16B). Therefore, it is not possible to establish an accurate binding site model for HHIP:GAG interactions. Data were analysed using a Langmuir 1:1 binding site model with non-specific binding consideration, but the actual stoichiometry of HHIP:GAG binding would be more complex. Nevertheless, K_D values from SPR experiments with HHIP and GAGs provide relative information about GAG interactions by motifs on distinct regions of the HHIP surface. Both HHIP-N and HHIP-C contain GAG binding sites, which is reflected in the higher affinity of HHIP FL ΔHx, in agreement with heparin chromatography results (Fig. 4.6).

A model for the assembly of HHIP at the cell surface is displayed in Figs. 4.16D-F. HS chains attached to core PGs at the cell surface display an array of lengths, from 40 to 300 sugars spreading along a length of ~20-150 nm. Syndecans and Glypicans, the two major cell surface PGs, possess 1 - 4 GAG attachment sites each (Sarrazin et al., 2011). HHIP assemblies formed solely via GAG binding would

require PG chains to be close together at the cell surface. Consequently, GAG chain availability due to presentation by PG molecules may regulate HHIP assembly size *in vivo*. This mode of oligomerisation cannot be validated without additional mutational data. HHIP also binds to CS (Fig. 4.7C), the *in vivo* outcome of which must also be considered. The fact that HS causes protein oligomerisation does not necessarily mean that CS also causes assembly in larger clusters in the same way (Coles et al., 2011; Whalen et al., 2013). Despite GAG chains being unbranched it has been shown that heparin fragments of 18 sugars and above display progressively larger degrees of bending, enabling potential oligomer formation continuous with crystallographic axes (Khan et al., 2010).

AUC data suggest that HHIP-N and HHIP-C both form oligomers, and that this is enhanced upon addition of 30-mer heparin (Figs. 4.15 and 4.16). In the absence of a HHIP FL ΔHx:heparin crystal structure or SV-AUC studies of this complex it is not possible to fully understand the role of HHIP-N within this assembly. Given the capacity of HHIP-N to form dimers (Figs. 4.14 and 4.15A), it is surprising that an equivalent shift to much larger molecular weight complexes seen with HHIP-C is not witnessed for HHIP-N (Fig. 4.15B). The favoured molecular weight of ~50 kDa implies coordination of a single heparin chain by 2 HHIP-N molecules. This may be explained by the HHIP-N:heparin interaction being more transient than suggested by the K_D values from SPR experiments, but this does not fully explain the relative stability of a ~50 kDa complex over higher molecular weight assemblies. 30-mer

heparin was used in an 8-fold molar excess relative to HHIP-N, and use of lower ratios may force more HHIP-N molecules onto a single heparin chain.

Taken together, a combination of protein- and GAG-induced oligomerisation modes impose the existence of high molecular weight clusters of HHIP in solution, which may regulate HHIP gradients and impact on Hh distribution in the extracellular matrix. Three separate clusters of predominantly basic residues within both N- and C-terminal regions of HHIP mediate these high molecular weight assemblies. For comparison, a combination of protein- and GAG-mediated contacts triggering the formation of large molecular weight complexes in solution has also been witnessed for CCL5, and is thought to be vital in regulation of its various cognate functions (Liang et al., 2016).

4.4.2 Consequences of GAG binding

The heparin binding site upon the Shh surface overlaps with that of the region utilised in binding to HHIP (Fig. 4.19A). Two of these residues, R124 and R154, are involved in electrostatic interactions with D377 and E380 of the helical extension of BL3 of HHIP-C, and heparin binding would disrupt this interaction (Bishop et al., 2009; Whalen et al., 2013). Whilst the effect of mutations to this ionic bond have not been analysed, it has been shown that the contact made by HHIP residue D383 with the Zn^{2+} ion coordinated by ShhN residues H135 and D148 is essential for the HHIP:ShhN interaction (Bishop et al., 2009; Bosanac et al., 2009). It is unlikely that heparin blocks the HHIP:HhN interaction, as HHIP-C binds to ShhN with a K_D of 1

nM (Bishop et al., 2009). For contrast, a K_D of 800 nM is observed in SPR binding experiments between heparin and HhN (Whalen et al., 2013). This may be enhanced *in vivo* however, due to an increase of the valence of HhN by lipid modifications. SPR experiments with GAGs may not give completely accurate values. The Zn^{2+}-binding region of HHIP-C contributes most of the affinity of interaction, and heparin potentially regulate HHIP-C:HhN affinity in particular cellular contexts (Bishop et al., 2009).

For comparison, no residues involved in coordination of either heparin or SOS by HHIP-C overlap with the ShhN interaction site (Fig. 4.19B). Therefore, HHIP may need to compete heparin off ShhN at the cell surface before a HHIP:ShhN complex can be formed. Whether a HHIP-C:heparin:HhN complex is capable of existing in solution requires further study, alongside how secreted HHIP prevents HhN:Ptc interactions.

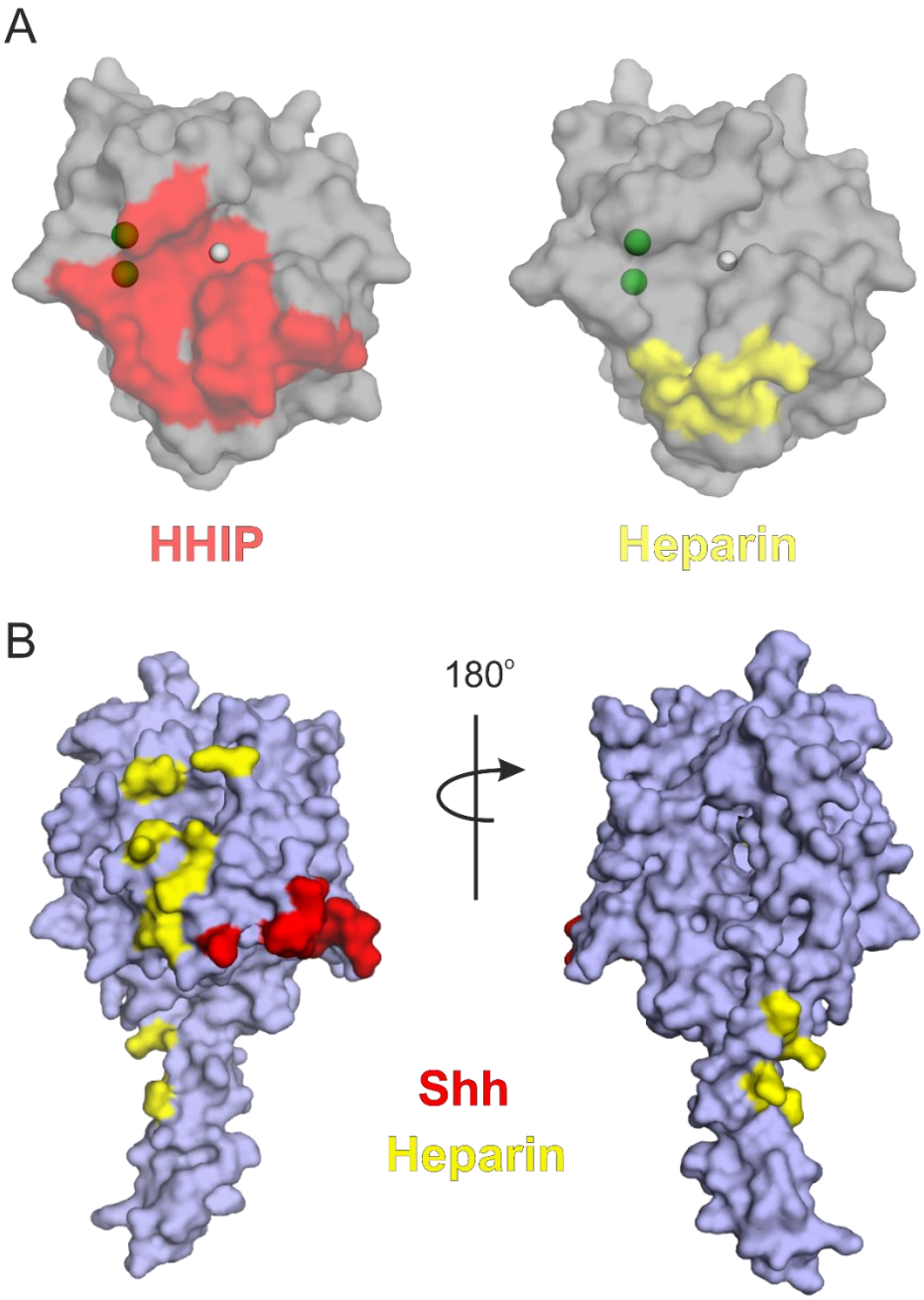


Figure 4.19 - Interface residues in interactions between HHIP, Shh and heparin. A) Surface residues utilised in the ShhN:HHIP-C (PDB ID: 2WFX) and ShhN:heparin (PDB ID: 4C4N) interactions, mapped onto the surface of the structure and coloured as indicated; B) Surface contact residues utilised in the HHIP-C:ShhN (red) and HHIP-C:heparin and HHIP-C:SOS complex structures (yellow).

In general, both heparin and the proteins to which it binds do not undergo large structural changes upon association. For example, heparin is required for both the formation of FGF dimers and subsequent receptor-binding events, and the structure of the FGF heparin-binding region is unchanged with and without heparin (Fig. 4.20A) (Pellegrini et al., 2000; Schlessinger et al., 2000; Waksman and Herr, 1998). Structures of ShhN in both CS- and heparin-bound states are unchanged when compared to the apo-form of Shh (Fig. 4.20B). Heparin and CS project in opposite directions in both ShhN:GAG complexes, but interactions are made with identical residues on the ShhN surface (Whalen et al., 2013). Thrombin is also structurally conserved between the Thrombin-apo, Thrombin:heparin and Thrombin:Nexin complex structures (Fig. 4.20C) (Carter et al., 2005; Li and Huntington, 2012). The HHIP-C surface represents a case similar to other GAG-bound structures, pre-formed for interaction with SOS and heparin (Fig. 4.12). In contrast to these three heparin-binding proteins, HHIP-N undergoes a novel disorder-to-order transition upon SOS binding (Fig. 4.20D).

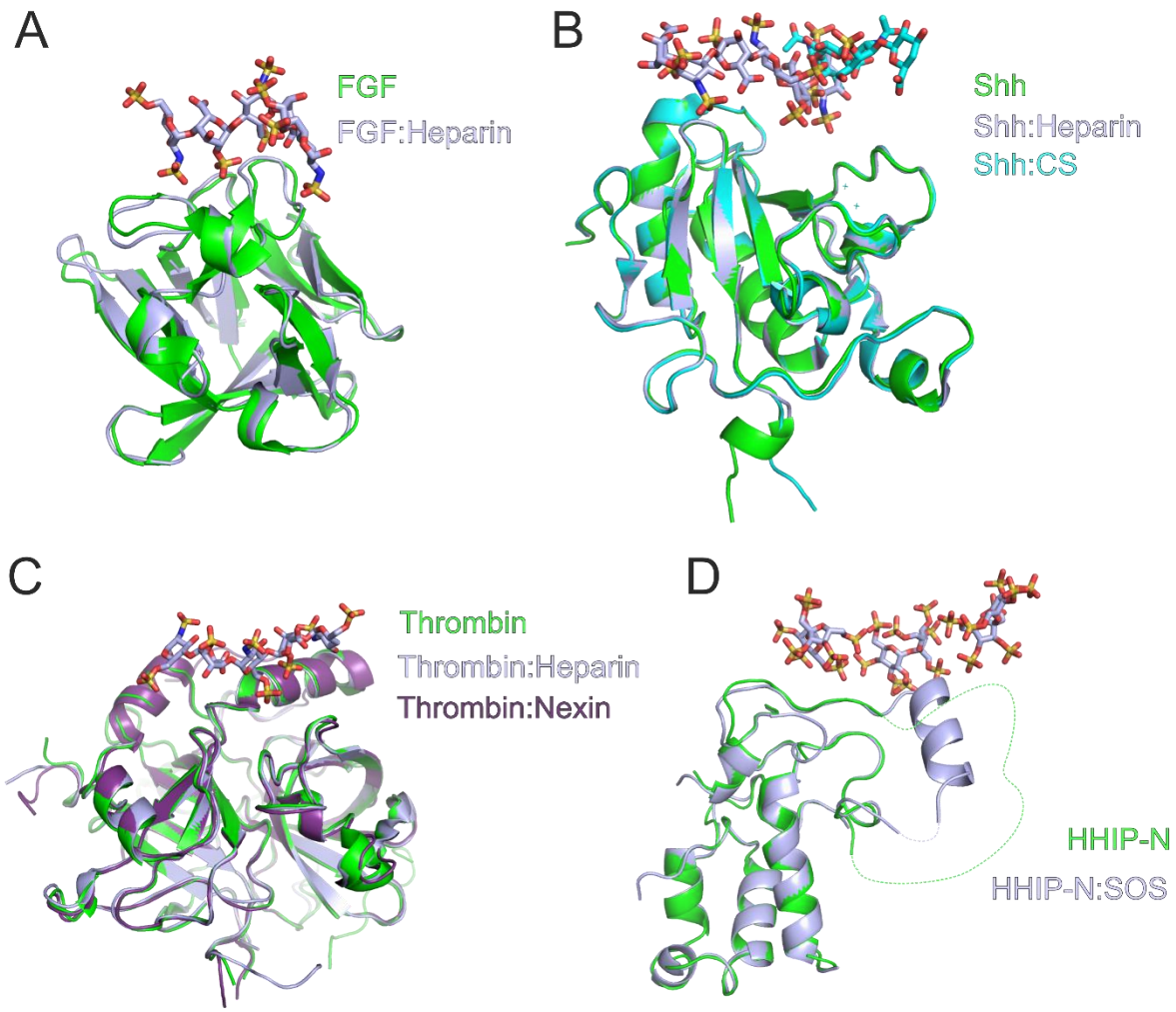


Figure 4.20 – Comparison of GAG-bound and GAG-free structures from the PDB. A) The structure of one protomer of apo-FGF (PDB ID: 2AFG) aligned with heparin-bound FGF (PDB ID: 1AXM; RMSD = 0.40); B) Shh (PDB ID: 1VHH) aligned with heparin-bound Shh (PDB ID: 4C4N; RMSD = 0.23) and CS-bound Shh (PDB ID: 4C4M; RMSD = 0.28); C) Thrombin (PDB ID: 2UUF) aligned with heparin-bound Thrombin (PDB ID: 1XMN; RMSD = 0.26) and Nexin-bound Thrombin (PDB ID: 4DY7; RMSD = 0.34); D) HHIP-N aligned with the HHIP-N:SOS complex structure. The disordered portion of the N-terminal GAG-binding sub-domain of HHIP-N is shown as a dashed green line.

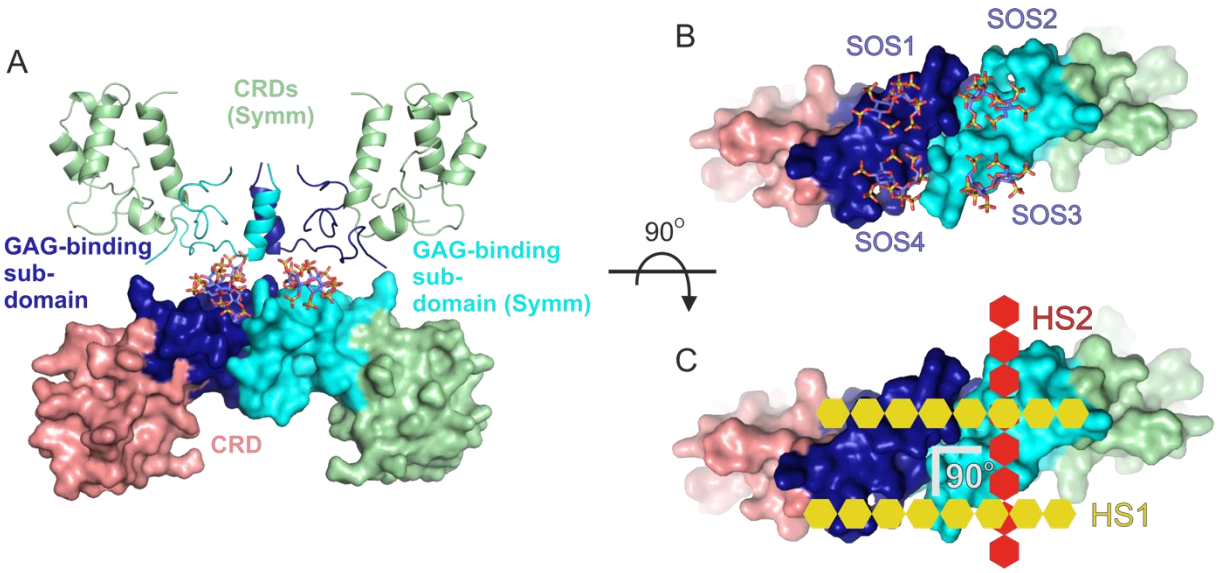


Figure 4.21 – HHIP-N:SOS dimer model from crystal packing – A) View of HHIP-N:SOS complexes from the crystallographic 4-fold axis. A symmetry-related molecule (surface representation) is coordinated by a protein:protein interaction. Two other symmetry-related molecules (cartoon representation) are coordinated via contacts made through SOS molecules. The HHIP-N molecule from the ASU is shown in blue and salmon, as in previous HHIP-N:SOS figures. For clarity, the GAG-binding subdomains of two symmetry-related molecules are shown in cyan, and one in dark blue; B) View from the crystallographic 2-fold axis. Two HHIP-N molecules form an interface not directly formed through protein:SOS interactions; C) Potential orientation of HS chains *in vivo*.

Comparison of the HHIP-N:SOS and HHIP-N-apo structures suggests that the folding of a structured GAG-binding sub-domain yields an interface between symmetry-related chains of 706 Å^2 (Fig. 4.21A). Whilst SOS stabilises the folding of this sub-domain and mediates contacts between other symmetry-related chains, this interface is made only through protein:protein contacts (Fig. 4.21B). The largest contact interface observed in the HHIP-N apo structure is 356 Å^2 (Krissinel and Henrick, 2007). The arrangement of SOS molecules in the HHIP-N:SOS complex structure gives insight into the arrangement of HHIP-N molecules on HS chains. Although each HHIP-N molecule binds 3 SOS molecules (Figs. 4.3B and C), this

would require a kinked HS chain which is not favourable (Mulloy and Forster, 2000). GAG chains therefore likely lie along the surface in one of two directions, separated by a 90^{o} angle (Fig. 4.21C). This may explain the absence of density for a heparin chain in HHIP-N:heparin complex crystals, which displayed equivalent unit cell dimensions as for the HHIP-N:SOS structure. Whilst it cannot be stated what the 'true' dimer interface of HHIP-N is, the HHIP-N:SOS complex crystal structure suggests that this protein:protein interface between HHIP-N molecules is stabilised by the presence of GAG molecules, and would explain the transient nature of the dimer peak observed in SEC-MALS experiments (Fig 4.14). The possibility that this occurs in solution is also supported by AUC (Fig. 4.15A and B). Therefore, the role of GAGs in relation to HHIP-N may be to stabilise an additional interface to HHIP via which to enhance protein clustering.

In summary, the dynamics of GAG-based association are different for both HHIP-N and HHIP-C. HHIP-C represents a much more typical mode of GAG interaction, with a binding surface pre-formed to perform this role. HHIP-N, however, undergoes a transition from disorder to order upon binding to a GAG molecule and this not only facilitates contacts between protein chains (as seen for the HHIP-C:GAG complex structures) but stabilises a protein:protein interface. This may also represent the dimeric contact formed transiently in the HHIP-N-apo state (Figs. 4.14 and 4.15A), the stabilisation of which by GAGs represents another potential physiological mode of HHIP regulation. Ultimately, this may play a role in the

spread of HHIP in the ECM as well as the establishing of HHIP gradients to fine-tune Hh signalling activation.

4.4.3 Further work

Further work toward understanding the regulation of vertebrate Hh signalling as a consequence of HHIP:GAG interactions would primarily benefit from investigating whether HHIP is capable of simultaneously binding both heparin and HhN ligands. Such an experiment would be achievable by injecting a pre-formed HHIP-C:HhN complex onto a heparin column and observing whether both proteins elute as a complex, and could be further supplemented by co-crystallisation experiments. Mutations to the ion-binding sites of HhN may give insight into how heparin regulates HHIP-C:HhN complex assembly (Bishop et al., 2009). Multimeric lipid-modified HhN potentiates signalling activation and GAG-bound HHIP assemblies may play a role in tight control of this process. Therefore, luciferase assays with HhN and HHIP-C pre-assembled into clusters using heparin oligosaccharides (as seen in AUC experiments) may aid in the further understanding of the physiological role of HHIP:GAG assemblies.

The assembly of HHIP in response to GAG binding requires further study. Given that AUC data of HHIP with 30-mer heparin gives rise to a range of high molecular weight species, the model of head-to-tail contacts between adjacent HHIP-C molecules facilitated by contacts with heparin is likely to be correct (Figs. 4.9 and 4.10). Further validation would entail mutation of sugar-binding residues from the

HHIP-C:GAG complex structures to acidic residues and repetition of AUC experiments. If the result of Fig. 4.16A can be reproduced with this mutant upon addition of heparin, then this would indicate that these are the true GAG interaction sites. Mutation of residues involved in the formation of protein:protein contacts and repetition of AUC experiments would verify if protein-only oligomerisation plays a role. This could be extended to single particle analysis by electron microscopy to analyse whether heparin induces a regularised array of HHIP-C molecules or if the species observed using AUC in this study are assembled in an anisotropic fashion. Once verified, studies of the development of mutants to the HHIP GAG-binding regions and oligomerisation surfaces may aid in further understanding of how these assembly modes regulate Hh signal activation.

It has not been possible to perform biophysical analysis of the Hx. It was shown in this study that the Hx is important for heparin binding by HHIP, but how this occurs requires further study (Figs. 4.2E-G and 4.18). Recombinant expression of this helix in isolation, purification and SEC-MALS would validate any propensity to form dimers, and GAG binding could be assayed further using the above SPR assay. Within the context of HHIP-C, proteolysis-susceptible arginine residues could be mutated to that of another polar residue (i.e. asparagine or glutamine) and SEC-MALS experiments utilised to assay whether this causes dimerisation. As discussed, HHIP-C dimerisation has only been detectable using AUC, which is more sensitive to detection of protein assemblies (Bishop et al., 2009). The relative arrangement of

HHIP-N and HHIP-C in the larger HHIP domain architecture are unknown. Although the HHIP-N construct comprises residues 39-209, residues 186-209 are disordered in the crystal structure. This represents the large amount of conformational space accessible to HHIP-N relative to HHIP-C, and may play a role in arrangement on HS chains at the cell surface. Solution studies of HHIP FL ΔHx via SAXS could enable modelling of envelopes available to the larger protein architecture and grant insight into the dynamics of GAG-induced complex assembly.

4.4.4 Concluding remarks

A wide variety of regulatory mechanisms are employed to control Hh signal transduction, and control of the distribution of the ligand itself is a proven method of modulation (Gallet et al., 2003; Whalen et al., 2013). The importance of HhN distribution within the context of appropriate signalling was discussed at length in section 1.4, and is contingent on a combination of regulatory mechanisms at the levels of morphogen modification, soluble and membrane-bound receptors and interactions with the ECM. Fine-tuning due to GAG molecules is still poorly understood beyond the level of protein oligomerisation and bioavailability, but its role is known to be instrumental in the control of secreted protein dynamics. From previous structural studies, HS chains are generally used for regulating the concentrations of secreted proteins at the cell surface (Fry, 1999; Whalen et al., 2013) or acting as a 'molecular staple' to bridge two proteins together (McLellan et al., 2006;

Pellegrini et al., 2000; Schlessinger et al., 2000). Regulation of assembly through GAG-mediated and protein:protein-based oligomerisation has been observed in structural studies of chemokine CCL5 in complex with heparin, forming super-helical rods. Whilst the true purpose of this is unknown, one hypothesis is to alter the distribution of CCL5 itself; another is to protect it from proteolysis (Liang et al., 2016). Given the notable susceptibility of HHIP to undergo inter-domain proteolysis (Bishop et al., 2009), this may play a role in increasing the stability of HHIP:GAG complexes by creating steric hindrance to proteases. GAG-induced association of HHIP may also serve to localise HHIP-N to areas of active Hh signalling to fulfil some currently uncharacterised signalling function linked to the binding of a small molecule by the CRD (see section 3.2). Additionally, whether CS prevents oligomerisation of HHIP as seen in other cases remains to be seen (Coles et al., 2011; Whalen et al., 2013). The influence of GAGs over HHIP assembly and even which sugars are present may add another element to the regulation of HhN gradients. Regulation of secretion and complex assembly by size represents the potential for a novel mode of vertebrate HhN antagonism; one which depends on the establishment of gradients of a secreted inhibitor capable of forming an array of super-clustered assemblies.

Chapter 5

Structural insights into the ROR2 extracellular domains

Chapter 5: Structural insights into the ROR2 extracellular domains

5.1 Summary

A number of receptors besides the GPCR-like Frizzed proteins are involved in the mediation of Wnt signal transduction outputs. ROR2, a receptor tyrosine kinase (RTK), possesses an ectodomain (ECD) comprising an N-terminal Ig domain, a CRD and a C-terminal juxtamembrane Kringle domain. Reports assume a role of ROR2 in non-canonical Wnt signalling, but the roles of the individual domains remain poorly characterised. In this chapter, the use of Pt-SAD combined with homology modelling enabled phasing and structure solution of the ROR2 CRD and Kringle domains (ROR2 CRD-Kr) using X-ray crystallography. Both domains form a structural unit with a small inter-domain interface. ROR2 constructs were poorly expressed in the absence of the Kringle domain, suggesting that this domain is extremely soluble and acts to stabilise the CRD. The closest structural homologue of the ROR2 CRD domain is that of MuSK, a related RTK. Structural comparison to other Fz-type CRDs suggests that the lack of a long kinked helix and an additional helical insert block the formation of a pre-formed surface 'groove,' potentially preventing palmitoleate (for Fz8) or cholesterol (for Smo) binding. This is supported by the absence of Wnt-interacting residues conserved in FZD receptors, the observation that the ROR2 CRD-Kr construct does not bind Wnt3a in a cell reporter assay. Solution studies of the ROR2 ECD using small angle X-ray scattering (SAXS) support a model consisting of a

flexible molecule in an extended conformation. A previously reported direct interaction with the BMPR1b (BMP Receptor 1b) ECD was not observed between purified proteins via various biophysical experiments.

5.2 Introduction

CRDs occur frequently within proteins involved in Wnt signalling, found in the 10 FZD receptors, the 5 inhibitory secreted sFRPs and co-receptors MuSK and the 2 ROR (**R**eceptor Tyrosine Kinase-like **O**rphan **R**eceptor) proteins (Niehrs, 2012). Whilst the activation of Wnt signalling by the ROR family is not well understood, it is generally accepted that they play a role in initiating non-canonical signalling pathways. This can occur through both FZD-dependent and FZD-independent mechanisms (Billiard et al., 2005; Grumolato et al., 2010).

ROR1/2 possess overlapping expression patterns in the developing embryo and are involved in the development of the nervous system, skeleton, cartilage, limbs, lungs and heart. ROR1 is largely expressed in the same regions as ROR2, albeit within a more restricted profile (Al-Shawi et al., 2001; Matsuda et al., 2001). *ROR2*$^{-/-}$ mutant embryos display severe skeletal and heart defects and die within 6 hours of birth due to respiratory failure. *ROR1*$^{-/-}$ embryos display respiration-based lethality but lack the skeletal and heart defects. ROR1/2 double knockout mutants exhibit a more severe phenotype more similar to ROR2-only knockouts than ROR1-only knockouts, suggesting that ROR1 is functionally redundant to some degree

compared with ROR2 (Nomi et al., 2001). ROR2 mutations in humans lead to autosomal recessive Robinow syndrome (RRS), characterised by severe skeletal dysplasia, heart defects and brachydactyly (Aglan et al., 2015). Aberrant signalling of both ROR1 and ROR2 is heavily linked to multiple forms of metastatic cancer. Therefore, prognosis and therapeutic targeting of these proteins represent prospective strategies for disease treatment as ROR1/2 expression is only detected in cancer cells in adult organisms (Borcherding et al., 2014; Rebagay et al., 2012).

ROR1/2 have a domain architecture comprising an N-terminal Immunoglobulin-like domain followed by a Fz-type CRD and a Kringle domain (Kr) (Fig. 5.1A). Many CRDs are found as part of a larger architecture, and there are few studies concerning their structural characterisation in a multi-domain context (Byrne et al., 2016; Gong et al., 2016). The ROR2 CRD has been shown to bind to multiple Wnt ligands with a preference for non-canonical Wnt5a (Billiard et al., 2005; Oishi et al., 2003). Both ROR1/2 also include a single TM helix and intracellular kinase and Ser/Thr-rich and Pro-rich domains. Kr domains are found in proteins including kremen, prothrombin, plasminogen and hepatocyte growth factor, and serve as protein-protein interaction motifs in these proteins (Deguchi et al., 1997; Law et al., 2013; Ultsch et al., 1998; Zebisch et al., 2016).

Several reports suggest alternative mechanisms for the activation of signalling by ROR2, which still require clarification. A summary of suggested interactions is displayed in Fig. 5.1B. Studies have indicated ROR2 as being involved in both

canonical- and non-canonical Wnt signalling pathways, with a more evidence for activation of the non-canonical Planar Cell Polarity (PCP) pathway (Grumolato et al., 2010; Liu et al., 2008; Mikels and Nusse, 2006). Furthermore, it has also been proposed that the ROR2 CRD forms a complex with the BMP receptor BMPR1b (Sammar et al., 2004). The potential for the Ig and Kr domains to interact with other proteins has not been investigated. In order to dissect the function of ROR2, constructs comprising the ectodomain (ECD) and CRD and Kr domains (CRD-Kr) were cloned, expressed and crystallised. Biophysical and cell reporter assays were used to investigate the ROR2 CRD:Wnt3a and ROR2 CRD:BMPR1b interactions.

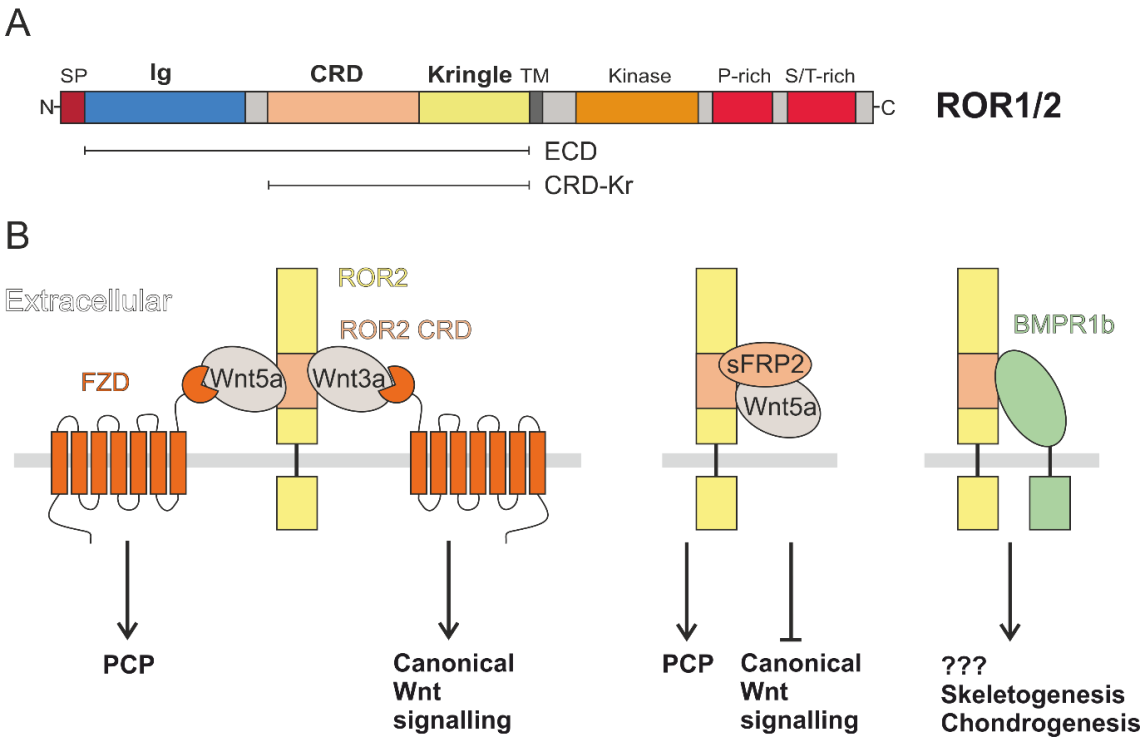


Figure 5.1 – Introduction to ROR2. A) Domain organisation of ROR1/2. Expression construct boundaries are indicated; B) Proposed signalling pathways activated by ROR2 through the CRD. This may comprise non-canonical signalling through a particular FZD and Wnt5a or canonical signalling utilising Wnt3a and some other particular FZD (left) (Grumolato et al., 2010; Mikels and Nusse, 2006). Signalling has also been suggested to progress irrespective of FZD incorporation; a Wnt ligand may activate signalling through ROR2 alone, and a complex with sFRP2 has also been linked to direct signal activation of the non-canonical PCP pathway (middle) (Brinkmann et al., 2016; Liu et al., 2008). A third hypothesised mode of signalling (right) comprises a complex with BMPR1b and activation of pathways involved in the development of the skeleton and cartilage (Sammar et al., 2004).

5.3 Results

5.3.1 Purification and crystallisation of ROR2

The ROR2 CRD-Kr could be purified to homogeneity in large quantities (Fig. 5.2A-B) and crystallised (Fig. 5.2C). Crystals often grew in clusters, which required separation via the use of an acupuncture needle (Fig. 5.2D). The ROR2 ECD could also be produced in large quantities and remained stable following deglycosylation with endoglycosidase EndoF1 (Fig. 5.2E and F). Despite approaches utilising lysine methylation and *in situ* proteolysis, crystals of the ROR2 ECD could not be produced (Dong et al., 2007; Walter et al., 2006). This is likely due to flexibility between the Ig domain and the CRD (see section 5.3.5). The purification of both BMPR1b and sFRP4, proteins used for additional experiments with ROR2, is displayed in Supplementary Figure 5.1.

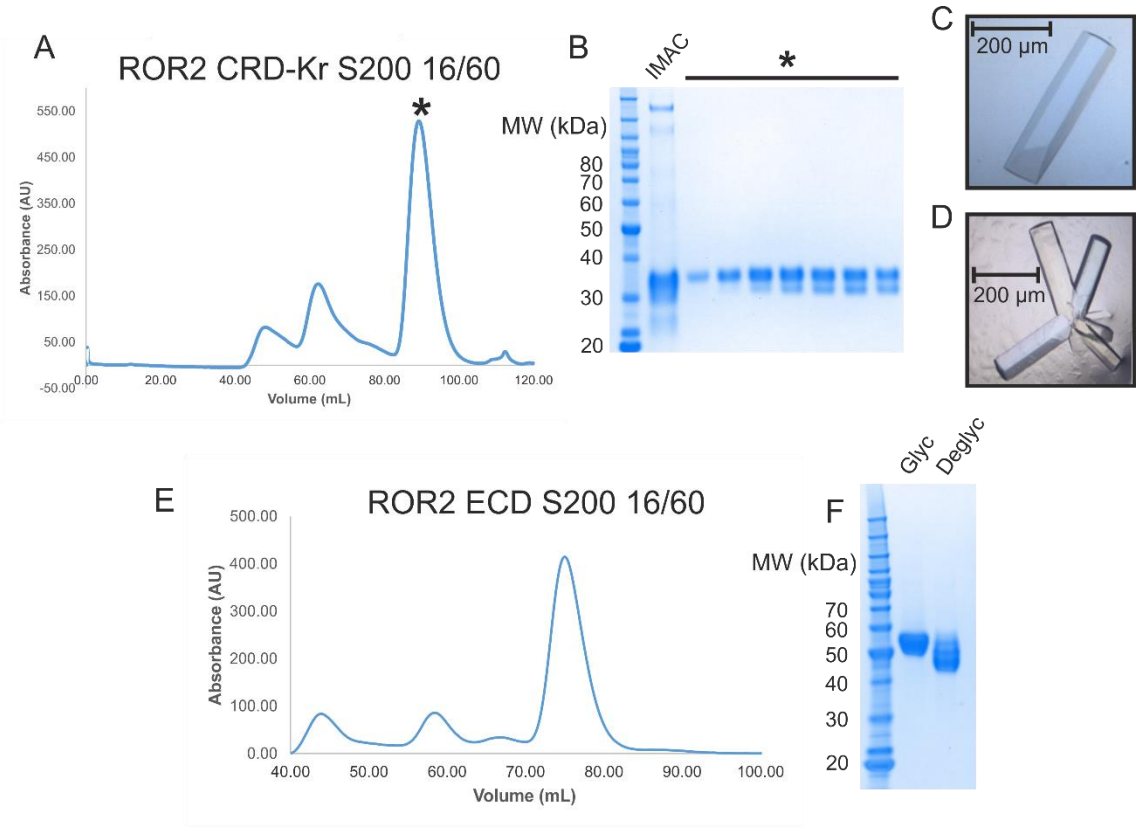


Figure 5.2 – Purification and crystallisation of ROR2 constructs. A-D) ROR2 CRD-Kr. A) UV trace from SEC for the purification of the CRD-Kr construct and B) analysis of SEC fractions by SDS-PAGE; Birefringent crystals of an appreciable size could be grown, either C) singly or D) in clusters; E-F) ROR2 ECD. E) UV trace from SEC for the purification of the ECD construct and F) purity assessment of SEC fractions by SDS-PAGE. 'Glyc' refers to the pooled peak prior to deglycosylation with EndoF1, whilst 'deglyc' refers to a sample deglycosylated for 2 hours at 37^{o}C and then purified via SEC. See Fig. 3.1A for a summary of purification strategies.

5.3.2 ROR2 CRD-Kr data collection

Diffraction data from ROR2 CRD-Kr crystals soaked with potassium hexachloroplatinate ($KPtCl_6$) was collected from 2 crystals on beamline I03 at a wavelength of 1.0500 Å. The two datasets were merged together using Xia2 (Winter, 2010), resulting in an anomalous slope of ~1.4 over all resolution shells (Table 5.1). This signal was present a resolution of ~3.7 Å, as estimated from the CCanom and SigAno statistics calculated using XSCALE (Kabsch, 2010) (Table 5.2). In addition, a native dataset used for structure refinement was collected at a wavelength of 0.9763 Å and diffracted to a resolution of 2.7 Å (Table 5.3, below).

Table 5.1

	ROR2 CRD-Kr Pt-SAD
Data collection	
Beamline	I03, DLS
Space group	$P3_121$
Unit-cell parameters	
a, b, c (Å)	106.1, 106.1, 42.2
α, β, γ (°)	90.0, 90.0, 120.0
No. of crystals / data sets	2 / 2)
Wavelength (Å)	1.0500
Resolution (Å)	53.0-3.0 (3.1-3.0)
No. of unique reflections	5599 (397)
Completeness (%)	99.9 (99.5)
Multiplicity	28.1 (24.3)
$I/\sigma(I)$	23.5 (6.4)
R_{merge} (%)	14.7 (59.6)
R_{pim} (%)	3.0 (12.3)
$CC_{1/2}$	1.0 (0.8)
Anomalous slope	1.4

Table 5.2

Resolution limit	CC_{anom}	SigAno
13.45	94*	6.02
9.51	92*	4.105
7.76	83*	3.468
6.72	87*	3.133
6.01	76*	2.828
5.49	68*	2.238
5.08	61*	1.897
4.75	54*	1.612
4.48	52*	1.475
4.25	30*	1.234
4.05	36*	1.181
3.88	21	1.036
3.73	32*	1.048
3.59	14	0.909
3.47	4	0.782
3.36	10	0.829
3.26	5	0.781
3.17	-3	0.778
3.09	-1	0.795
3.01	-9	0.725

Tables 5.1 and 5.2 – SAD data collection for ROR2 CRD-Kr soaked with $KPtCl_6$. A) Data collection statistics. Values in parentheses correspond to the highest resolution shell; B) Statistics for the scaling of Bijvoet pairs under anomalous scattering conditions, calculated from the program XSCALE as a function of resolution (Kabsch, 2010). CC_{anom} – % correlation between random half-sets of anomalous intensity differences. SigAno - (|F(+)-F(-)|/σ; mean anomalous difference versus standard deviation of anomalous intensity. A dashed red line indicates the resolution to which an anomalous signal is detected.

5.3.3 Structure solution and refinement of ROR2 CRD-Kr

The ROR2 CRD-Kr construct crystallised as a monomer in the ASU. An initial structure was solved from $KPtCl_6$-soaked crystals, which identified 5 heavy atom sites. Using the best density modified map (FOM = 0.36), autobuilding resulted in a model for almost the entire CRD region (residues 173-305 built, 200-204 missing; Cα RMSD versus final CRD = 0.64) (Fig. 5.3A). However, the Kr domain could not be traced automatically. To overcome this problem, the auto-traced model of the CRD was used as a molecular replacement (MR) search model against higher resolution native data (2.7 Å) (Fig 5.3B). This solution was fixed and a second search using an ensemble of three homology models for the Kr domain was performed (Raman et al., 2009). This resulted in TFZ and LLG scores of 19 and 424 indicating an improved solution, as could be seen from a readily-interpretable 2Fo-Fc map (Fig. 5.3C). In contrast, performing the same search with Kringle 5 of Plasminogen (44% sequence identity to the ROR2 Kr domain) resulted in worse MR scores and poorer maps (Fig. 5.3D). Manual rebuilding and 40 cycles of jelly-body refinement using *Refmac5 (*Murshudov *et al., 2011)* generated a model with an R_{work}/R_{free} = 31%/37%, which was subsequently refined to R_{work}/R_{free} = 24%/26% (Fig 5.3E). Auto-tracing of the first experimental map gave a high-quality model of the CRD, as seen by the similarity between this domain in the initial and final models. The use of homology modelling was necessary for building the Kr domain, however, and this deviates substantially from the initial to the final model (Fig 5.3F).

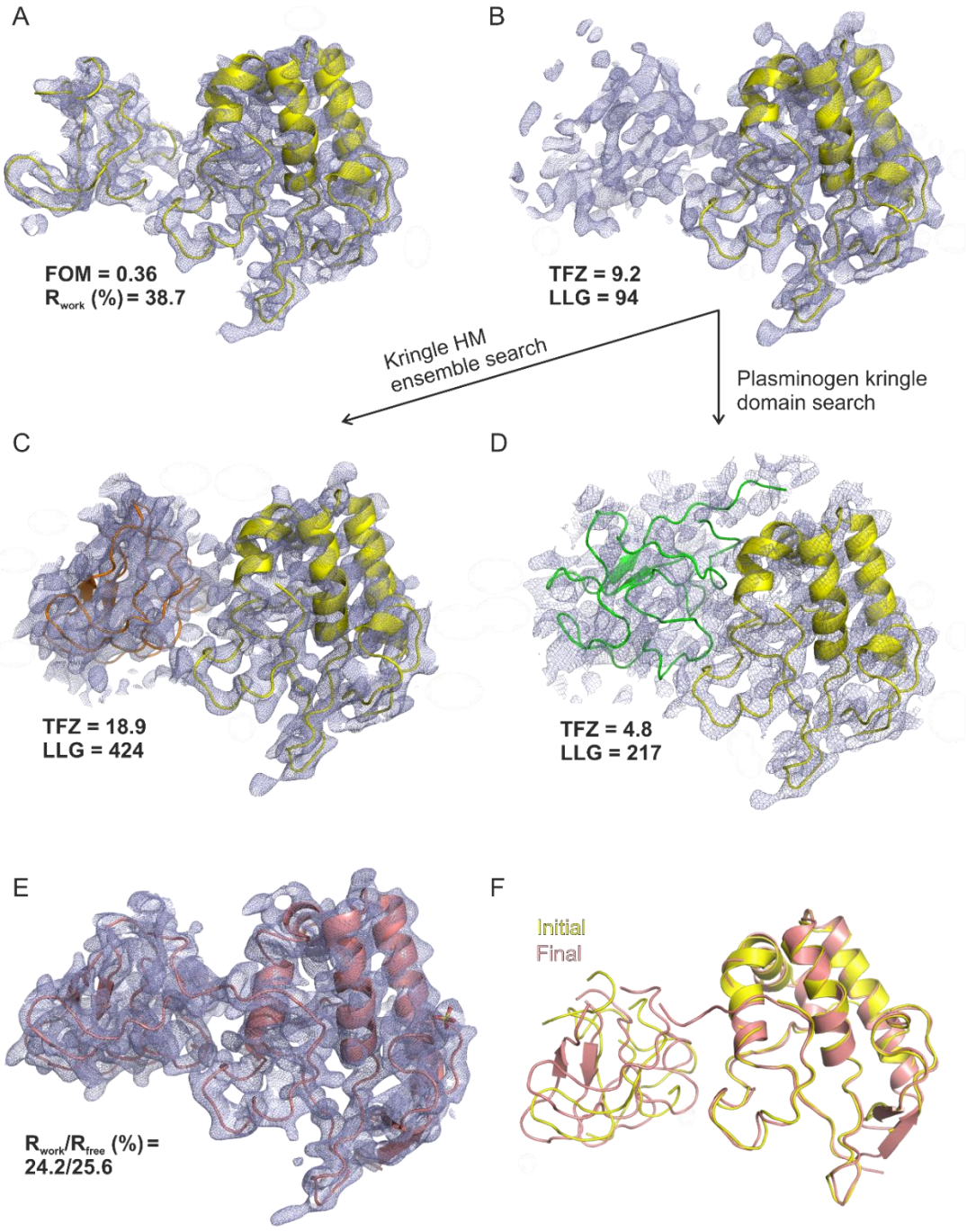


Figure 5.3 – Structure solution and refinement of the ROR2 CRD and Kringle domain. 2Fo-Fc maps for each step are shown as a blue mesh, with partial models produced at each step coloured in yellow; A) First experimentally phased map and subsequent partial model from auto-tracing; B) MR output with only the previously traced CRD as a search model. Density for the Kr domain is visible; C) MR search for an ensemble of homology models (HM) with the CRD already placed; D) MR search for the Plasminogen Kringle 5 domain (PDB ID: 5HPG) with the CRD solution fixed; E) Final model (salmon); F) Structural alterations in the initial model (yellow) compared to the final refined model (salmon).

Model building and refinement of ROR2 CRD-Kr was further facilitated via S-SAD data collection on beamline I23 (DLS) (see Supplementary Table 5.1 for SHELXD statistics). In-vacuum crystallography combined with a curved detector array enables the collection of data using a wavelength of 2.48 Å (5 keV; f'' = 1.31 versus f'' = 0.72 at 1.77 Å), reducing the volume of data required for S marker atom location (Wagner et al., 2016) (see section 3.4.3). 14 Sulphur sites are visible in the final CRD-Kr structure (Fig. 5.4A). Two SO_4^{2-} ions are visible (A and B), four methionines (M1-M4), seven unresolved disulphide bonds (I, II, III, V, VI, VII and VIII) and one resolved disulphide bond (IV). Adjacent disulphide bonds VII and VIII in the Kr domain cannot be resolved from one another in the anomalous difference map.

Disulphide bond VI is formed between the chain extending from the C-terminus of the CRD and the C-terminus of the Kr domain. The 2Fo-Fc electron density map was noisy in this region, making chain tracing difficult. A strong peak in the anomalous difference map was observable in this region, enabling placement of residues C316 and C394. Residues 312-315 from the C-terminus of the CRD are disordered in the final structure, and the only residue adjacent to C316 that could be built was Y317. The remainder of the residues preceding the Kr domain (318-320) were also missing from the electron density. The same is true for the residues at the very C-terminus of the Kr domain, with P392 and S393 not being accounted for in the 2Fo-Fc map (Fig. 5.3B). As a result, disulphide-bonded residues C316-C394 and

respective adjacent residues Y317 (extends from the CRD to the Kr domain) and S395 (extends towards the Kr domain C-terminus) could be placed in the map. Two sulphate ions were also placed using information from peaks in both the anomalous and 2Fo-Fc maps, the residues in proximity to these peaks and the crystallisation condition (~1.5 M $LiSO_4$). The sulphate at site B is coordinated by the backbone NH group of G360 and side chain of S352 in the Kr domain (Fig. 5.3C), whilst the sulphate at site A forms hydrogen bonds with three arginine residues (Fig. 5.3D).

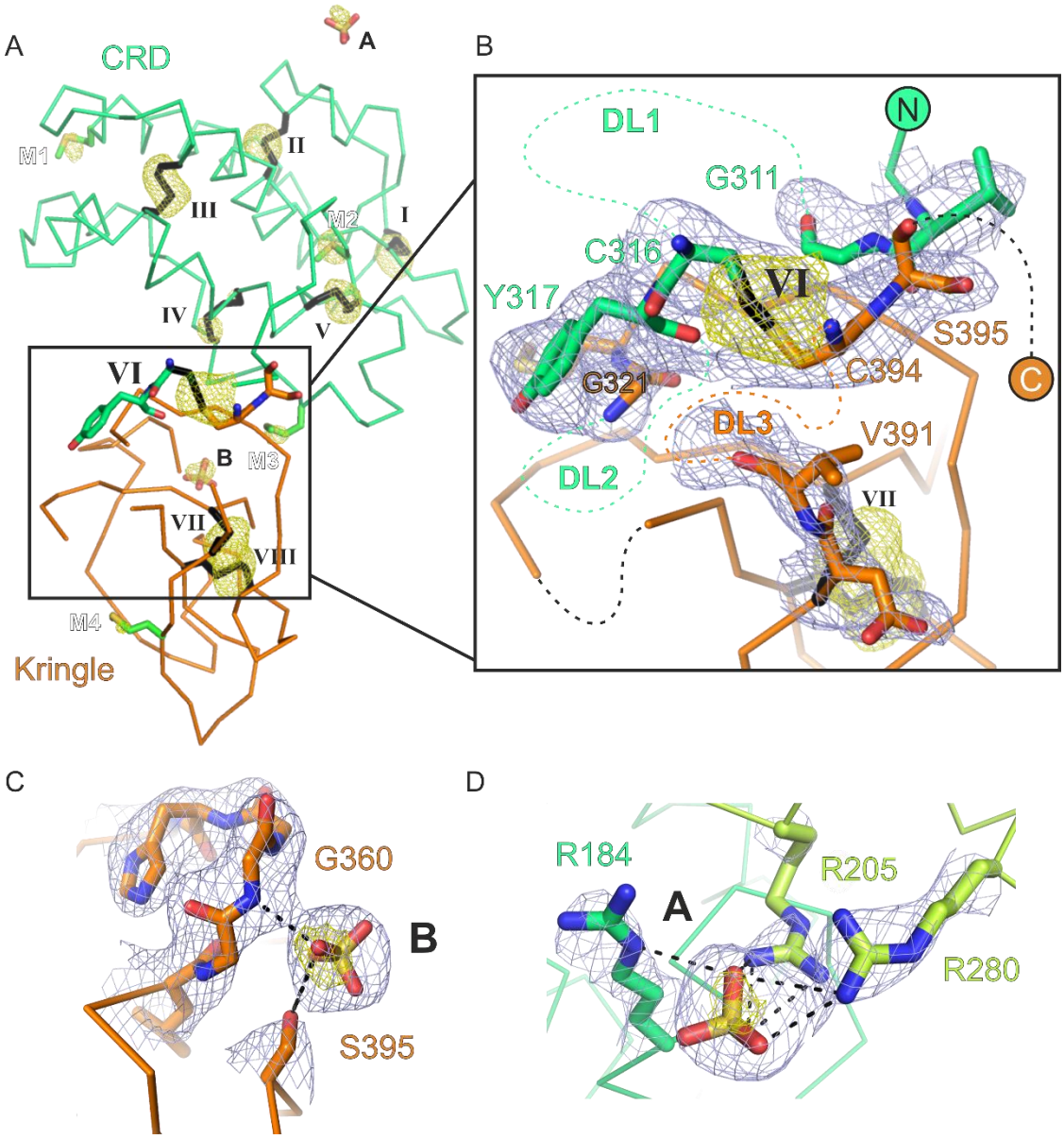


Figure 5.4 – Use of Sulphur-SAD data in model building and refinement. A) Sulphur anomalous difference map contoured at 3σ calculated for the final refined ROR2 CRD-Kr model using Phenix (Adams et al., 2010). Disulphide bonds are numbered with Roman numerals (I-VIII). Two sulphate ions are labelled A and B, and four methionine residues as M1-M4; B) Close up of site XI (disulphide VI), including anomalous difference density contoured at 3σ (yellow) and 2Fo-Fc density contoured at 1σ (blue). Residues are numbered and coloured as in A). Disordered loops (DL1-3) are shown as dashed lines and are with domain-specific colouring, with additional disordered loops in black; C) The sulphate coordinated by the Kringle domain (B), with both anomalous and 2Fo-Fc density contoured and coloured as in B); D) The sulphate coordinated by 2 symmetry-related CRDs (A), with associated anomalous and 2Fo-Fc density coloured as previous.

5.3.4 Structure of the ROR2 CRD and Kr domains

The ROR2 CRD-Kr crystal structure is shown in Fig 5.5, with the corresponding data collection and refinement statistics in Table 5.3. The CRD comprises a fold comparable to the CRDs of MuSK and Fz8 (see section 5.3.5), composed of 5 α-helices (α1-5) and a β-sheet (strands β1 and β2). The Kr domain lacks significant secondary structure, displaying one short β-sheet. This lack of secondary structural elements has been observed in other Kr domains (Deguchi et al., 1997; Ultsch et al., 1998; Zebisch et al., 2016). The CRD is stabilised by 5 disulphide bonds, located between β1 and a loop extending from α2 (I), the second-longest helix α2 and the loop preceding longest helix α1 (II), helix α2 and the turn between helices α3/4 (III), long loops following α2 and α3 (IV) and α3 itself and the loop extending from α5 (V). The Kr domain is stabilised by three disulphide bonds. The first (VI) is formed between the very N-terminus and the C-terminus of the domain. Two disulphide bonds, VII and VIII, are also found within the Kr domain (Fig. 5.5A). Overall, the two domains exhibit a compact structural unit, closely associated with one another (Fig. 5.5B). B-factors are highest in the region surrounding disulphide VI, as well as on the associated face of the Kr domain (Fig. 5.5C).

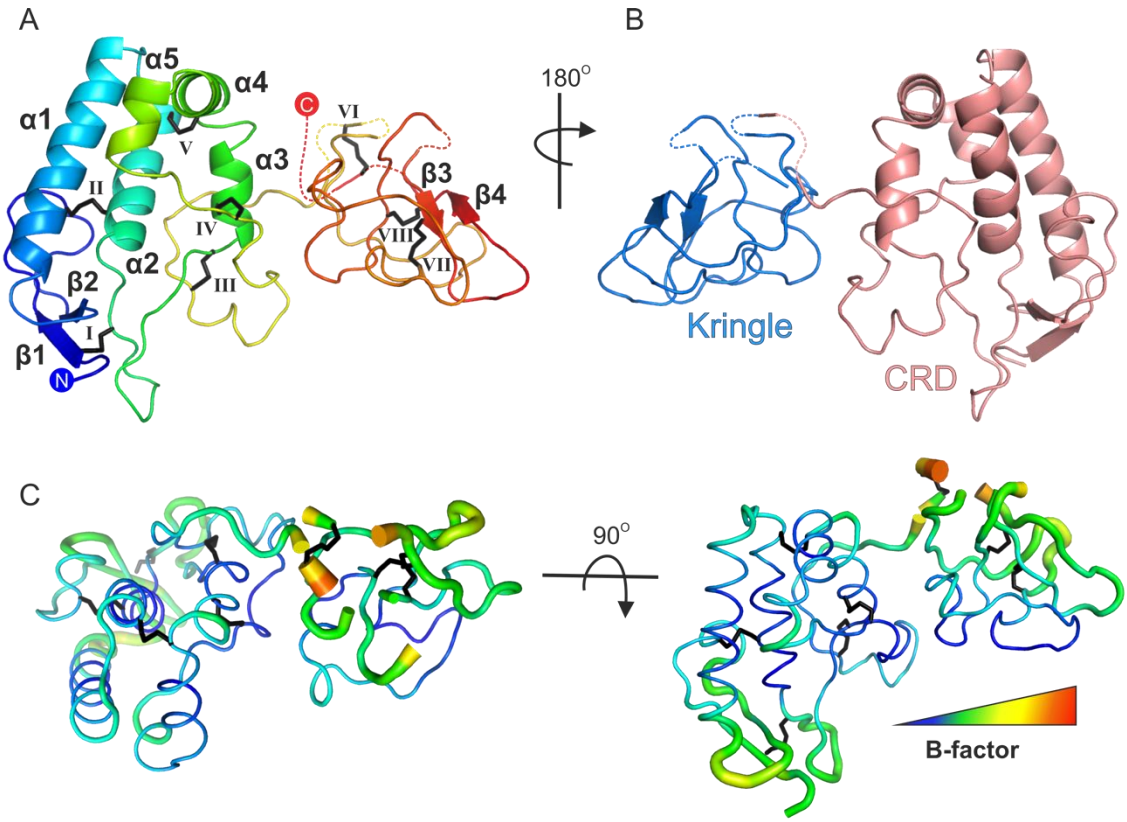


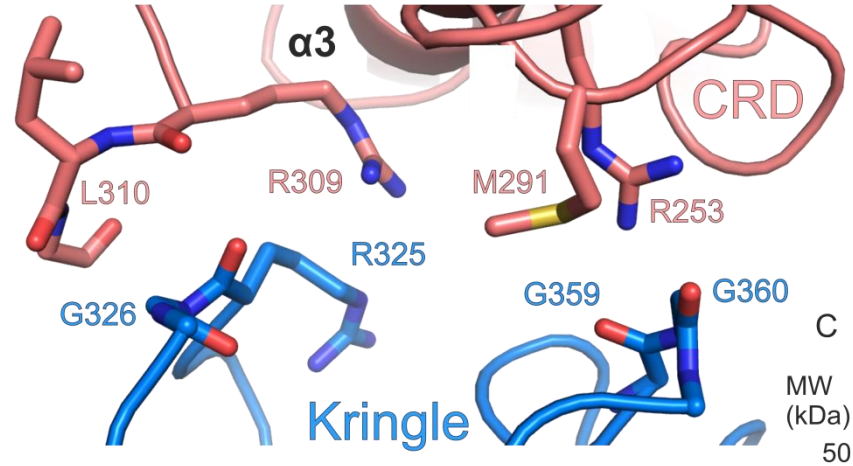
Figure 5.5 – ROR2 CRD-Kr crystal structure. A) α-helices and β-strands are numbered α1-5 and β1-4 according to their position in sequence. Disulphide bonds are numbered using Roman numerals. ROR2 CRD-Kr is coloured in a rainbow representation (blue to red) from the N- to the C-terminus. Disordered regions are shown as dashed lines; B) Two-domain representation of ROR2 CRD-Kr. Domains are shown in salmon (CRD) and blue (Kr). Disordered residues are displayed as dashed lines corresponding to their domain; C) B-factors mapped onto the structure as a rainbow (blue to red) and represented as backbone thickness. The structure is shown from a top-down view (left) and from the same view as in A (right).

	ROR2 CRD-Kr (Native)
Data collection	
Beamline	I03, DLS
Space group	P 3_221
Unit-cell parameters	
a, b, c (Å)	109.6, 109.6, 45.0
α, β, γ (°)	90.0, 90.0, 120.0
No. of crystals / data sets	1 / 1
Wavelength (Å)	0.9763
Resolution (Å)	54.8-2.7 (2.75-2.7)
No. of unique reflections	8699 (443)
Completeness (%)	99.4 (100.0)
Multiplicity	9.9 (8.7)
$I/\sigma(I)$	16.3 (1.0)
R_{merge} (%)	7.8 (232.4)
R_{pim} (%)	2.6 (8.3)
$CC_{1/2}$	1.0 (0.5)
Refinement	
No. reflections (test set)	8688 (430)
R_{work}/R_{free}	24.2 / 25.6
No. atoms	
Protein	1670
Ligand	10
Mean B factor (Å)	
Protein	112.0
Ligand	137.1
RMSD bond lengths (Å)	0.008
RMSD bond angles (Å)	0.91
Ramachandran plot (%)	
Favoured	96.1
Allowed	3.4
Outliers	0.5

Table 5.3 – ROR2 CRD-Kr native data collection and refinement statistics. Values in parenthesis correspond to the highest resolution shell unless otherwise stated.

From the ROR2 CRD-Kr structure, it appears that the two domains possess a contact interface. This was analysed using PDBsum, and appears to comprise an interface composed mostly of Van der Waals interactions, with a single hydrogen bond between R253 and the carbonyl oxygen of G360 (Laskowski, 2001). A number of weak contacts are made between the side chains of R309 and R325. Of note is the number of contacts dependent on interactions with the backbone of the Kr domain. The interface between the two domains buries ~354 $Å^2$ per domain with reasonable shape complementarity (Sc = 0.68) (Fig. 5.6A) (Krissinel and Henrick, 2007; Lawrence and Colman, 1993). It is also possible that further interactions may be mediated by water molecules, although none are accounted for in the observed electron density map. The atomic contacts and the binding distances are detailed in Fig. 5.6B. Analysis of recombinant expression and secretion of ROR2 constructs containing the CRD domain show increased expression with constructs comprising the CRD-Kr region (Fig. 5.6C). This suggests a potentially stabilising role for the Kr domain, involved in forming the observed CRD-Kr structural unit.

A



B

Residue 1	Atom 1		Atom 2	Residue 2	Distance (Å)
ARG 253	NH2	⟷	C	GLY 359	3.61
ARG 253	NH2	⟷	O	GLY 359	3.24
MET 291	Cγ	⟷	O	GLY 360	3.38
MET 291	Sδ	⟷	CA	GLY 360	3.87
ARG 309	Cα	⟷	O	ARG 325	3.77
ARG 309	O	⟷	O	ARG 325	3.15
ARG 309	Cζ	⟷	CG	ARG 325	3.74
ARG 309	Cζ	⟷	CD	ARG 325	3.89
ARG 309	NH1	⟷	CB	ARG 325	3.71
ARG 309	NH2	⟷	CG	ARG 325	3.63
ARG 309	NH2	⟷	CD	ARG 325	3.40
ARG 309	NH2	⟷	NE	ARG 325	3.68
LEU 310	O	⟷	CA	GLY 326	3.63

C

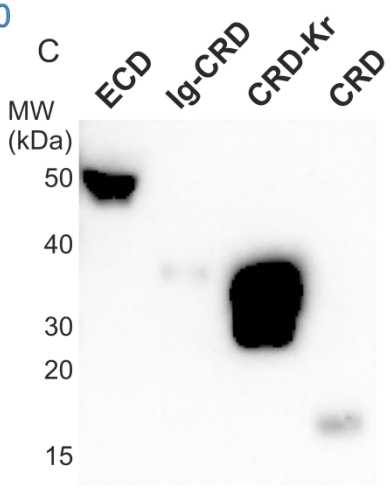


Figure 5.6 – Interface between the CRD and Kr domains. A) Close-up of the interface formed between the ROR2 CRD and Kr domains, with side chains and backbone coloured in salmon (CRD) and blue (Kr domain); B) Residues and specific atomic contacts involved in the interface, as calculated using PDBsum (Laskowski, 2001); C) Secretion of different ROR2 constructs containing the CRD following transient expression in HEK293T cells. Equal volumes of HEK293T cell supernatant into which the indicated constructs had been secreted for the same time period were loaded onto an SDS-PAGE gel and subjected to western blot analysis.

Despite being the only residue to contribute a side chain hydrogen bond to the interface between the ROR2 CRD and the Kr domain, an R253 mutation in RRS patients has not been identified to date. Published RRS mutants were cloned into the ROR2 CRD-Kr background and analysed by western blot to test for secretion (Aglan et al., 2015). No mutant is secreted, suggesting that the absence of ROR2 at the cell

surface is responsible for the disease and not the loss of interaction with a binding partner in these mutants (Fig. 5.7A). The R366W mutation is located in a region near to the CRD-Kr interface (Fig. 5.7B). Additionally, four mutations concern the removal or addition of a cysteine residue, which is likely to impair folding. In summary, published mutations that lead to RRS are likely to be attributed to the absence of ROR2 at the cell surface rather than the loss of an interaction mapping to one of these particular residues. Whilst expression studies suggest that the CRD-Kr interface is important for the stability of ROR2, mutations have not yet been identified which may disrupt this to cause RRS.

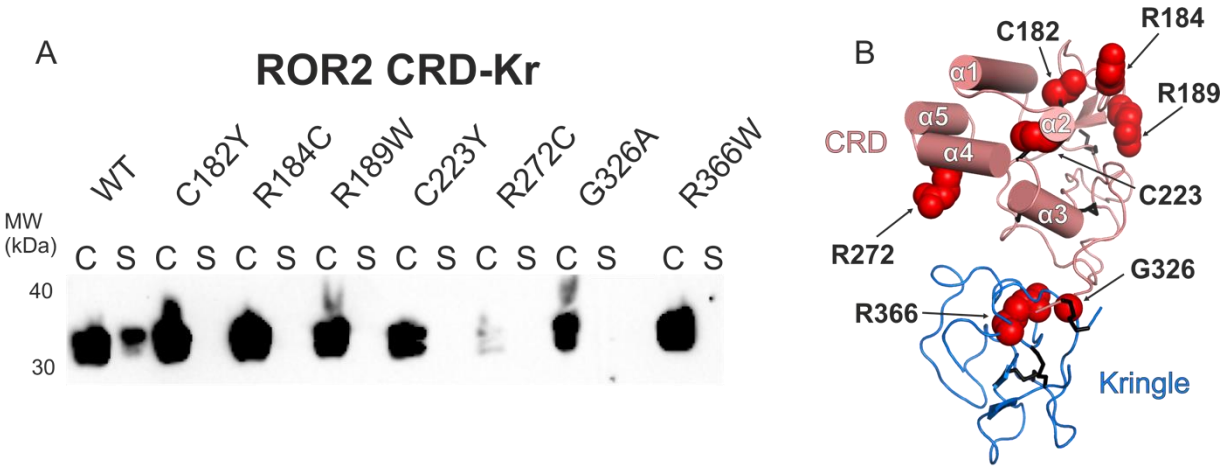


Figure 5.7 – Secretion test of ROR2 CRD-Kr RRS disease mutants – A) RRS mutants were cloned into the ROR2 CRD-Kr construct background. Equal volumes of HEK293T cell suspension ('C') and supernatant ('S') of the mutants indicated were analysed by western blot (WT - wild-type); B) Mutations mapped onto the ROR2 CRD-Kr structure. Side chains are shown as red spheres. The ROR2 domains are coloured as in Fig 5.5B.

5.3.5 Structural analysis

Structural phylogenetic analysis of the ROR2 CRD places it within the category of 'grooves' alongside Fz-type CRDs found in Wnt signalling and Hh signalling effector protein Smo (Fig. 5.8A). Structural alignment with the closely related CRD from MuSK (RMSD = 0.87 for 111 Cα positions; Supplementary Table 5.2) displays structural similarity between the two proteins, with α1 to α3 and the connecting loops in roughly equivalent positions. Helices α4 and α5 in ROR2 are replaced with a longer helix passing across the top of the structure in MuSK (Fig. 5.8B).

Comparison with Wnt-bound Fz8 (Fz8-PAM) (RMSD = 1.22 for 81 Cα positions) displays divergence between the positions of the helices. Helices α1 and α2 are in similar positions to that of Fz8-PAM, whilst a third α helix of Fz8 projects in a different direction over the location of ROR2 helices α3-5 (Fig. 5.8C). This third helix of Fz8 forms a groove with α1, within which the palmitoleate of Wnt is bound. Helix α5 of the ROR2 CRD projects across this third helix of Fz8 and into the position of this groove. Further comparison with the Smo CRD in complex with cholesterol (RMSD = 1.29 for 78 Cα positions) exhibits an equivalent situation, with helix α5 of the ROR2 CRD sitting across the ligand-binding groove (Fig. 5.8D). As with MuSK and Fz8, α1 and α2 are well-aligned between the ROR2 CRD and Smo-CLR. A helix of Smo-CRD which aligns with α4 of ROR2 is used in cholesterol binding. As with

Fz8, a contact between this helix and α1 of Smo-CRD forms the cholesterol-binding groove and the position of this groove is occupied by ROR2 helix α5.

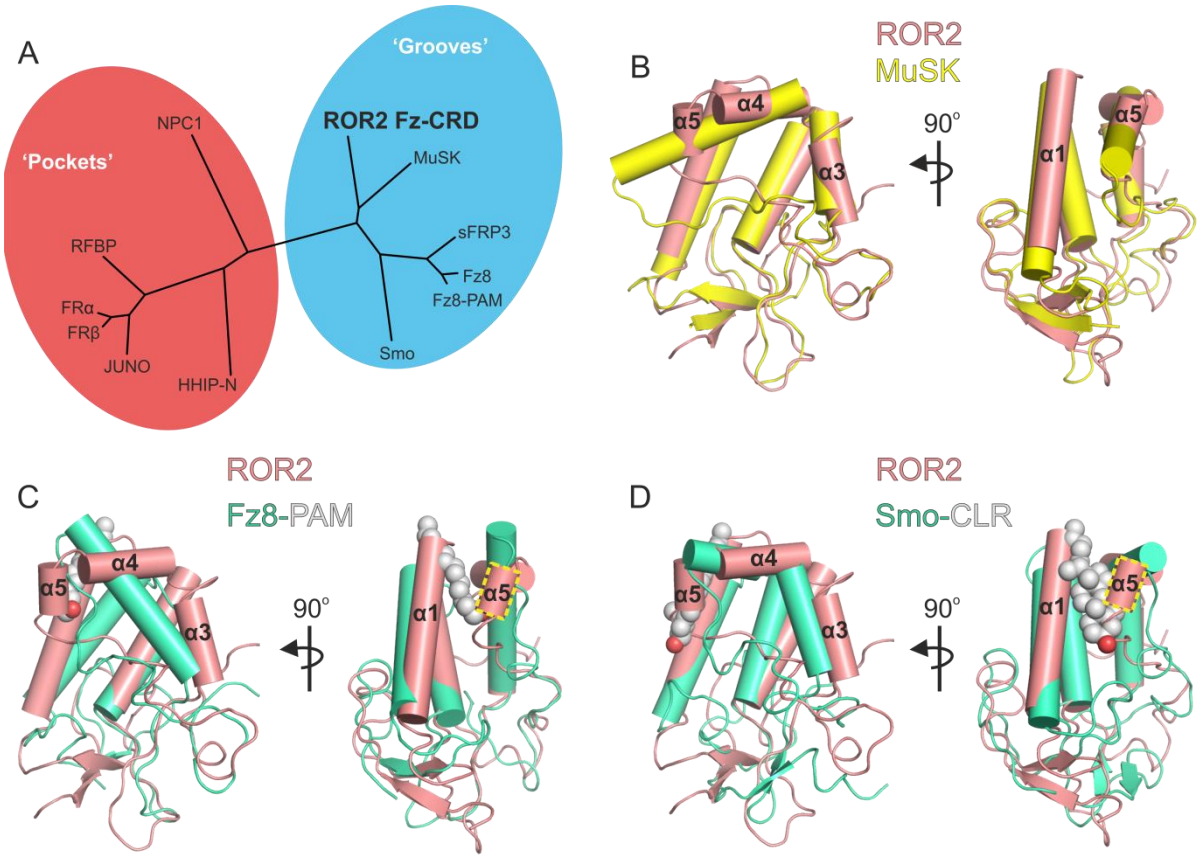


Figure 5.8 – Structural phylogenetic analysis of ROR2 CRD and superposition with structurally-related proteins. A) Structural phylogeny of CRDs including ROR2, based on Cα RMSD values (see Supplementary Table 5.2); B) Superposition of ROR2 CRD and the CRD of structurally related homologue MuSK (PDB ID: 3HKL); C) Superposition of ROR2 CRD upon Fz8 bound to Wnt8 (PDB ID: 4F0A). The palmitoleate from Wnt8 (PAM) is displayed in sphere representation; D) Structural alignment of the ROR2 CRD with the cholesterol (CLR)-bound CRD of Smoothened (PDB ID: 5L7D). CLR is depicted in sphere representation. In B-D, ROR2 helices are labelled α1 and α3-5, and disulphides have been omitted for ease of viewing. In C-D the α5 helix is highlighted with a dashed yellow box.

The ROR2 CRD does not possess a visible hydrophobic groove for the binding of a lipid-modified Wnt ligand. In Fz8, this groove is present in approximately

equivalent positions in both the apo- and Wnt-bound structures (Figs. 5.9A-B), the cholesterol-bound SMO CRD (Fig. 5.9C) and in the ligand-free sFRP3 (Fig. 5.9D). Neither the ROR2 CRD structure nor the MuSK CRD structure displays obvious hydrophobic grooves present for the recognition of palmitoleated-Wnt ligands (Fig. 5. 9E-F).

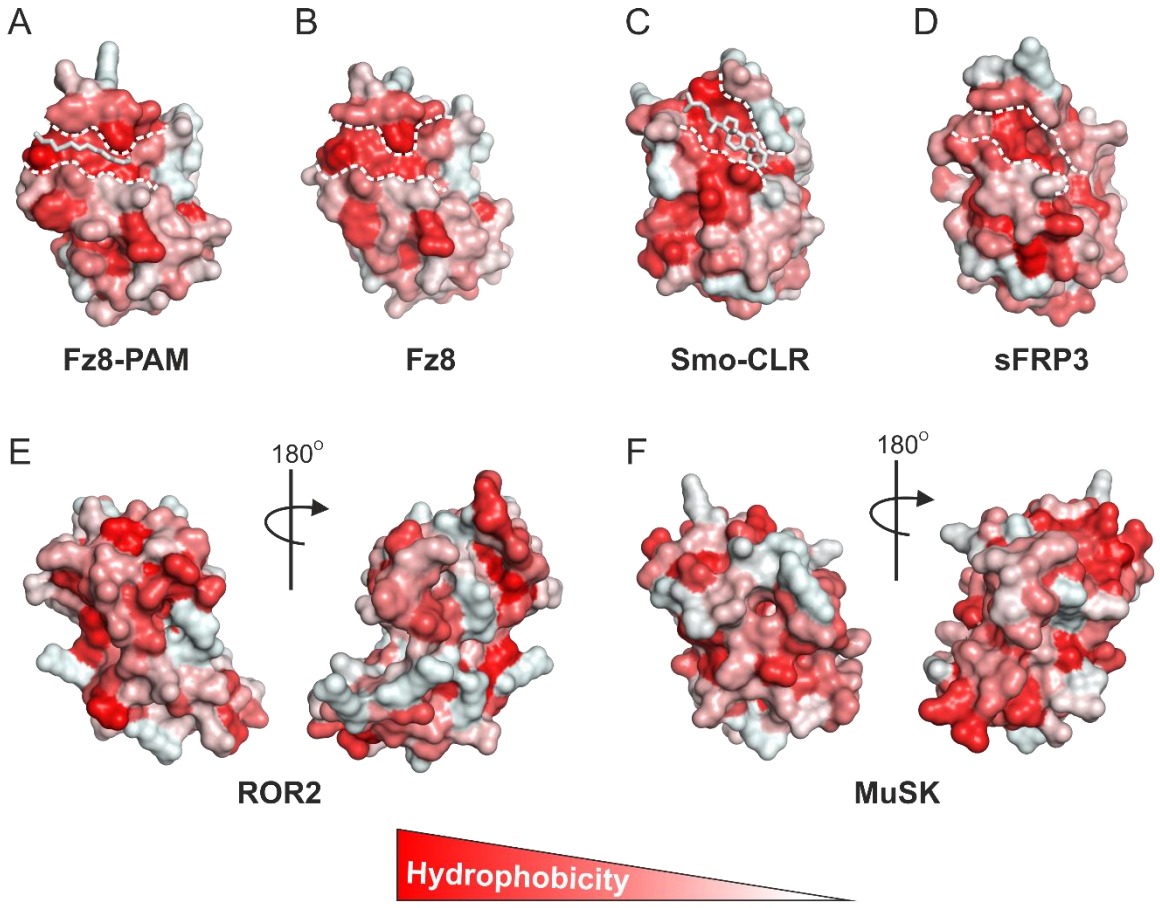


Figure 5.9 – Visualisation of hydrophobic surfaces of Fz-type CRDs. A) Fz8 with the palmitoleate from Wnt8 shown in stick representation (PDB ID: 4F0A); B) Fz8 in the apo-form (PDB ID: 1IJY); C) Smo bound to cholesterol (CLR; PDB ID: 5L7D); D) sFRP3 (PDB ID: 1IJX). All grooves in A-D are indicated by white dashed lines; E) The CRD of ROR2 shown in equivalent orientation to the structures in A-D (left) and rotated through 180°; F) The CRD domain of MuSK (PDB ID: 3HKL) oriented as in E).

Structural phylogenetic analysis of the ROR2 Kr domain does not represent the same level of structural deviation as found by comparison of CRDs (Fig. 5.10A and Supplementary Table 5.3). The most dissimilar Kr domain structurally characterised is found in Kremen (RMSD = 0.71 for 57 Cα positions). For comparison, an RMSD of 2.77 is observed when comparing the ROR2 CRD to FRα (for 51 Cα positions). Structural superposition of the ROR2 Kr domain with both the most (Fig. 5.10B) and the least divergent (Fig. 5.10C) Kr domains analysed displays conservation of the positions of disulphide bonds VII and VII from ROR2 as well as the β-sheet. Additionally, the loops are in roughly similar positions. Disulphide VI from ROR2 does not align with the equivalent disulphide in either superposition.

In the HGF NK-1 (N-terminal and Kr domain fragment-1) structure, the Kr domain occupies a similar position relative to the N-terminal domain as the Kr domain in ROR2 relative to the CRD (Fig. 5.10D). This face of the Kr domain forms a range of interactions with the N-domain of HGF. The Kr domain of NK-1 buries ~450 Å^2 per domain with 17 ordered water molecules and two direct inter-domain hydrogen bonds are visible (Ultsch et al., 1998). The ROR2 Kr domain buries ~354 Å^2 of the CRD surface (Krissinel and Henrick, 2007), and therefore this face of the Kr domain is likely involved in the formation of a protein:protein interface.

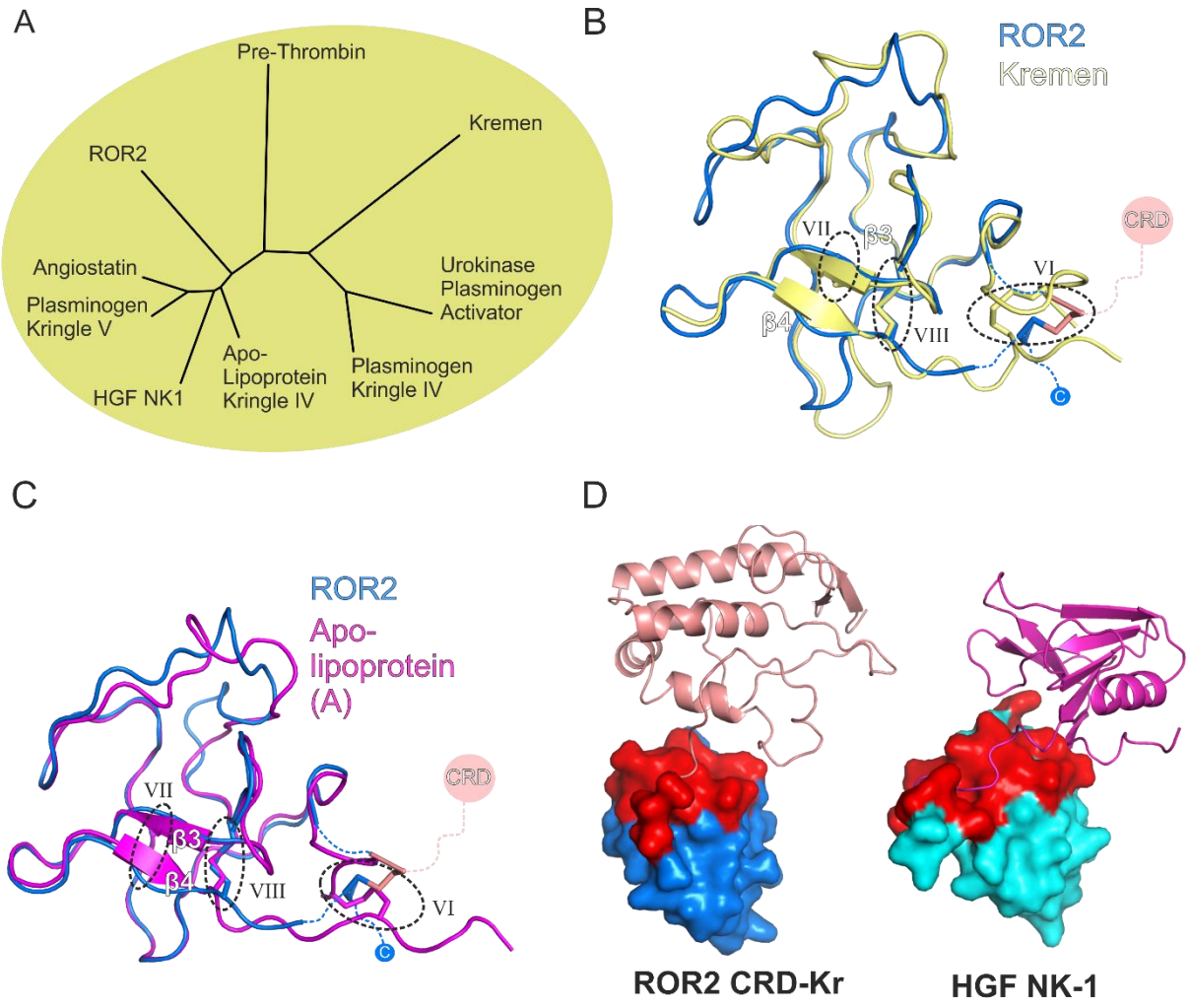


Figure 5.10 - Structural analysis of the ROR2 Kringle domain. A) Structural phylogeny of Kr domains related to that of ROR2 (see Supplementary Table 5.3); B) Superposition of the ROR2 Kr domain (blue) and the most distantly-related Kr domain analysed (Kremen, yellow; PDB ID: 5FWS). Disulphides are circled with dashed lines and annotated with Roman numerals and β-strands are numbered as in Fig. 5.5; C) Superposition of the ROR2 Kr with the most closely structurally-related Kr domain (Apo-lipoprotein Kr domain IV, magenta; PDB ID: 1I71). In B and C, the relative position of the ROR2 CRD is indicated by a dashed line, and the cysteine in disulphide bond VI from the C-terminus of the CRD is shown in salmon. Disordered loops within the ROR2 Kr domain are shown as dashed blue lines; D) Comparison of ROR2 CRD-Kr with HGF NK-1 (PDB ID: 1BHT). Kr domains from both structures are shown in surface representation and interface residues are coloured red.

5.3.5 Biophysical characterisation of ROR2 constructs

To characterise the oligomeric state in solution, the ROR2 ECD was characterised using MALS. Species eluting with molecular weights of approximately 47.5 and 45.4 kDa for the native purified protein and deglycosylated sample respectively are observed in MALS experiments (compared with the predicted molecular weight of 42.8 kDa for the protein alone). This indicates that no domain in the ROR2 ECD mediates oligomerisation (Fig. 5. 11).

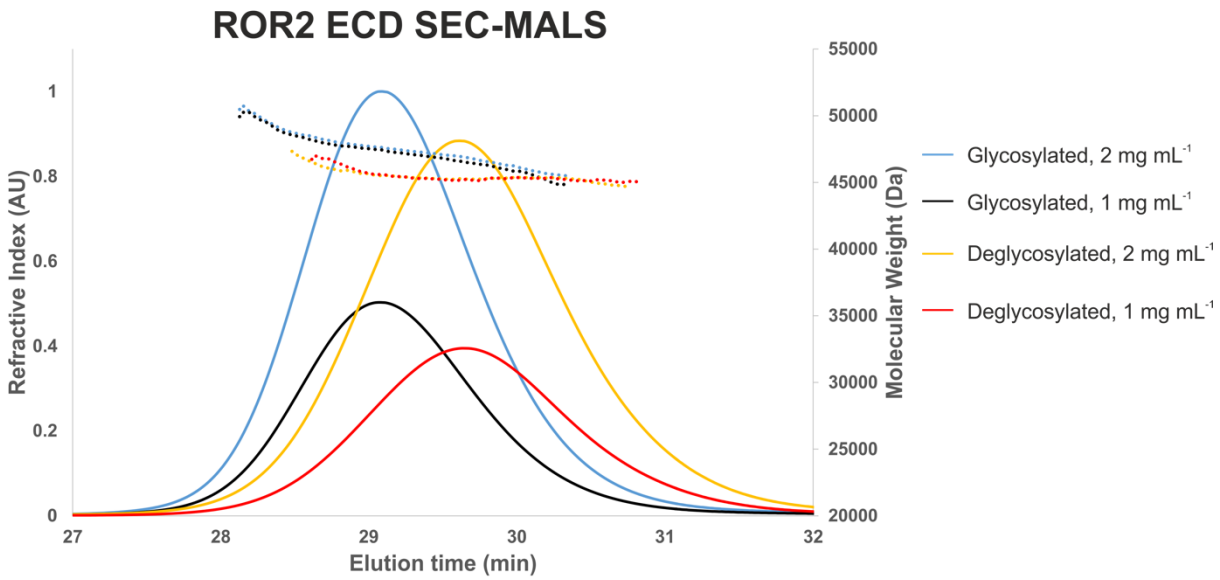


Figure 5.11 – SEC-MALS analysis of the ROR2 ECD. Samples treated with EndoF1 (deglycosylated) and without ndoF1 (glycosylated) were separated utilising an S200 10/30 column and eluate analysed via MALS (Concentration range 46-23 μM).

In the absence of crystals of the ROR2 ECD, small-angle X-ray scattering (SAXS) experiments were performed in order to study the arrangement of the Ig

domain relative to the CRD and the Kr domain. Attempting to fit the crystal structure of the ROR2 CRD-Kr region to the experimental data collected from the ROR2 ECD gave a χ^2 value of 3.35 (Fig. 5.12A). Adding a homology model of the N-terminal Ig domain and modelling the residues between this domain and the CRD produced an improvement in the fit and a drop in this value to 1.79 (Fig. 5.12B). Modelling of glycosylation sites in this construct was performed alongside a minimal ensemble search, which gave two extended models of ROR2 differing in linker residue orientations, thus leading to a decrease in the χ^2 value to 1.39 (relative populations of 54% and 46% of the two models) (Fig. 5.12C). For comparison, a glycosylated model of the ROR2 ECD in which the N-terminal Ig domain is brought into close contact with the CRD ('non-extended') leads to a poorer fit to the data, with an increase in χ^2 to 3.10 (Fig. 5.12D). As a control, a theoretical curve was also calculated for the ECD of TrkA (37.8 kDa, 6 N-linked glycosylation sites; 1 Leucine-rich repeat, 2 Ig domains), (Fig. 5.12E). Comparison of fitting residuals between the glycosylated and extended ensemble of ROR2 with the non-extended glycosylated model exhibits smaller deviation between modelled and experimental values for the extended model (Fig. 5.12F). Therefore, the ROR2 ECD is found in an extended conformation and not in a non-elongated 'ring-like' one.

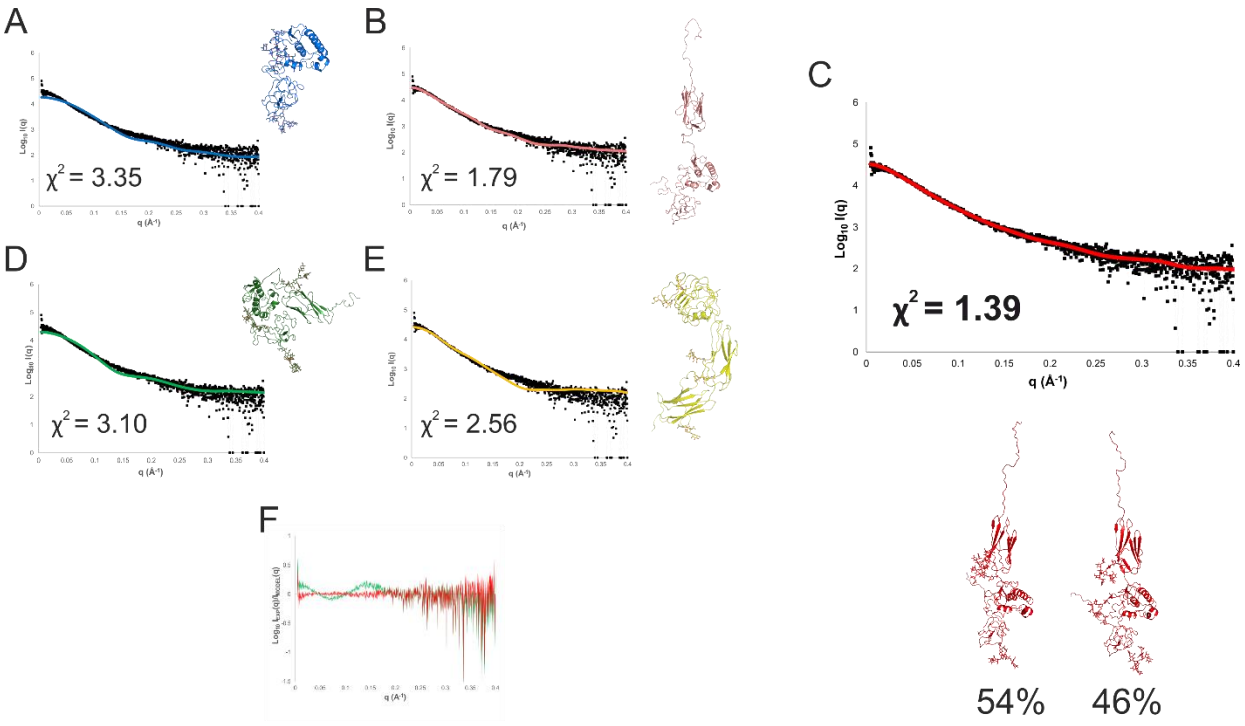


Figure 5.12 – SAXS analysis of the ROR2 ECD; SAXS data were collected from the ROR2 ECD. A number of models comprising A) ROR2 CRD-Kr); B) ROR2 ECD, no glycosylation; C) extended ROR2 ECD ensemble (glycosylated; relative proportions of two different models are displayed beneath the curve) and D) non-extended ROR2 ECD (glycosylated) were prepared using ALLOSMOD (Weinkam et al., 2012), alongside E) the crystal structure of the TrkA ECD (PDB ID: 2IFG). From these models, theoretical SAXS curves were calculated. These were compared to an experimental SAXS curve collected for the glycosylated ROR2 ECD. χ^2 values were then calculated to analyse the goodness of fit to the experimental data, and are indicated beneath the curves. In each case, black dots correspond to the experimental curves, whilst the coloured line represents the curve calculated from the model; F) Fitting residuals for the models from C (red line) and D (green line), compared with the experimental ROR2 ECD SAXS curve.

5.3.6 Binding experiments with ROR2

ROR2 CRD:Wnt interactions have been reported to be involved in both canonical and non-canonical Wnt signalling pathways, and therefore binding of the ROR2 CRD-Kr construct to canonical pathway activator Wnt3a was assayed using a cell-based reporter assay (Li et al., 2008; Oishi et al., 2003). Addition of the ROR2 CRD-Kr does not significantly alter Wnt-mediated luciferase activity (over the range 1 μM to 125 nM), implying no interaction between the ROR2 CRD and Wnt3a, or that this interaction is compatible with simultaneous FZD:Wnt3a binding. For comparison, the Wnt-binding signalling antagonist sFRP4 antagonises Wnt3a-induced canonical Wnt signalling in a dose-dependent fashion over the range 0.5 - 1 μM (Fig. 5.13).

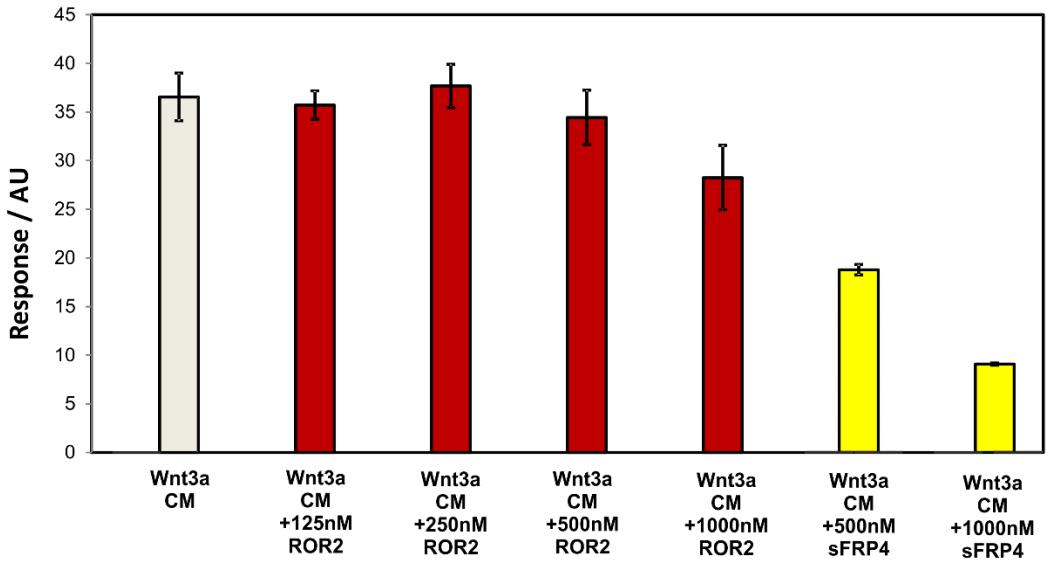


Figure 5.13 – Wnt3a response assay. Grey bar – Wnt3a conditioned media (CM); red bar – titration of purified ROR2 CRD-Kr into Wnt3a CM; yellow bars – positive control titration of sFRP3 into Wnt3a CM. Error bars represent one standard deviation (AU – Absorbance units).

An interaction between the ROR2 CRD domain and the ECD of BMP receptor BMPR1b has been reported (Sammar et al., 2009), and biophysical assays were used to verify this interaction using purified proteins. The interaction was probed utilising an SPR equilibrium binding experiment. Constructs comprising ROR2 ECD, CRD-Kr and CRD were immobilised at ~300 RU on the sensor chip. Experiments with BMPR1b injected at a concentration of 200 μM did not display visible binding (Fig. 5.14A). MALS experiments involving a titration of the BMPR1b ECD into ROR2 CRD-Kr did not cause a shift to a MW higher than that of the ROR2 CRD-Kr construct injected alone (Fig. 5.14B). In order to analyse whether the interaction is mediated by some other component present in cell culture medium, the ROR2 ECD (His_6-tagged) and BMPR1b ECD (1D4-tagged) were co-transfected into HEK 293T cells and purified via both IMAC and SEC. The resultant peaks were analysed via SDS-PAGE and did not contain BMPR1b (Fig. 5.14C-E). These data suggest that the ROR2 CRD does not interact with the BMPR1b ECD and is unlikely to bind to Wnt3a in the same manner as FZD.

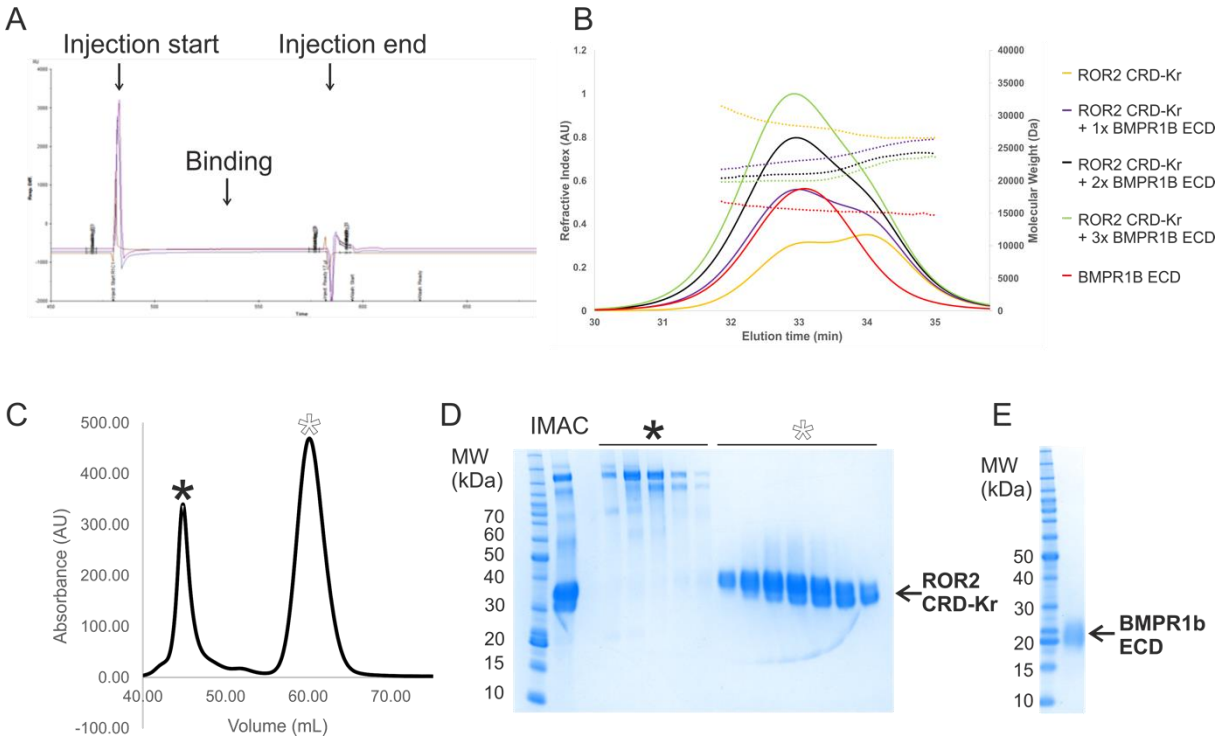


Figure 5.14 – Biophysical and cellular study of the proposed ROR2 CRD:BMPR1b ECD interaction. A) SPR experiment with ROR2 ECD, CRD-Kr and CRD as ligand and BMPR1b injected at 200 μM final concentration. The absence of an increase in response units over the course of the analyte injection process ('binding') indicates the absence of a ligand:analyte interaction; B) SEC-MALS experiment with injections comprising ROR2 CRD-Kr alone (1 mg mL^{-1}; 37 μM), BMPR1b ECD titrated at 1:1, 2:1 and 3:1 molar ratios into ROR2 CRD-Kr and BMPR1b ECD alone (1.5 mg mL^{-1}; 106 μM). The ROR2 CRD-Kr region showed an experimental molecular weight of 27.8 kDa (theoretical molecular weight: 26.7 kDa protein plus 2 predicted N-linked glycosylation sites); C) UV trace for purification of co-expressed ROR2 CRD-Kr (His_6-tagged) and BMPR1b (1D4-tagged) by SEC (S200 16/60) following an IMAC step; D) Molecular weight analysis of peaks from C) by SDS-PAGE, with peaks indicated using black and white asterisks; E) SDS-PAGE of purified BMPR1b ECD for reference.

5.4 Discussion

Within this work, structural studies of the Wnt signalling regulator ROR2 have been performed. The CRD-Kr region has been structurally characterised utilising an approach combining Pt-SAD phasing and molecular replacement. Both domains form an associated structural unit. This structure represents the first example of a CRD determined from a multi-domain protein alongside an adjacent domain. The ROR2 CRD belongs to the 'groove' sub-group of CRDs, and possesses a helical topology maintained by 5 disulphide bonds similar to that of MuSK. A contact interface is observed between the CRD and the Kr domain, forming a structural unit. This arrangement of the Kr domain relative to other domains has been observed in other structurally-characterised proteins, such as the NK-1 fragment of HGF (Ultsch et al., 1998). SAXS studies revealed that the Ig domain is flexible relative to the structural unit of the CRD-Kr region. This causes the ECD to adopt an elongated arrangement in solution similar to other multi-domain proteins such as RPTPσ and TrkA (Coles et al., 2014; Wehrman et al., 2007)

5.4.1 Functional insights from the ROR2 CRD structure

The molecular nature of ROR2 signal transduction has yet to be fully delineated. Studies have presented evidence for direct interactions between the ROR2 CRD and canonical Wnts 1, 3a and 8a (Billiard et al., 2005) and non-canonical Wnts 5a and 11 (Hikasa et al., 2002). Additionally, it has been shown that Wnt can initiate signalling through ROR2 alone (Liu et al., 2008), via recruitment of ROR2 and a FZD receptor (Grumolato et al., 2010) or through an even larger architecture comprising ROR2, FZD and LRP5 (Li et al., 2008). Other studies have suggested that ROR2 functions exclusively through non-canonical signalling via Wnt5a binding, and does not physically interact with Wnt3a (Grumolato et al., 2010; Oishi et al., 2003). ROR2 and Wnt5a knockout organisms display overlapping phenotypes (Oishi et al., 2003). Stimulation of FZD4- and LRP5-expressing cells with Wnt5a activates canonical signalling, representing the ability of Wnt5a to activate different signalling pathways depending on receptor context (Mikels and Nusse, 2006). To add to this, the ROR2 ECD is reported to interact with the ECD of PTK7, another RTK involved in the activation of non-canonical signalling by Wnt5a (Martinez et al., 2015). Therefore, the role of ROR2 signalling in response to Wnt ligands is poorly understood but appears to vary depending on the presence of various co-receptors.

A cell-based reporter assay shows no modulation of Wnt3a signal transduction via ROR2 CRD-Kr, whilst known antagonist sFRP4 demonstrated signal inhibition (Fig. 5.12), confirming previous results (Grumolato et al., 2010; Oishi et al.,

2003). The structure of the ROR2 CRD does not present an obvious Wnt-binding region. The 'lipid thumb' of Wnt8 is bound within a groove on the Fz8 surface, two regions which are well conserved across Wnt ligands and the CRDs of their FZD receptors (Janda et al., 2012). Besides P130 (P275 in ROR2), the residues in this site are not conserved in ROR2. Additionally, ROR2 possesses two helices (α1 and α5), which occupy the 'groove' seen in other FZD CRDs. Specifically, helix α5 (highlighted with dashed white lines) represents an insertion relative to Fz-type CRDs from FZDs clashes with any potential 'groove' (Fig. 5.15A). In comparison to the Wnt8:Fz8 complex structure, the 'index finger'-binding site of ROR2 is more similar in structure to that of Fz8 (Fig. 5.15B) but shows low sequence conservation (Fig. 5.15C). Wnt ligands are also well conserved in the 'index finger', implying a similar mode of binding between Wnt ligands and FZD CRDs as in the Wnt8:Fz8 complex (Fig. 5.15D) (Janda et al., 2012). However, the absence of conservation between FZD receptor CRDs and ROR2 at the hypothesised Wnt recognition sites render it unlikely that the ROR2 CRD structure represents a pre-formed Wnt-binding surface. This is in contrast to Fz8 structures solved in the absence and presence of Wnt, which does not require remodelling in order to interact with the Wnt ligand (Dann et al., 2001; Janda et al., 2012). Fz4 also closely resembles the structure of Fz8 in the absence of a binding partner, as does structurally analogous Wnt antagonist sFRP3 (Chang et al., 2015; Dann et al., 2001). A recent report found that ligand binding to the human Smo CRD triggers a conformational change relative to its apo-form (Huang et al., 2016); this is at odds with the structure of zebrafish apo-Smo, however, and this study also does

not take the 7TM region of Smo into account (Byrne et al., 2016; Nachtergaele et al., 2013). Visible hydrophobic surface grooves are seen for other Fz-type CRDs for which ligands have been identified (Fig. 5.9A-D), whereas no such regions are seen for the ROR2 or MuSK CRDs (Fig. 5.9E and F).

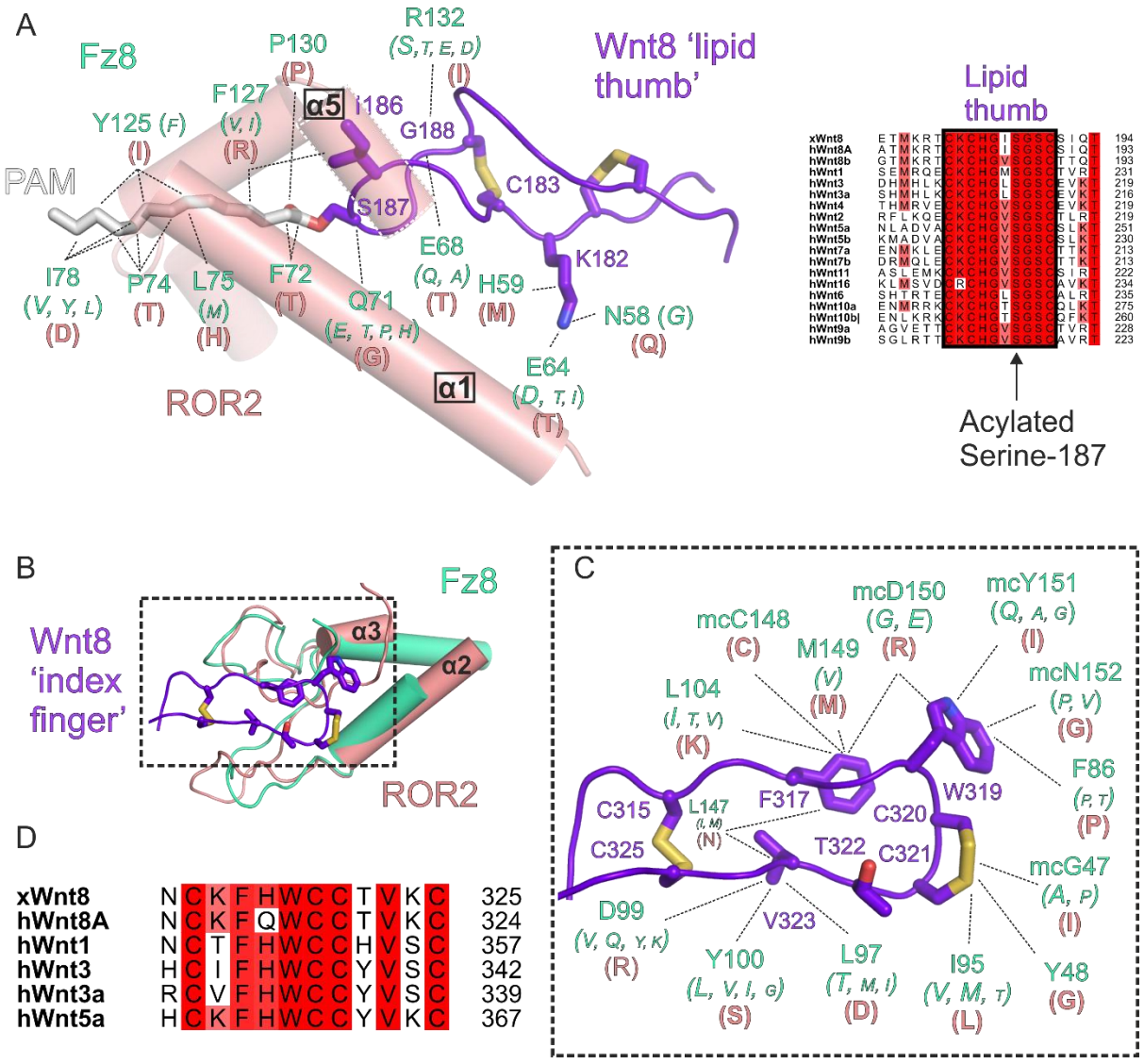


Figure 5.15 – Structural analysis of Wnt-binding interfaces on the ROR2 CRD. A) (Left) Close-up view of the 'lipid-thumb' interface of Fz8-Wnt8 aligned with the ROR2 Fz-type CRD. Residues from mFz8 are shown in green and alternative residues in other human FZD proteins are shown in parenthesis, with character size representing abundance. Black dashed lines indicate non-covalent interactions. Salmon characters in parenthesis indicate ROR2 residues at these positions. ROR2 CRD helix α1 is labelled and 'helical insert' α5 is indicated with white dashed lines. Wnt8 lipid thumb residues are displayed in purple. (Right) Alignment of Wnt lipid thumb sequences; B) Wnt8:Fz8 interaction site 2, utilising the Wnt 'index finger' region. Fz8 and ROR2 are aligned to show the structural similarity in these regions; C) Close-up of interactions with the Wnt8 'index finger' annotated as in A). 'mc' refers to main-chain contacts; D) Sequence alignment of Wnt 'index finger' regions between proteins reported to bind ROR2 in the literature.

The arrangement of helices in the MuSK CRD is similar to that of ROR2. The long helix α1 clashes with palmitoleate in both ROR2 and MUSK CRD structures, whilst this helix is shifted in the Wnt8:Fz8 to permit formation of a hydrophobic groove and enable binding to palmitoleate (Figs. 5.8D and 5.9A-B). Helices α4 and α5 in ROR2 align with a longer helix (α4 in MuSK) and project into the possible cholesterol- or palmitoleate-binding site (Fig. 5.16A). For the helices of the ROR2 CRD, this α3-α5 region displays the highest crystallographic B-factors, which represents greater flexibility in these regions and the potential for ligand-induced reorganisation. The same can be said for the MuSK CRD, with B-factors being highest in α1 and α4 (Fig. 5.16B). There are no disulphide bonds stabilising the regions occluding ligand-binding in either ROR2 or MuSK. Therefore, these regions may be remodelled upon Wnt binding, representing a novel transition in a Fz-type CRD in order to form a ligand-binding groove. As discussed, physical interactions between ROR2/MuSK and Wnt ligands have been mapped to their CRDs (Oishi et al., 2003; Jing et al., 2009; Zhang et al., 2012). These interactions are linked to non-canonical signalling, and a structural basis for how ligand-induced activation occurs requires further study.

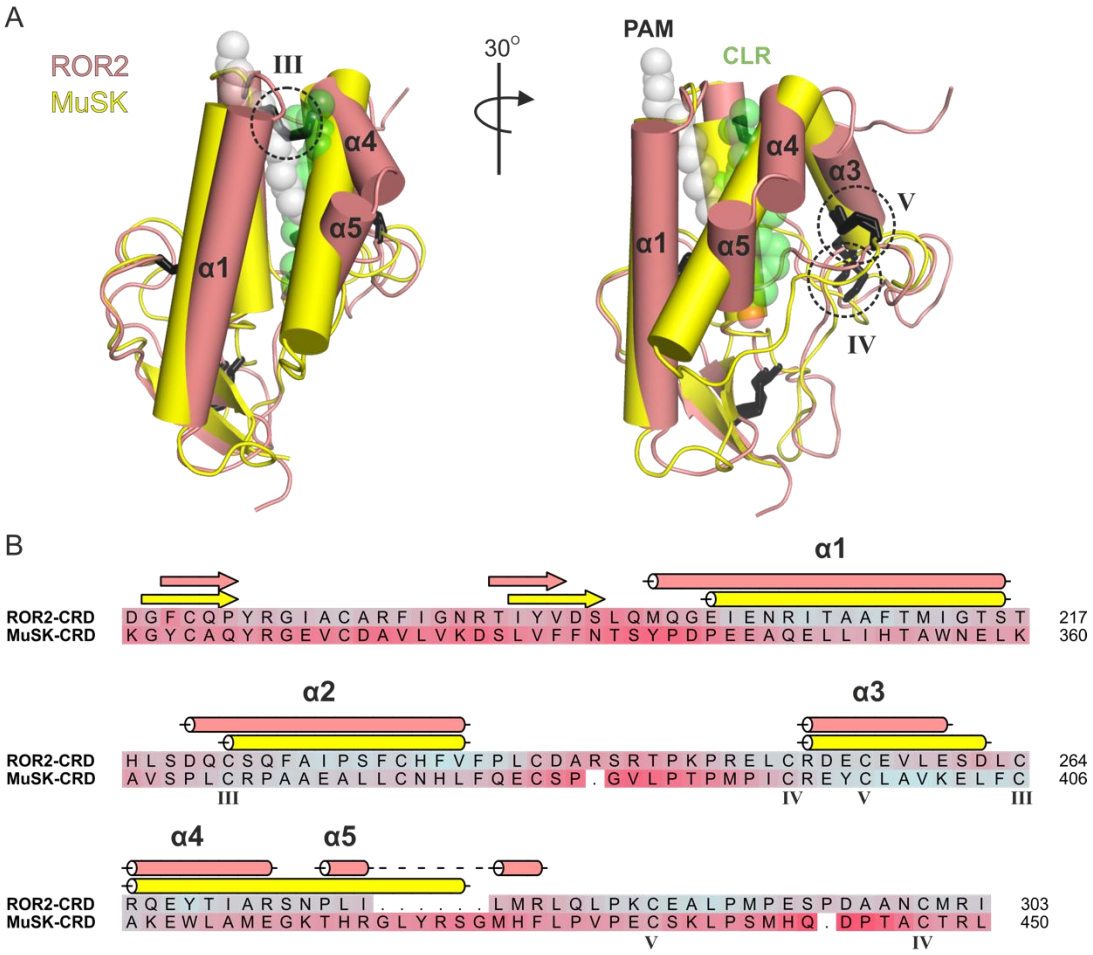


Figure 5.16 – ROR2/MuSK CRD comparison. A) Structural alignment of the ROR2 and MuSK (PDB ID: 3HKL) CRDs, with helices numbered α1-5 and notable disulphides numbered with Roman numerals and circled with dashed lines. Palmitoleate (PAM) and cholesterol (CLR) from Fz8 (PDB ID: 4F0A) and Smoothened (PDB ID: 5L7D) are shown as transparent spheres; B) Sequence alignment of ROR2 and MuSK CRDs. Secondary structure is displayed above the alignment and is coloured as in A), with α1-5 numbered for ROR2. Disulphide bonds indicated in A) are numbered beneath the sequences as Roman numerals. Crystallographic B-factors are plotted in pale blue (low) to red (high) over the sequences.

The CRDs of both FZD2 and sFRP2 have been reported to form complexes with the ROR2 ECD, with this activity in the latter case being attributed to the ROR2 CRD (Brinkmann et al., 2016; Oishi et al., 2003). The ASUs of the Fz8 and sFRP3 crystal structures were witnessed to contain dimers of both proteins, leading to speculation that dimerisation of Fz-type CRD-containing proteins was important for function (Dann et al., 2001). Whilst no element of the ROR2 ECD is seen to mediate oligomerisation (Figs. 5.11) and the solution state of sFRP3 and Fz8 has not been investigated, the formation of CRD heterodimers may contribute to their functions *in vivo*.

In conclusion, there are many candidate binding partners for the ROR2 CRD and Wnt may not be the only ligand. Furthermore, no obvious groove for the binding of the lipid modification on Wnt ligands exists upon the surfaces of either the ROR2 or MuSK CRDs and therefore an as-yet unseen conformational change would be required for an interaction structurally comparable to that of Fz8 with Wnt8 (Janda et al., 2012). A lack of surface conservation between ROR2 from different species and homologous Fz-type CRDs renders prediction of receptor-binding elements in the CRD difficult.

5.4.2 The roles of the ROR2 Ig and Kringle domains

The ROR2 ECD is a modular unit, comprising an Ig domain, a CRD and a Kr domain (Fig. 5.1A). All studies up to this point have assumed that the CRD is responsible for all ROR2 interactions *in vivo*. Whilst it is possible to produce variants

of the ECD lacking the Kr domain, the expression levels of such constructs are severely impaired (Fig. 5.6C). Based on the structure presented here, the ROR2 CRD-Kr region is seen to fold into a structural unit potentially explaining this increased stability (Figs. 5.5A-B and 5.6A). The Kr domain of NK-1 fulfils a similar structural role to the N-domain, forming a contact interface between the two domains (Fig. 5.9D). Additionally, this domain forms contacts with the HGF receptor MET (Ultsch et al., 1998). The ROR2 CRD-Kr interface is not as extensive however, and is dominated by weak interactions.

Drosophila ROR2 is only composed of the CRD and the Kr domain, and is only expressed in the developing nervous systems and not the limbs as in vertebrate (Al-Shawi et al., 2001; Matsuda et al., 2001; Oishi, 1997). The *C. elegans* ROR homologue, CAM-1, displays the equivalent three-domain architecture of human ROR2 and binds to Wnt homologues (Green et al., 2007). Monomeric species of purified ROR2 CRD-Kr and ECD constructs are observed at the concentrations analysed, suggesting that no ROR2 domain facilitates oligomerisation (Figs. 5.10 and 5.13B). This suggests that the Ig domains of human ROR2 and *C. elegans* CAM-1 may have been retained to interact with some unidentified ligand. The presence of the Kr domain in dROR2 implies an essential role across species, be it in CRD stabilisation (Fig. 5.6A) or ligand binding. However, without a system with which to functionally analyse ROR2 it is not possible to identify the functional relevance of the CRD-Kr structural unit.

Solution studies of the ROR2 ECD indicate a flexible molecule in an elongated conformation. It was not possible to model the ROR2 ECD to the extent of a perfect fit, likely due to poor data quality at high q angles, poor buffer subtraction and availability of molecular dynamics simulation tools. However, the improvement in model fit upon the addition of the Ig domain (Figs. 5.12A-B) and the inclusion of glycosylation sites and relative orientation of the modelled Ig domain (Fig. 5.12C) gives insight into the overall architecture of the ROR2 ECD. Solution studies utilising a model in which the linker between the Ig domain and the CRD had been kinked, resulting in a more globular, ring-like structure gave a poorer fit (Fig. 5.12D). This suggests that an extended conformation at the N-terminus as being closer to the true solution state. Ig domains are found in a number of elongated receptor molecules, and typically exist as protein-protein interaction interfaces or as structural components to elongate molecules and facilitate cell-cell adhesion. Axonal and synaptic protein RPTPσ, for example, possesses an ECD with 3 N-terminal Ig domains and 5-9 FNIII domains. A great deal of flexibility is displayed by this protein, which projects across the synaptic cleft to interact with other proteins (Coles et al., 2011, 2014). It is therefore possible that the Ig domain of human ROR2 interacts with another protein *in vivo*. The large amount of conformational space explored by this Ig domain may be required to facilitate *trans*-interactions.

The Fz-type CRD is a common motif involved in the regulation of morphogen signalling, particularly by the Wnt pathway. Fz-type CRDs are typically found at the

N-termini of homologous proteins, except for in ROR1/2 and MuSK. Further parallels can be drawn to MuSK, which has an ECD comprising three N-terminal Ig domains and a juxtamembrane CRD. Whilst the CRD has been shown to bind to Wnt ligands, the first Ig domain has been reported to bind LRP4 and the third Ig domain to BMP ligands (Yilmaz et al., 2016; Zhang et al., 2011). Although the reported ROR2 CRD:BMPR1b interaction could not be replicated *in vitro* (Fig. 5.14), these results do not preclude such an arrangement *in vivo*. GDF-5, related to BMPs, was reported to enhance ROR2:BMPR1b cross-linking and, given the MuSK:BMP interaction, may represent an interaction partner for ROR2 (Sammar et al., 2004).

In conclusion, the ROR2 ECD has several domains, each of which is likely involved in some aspect of signalling. Interactions with Wnt ligands through the central CRD represent the best understood binding event performed by ROR2, but it is unlikely that this represents the full picture. Other co-receptors and ligands may play a vital role within the cellular context, and interactions may be further extended to the Ig and Kr domains.

5.4.3 Further work

The insolubility of Wnt ligands represents the major issue in studying receptor interactions. Endogenous proteins can influence cellular assays and results from these do not necessarily reflect direct binding between two proteins. Production of large enough quantities for cellular assays is achievable, but is not sufficient for biophysical assays. Therefore, a focus for the study of ROR2 entails large-scale of

Wnt ligands for direct binding assays. Binding assays via SPR, ITC or ELISA with both Wnt ligands and proteins already in production (BMPs 2, 4 and 7, GDF-5, sFRPs, LRP5/6 and CRDs from FZD members) must be performed with ROR2 to validate interactions and identify the domain(s) required. This may be supplemented by molecular dynamics simulations to analyse whether α5 may change conformation to yield a lipid-binding groove.

Signal activation by Wnt5a antagonises signalling activation by Wnt3a. This is dependent on signalling through ROR2 (Mikels and Nusse, 2006). As discussed, Wnt5a is likely to be a true ligand for ROR2, and a system to study this interaction could be used to test ROR2 mutations. This could include mutations to disrupt the CRD-Kr interface, or test whole domain deletions. Additionally, a chimera replacing ROR2 α3-5 with the palmitoleate-binding α-helix from the comparable of Fz8 could be used to test for Wnt3a binding and signalling ability, which may give insight into the evolution of CRD:Wnt interactions.

The roles of all three extracellular domains of ROR2 are currently poorly understood, but there are at least candidate ligands for interaction with the CRD. Therefore, tandem affinity purification and MS analysis of the Ig and Kr domains may identify potential interacting species. SAXS analysis shows that the Ig domain of ROR2 is not closely associated with the CRD in the same fashion as the Kr domain, and therefore it may be possible to produce this domain in isolation in large quantities for crystallographic analysis. Beyond this, crystallisation of the core Ig

domain and collection of higher quality SAXS data of the full ECD (via dialysis of samples to gain a higher quality buffer subtraction, use of sodium nitrate in all buffers to reduce radiation damage (Kmetko et al., 2011) and higher protein concentration to give higher intensity data in the high q angle region) would aid in the construction of a more complete picture of the domain arrangement in solution.

5.4.4 Concluding remarks

Of the 5 Wnt co-receptors with intracellular RTK domains (ROR1/2, Ryk, MuSK and PTK7), 4 are thought to be inactive 'pseudokinases' and the reason for this is unclear (Mendrola et al., 2013). RTK signalling requires the formation of a symmetric complex in which a dimeric ligand brings together two receptors at the cell surface and causes trans-phosphorylation of the intracellular kinase domains (Lemmon and Schlessinger, 2010). The structure of the ROR2 kinase domain is continuous with an inactive enzyme, with a tyrosine residue occupying the ATP-binding site (Mendrola et al., 2013). The intracellular region of ROR2 is required to activate Wnt5a-mediated inhibition of canonical signalling, but tyrosine kinase activity is not required for activation of non-canonical signalling (Grumolato et al., 2010; Mikels and Nusse, 2006). These intracellular kinase domains may be phosphorylated via different effectors to enable interactions with various partners and integrate signals into several signal transduction pathways as a consequence of different extracellular modulators.

The lack of variation in Wnt 'lipid thumb' and 'index finger' regions (Fig. 5. 14A and D) means that there is unlikely to be a strict 'Wnt receptor code.' Wnt ligands are known to signal through an asymmetric Wnt:FZD:co-receptor platform (Janda et al., 2012). It is possible that ROR2 utilises a third conserved region of Wnt ligands, found on the convex face halfway between the two Fz-binding sites. This likely represents the a co-receptor-binding site, and is a potential LRP5/6-/PTK7-/Ryk-/ROR-binding region in other modes of Wnt signalling (Janda et al., 2012) (Fig. 5.17A, dashed oval). ROR2 binding to Wnt5a is known to require the CRD (Oishi et al., 2003). There is sequence homology between both FZD receptor CRD Wnt-binding sites and in the receptor binding elements of Wnt ligands themselves (Fig. 5.15). In the absence of structural characterisation of the ROR2 CRD:Wnt interaction, it must be assumed that this binding mode is equivalent to the Wnt8:Fz8 interaction (Janda et al., 2012). As discussed, this would require conformational rearrangement of the ROR2 CRD. This raises the question of whether Wnt ligands are capable of acting as dimers (or higher-order oligomers) and if the true role of receptors such as ROR2 is to undergo ligand-induced dimerisation (Fig. 5.17B). Further structural characterisation of 'non-canonical' Wnts and their receptor complexes will grant insight into the likely mechanics utilised by co-receptors such as ROR2 in cellular signal transduction.

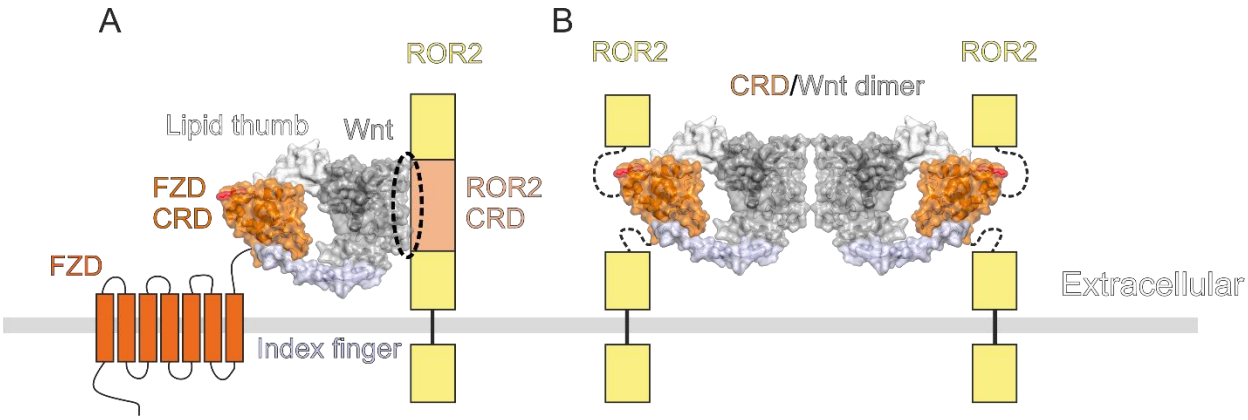


Figure 5.17 – Potential modes of ROR2:Wnt interaction. A) The ROR2 CRD, like the LRP5/6 co-receptor, uses an alternative Wnt interface to the FZD receptor CRD (dashed oval); B) The ROR2 CRD:Wnt interaction is the same as the Wnt8:Fz8 interaction, and signalling is activated via a dimeric Wnt ligand bringing together two ROR2 molecules. The ROR2 CRD:Wnt interaction is represented using the structure of the Wnt8:Fz8 complex (PDB code: 4F0A).

Chapter 6

Summary and concluding remarks

Chapter 6: Summary and concluding remarks

Morphogenesis is a process vital for the establishment of multicellular life, found in higher eukaryotes such as humans down to flies and worms (Perrimon et al., 2012). Morphogens do not act in isolation at any one time, and are subject to tight spatial and temporal control mechanisms. Dysregulation of signalling processes is linked to a number of disease states. A host of receptors are required for appropriate patterning and these are composed of a variety of domains. An extracellular domain common to receptors and modulators within morphogen signalling is the Cysteine-Rich Domain (CRD), a tightly disulphide-bonded α-helical bundle typically consisting of 150 – 250 amino acids (Bazan and de Sauvage, 2009). These are found individually and also as part of larger multi-domain architectures (Bishop et al., 2009). The binding partners of CRDs are diverse, ranging from small molecules to entire proteins (Janda et al., 2012; Wibowo et al., 2013).

Thus far, two CRD-containing proteins within a multi-domain architecture have been characterised structurally: Smo, a cholesterol-binding 7TM-containing Hh effector (Byrne et al., 2016) and NPC1, a lysosomal transmembrane pump with three lumenal domains (Gong et al., 2016). In both cases, interactions between the CRD and other domains of the protein are functionally important, and these are also linked to small molecule-binding activities.

6.1 HHIP and GAG binding

HHIP is a Hh signalling antagonist that comprises an N-terminal pocket-type CRD attached by a flexible linker to the C-terminal β-propeller domain to which HhN binds (Bishop et al., 2009). Secretion has been reported for HHIP, but regulation of this process was uncharacterised prior to this work (Coulombe et al., 2004).

GAG binding sites upon HHIP were investigated, identifying discrete interaction sites in HHIP-N and HHIP-C. This is the first example of a secreted inhibitor of Hh signalling, whilst secreted antagonists of other morphogens in other signalling systems (Wnt, BMP) are well-known (Kakugawa et al., 2015; Nolan et al., 2013). HHIP-N comprises a novel fold with a distinct N-terminal GAG-binding sub-domain linked by disulphide bonds to a classic α-helical CRD motif. Mutational data verify the basic residues that interact with GAG mimic SOS in the HHIP-N:SOS complex structure as functionally relevant. This represents a novel function for a CRD family member.

GAG-binding motifs are also observed on two distinct regions of the HHIP-C surface. Additionally, proteolysis of the domain linker between HHIP-N and HHIP-C may represent a mechanism to control the spread of HHIP molecules through reduction of HHIP cell surface binding affinity. HHIP undergoes oligomerisation through both protein-protein and GAG-mediated interactions. Sites for HHIP oligomerisation are found within both HHIP-N and HHIP-C, as well as a proposed leucine zipper-like dimerisation motif in the C-terminal helix.

To summarise, HHIP is a multi-domain protein with a characterised function in HhN ligand binding. Regulation of HHIP assemblies by oligomerisation through the use of domain:domain contacts as well as cooperative GAG binding is likely to represent strategies employed by HHIP to generate temporally and spatially correct gradients of active Hh signalling. This also suggests that HHIP acts as a 'counterpole morphogen,' forming gradients of a secreted signalling antagonist.

6.2 ROR2

The structure of the ROR2 CRD and the Kringle domain (Kr) is the first high resolution crystal structure of a CRD with a contiguous globular domain. Additionally, CRDs are positioned at the very N-terminus of almost all proteins in which they are found, and much of their structural characterisation to this point has focused on such examples. MuSK is comparable to ROR2 in that the CRD is not positioned at the N-terminus but is preceded by three Ig-domains and is located adjacent to the plasma membrane. Therefore, the analysis of ROR2 in this study concerns the characterisation of the first CRD in an intra-domain position within a receptor ECD, and biophysical studies of all three domains of ROR2 were performed within this study.

The ROR2 CRD comprises the standard α-helical fold stabilised by 5 disulphide bonds, comprising a similar arrangement to other Fz-type CRDs such as MusK, Fz8 and Smo. Unlike apo-structures of Fz8, Fz4, Smo and sFRP3, no

hydrophobic groove is visible on the surface of the ROR2 CRD. The most structurally similar Fz-type CRD to ROR2, MuSK, also lacks such a lipid-binding groove (Cα RMSD of 1.1 over 111 residues). This pre-formed groove is present in apo-structures in order to prepare the Fz-type CRDs for a ligand binding event. For ROR2 to directly interact with a Wnt ligand in the same way as in the Wnt8:Fz8 complex structure, a conformational change would be required. Such an event has not yet observed for a groove-type CRD; the Fz8 CRD does not undergo structural reorganisation to bind the Wnt ligand (Dann et al., 2001; Janda et al., 2012), and the Smo CRD does not undergo structural reorganisation upon cholesterol recognition (Byrne et al., 2016; Luchetti et al., 2016; Nachtergaele et al., 2013). Therefore, there is a strong possibility that the ROR2 CRD ether does not represent a Wnt-binding domain, or that ROR2 binds to an alternative and uncharacterised site on the Wnt ligand to the one observed in the Fz8:Wnt8 complex structure. Therefore, the ROR2 CRD structure gives novel insights into the role of co-receptors in Wnt signalling.

The role of the ROR2 Kr domain in the three-dimensional arrangement of the ECD is similar to that of other Kr domain-containing proteins. The Kr domain forms a domain:domain contact interface with a high degree of shape complementarity (Sc = 0.68). This may be a driving force in this interaction, as the remainder of the interface is predominantly constituted by weak Van der Waals interactions and only a single hydrogen bond. Secretion of ROR2 constructs is severely reduced when the Kr domain is missing, suggesting that this domain:domain interaction plays a role in stabilising the CRD and that this arrangement is present in solution. This may

additionally play a role in signalling, forming an interaction interface. It is also possible that the positioning of the Kr domain on the CRD represents a transient interaction, given the small size of the interface and the weakness of the interactions. The domain-domain linker also displays the highest B-factors for the CRD-Kr structure. Flexibility at this site may reflect movement of the Kr domain relative to the CRD. The Kr domain may not always be bound to the CRD, and its role may be more linked to protein folding in the endoplasmic reticulum rather than to constant stabilisation of the CRD.

Furthermore, the Ig domain at the N-terminus of ROR2 places the receptor ECD in an extended conformation, and is absent in fly ROR genes. This suggests that the ROR2 Ig domain may form interactions with an unknown protein in other species; the third Ig domain of related receptor MuSK is dimerised by BMP ligands, and this may reflect the true role of ROR2 as an RTK. The expression profile of fly ROR2 is limited to the nervous system (Oishi, 1997), while vertebrate ROR2 expression is more widespread and is also observed within the developing limbs and skeleton (Al-Shawi et al., 2001; Matsuda et al., 2001; Stricker et al., 2006). The Ig domain in vertebrate ROR2 may therefore bind some ligand involved in vertebrate limb development which is not required for development of the fly wing or leg. This would require further study of the Ig domain however, as the *C. elegans* ROR homologue CAM-1 contains all three domains (Green et al., 2007). Therefore, a role for the ROR2 Ig domain is purely speculative without identification of a binding partner.

In conclusion, crystallographic studies have identified a novel contact between a CRD and a contiguous domain in ROR2. Evidence presented here suggests that any ROR2:Wnt interaction is not compatible with the method of FZD:Wnt ligand interaction (Janda et al., 2012). Much evidence has linked the ROR2 CRD to Wnt signalling, and therefore it is possible that ROR2 is accommodated into a ternary complex with FZD and Wnt in place of LRP5/6. The presence of other domains in ROR2 may be linked to the recognition of ligands other than Wnt. A helical insertion in the ROR2 CRD requires a conformational change in the structure in order to bind Wnt ligands in the same way as Fz8 (Janda et al., 2012).

6.3 The CRD in ligand binding

The first disordered loop (DL1) from HHIP-N represents a motif which occupies the same location in space as ligand-binding elements in other CRDs (pocket- and groove-type), as well as being aligned based on sequence. The second disordered loop (DL2) seems to be less well conserved on sequence but which may also be found within additional ligand-binding elements in other pocket-type CRDs.

The relationship between Hh signalling and small molecules has been discussed. Recent evidence supports a model in which cholesterol is an activator of Hh signalling, which is the result of cholesterol binding within a groove of the Smo CRD (Byrne et al., 2016; Huang et al., 2016). This contacts elements of the GPCR-like 7TM region and mutations to residues involved in this reduces Hh signalling

activation (Byrne et al., 2016). Comparisons with all pocket-type CRDs places DL1 of HHIP-N in the same structural location as ligand- or receptor-binding elements. The same can be said of palmitoleate- and cholesterol-binding residues in Fz8 and Smo. One explanation for a role of HHIP-N DL1 and DL2 may be in the binding of the palmitoyl moiety on attached to HhN, like the proposed role of the hydrophobic groove seen at the surface of sFRP3. A cellular reporter assay using a HHIP-N CRD-Fc fusion and conditioned media containing lipid-modified HhN did not report a direct interaction between HHIP-N and HhN (Bosanac et al., 2009). Direct binding assays with lipid-modified HhN are required to verify this finding. Another possible role of DL1 and DL2 is in the binding of cholesterol or some other Smo agonist molecule. Preliminary results suggest that HHIP-N displays binding to cholesterol-coupled beads (section 3.4.4), which requires further experiments for verification. Nevertheless, this may represent a true function for the HHIP-N CRD region as a sterol depletion would enable HHIP to further antagonise Hh signalling beyond preventing HhN:Ptc interaction. The membranes of cilia, to which Smo is recruited upon activation of Hh signalling, are sterol-enriched relative to other cellular membranes (Emmer et al., 2010; Rohatgi et al., 2007). This has been suggested to be linked to lipid rafts, to which the trafficking of particular membrane complexes has been linked. Ptc is observed to localise to the ciliary base in order to prevent entry of Smo to the cilium (Rohatgi et al., 2007). Therefore, this proposed function of HHIP-N may act as a backup to prevent aberrant signal activation from Smo-activating molecules following HhN removal.

Structural superposition of the ROR2 CRD to the HHIP-N CRD region (Cα RSMD = 2.3 for 54 Cα positions) displays alignment of only the α1 helix. In other CRDs, a stretch of sequence adjacent to the α1 helix of HHIP-N makes contacts with a small molecule ligand (as observed in the FRα/β, NPC1, RDBP, Fz8 and Smo ligand-bound structures; see section 3.3.5), and is likely to occupy the same location as DL1 of HHIP-N. Such a secondary structural element is missing from the ROR2 CRD (Fig. 6.1A). Sequence alignment of the HHIP-N and ROR2 CRDs also indicates that this functional sequence motif is lacking in ROR2 (Fig. 6.2B). This may represent further evidence for an absence of a direct CRD:Wnt interaction in ROR2.

In summary, the true roles for the HHIP-N and ROR2 CRDs require further experiments. Disordered loops DL1 and DL2 of HHIP-N represent sequence elements aligned with ligand-binding elements of other CRDs and their absence in ROR2 may reinforce the idea that it does not bind a palmitoylated Wnt ligand. This is further supported by the absence of a pre-formed hydrophobic groove on the ROR2 CRD surface.

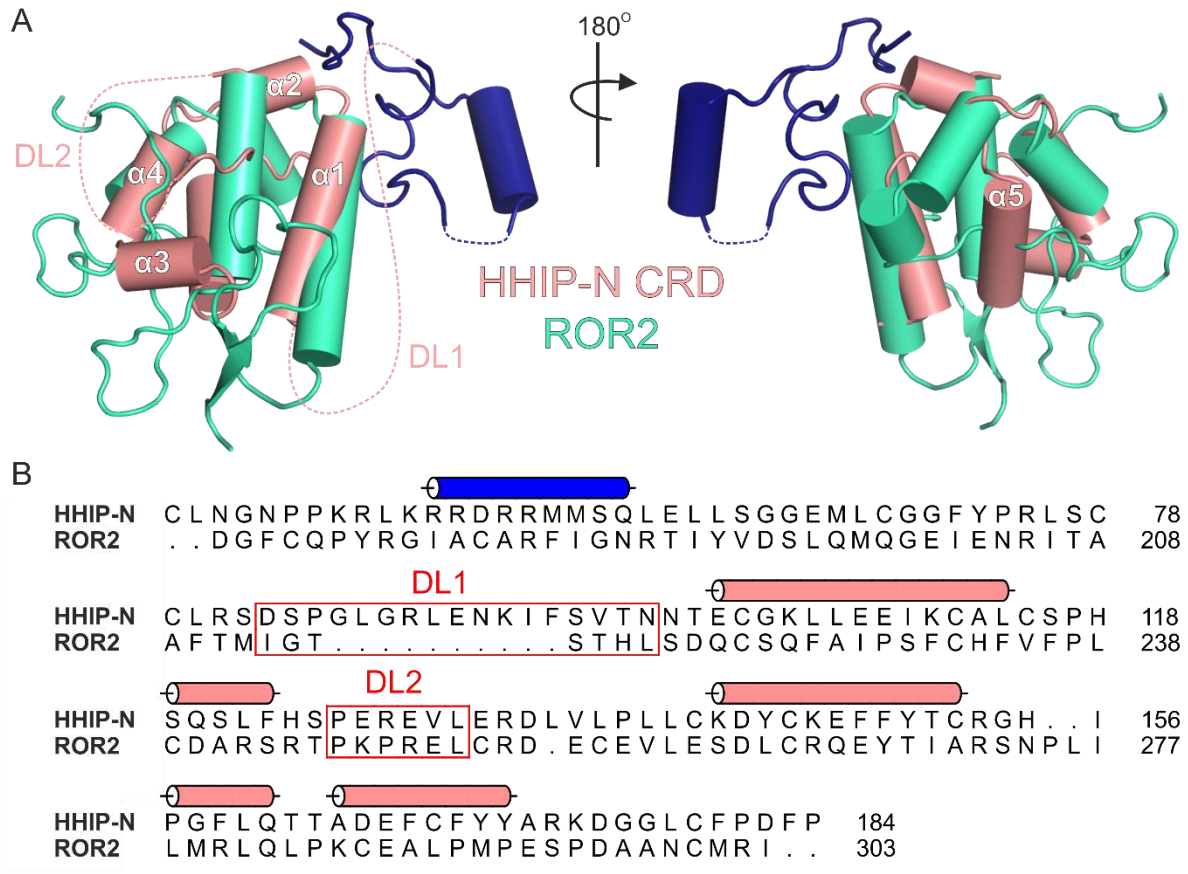


Figure 6.1 – Comparison of HHIP-N and ROR2 CRDs. A) Structural superposition, with CRDs coloured as annotated in the figure. The HHIP-N GAG-binding sub-domain is coloured in blue and DL1 and DL2 are shown as dashed lines. Helices α1-5 are labelled for ROR2; B) Sequence alignment, with DL1 and DL2 of HHIP-N outlined in boxes. The secondary structure of HHIP-N is displayed above the sequences.

6.4 Concluding remarks

A summary of structural and biophysical characterisation of proteins studied in this thesis is displayed in Fig. 6.2. HHIP is a multi-domain protein, with surface clusters of residues involved in GAG binding. Cleavage of a linker between these domains may represent a mechanism to uncouple HHIP-N and HHIP-C, and may represent a mechanism to limit the spatial regulation of Hh signalling. HHIP-C oligomer size is regulated through interactions with the GAG layer, and how HHIP-

N modulates these assemblies is unknown. Therefore, the pocket-type CRD of HHIP-N may act to regulate secretion of HHIP and consequently the range of direct HhN antagonism by HHIP-C by reducing affinity for the cell surface. Another role may be the release of HHIP-N in the vicinity of cells in which the bulk of Hh signalling activity has been inhibited via the removal of HhN by HHIP-C, and fine-tuning of the pathway by the removal of particular sterols by the CRD is required (Fig. 6.2A).

Fz-, or groove-type, CRDs display roles in small molecule binding, protein-protein interaction or a combination of the two (Fig 6.2B). The ROR2 CRD has a Fz-type fold comparable to Fz8, which forms a structural unit with the C-terminal Kr domain. A helical insertion (α5) precludes the existence of a hydrophobic groove as seen in related Fz-type CRDS. How the Ig domain of the ROR2 ECD is related to its function is unknown, but may potentially interact with another protein in order to integrate ROR2 into different signal transduction pathways (Fig. 6.2C).

In summary, receptors involved in the Hh and Wnt pathways have been characterised in this study. Whilst largely unrelated, both signalling pathways entail lipid-modified ligands and rely on proteins with CRDs for activation and regulation. X-ray crystallographic data was combined with other biophysical techniques in order to investigate modes of regulation by the proteins in question as thoroughly as possible. Taken together, these results give new insights into the extracellular regulation of signal transduction processes essential in organismal development and linked to a number of disease states.

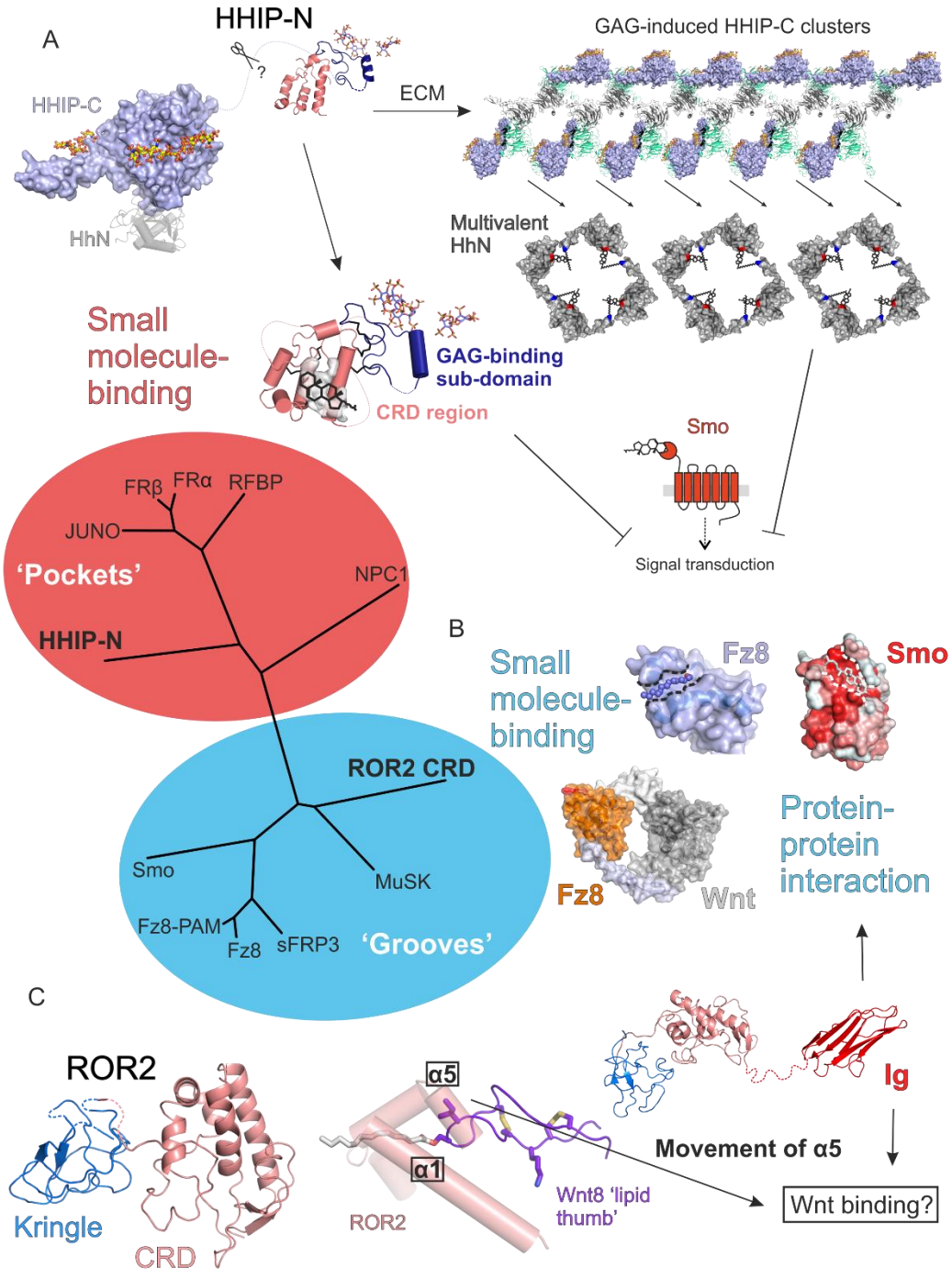


Figure 6.2 – Summary of thesis results. A) HHIP is a multi-domain protein, regulated by oligomerisation. Pocket-type CRDs such as HHIP-N are involved in small molecule binding, and this may add an extra element to control of Hh signalling by HHIP. Proteolysis may also influence these processes; B) Fz-type CRDs have roles in the binding of small molecules and to other proteins; C) The α-helical ROR2 CRD forms a structural unit with the Kr domain, and is unlikely to represent a Wnt-binding interface. The ROR2 ECD possesses an elongated conformation and potentially interacts with currently unknown ligands.

Bibliography

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Agirre, J., Davies, G., Wilson, K., and Cowtan, K. (2015). Carbohydrate anomalies in the PDB. Nat. Chem. Biol. *11*, 303–303.

Aglan, M., Amr, K., Ismail, S., Ashour, A., Otaify, G.A., Mehrez, M.A.I., Aboul-Ezz, E.H.A., El-Ruby, M., Mazen, I., Abdel-Hamid, M.S., et al. (2015). Clinical and molecular characterization of seven Egyptian families with autosomal recessive robinow syndrome: Identification of four novel *ROR2* gene mutations. Am. J. Med. Genet. A. *167*, 3054–3061.

Akey, D.L., Brown, W.C., Konwerski, J.R., Ogata, C.M., and Smith, J.L. (2014). Use of massively multiple merged data for low-resolution S-SAD phasing and refinement of flavivirus NS1. Acta Crystallogr. D Biol. Crystallogr. *70*, 2719–2729.

Allen, B.L., Tenzen, T., and McMahon, A.P. (2007). The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development. Genes Dev. *21*, 1244–1257.

Allen, B.L., Song, J.Y., Izzi, L., Althaus, I.W., Kang, J.-S., Charron, F., Krauss, R.S., and McMahon, A.P. (2011). Overlapping roles and collective requirement for the coreceptors GAS1, CDO, and BOC in SHH pathway function. Dev. Cell *20*, 775–787.

Almén, M., Nordström, K.J., Fredriksson, R., and Schiöth, H.B. (2009). Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biol. *7*, 50.

Al-Shawi, R., Ashton, S.V., Underwood, C., and Simons, J.P. (2001). Expression of the Ror1 and Ror2 receptor tyrosine kinase genes during mouse development. Dev. Genes Evol. *211*, 161–171.

Amit, S. (2002). Axin-mediated CKI phosphorylation of beta -catenin at Ser 45: a molecular switch for the Wnt pathway. Genes Dev. *16*, 1066–1076.

Angers, S., Thorpe, C.J., Biechele, T.L., Goldenberg, S.J., Zheng, N., MacCoss, M.J., and Moon, R.T. (2006). The KLHL12–Cullin-3 ubiquitin ligase negatively regulates the Wnt–β-catenin pathway by targeting Dishevelled for degradation. Nat. Cell Biol. *8*, 348–357.

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr. D Biol. Crystallogr. *62*, 1243–1250.

Ashe, H.L. (2006). The interpretation of morphogen gradients. Development *133*, 385–394.

Avaron, F., Smith, A., and Akimenko, M.-A. (2007). Sonic Hedgehog Signalling in the Developing and Regenerating Fins of Zebrafish. Shh and Gli Signalling and Development, (New York, NY: Springer New York), pp. 93–106.

Aydin, H., Sultana, A., Li, S., Thavalingam, A., and Lee, J.E. (2016). Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex. Nature *534*, 562-565.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. *98*, 10037–10041.

Banerjee, S., Gordon, L., Donn, T.M., Berti, C., Moens, C.B., Burden, S.J., and Granato, M. (2011). A novel role for MuSK and non-canonical Wnt signaling during segmental neural crest cell migration. Development *138*, 3287–3296.

Bänziger, C., Soldini, D., Schütt, C., Zipperlen, P., Hausmann, G., and Basler, K. (2006). Wntless, a Conserved Membrane Protein Dedicated to the Secretion of Wnt Proteins from Signaling Cells. Cell *125*, 509–522.

Bazan, J.F., and de Sauvage, F.J. (2009). Structural ties between cholesterol transport and morphogen signaling. Cell *138*, 1055–1056.

Beachy, P.A., Hymowitz, S.G., Lazarus, R.A., Leahy, D.J., and Siebold, C. (2010). Interactions between Hedgehog proteins and their binding partners come into view. Genes Dev. *24*, 2001–2012.

Bell, C.H., Healey, E., van Erp, S., Bishop, B., Tang, C., Gilbert, R.J.C., Aricescu, A.R., Pasterkamp, R.J., and Siebold, C. (2013). Structure of the Repulsive Guidance Molecule (RGM)-Neogenin Signaling Hub. Science *341*, 77–80.

Bianchi, E., Doe, B., Goulding, D., and Wright, G.J. (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. Nature *508*, 483–487.

Bibby, J., Keegan, R.M., Mayans, O., Winn, M.D., and Rigden, D.J. (2012). *AMPLE*: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed *ab initio* models. Acta Crystallogr. D Biol. Crystallogr. *68*, 1622–1631.

Billiard, J., Way, D.S., Seestaller-Wehr, L.M., Moran, R.A., Mangine, A., and Bodine, P.V.N. (2005). The Orphan Receptor Tyrosine Kinase Ror2 Modulates Canonical Wnt Signaling in Osteoblastic Cells. Mol. Endocrinol. *19*, 90–101.

Bishop, B., Aricescu, A.R., Harlos, K., O'Callaghan, C.A., Jones, E.Y., and Siebold, C. (2009). Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. Nat. Struct. Mol. Biol. *16*, 698–703.

Blassberg, R., Macrae, J.I., Briscoe, J., and Jacob, J. (2016). Reduced cholesterol levels impair Smoothened activation in Smith–Lemli–Opitz syndrome. Hum. Mol. Genet. *25*, 693–705.

Bond, C.S., and Schüttelkopf, A.W. (2009). *ALINE*: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr. D Biol. Crystallogr. *65*, 510–512.

Borcherding, N., Kusner, D., Liu, G.-H., and Zhang, W. (2014). ROR1, an embryonic protein with an emerging role in cancer biology. Protein Cell *5*, 496–502.

Bosanac, I., Maun, H.R., Scales, S.J., Wen, X., Lingel, A., Bazan, J.F., de Sauvage, F.J., Hymowitz, S.G., and Lazarus, R.A. (2009). The structure of SHH in complex with HHIP reveals a recognition role for the Shh pseudo active site in signaling. Nat. Struct. Mol. Biol. *16*, 691–697.

Bourhis, E., Tam, C., Franke, Y., Bazan, J.F., Ernst, J., Hwang, J., Costa, M., Cochran, A.G., and Hannoush, R.N. (2010). Reconstitution of a Frizzled8-Wnt3a-LRP6 Signaling Complex Reveals Multiple Wnt and Dkk1 Binding Sites on LRP6. J. Biol. Chem. *285*, 9172–9179.

Bourikas, D., Pekarik, V., Baeriswyl, T., Grunditz, Å., Sadhu, R., Nardó, M., and Stoeckli, E.T. (2005). Sonic hedgehog guides commissural axons along the longitudinal axis of the spinal cord. Nat. Neurosci. *8*, 297–304.

Brinkmann, E.-M., Mattes, B., Kumar, R., Hagemann, A.I.H., Gradl, D., Scholpp, S., Steinbeisser, H., Kaufmann, L.T., and Özbek, S. (2016). Secreted Frizzled-related Protein 2 (sFRP2) Redirects Non-canonical Wnt Signaling from Fz7 to Ror2 during Vertebrate Gastrulation. J. Biol. Chem. *291*, 13730–13742.

Briscoe, J., and Small, S. (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. Development *142*, 3996–4009.

Briscoe, J., and Thérond, P.P. (2013). The mechanisms of Hedgehog signalling and its roles in development and disease. Nat. Rev. Mol. Cell Biol. *14*, 418–431.

Brown, P.H., and Schuck, P. (2006). Macromolecular Size-and-Shape Distributions by Sedimentation Velocity Analytical Ultracentrifugation. Biophys. J. *90*, 4651–4661.

Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Bryja, V., Andersson, E.R., Schambony, A., Esner, M., Bryjova, L., Biris, K.K., Hall, A.C., Kraft, B., Cajanek, L., Yamaguchi, T.P., et al. (2008). The Extracellular Domain of Lrp5/6 Inhibits Noncanonical Wnt Signaling In Vivo. Mol. Biol. Cell *20*, 924–936.

Byrne, E.F.X., Sircar, R., Miller, P.S., Hedger, G., Luchetti, G., Nachtergaele, S., Tully, M.D., Mydock-McGrane, L., Covey, D.F., Rambo, R.P., et al. (2016). Structural basis of Smoothened regulation by its extracellular domains. Nature *535*, 517-522.

Callejo, A., Torroja, C., Quijada, L., and Guerrero, I. (2006). Hedgehog lipid modifications are required for Hedgehog stabilization in the extracellular matrix. Dev. Camb. Engl. *133*, 471–483.

Capurro, M.I., Xiang, Y.-Y., Lobe, C., and Filmus, J. (2005). Glypican-3 promotes the growth of hepatocellular carcinoma by stimulating canonical Wnt signaling. Cancer Res. *65*, 6245–6254.

Capurro, M.I., Xu, P., Shi, W., Li, F., Jia, A., and Filmus, J. (2008). Glypican-3 Inhibits Hedgehog Signaling during Development by Competing with Patched for Hedgehog Binding. Dev. Cell *14*, 700–711.

Carlsson, P., and Kjellén, L. (2012). Heparin biosynthesis. Handb. Exp. Pharmacol. *207*, 23–41.

Carter, W.J., Cama, E., and Huntington, J.A. (2005). Crystal Structure of Thrombin Bound to Heparin. J. Biol. Chem. *280*, 2745–2749.

Carugo, O., and Carugo, K.D. (2005). When X-rays modify the protein structure: radiation damage at work. Trends Biochem. Sci. *30*, 213–219.

Cavallo, R.A., Cox, R.T., Moline, M.M., Roose, J., Polevoy, G.A., Clevers, H., Peifer, M., and Bejsovec, A. (1998). Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature *395*, 604–608.

Chamoun, Z., Mann, R.K., Nellen, D., von Kessler, D.P., Bellotto, M., Beachy, P.A., and Basler, K. (2001). Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. Science *293*, 2080–2084.

Chang, T.-H., Hsieh, F.-L., Zebisch, M., Harlos, K., Elegheert, J., and Jones, E.Y. (2015). Structure and functional properties of Norrin mimic Wnt for signalling with Frizzled4, Lrp5/6, and proteoglycan. eLife *4*.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I., et al. (2007). Glycoprotein structural genomics: solving the glycosylation problem. Struct. Lond. Engl. 1993 *15*, 267–273.

Chazotte, B. (2011). Labeling nuclear DNA with hoechst 33342. Cold Spring Harb. Protoc. *2011*, pdb.prot5557.

Chen, C., Ke, J., Zhou, X.E., Yi, W., Brunzelle, J.S., Li, J., Yong, E.-L., Xu, H.E., and Melcher, K. (2013). Structural basis for molecular recognition of folic acid by folate receptors. Nature *500*, 486–489.

Chen, J.K., Taipale, J., Cooper, M.K., and Beachy, P.A. (2002). Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. Genes Dev. *16*, 2743–2748.

Chen, M.-H., Li, Y.-J., Kawakami, T., Xu, S.-M., and Chuang, P.-T. (2004). Palmitoylation is required for the production of a soluble multimeric Hedgehog protein complex and long-range signaling in vertebrates. Genes Dev. *18*, 641–659.

Chen, S., Bubeck, D., MacDonald, B.T., Liang, W.-X., Mao, J.-H., Malinauskas, T., Llorca, O., Aricescu, A.R., Siebold, C., He, X., et al. (2011a). Structural and functional studies of LRP6 ectodomain reveal a platform for Wnt signaling. Dev. Cell *21*, 848–861.

Chen, X., Tukachinsky, H., Huang, C.-H., Jao, C., Chu, Y.-R., Tang, H.-Y., Mueller, B., Schulman, S., Rapoport, T.A., and Salic, A. (2011b). Processing and turnover of the Hedgehog protein in the endoplasmic reticulum. J. Cell Biol. *192*, 825–838.

Chuang, P.-T. (2003). Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. Genes Dev. *17*, 342–347.

Chuang, P.T., and McMahon, A.P. (1999). Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. Nature *397*, 617–621.

Ciani, L., and Salinas, P.C. (2005). Signalling in neural development: WNTS in the vertebrate nervous system: from patterning to neuronal connectivity. Nat. Rev. Neurosci. *6*, 351–362.

Cole, J.L., Lary, J.W., P. Moody, T., and Laue, T.M. (2008). Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. In Methods in Cell Biology, (Elsevier), pp. 143–179.

Coles, C.H., Shen, Y., Tenney, A.P., Siebold, C., Sutton, G.C., Lu, W., Gallagher, J.T., Jones, E.Y., Flanagan, J.G., and Aricescu, A.R. (2011). Proteoglycan-specific molecular switch for RPTPσ clustering and neuronal extension. Science *332*, 484–488.

Coles, C.H., Mitakidis, N., Zhang, P., Elegheert, J., Lu, W., Stoker, A.W., Nakagawa, T., Craig, A.M., Jones, E.Y., and Aricescu, A.R. (2014). Structural basis for extracellular cis and trans RPTPσ signal competition in synaptogenesis. Nat. Commun. *5*, 5209.

Cooper, M.K., Wassif, C.A., Krakowiak, P.A., Taipale, J., Gong, R., Kelley, R.I., Porter, F.D., and Beachy, P.A. (2003). A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. Nat. Genet. *33*, 508–513.

Copp, A.J., and Greene, N.D. (2009). Genetics and development of neural tube defects. J. Pathol. *220*, 217-230.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. *16*, 10881–10890.

Coulombe, J., Traiffort, E., Loulier, K., Faure, H., and Ruat, M. (2004). Hedgehog interacting protein in the mature brain: membrane-associated and soluble forms. Mol. Cell. Neurosci. *25*, 323–333.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011.

Cowtan, K. (2010). Recent developments in classical density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 470–478.

Dann, C.E., Hsieh, J.C., Rattner, A., Sharma, D., Nathans, J., and Leahy, D.J. (2001). Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. Nature *412*, 86–90.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, W375–W383.

Debreczeni, J.E., Bunkóczi, G., Girmann, B., and Sheldrick, G.M. (2003). In-house phase determination of the lima bean trypsin inhibitor: a low-resolution sulfur-SAD case. Acta Crystallogr. D Biol. Crystallogr. *59*, 393–395.

Deguchi, H., Takeya, H., Gabazza, E.C., Nishioka, J., and Suzuki, K. (1997). Prothrombin kringle 1 domain interacts with factor Va during the assembly of prothrombinase complex. Biochem. J. *321 (Pt 3)*, 729–735.

Dejima, K., Kanai, M.I., Akiyama, T., Levings, D.C., and Nakato, H. (2011). Novel contact-dependent bone morphogenetic protein (BMP) signaling mediated by heparan sulfate proteoglycans. J. Biol. Chem. *286*, 17103–17111.

Dhoot, G.K., Gustafsson, M.K., Ai, X., Sun, W., Standiford, D.M., and Emerson, C.P. (2001). Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase. Science *293*, 1663–1666.

Dijksterhuis, J.P., Petersen, J., and Schulte, G. (2014). WNT/Frizzled signalling: receptor-ligand selectivity with focus on FZD-G protein signalling and its physiological relevance: IUPHAR Review 3. Br. J. Pharmacol. *171*, 1195–1209.

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., et al. (2007). In situ proteolysis for protein crystallization and structure determination. Nat. Methods *4*, 1019–1021.

Dorn, K.V., Hughes, C.E., and Rohatgi, R. (2012). A Smoothened-Evc2 Complex Transduces the Hedgehog Signal at Primary Cilia. Dev. Cell *23*, 823–835.

Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. Transgenic Res. *14*, 477–482.

Eaton, S. (2008). Multiple roles for lipids in the Hedgehog signalling pathway. Nat. Rev. Mol. Cell Biol. *9*, 437–445.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., and McMahon, A.P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417–1430.

El Omari, K., Iourin, O., Kadlec, J., Fearn, R., Hall, D.R., Harlos, K., Grimes, J.M., and Stuart, D.I. (2014). Pushing the limits of sulfur SAD phasing: de novo structure solution of the N-terminal domain of the ectodomain of HCV E1. Acta Crystallogr. D Biol. Crystallogr. *70*, 2197–2203.

Emmer, B.T., Maric, D., and Engman, D.M. (2010). Molecular mechanisms of protein and lipid targeting to ciliary membranes. J. Cell Sci. *123*, 529–536.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Endoh-Yamagami, S., Evangelista, M., Wilson, D., Wen, X., Theunissen, J.-W., Phamluong, K., Davis, M., Scales, S.J., Solloway, M.J., de Sauvage, F.J., et al. (2009). The Mammalian Cos2 Homolog Kif7 Plays an Essential Role in Modulating Hh Signal Transduction during Development. Curr. Biol. *19*, 1320–1326.

Evangelista, M., Tian, H., and de Sauvage, F.J. (2006). The Hedgehog Signaling Pathway in Cancer. Clin. Cancer Res. *12*, 5924–5928.

Evans, P., and McCoy, A. (2008). An introduction to molecular replacement. Acta Crystallogr. D Biol. Crystallogr. *64*, 1–10.

Fanto, M., Weber, U., Strutt, D.I., and Mlodzik, M. (2000). Nuclear signaling by Rac and Rho GTPases is required in the establishment of epithelial planar polarity in the Drosophila eye. Curr. Biol. *10*, 979–988.

Filmus, J., Shi, W., Wong, Z.M., and Wong, M.J. (1995). Identification of a new membrane-bound heparan sulphate proteoglycan. Biochem. J. *311 (Pt 2)*, 561–565.

Foadi, J., Aller, P., Alguel, Y., Cameron, A., Axford, D., Owen, R.L., Armour, W., Waterman, D.G., Iwata, S., and Evans, G. (2013). Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *69*, 1617–1632.

Frantz, C., Stewart, K.M., and Weaver, V.M. (2010). The extracellular matrix at a glance. J. Cell Sci. *123*, 4195–4200.

Fry, E.E. (1999). The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex. EMBO J. *18*, 543–554.

Fuerer, C., Habib, S.J., and Nusse, R. (2009). A study on the interactions between heparan sulfate proteoglycans and Wnt proteins. Dev. Dyn. *239*, 184-190.

Gage, F.H., and Temple, S. (2013). Neural Stem Cells: Generating and Regenerating the Brain. Neuron *80*, 588–601.

Gallet, A., Rodriguez, R., Ruel, L., and Therond, P.P. (2003). Cholesterol modification of hedgehog is required for trafficking and movement, revealing an asymmetric cellular response to hedgehog. Dev. Cell *4*, 191–204.

Gao, X., and Hannoush, R.N. (2014). Single-cell imaging of Wnt palmitoylation by the acyltransferase porcupine. Nat. Chem. Biol. *10*, 61–68.

Gildea, R.J., Waterman, D.G., Parkhurst, J.M., Axford, D., Sutton, G., Stuart, D.I., Sauter, N.K., Evans, G., and Winter, G. (2014). New methods for indexing multi-lattice diffraction data. Acta Crystallogr. D Biol. Crystallogr. *70*, 2652–2666.

Gong, X., Qian, H., Zhou, X., Wu, J., Wan, T., Cao, P., Huang, W., Zhao, X., Wang, X., Wang, P., et al. (2016). Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. Cell *165*, 1467–1478.

Gonnissen, A., Isebaert, S., and Haustermans, K. (2015). Targeting the Hedgehog signaling pathway in cancer: beyond Smoothened. Oncotarget *6*, 13899–13913.

Gould, S.E., Upholt, W.B., and Kosher, R.A. (1992). Syndecan 3: a member of the syndecan family of membrane-intercalated proteoglycans that is expressed in high amounts at the onset of chicken limb cartilage differentiation. Proc. Natl. Acad. Sci. U. S. A. *89*, 3271–3275.

Green, J., Nusse, R., and van Amerongen, R. (2014). The role of Ryk and Ror receptor tyrosine kinases in Wnt signal transduction. Cold Spring Harb. Perspect. Biol. *6*.

Green, J.L., Inoue, T., and Sternberg, P.W. (2007). The C. elegans ROR receptor tyrosine kinase, CAM-1, non-autonomously inhibits the Wnt pathway. Development *134*, 4053–4062.

Grumolato, L., Liu, G., Mong, P., Mudbhary, R., Biswas, R., Arroyave, R., Vijayakumar, S., Economides, A.N., and Aaronson, S.A. (2010). Canonical and noncanonical Wnts use a common mechanism to activate completely unrelated coreceptors. Genes Dev. *24*, 2517–2530.

Guttman, M., Weinkam, P., Sali, A., and Lee, K.K. (2013). All-Atom Ensemble Modeling to Analyze Small-Angle X-Ray Scattering of Glycosylated Proteins. Structure *21*, 321–331.

Han, L., Nishimura, K., Sadat Al Hosseini, H., Bianchi, E., Wright, G.J., and Jovine, L. (2016). Divergent evolution of vitamin B9 binding underlies Juno-mediated adhesion of mammalian gametes. Curr. Biol. *26*, R100-101.

Hausmann, G., von Mering, C., and Basler, K. (2009). The Hedgehog Signaling Pathway: Where Did It Come From? PLoS Biol. *7*, e1000146.

Heckman, K.L., and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. Nat. Protoc. *2*, 924–932.

Hedman, K., Kurkinen, M., Alitalo, K., Vaheri, A., Johansson, S., and Höök, M. (1979). Isolation of the pericellular matrix of human fibroblast cultures. J. Cell Biol. *81*, 83–91.

Hendrickson, W.A. (2014). Anomalous diffraction in crystallographic phase evaluation. Q. Rev. Biophys. *47*, 49–93.

Hikasa, H., Shibata, M., Hiratani, I., and Taira, M. (2002). The Xenopus receptor tyrosine kinase Xror2 modulates morphogenetic movements of the axial mesoderm and neuroectoderm via Wnt signaling. Dev. Camb. Engl. *129*, 5227–5239.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Holtz, A.M., Griffiths, S.C., Davis, S.J., Bishop, B., Siebold, C., and Allen, B.L. (2015). Secreted HHIP1 interacts with heparan sulfate and regulates Hedgehog ligand localization and function. J. Cell Biol. *209*, 739–758.

Howarth, M., and Ting, A.Y. (2008). Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. Nat. Protoc. *3*, 534–545.

Huang, P., Nedelcu, D., Watanabe, M., Jao, C., Kim, Y., Liu, J., and Salic, A. (2016). Cellular Cholesterol Directly Activates Smoothened in Hedgehog Signaling. Cell *166*, 1176–1187.

Humke, E.W., Dorn, K.V., Milenkovic, L., Scott, M.P., and Rohatgi, R. (2010). The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. Genes Dev. *24*, 670–682.

Inestrosa, N.C., and Varela-Nallar, L. (2014). Wnt signaling in the nervous system and in Alzheimer's disease. J. Mol. Cell Biol. *6*, 64–74.

Iozzo, R.V., and Schaefer, L. (2015). Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. Matrix Biol. *42*, 11–55.

Jackman, A. (2004). Antifolates targeted specifically to the folate receptor. Adv. Drug Deliv. Rev. *56*, 1111–1125.

Janda, C.Y., and Garcia, K.C. (2015). Wnt acylation and its functional implication in Wnt signalling regulation. Biochem. Soc. Trans. *43*, 211–216.

Janda, C.Y., Waghray, D., Levin, A.M., Thomas, C., and Garcia, K.C. (2012). Structural basis of Wnt recognition by Frizzled. Science *337*, 59–64.

Jia, J., Zhang, L., Zhang, Q., Tong, C., Wang, B., Hou, F., Amanai, K., and Jiang, J. (2005). Phosphorylation by Double-Time/CKIε and CKIα Targets Cubitus Interruptus for Slimb/β-TRCP-Mediated Proteolytic Processing. Dev. Cell *9*, 819–830.

Jing, L., Lefebvre, J.L., Gordon, L.R., and Granato, M. (2009). Wnt Signals Organize Synaptic Prepattern and Axon Guidance through the Zebrafish unplugged/MuSK Receptor. Neuron *61*, 721–733.

Johnson, T., Ouhtit, A., Gaur, R., Fernando, A., Schwarzenberger, P., Su, J., Ismail, M.F., El-Sayyad, H.I., Karande, A., Elmageed, Z.A., et al. (2009). Biochemical characterization of riboflavin carrier protein (RCP) in prostate cancer. Front. Biosci. Landmark Ed. *14*, 3634–3640.

Jones, L.L., Margolis, R.U., and Tuszynski, M.H. (2003). The chondroitin sulfate proteoglycans neurocan, brevican, phosphacan, and versican are differentially regulated following spinal cord injury. Exp. Neurol. *182*, 399–411.

Kabsch, W. (2010). *XDS*. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kakugawa, S., Langton, P.F., Zebisch, M., Howell, S.A., Chang, T.-H., Liu, Y., Feizi, T., Bineva, G., O'Reilly, N., Snijders, A.P., et al. (2015). Notum deacylates Wnt proteins to suppress signalling activity. Nature *519*, 187–192.

Kalus, I., Rohn, S., Puvirajesinghe, T.M., Guimond, S.E., Eyckerman-Kölln, P.J., ten Dam, G., van Kuppevelt, T.H., Turnbull, J.E., and Dierks, T. (2015). Sulf1 and Sulf2 Differentially Modulate Heparan Sulfate Proteoglycan Sulfation during Postnatal Cerebellum Development: Evidence for Neuroprotective and Neurite Outgrowth Promoting Functions. PLOS ONE *10*, e0139853.

Kanato, Y., Ono, S., Kitajima, K., and Sato, C. (2009). Complex Formation of a Brain-Derived Neurotrophic Factor and Glycosaminoglycans. Biosci. Biotechnol. Biochem. *73*, 2735–2741.

Karplus, P.A., and Diederichs, K. (2012). Linking Crystallographic Model and Data Quality. Science *336*, 1030–1033.

Karplus, P.A., and Diederichs, K. (2015). Assessing and maximizing data quality in macromolecular crystallography. Curr. Opin. Struct. Biol. *34*, 60–68.

Kawahira, H. (2003). Combined activities of hedgehog signaling inhibitors regulate pancreas development. Development *130*, 4871–4879.

Khan, S., Gor, J., Mulloy, B., and Perkins, S.J. (2010). Semi-Rigid Solution Structures of Heparin by Constrained X-ray Scattering Modelling: New Insight into Heparin–Protein Complexes. J. Mol. Biol. *395*, 504–521.

Kim, J., Kato, M., and Beachy, P.A. (2009). Gli2 trafficking links Hedgehog-dependent activation of Smoothened in the primary cilium to transcriptional activation in the nucleus. Proc. Natl. Acad. Sci. *106*, 21666–21671.

Kim, M.-S., Saunders, A.M., Hamaoka, B.Y., Beachy, P.A., and Leahy, D.J. (2011). Structure of the protein core of the glypican Dally-like and localization of a region important for hedgehog signaling. Proc. Natl. Acad. Sci. U. S. A. *108*, 13112–13117.

Kim, N., Stiegler, A.L., Cameron, T.O., Hallock, P.T., Gomez, A.M., Huang, J.H., Hubbard, S.R., Dustin, M.L., and Burden, S.J. (2008). Lrp4 Is a Receptor for Agrin and Forms a Complex with MuSK. Cell *135*, 334–342.

Kleywegt, G.J. (2000). Validation of protein crystal structures. Acta Crystallogr. D Biol. Crystallogr. *56*, 249–265.

Kleywegt, G.J., and Read, R.J. (1997). Not your average density. Structure *5*, 1557–1569.

Kmetko, J., Warkentin, M., Englich, U., and Thorne, R.E. (2011). Can radiation damage to protein crystals be reduced using small-molecule compounds? Acta Crystallogr. D Biol. Crystallogr. *67*, 881–893.

Kreuger, J., and Kjellen, L. (2012). Heparan Sulfate Biosynthesis: Regulation and Variability. J. Histochem. Cytochem. *60*, 898–907.

Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. J. Mol. Biol. *372*, 774–797.

Krylov, D., and Vinson, C.R. (2001). Leucine Zipper. eLS. .

Kuhl, M. (2000). Ca2+/Calmodulin-dependent Protein Kinase II Is Stimulated by Wnt and Frizzled Homologs and Promotes Ventral Cell Fates in Xenopus. J. Biol. Chem. *275*, 12701–12711.

Kuwabara, P.E., and Labouesse, M. (2002). The sterol-sensing domain: multiple families, a unique role? Trends Genet. *18*, 193–201.

Kwon, H.J., Abi-Mosleh, L., Wang, M.L., Deisenhofer, J., Goldstein, J.L., Brown, M.S., and Infante, R.E. (2009). Structure of N-terminal domain of NPC1 reveals distinct subdomains for binding and transfer of cholesterol. Cell *137*, 1213–1224.

Kwong, L., Bijlsma, M.F., and Roelink, H. (2014). Shh-mediated degradation of Hhip allows cell autonomous and non-cell autonomous Shh signalling. Nat. Commun. *5*, 4849.

Laskowski, R.A. (2001). PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res. *29*, 221–222.

Laskowski, R.A., Watson, J.D., and Thornton, J.M. (2005). ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res. *33*, W89–W93.

Laughlin, J.D., Ha, T.S., Jones, D.N.M., and Smith, D.P. (2008). Activation of Pheromone-Sensitive Neurons Is Mediated by Conformational Activation of Pheromone-Binding Protein. Cell *133*, 1255–1265.

Law, R.H., Abu-Ssaydeh, D., and Whisstock, J.C. (2013). New insights into the structure and function of the plasminogen/plasmin system. Curr. Opin. Struct. Biol. *23*, 836–841.

Lawrence, M.C., and Colman, P.M. (1993). Shape Complementarity at Protein/Protein Interfaces. J. Mol. Biol. *234*, 946–950.

Lee, H.-J., Bao, J., Miller, A., Zhang, C., Wu, J., Baday, Y.C., Guibao, C., Li, L., Wu, D., and Zheng, J.J. (2015). Structure-based Discovery of Novel Small Molecule Wnt Signaling Inhibitors by Targeting the Cysteine-rich Domain of Frizzled. J. Biol. Chem. *290*, 30596–30606.

Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. Cell *141*, 1117–1134.

Li, W., and Huntington, J.A. (2012). Crystal structures of protease nexin-1 in complex with heparin and thrombin suggest a 2-step recognition mechanism. Blood *120*, 459–467.

Li, C., Chen, H., Hu, L., Xing, Y., Sasaki, T., Villosis, M.F., Li, J., Nishita, M., Minami, Y., and Minoo, P. (2008). Ror2 modulates the canonical Wnt signaling in lung epithelial cells through cooperation with Fzd2. BMC Mol. Biol. *9*, 11.

Li, Z.J., Nieuwenhuis, E., Nien, W., Zhang, X., Zhang, J., Puviindran, V., Wainwright, B.J., Kim, P.C.W., and Hui, C. -c. (2012). Kif7 regulates Gli2 through Sufu-dependent and -independent functions during skin development and tumorigenesis. Development *139*, 4152–4161.

Liang, W.G., Triandafillou, C.G., Huang, T.-Y., Zulueta, M.M.L., Banerjee, S., Dinner, A.R., Hung, S.-C., and Tang, W.-J. (2016). Structural basis for oligomerization and glycosaminoglycan binding of CCL5 and CCL3. Proc. Natl. Acad. Sci. *113*, 5000–5005.

Liem, K.F., He, M., Ocbina, P.J.R., and Anderson, K.V. (2009). Mouse Kif7/Costal2 is a cilia-associated protein that regulates Sonic hedgehog signaling. Proc. Natl. Acad. Sci. *106*, 13377–13382.

Lin, X., and Perrimon, N. (1999). Dally cooperates with Drosophila Frizzled 2 to transduce Wingless signalling. Nature *400*, 281–284.

Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancia, F., and Hendrickson, W.A. (2012). Structures from anomalous diffraction of native biological macromolecules. Science *336*, 1033–1037.

Liu, Q., Liu, Q., and Hendrickson, W.A. (2013). Robust structural analysis of native biological macromolecules from multi-crystal anomalous diffraction data. Acta Crystallogr. D Biol. Crystallogr. *69*, 1314–1332.

Liu, Y., Rubin, B., Bodine, P.V.N., and Billiard, J. (2008). Wnt5a induces homodimerization and activation of Ror2 receptor tyrosine kinase. J. Cell. Biochem. *105*, 497–502.

Liu, Z.-J., Chen, L., Wu, D., Ding, W., Zhang, H., Zhou, W., Fu, Z.-Q., and Wang, B.-C. (2011). A multi-dataset data-collection strategy produces better diffraction data. Acta Crystallogr. A *67*, 544–549.

Long, J., Tokhunts, R., Old, W.M., Houel, S., Rodgriguez-Blanco, J., Singh, S., Schilling, N., Capobianco, A.J., Ahn, N.G., and Robbins, D.J. (2015). Identification of a family of fatty-acid-speciated sonic hedgehog proteins, whose members display differential biological properties. Cell Rep. *10*, 1280–1287.

Luchetti, G., Sircar, R., Kong, J.H., Nachtergaele, S., Sagner, A., Byrne, E.F., Covey, D.F., Siebold, C., and Rohatgi, R. (2016). Cholesterol activates the G-protein coupled receptor Smoothened to promote Hedgehog signaling. eLife *5*.

Luna-Ulloa, L.B., Hernández-Maqueda, J.G., Castañeda-Patlán, M.C., and Robles-Flores, M. (2011). Protein kinase C in Wnt signaling: Implications in cancer initiation and progression. IUBMB Life *63*, 915–921.

MacDonald, B.T., and He, X. (2012). Frizzled and LRP5/6 Receptors for Wnt/ -Catenin Signaling. Cold Spring Harb. Perspect. Biol. *4*, a007880–a007880.

MacDonald, B.T., Tamai, K., and He, X. (2009). Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. Dev. Cell *17*, 9–26.

Malinauskas, T., Aricescu, A.R., Lu, W., Siebold, C., and Jones, E.Y. (2011). Modular mechanism of Wnt signaling inhibition by Wnt inhibitory factor 1. Nat. Struct. Mol. Biol. *18*, 886–893.

Manon-Jensen, T., Multhaupt, H.A.B., and Couchman, J.R. (2013). Mapping of matrix metalloproteinase cleavage sites on syndecan-1 and syndecan-4 ectodomains. FEBS J. *280*, 2320–2331.

Mao, B., Wu, W., Davidson, G., Marhold, J., Li, M., Mechler, B.M., Delius, H., Hoppe, D., Stannek, P., Walter, C., et al. (2002). Kremen proteins are Dickkopf receptors that regulate Wnt/β-catenin signalling. Nature *417*, 664–667.

Martinelli, D.C., and Fan, C.-M. (2009). A sonic hedgehog missense mutation associated with holoprosencephaly causes defective binding to GAS1. J. Biol. Chem. *284*, 19169–19172.

Martinez, S., Scerbo, P., Giordano, M., Daulat, A.M., Lhoumeau, A.-C., Thome, V., Kodjabachian, L., and Borg, J.-P. (2015). The PTK7 and ROR2 receptors interact in the vertebrate WNT/PCP pathway. J. Biol. Chem. *290*, 30562-30572.

Matsuda, T., Nomi, M., Ikeya, M., Kani, S., Oishi, I., Terashima, T., Takada, S., and Minami, Y. (2001). Expression of the receptor tyrosine kinase genes, Ror1 and Ror2, during mouse development. Mech. Dev. *105*, 153–156.

Maun, H.R., Wen, X., Lingel, A., de Sauvage, F.J., Lazarus, R.A., Scales, S.J., and Hymowitz, S.G. (2010). Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. J. Biol. Chem. *285*, 26570–26580.

Maurel, P., Rauch, U., Flad, M., Margolis, R.K., and Margolis, R.U. (1994). Phosphacan, a chondroitin sulfate proteoglycan of brain that interacts with neurons and neural cell-adhesion molecules, is an extracellular variant of a receptor-type protein tyrosine phosphatase. Proc. Natl. Acad. Sci. U. S. A. *91*, 2512–2516.

McCabe, J.M., and Leahy, D.J. (2015). Smoothened Goes Molecular: New Pieces in the Hedgehog Signaling Puzzle. J. Biol. Chem. *290*, 3500–3507.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). *Phaser* crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

McLellan, J.S., Yao, S., Zheng, X., Geisbrecht, B.V., Ghirlando, R., Beachy, P.A., and Leahy, D.J. (2006). Structure of a heparin-dependent complex of Hedgehog and Ihog. Proc. Natl. Acad. Sci. *103*, 17208–17213.

McLellan, J.S., Zheng, X., Hauk, G., Ghirlando, R., Beachy, P.A., and Leahy, D.J. (2008). The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. Nature *455*, 979–983.

McNicholas, S., and Agirre, J. (2017). Glycoblocks: a schematic three-dimensional representation for glycans and their interactions. Acta Crystallogr. Sect. Struct. Biol. *73*.

Mehlen, P., Mille, F., and Thibert, C. (2005). Morphogens and cell survival during development. J. Neurobiol. *64*, 357–366.

Mendrola, J.M., Shi, F., Park, J.H., and Lemmon, M.A. (2013). Receptor tyrosine kinases with intracellular pseudokinase domains. Biochem. Soc. Trans. *41*, 1029–1036.

Mikami, T., and Kitagawa, H. (2013). Biosynthesis and function of chondroitin sulfate. Biochim. Biophys. Acta BBA - Gen. Subj. *1830*, 4719–4733.

Mikels, A.J., and Nusse, R. (2006). Purified Wnt5a Protein Activates or Inhibits β-Catenin–TCF Signaling Depending on Receptor Context. PLoS Biol. *4*, e115.

Monaco, H.L. (1997). Crystal structure of chicken riboflavin-binding protein. EMBO J. *16*, 1475–1483.

Mulloy, B., and Forster, M.J. (2000). Conformation and dynamics of heparin and heparan sulfate. Glycobiology *10*, 1147–1156.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). *REFMAC* 5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367.

Myers, B.R., Sever, N., Chong, Y.C., Kim, J., Belani, J.D., Rychnovsky, S., Bazan, J.F., and Beachy, P.A. (2013). Hedgehog Pathway Modulation by Multiple Lipid Binding Sites on the Smoothened Effector of Signal Response. Dev. Cell *26*, 346–357.

Nachtergaele, S., Whalen, D.M., Mydock, L.K., Zhao, Z., Malinauskas, T., Krishnan, K., Ingham, P.W., Covey, D.F., Siebold, C., and Rohatgi, R. (2013). Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. eLife *2*.

Nicholls, R.A., Fischer, M., McNicholas, S., and Murshudov, G.N. (2014). Conformation-independent structural comparison of macromolecules with ProSMART. Acta Crystallogr. D Biol. Crystallogr. *70*, 2487–2499.

Niehrs, C. (2012). The complex world of WNT receptor signalling. Nat. Rev. Mol. Cell Biol. *13*, 767–779.

Nile, A.H., and Hannoush, R.N. (2016). Fatty acylation of Wnt proteins. Nat. Chem. Biol. *12*, 60–69.

Nolan, K., Kattamuri, C., Luedeke, D.M., Deng, X., Jagpal, A., Zhang, F., Linhardt, R.J., Kenny, A.P., Zorn, A.M., and Thompson, T.B. (2013). Structure of Protein Related to Dan and Cerberus: Insights into the Mechanism of Bone Morphogenetic Protein Antagonism. Structure *21*, 1417–1429.

Nomi, M., Oishi, I., Kani, S., Suzuki, H., Matsuda, T., Yoda, A., Kitamura, M., Itoh, K., Takeuchi, S., Takeda, K., et al. (2001). Loss of mRor1 Enhances the Heart and Skeletal Abnormalities in mRor2-Deficient Mice: Redundant and Pleiotropic Functions of mRor1 and mRor2 Receptor Tyrosine Kinases. Mol. Cell. Biol. *21*, 8329–8335.

O'callaghan, C.A., Byford, M.F., Wyer, J.R., Willcox, B.E., Jakobsen, B.K., McMichael, A.J., and Bell, J.I. (1999). BirA enzyme: production and application in the study of membrane receptor-ligand interactions by site-specific biotinylation. Anal. Biochem. *266*, 9–15.

Ochi, H., Pearson, B.J., Chuang, P.-T., Hammerschmidt, M., and Westerfield, M. (2006). Hhip regulates zebrafish muscle development by both sequestering Hedgehog and modulating localization of Smoothened. Dev. Biol. *297*, 127–140.

Ohto, U., Ishida, H., Krayukhina, E., Uchiyama, S., Inoue, N., and Shimizu, T. (2016). Structure of IZUMO1-JUNO reveals sperm-oocyte recognition during mammalian fertilization. Nature *534*, 566–569.

Oishi, I. (1997). A Novel Drosophila Receptor Tyrosine Kinase Expressed Specifically in the Nervous System. UNIQUE STRUCTURAL FEATURES AND IMPLICATION IN DEVELOPMENTAL SIGNALING. J. Biol. Chem. *272*, 11916–11923.

Oishi, I., Suzuki, H., Onishi, N., Takada, R., Kani, S., Ohkawara, B., Koshida, I., Suzuki, K., Yamada, G., Schwabe, G.C., et al. (2003). The receptor tyrosine kinase Ror2 is involved in non-canonical Wnt5a/JNK signalling pathway. Genes Cells *8*, 645–654.

Pape, T., and Schneider, T.R. (2004). *HKL2MAP*: a graphical user interface for macromolecular phasing with *SHELX* programs. J. Appl. Crystallogr. *37*, 843–844.

Patel, D., Bauman, J.D., and Arnold, E. (2014). Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. Prog. Biophys. Mol. Biol. *116*, 92–100.

Pei, J., and Grishin, N.V. (2012). Cysteine-rich domains related to Frizzled receptors and Hedgehog-interacting proteins. Protein Sci. Publ. Protein Soc. *21*, 1172–1184.

Pelikan, M., Hura, G., and Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen. Physiol. Biophys. *28*, 174–189.

Pellegrini, L., Burke, D.F., von Delft, F., Mulloy, B., and Blundell, T.L. (2000). Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature *407*, 1029–1034.

Pepinsky, R.B., Zeng, C., Wen, D., Rayhorn, P., Baker, D.P., Williams, K.P., Bixler, S.A., Ambrose, C.M., Garber, E.A., Miatkowski, K., et al. (1998). Identification of a palmitic acid-modified form of human Sonic hedgehog. J. Biol. Chem. *273*, 14037–14045.

Perrimon, N., Pitsouli, C., and Shilo, B.-Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harb. Perspect. Biol. *4*, a005975.

Peters, C., Wolf, A., Wagner, M., Kuhlmann, J., and Waldmann, H. (2004). The cholesterol membrane anchor of the Hedgehog protein confers stable membrane association to lipid-modified proteins. Proc. Natl. Acad. Sci. *101*, 8531–8536.

Petrova, R., and Joyner, A.L. (2014). Roles for Hedgehog signaling in adult organ homeostasis and repair. Development *141*, 3445–3457.

Presto, J., Thuveson, M., Carlsson, P., Busse, M., Wilen, M., Eriksson, I., Kusche-Gullberg, M., and Kjellen, L. (2008). Heparan sulfate biosynthesis enzymes EXT1 and EXT2 affect NDST1 expression and heparan sulfate sulfation. Proc. Natl. Acad. Sci. *105*, 4751–4756.

Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., et al. (2009). Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins Struct. Funct. Bioinforma. *77*, 89–99.

Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496*, 477–481.

Rebagay, G., Yan, S., Liu, C., and Cheung, N.-K. (2012). ROR1 and ROR2 in Human Malignancies: Potentials for Targeted Therapy. Front. Oncol. *2*.

Regard, J.B., Zhong, Z., Williams, B.O., and Yang, Y. (2012). Wnt signaling in bone development and disease: making stronger bone with Wnts. Cold Spring Harb. Perspect. Biol. *4*.

Resh, M.D. (2013). Covalent lipid modifications of proteins. Curr. Biol. *23*, R431–R435.

Riffel, N., Harlos, K., Iourin, O., Rao, Z., Kingsman, A., Stuart, D., and Fry, E. (2002). Atomic Resolution Structure of Moloney Murine Leukemia Virus Matrix Protein and Its Relationship to Other Retroviral Matrix Proteins. Structure *10*, 1627–1636.

Ring, A.M., Manglik, A., Kruse, A.C., Enos, M.D., Weis, W.I., Garcia, K.C., and Kobilka, B.K. (2013). Adrenaline-activated structure of β2-adrenoceptor stabilized by an engineered nanobody. Nature *502*, 575–579.

Robbins, D.J., Fei, D.L., and Riobo, N.A. (2012). The Hedgehog signal transduction network. Sci. Signal. *5*, re6.

Rohatgi, R., Milenkovic, L., and Scott, M.P. (2007). Patched1 regulates hedgehog signaling at the primary cilium. Science *317*, 372–376.

Rothbächer, U., Laurent, M.N., Deardorff, M.A., Klein, P.S., Cho, K.W.Y., and Fraser, S.E. (2000). Dishevelled phosphorylation, subcellular localization and multimerization regulate its role in early embryogenesis. EMBO J. *19*, 1010–1022.

Rupp, B. (2009). Biomolecular crystallography: principles, practice, and application to structural biology (Garland Science).

Šali, A., and Blundell, T.L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. J. Mol. Biol. *234*, 779–815.

Sammar, M., Stricker, S., Schwabe, G.C., Sieber, C., Hartung, A., Hanke, M., Oishi, I., Pohl, J., Minami, Y., Sebald, W., et al. (2004). Modulation of GDF5/BRI-b signalling through interaction with the tyrosine kinase receptor Ror2: Ror2 modulates GDF5 signalling. Genes Cells *9*, 1227–1238.

Sammar, M., Sieber, C., and Knaus, P. (2009). Biochemical and functional characterization of the Ror2/BRIb receptor complex. Biochem. Biophys. Res. Commun. *381*, 1–6.

Sanders, T.A., Llagostera, E., and Barna, M. (2013). Specialized filopodia direct long-range transport of SHH during vertebrate tissue patterning. Nature *497*, 628–632.

Sarrazin, S., Lamanna, W.C., and Esko, J.D. (2011). Heparan Sulfate Proteoglycans. Cold Spring Harb. Perspect. Biol. *3*, a004952–a004952.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schlessinger, J., Plotnikov, A.N., Ibrahimi, O.A., Eliseenkova, A.V., Yeh, B.K., Yayon, A., Linhardt, R.J., and Mohammadi, M. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Mol. Cell *6*, 743–750.

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. Biophys. J. *105*, 962–974.

Sekulic, A., Migden, M.R., Oro, A.E., Dirix, L., Lewis, K.D., Hainsworth, J.D., Solomon, J.A., Yoo, S., Arron, S.T., Friedlander, P.A., et al. (2012). Efficacy and safety of vismodegib in advanced basal-cell carcinoma. N. Engl. J. Med. *366*, 2171–2179.

Sharpe, H.J., Wang, W., Hannoush, R.N., and de Sauvage, F.J. (2015). Regulation of the oncoprotein Smoothened by small molecules. Nat. Chem. Biol. *11*, 246–255.

Sheldrick, G.M. (2010). Experimental phasing with *SHELXC* / *D* / *E*: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 479–485.

Shen, M., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. Protein Sci. *15*, 2507–2524.

Stiegler, A.L., Burden, S.J., and Hubbard, S.R. (2009). Crystal structure of the frizzled-like cysteine-rich domain of the receptor tyrosine kinase MuSK. J. Mol. Biol. *393*, 1–9.

Stricker, S., Verhey Van Wijk, N., Witte, F., Brieske, N., Seidel, K., and Mundlos, S. (2006). Cloning and expression pattern of chicken *Ror2* and functional characterization of truncating mutations in Brachydactyly type B and Robinow syndrome. Dev. Dyn. *235*, 3456–3465.

Strutt, H., Thomas, C., Nakano, Y., Stark, D., Neave, B., Taylor, A.M., and Ingham, P.W. (2001). Mutations in the sterol-sensing domain of Patched suggest a role for vesicular trafficking in Smoothened regulation. Curr. Biol. *11*, 608–613.

Stuart, D.I., Levine, M., Muirhead, H., and Stammers, D.K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. J. Mol. Biol. *134*, 109–142.

Suhre, K., and Sanejouand, Y.-H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. *32*, W610–W614.

Svensson, G., Awad, W., Håkansson, M., Mani, K., and Logan, D.T. (2012). Crystal structure of N-glycosylated human glypican-1 core protein: structure of two loops evolutionarily conserved in vertebrate glypican-1. J. Biol. Chem. *287*, 14040–14051.

Taipale, J., Cooper, M.K., Maiti, T., and Beachy, P.A. (2002). Patched acts catalytically to suppress the activity of Smoothened. Nature *418*, 892–897.

Takada, R., Satomi, Y., Kurata, T., Ueno, N., Norioka, S., Kondoh, H., Takao, T., and Takada, S. (2006). Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. Dev. Cell *11*, 791–801.

Tanaka, Y., Okada, Y., and Hirokawa, N. (2005). FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left–right determination. Nature *435*, 172–177.

Tarentino, A.L., and Plummer, T.H. (1994). [4] Enzymatic deglycosylation of asparagine-linked glycans: Purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. Methods in Enzymology pp. 44–57.

Taylor, F.R., Wen, D., Garber, E.A., Carmillo, A.N., Baker, D.P., Arduini, R.M., Williams, K.P., Weinreb, P.H., Rayhorn, P., Hronowski, X., et al. (2001). Enhanced Potency of Human Sonic Hedgehog by Hydrophobic Modification. Biochemistry *40*, 4359–4371.

Tegoni, M., Campanacci, V., and Cambillau, C. (2004). Structural aspects of sexual attraction and chemical communication in insects. Trends Biochem. Sci. *29*, 257–264.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J., and Adams, P.D. (2008). Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. Acta Crystallogr. D Biol. Crystallogr. *64*, 61–69.

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.-W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the *PHENIX AutoSol* wizard. Acta Crystallogr. D Biol. Crystallogr. *65*, 582–601.

Terwilliger, T.C., Bunkóczi, G., Hung, L.W., Zwart, P.H., Smith, J.L., Akey, D.L., and Adams, P.D. (2016). Can I solve my structure by SAD phasing? Anomalous signal in SAD phasing. Acta Crystallogr. Sect. Struct. Biol. *72*, 346–358.

The, I., Bellaiche, Y., and Perrimon, N. (1999). Hedgehog movement is regulated through tout velu-dependent synthesis of a heparan sulfate proteoglycan. Mol. Cell *4*, 633–639.

Thelin, M.A., Bartolini, B., Axelsson, J., Gustafsson, R., Tykesson, E., Pera, E., Oldberg, Å., Maccarana, M., and Malmstrom, A. (2013). Biological functions of iduronic acid in chondroitin/dermatan sulfate. FEBS J. *280*, 2431–2446.

Theocharis, A.D., Skandalis, S.S., Gialeli, C., and Karamanos, N.K. (2016). Extracellular matrix structure. Adv. Drug Deliv. Rev. *97*, 4–27.

Thomas, G. (2002). Furin at the cutting edge: From protein traffic to embryogenesis and disease. Nat. Rev. Mol. Cell Biol. *3*, 753–766.

Trott, O., and Olson, A.J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. *31*, 455-461.

Tukachinsky, H., Lopez, L.V., and Salic, A. (2010). A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu–Gli protein complexes. J. Cell Biol. *191*, 415–428.

Tukachinsky, H., Kuzmickas, R.P., Jao, C.Y., Liu, J., and Salic, A. (2012). Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand. Cell Rep. *2*, 308–320.

Tukachinsky, H., Petrov, K., Watanabe, M., and Salic, A. (2016). Mechanism of inhibition of the tumor suppressor Patched by Sonic Hedgehog. Proc. Natl. Acad. Sci. *113*, E5866-E5875.

Tyson, J.J., and Novak, B. (2014). Control of cell growth, division and death: information processing in living cells. Interface Focus *4*, 20130070–20130070.

Ueno, M., Yamada, S., Zako, M., Bernfield, M., and Sugahara, K. (2001). Structural characterization of heparan sulfate and chondroitin sulfate of syndecan-1 purified from normal murine mammary gland epithelial cells. Common phosphorylation of xylose and differential sulfation of galactose in the protein linkage region tetrasaccharide sequence. J. Biol. Chem. *276*, 29134–29140.

Ultsch, M., Lokker, N.A., Godowski, P.J., and de Vos, A.M. (1998). Crystal structure of the NK1 fragment of human hepatocyte growth factor at 2.0 A resolution. Struct. Lond. Engl. 1993 *6*, 1383–1393.

Van Der Merwe, P.A. Surface plasmon resonance.

Varjosalo, M., and Taipale, J. (2008). Hedgehog: functions and mechanisms. Genes Dev. *22*, 2454–2472.

Veverka, V., Henry, A.J., Slocombe, P.M., Ventom, A., Mulloy, B., Muskett, F.W., Muzylak, M., Greenslade, K., Moore, A., Zhang, L., et al. (2009). Characterization of the Structural Features and Interactions of Sclerostin: MOLECULAR INSIGHT INTO A KEY REGULATOR OF Wnt-MEDIATED BONE FORMATION. J. Biol. Chem. *284*, 10890–10900.

Vlodavsky, I., Iozzo, R.V., and Sanderson, R.D. (2013). Heparanase: Multiple functions in inflammation, diabetes and atherosclerosis. Matrix Biol. *32*, 220–222.

Wagner, A., Duman, R., Henderson, K., and Mykhaylyk, V. (2016). In-vacuum long-wavelength macromolecular crystallography. Acta Crystallogr. Sect. Struct. Biol. *72*, 430–439.

Waksman, G., and Herr, A.B. (1998). New insights into heparin-induced FGF oligomerization. Nat. Struct. Biol. *5*, 527–530.

Walter, T.S., Diprose, J.M., Mayo, C.J., Siebold, C., Pickford, M.G., Carter, L., Sutton, G.C., Berrow, N.S., Brown, J., Berry, I.M., et al. (2005). A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. Acta Crystallogr. D Biol. Crystallogr. *61*, 651–657.

Walter, T.S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. Structure *14*, 1617–1622.

Walter, T.S., Ren, J., Tuthill, T.J., Rowlands, D.J., Stuart, D.I., and Fry, E.E. (2012). A plate-based high-throughput assay for virus stability and vaccine formulation. J. Virol. Methods *185*, 166–170.

Wang, C., Wu, H., Katritch, V., Han, G.W., Huang, X.-P., Liu, W., Siu, F.Y., Roth, B.L., Cherezov, V., and Stevens, R.C. (2013). Structure of the human smoothened receptor bound to an antitumour agent. Nature *497*, 338–343.

Wang, C., Wu, H., Evron, T., Vardy, E., Han, G.W., Huang, X.-P., Hufeisen, S.J., Mangano, T.J., Urban, D.J., Katritch, V., et al. (2014). Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. Nat. Commun. *5*, 4355.

Wang, Q., Symes, A.J., Kane, C.A., Freeman, A., Nariculam, J., Munson, P., Thrasivoulou, C., Masters, J.R.W., and Ahmed, A. (2010). A Novel Role for Wnt/Ca2+ Signaling in Actin Cytoskeleton Remodeling and Cell Motility in Prostate Cancer. PLoS ONE *5*, e10456.

Wang, X.F., Lin, H.Y., Ng-Eaton, E., Downward, J., Lodish, H.F., and Weinberg, R.A. (1991). Expression cloning and characterization of the TGF-beta type III receptor. Cell *67*, 797–805.

Watanabe, H., Yamada, Y., and Kimata, K. (1998). Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cartilage structure and function. J. Biochem. (Tokyo) *124*, 687–693.

Watanabe, K., Yamada, H., and Yamaguchi, Y. (1995). K-glypican: a novel GPI-anchored heparan sulfate proteoglycan that is highly expressed in developing brain and kidney. J. Cell Biol. *130*, 1207–1218.

Wawrzak, D., Métioui, M., Willems, E., Hendrickx, M., de Genst, E., and Leyns, L. (2007). Wnt3a binds to several sFRPs in the nanomolar range. Biochem. Biophys. Res. Commun. *357*, 1119–1123.

Wehrman, T., He, X., Raab, B., Dukipatti, A., Blau, H., and Garcia, K.C. (2007). Structural and Mechanistic Insights into Nerve Growth Factor Interactions with the TrkA and p75 Receptors. Neuron *53*, 25–38.

Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Rose, J., Ebihara, A., Kuramitsu, S., et al. (2014). Fast native-SAD phasing for routine macromolecular structure determination. Nat. Methods. *12*, 131-133.

Weinkam, P., Pons, J., and Sali, A. (2012). Structure-based model of allostery predicts coupling between distant sites. Proc. Natl. Acad. Sci. *109*, 4875–4880.

Weiss, M.S. (2001). Global indicators of X-ray data quality. J. Appl. Crystallogr. *34*, 130–135.

Wen, J., Xiao, J., Rahdar, M., Choudhury, B.P., Cui, J., Taylor, G.S., Esko, J.D., and Dixon, J.E. (2014). Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. Proc. Natl. Acad. Sci. *111*, 15723–15728.

Whalen, D.M., Malinauskas, T., Gilbert, R.J.C., and Siebold, C. (2013). Structural insights into proteoglycan-shaped Hedgehog signaling. Proc. Natl. Acad. Sci. U. S. A. *110*, 16420–16425.

Wibowo, A.S., Singh, M., Reeder, K.M., Carter, J.J., Kovach, A.R., Meng, W., Ratnam, M., Zhang, F., and Dann, C.E. (2013). Structures of human folate receptors reveal biological trafficking states and diversity in folate and antifolate recognition. Proc. Natl. Acad. Sci. U. S. A. *110*, 15180–15188.

Willert, K., Shibamoto, S., and Nusse, R. (1999). Wnt-induced dephosphorylation of axin releases beta-catenin from the axin complex. Genes Dev. *13*, 1768–1773.

Willert, K., Brown, J.D., Danenberg, E., Duncan, A.W., Weissman, I.L., Reya, T., Yates, J.R., and Nusse, R. (2003). Wnt proteins are lipid-modified and can act as stem cell growth factors. Nature *423*, 448–452.

Wilson, N.H., and Stoeckli, E.T. (2013). Sonic hedgehog regulates its own receptor on postcrossing commissural axons in a glypican1-dependent manner. Neuron *79*, 478–491.

Wilusz, R.E., Sanchez-Adams, J., and Guilak, F. (2014). The structure and function of the pericellular matrix of articular cartilage. Matrix Biol. *39*, 25–32.

Winter, G. (2010). X*ia2*: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. *43*, 186–190.

Wolpert, L. (1996). One hundred years of positional information. Trends Genet. *12*, 359–364.

Wyatt, P.J. (1993). Light scattering and the absolute characterization of macromolecules. Anal. Chim. Acta *272*, 1–40.

Xu, S., Benoff, B., Liou, H.-L., Lobel, P., and Stock, A.M. (2007). Structural basis of sterol binding by NPC2, a lysosomal protein deficient in Niemann-Pick type C2 disease. J. Biol. Chem. *282*, 23525–23531.

Yam, P.T., and Charron, F. (2013). Signaling mechanisms of non-conventional axon guidance cues: the Shh, BMP and Wnt morphogens. Curr. Opin. Neurobiol. *23*, 965–973.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics *21*, 3369–3376.

Yilmaz, A., Kattamuri, C., Ozdeslik, R.N., Schmiedel, C., Mentzer, S., Schorl, C., Oancea, E., Thompson, T.B., and Fallon, J.R. (2016). MuSK is a BMP co-receptor that shapes BMP responses and calcium signaling in muscle cells. Sci. Signal. *9*, ra87-ra87.

Zebisch, M., and Jones, E.Y. (2015). Crystal structure of R-spondin 2 in complex with the ectodomains of its receptors LGR5 and ZNRF3. J. Struct. Biol. *191*, 149–155.

Zebisch, M., Jackson, V.A., Zhao, Y., and Jones, E.Y. (2016). Structure of the Dual-Mode Wnt Regulator Kremen1 and Insight into Ternary Complex Formation with LRP6 and Dickkopf. Structure *24*, 1599–1605.

Zechner, D., Müller, T., Wende, H., Walther, I., Taketo, M.M., Crenshaw, E.B., Treier, M., Birchmeier, W., and Birchmeier, C. (2007). Bmp and Wnt/β-catenin signals control expression of the transcription factor Olig3 and the specification of spinal cord neurons. Dev. Biol. *303*, 181–190.

Zeng, X., Huang, H., Tamai, K., Zhang, X., Harada, Y., Yokota, C., Almeida, K., Wang, J., Doble, B., Woodgett, J., et al. (2007). Initiation of Wnt signaling: control of Wnt coreceptor Lrp6 phosphorylation/activation via frizzled, dishevelled and axin functions. Development *135*, 367–375.

Zhang, B., Liang, C., Bates, R., Yin, Y., Xiong, W.-C., and Mei, L. (2012). Wnt proteins regulate acetylcholine receptor clustering in muscle cells. Mol. Brain *5*, 7.

Zhang, W., Coldefy, A.-S., Hubbard, S.R., and Burden, S.J. (2011). Agrin Binds to the N-terminal Region of Lrp4 Protein and Stimulates Association between Lrp4 and the First Immunoglobulin-like Domain in Muscle-specific Kinase (MuSK). J. Biol. Chem. *286*, 40624–40630.

Zhao, Y., Ren, J., Harlos, K., and Stuart, D.I. (2016). Structure of glycosylated NPC1 luminal domain C reveals insights into NPC2 and Ebola virus interactions. FEBS Lett. *590*, 605–612.

Zong, Y., Zhang, B., Gu, S., Lee, K., Zhou, J., Yao, G., Figueiredo, D., Perry, K., Mei, L., and Jin, R. (2012). Structural basis of agrin-LRP4-MuSK signaling. Genes Dev. *26*, 247–258.

Supplementary Information

Abbreviations (not expanded upon in the text)

μg – Microgram

μg – Microgram

AU – Absorbance Unit

CCP4 – Collaborative Computing project 4

cDNA – Complementary Deoxyribonucleic Acid

CM5 – Carboxymethyl 5

Da – Dalton

DDR – Discoidin Domain Receptor

EndoF1 – Endo-*N*-acetylglucosaminidase F1

Fn – Fibronectin

GPI – Glycosylphosphodylinositol

GDNFR – Glial-Derived Neurotrophic Factor Receptor

HEK293T – Human Embryonic Kidney 293 cell line expressing SV40-type T antigen

Ig – Immunoglobulin

His_6-tag – Hexahistidine tag

IMAC – Immobilised Metal Affinity Chromatography

LB – Luria-Bertani Media

mg – Milligram

mM – Millimolar

M – Molar

NCS – Non-Crystallographic Symmetry

nL – Nanolitre

nM – Nanomolar

PCR – Polymerase Chain Reaction

PEG – Polyethylene Glycol

pmol – picomole

v/v – Volume by volume

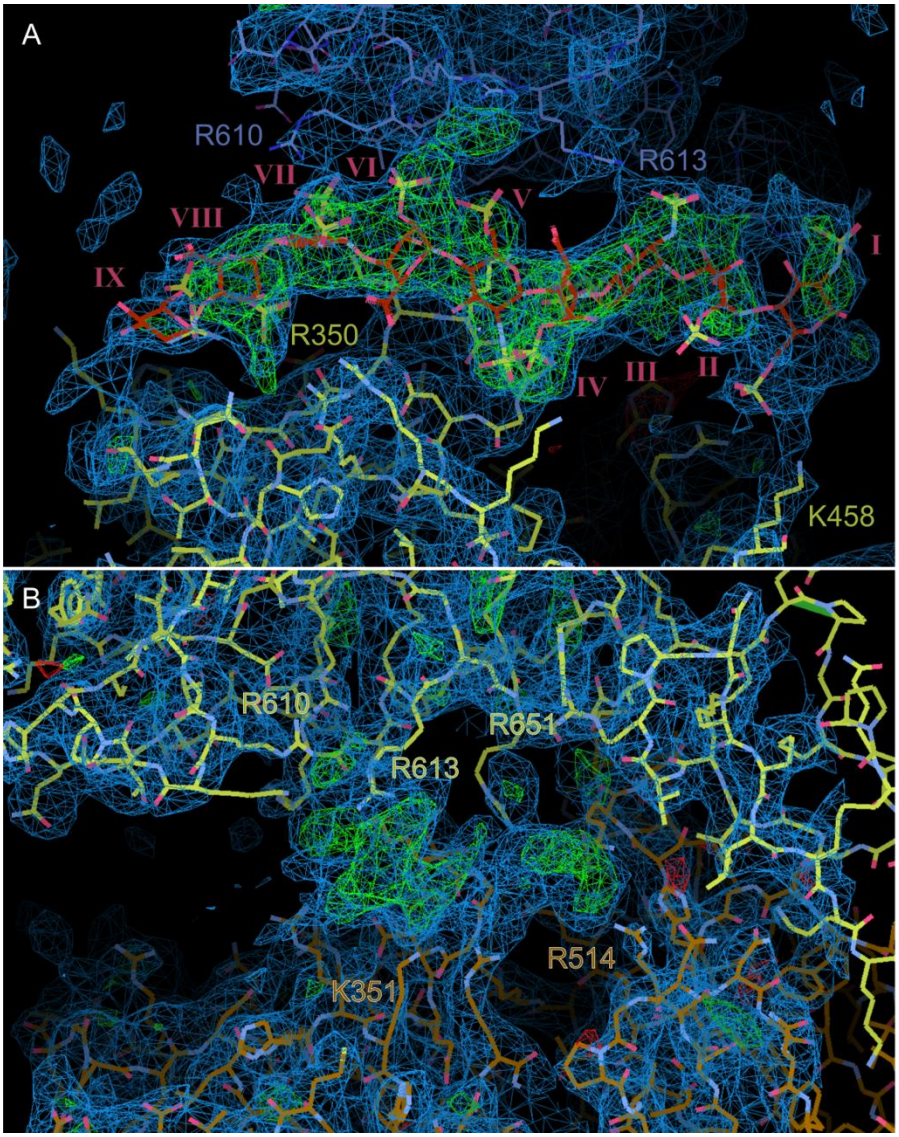
Supplementary Materials

Protein	Fz8-PAM (4F0A)[A]	Smo (5L7D)	Fz8 (1IJY)	sFRP3 (1IJX)	MuSK (3HKL)	NPC1 (3GKI)	RFBP (not in PDB)	FRα (4LRH)	JUNO (5EJN)	FRβ (4KMZ)
HHIP-N CRD	2.20[B] 38[C] 14.94[D]	2.35 41 13.57	2.27 42 14.35	2.24 41 14.36	1.84 51 23.13	1.99 61 19.7	1.53 71 34.69	1.53 66 34.33	1.54 63 32.47	1.55 66 34.83
Fz8-PAM (4F0A)		1.00 83 43.64	0.16 117 102.54	0.40 106 79.01	1.30 78 38.33	2.17 61 21.36	2.54 40 13.35	2.76 44 12.16	2.51 44 13.31	2.43 43 14.14
Smo (5L7D)			1.09 84 43.18	1.04 80 44.60	1.26 78 40.84	2.32 54 19.01	2.68 46 11.24	2.62 46 13.73	2.64 44 12.47	2.49 48 15.65
Fz8 (1IJY)				0.45 108 78.07	1.35 76 36.13	2.14 62 22.27	2.50 45 15.00	2.66 43 13.41	2.38 45 15.40	2.53 43 14.34
sFRP3 (1IJX)					1.36 79 36.65	2.41 60 17.28	2.31 45 17.60	2.59 47 14.13	2.49 39 13.87	2.47 46 15.62
MuSK (3HKL)						2.14 67 22.48	2.25 55 20.03	2.30 55 15.14	2.16 53 19.21	2.13 53 23.24
NPC1 (3GKI)							1.73 82 35.46	1.95 89 31.57	1.74 81 36.37	1.74 90 30.58
RFBP (not in PDB)								0.68 160 104.06	0.85 138 80.30	0.66 157 105.39
FRα (4LRH)									0.42 161 125.21	0.16 194 176.05
JUNO (5EJN)										0.46 157 120.83

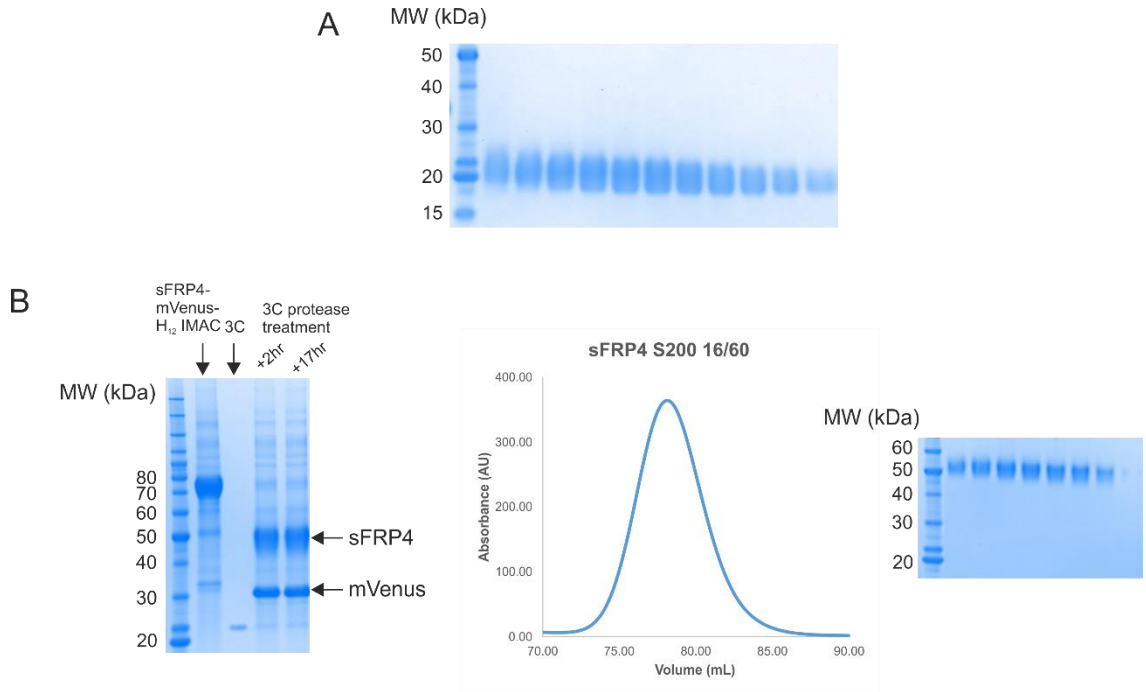
Supplementary Table 3.1 – Evolutionary analysis of CRD structures. [A]PDB accession codes are shown in parenthesis; [B]Root-mean Square Deviation (RMSD) values were calculated for equivalent Cα atom positions; [C]Number of equivalent Cα positions utilised in RMSD calculation; [D]Summed structural correlation (total probability) values calculated using SHP. The phylogenetic tree for analysed CRDs (Fig. 3.6A) was arranged with these summed structural correlation values to construct a distance matrix.

Construct name/number	Residue boundaries
HHIP-L1	
FL	20-782
CRD-β-propeller	20-605
CRD	20-165
1	20-180
2	20-177
3	20-161
4	20-159
5	20-151
6	23-180
7	23-177
8	23-161
9	23-159
10	23-151
11	23-165
12	23-605
13	668-782
HHIP-L2	
CRD-β-propeller	42-621
CRD	42-190
1	42-203
2	42-200
3	42-184
4	42-182
5	42-175
6	45-203
7	45-200
8	45-184
9	45-182
10	42-175
11	42-621
12	45-190

Supplementary Table 3.2 – Residue boundaries for HHIP-like expression constructs (corresponding to blots from Fig. 3.10D-F).



Supplementary Figure 4.1 – HHIP-C:heparin complex ligand-bound sites. A) View of the 9-mer heparin molecule visible (red), bound between HHIP-C chain A in the ASU (β-propeller; yellow) and a symmetry-related chain (EGF repeats; blue). The $2F_o$-F_c (1.0 σ, blue) and F_o-F_c (3.0 σ, green) electron density maps calculated for the final heparin molecule (occupancies set to zero) are displayed. Sugar residues are numbered using Roman numerals (odd, O2-Sulphoiduronic acid; even, N,O6-disulphoglucosamine). Residues coordinating heparin are labelled (see Fig. 4.11A); B) Equivalent site within the ASU. The EGF repeats of chain A (yellow) contact the β-propeller of a second chain B (brown). $2F_o$-F_c (1.0 σ, blue) and F_o-F_c (3.0 σ, green and red) electron density maps calculated for the final model are displayed, and a central region of density is observed that likely corresponds to a poorly-ordered heparin molecule. Residues seen to coordinate heparin and SOS are highlighted (see Figs. 4.10C and 4.11A-B).



Supplementary Figure 5.1 – Purification of proteins used for experiments in the study of ROR2. A) SDS-PAGE gel showing the peak fraction for purification of the BMPR1b ectodomain; B) Purification of sFRP4. sFRP4 was expressed as a sFRP4-mVenus-H_{12} fusion construct with a 3C protease site between both proteins and purified via IMAC (lane 2, left). To find an optimal time for cleavage, 3C protease was incubated with sFRP4-mVenus-H_{12} at 25^{o}C for a period of 2 hours (lane 4) and 17 hours (lane 5). 3C protease was also run on this gel as a control (lane 3). Total purified protein was cleaved via incubation of sFRP4-mVenus-H_{12} with 3C protease at a ratio of 100:1 (w/w) for 2 hours at 25^{o}C and sFRP4 was purified away from the protease and mVenus tag via the use of a second IMAC step. The IMAC flow-through was collected, concentrated and injected onto an S200 16/60 column. The SEC trace for this purification step is shown (middle) and purity was assessed via SDS-PAGE (right).

Resolution shell	**16.07**	**9.89**	**7.45**	**6.09**	**5.21**	**4.59**	**4.12**	**3.75**	**3.45**	**3.21**	**3.00**
Completeness (%)	96.2	99.3	99.6	99.7	99.8	100.0	99.9	99.9	100.0	99.9	99.9
SigAno	3.32	3.26	2.86	2.07	1.51	1.06	0.85	0.75	0.76	0.58	0.61
CCanom	85.6	97.7	94.9	82.6	59.6	31.9	11.7	-3.9	-5.0	-23.1	-4.3

Supplementary Table 5.1 – SHELXC output for ROR2 CRD-Kr S-SAD data collected on beamline I23. Data are presented on a per-resolution shell basis. SigAno and CCanom were calculated as outlined in the legend for Fig. 3.2.

Protein	Fz8-PAM (4F0A)	Smo (5L7D)	Fz8 (1IJY)	sFRP3 (1IJX)	MuSK (3HKL)	NPC1 (3GKI)	RFBP (not in PDB)	FRα (4LRH)	JUNO (5EJN)	FRβ (4KMZ)	ROR2 CRD
HHIP-N CRD	2.20 38 14.94	2.35 41 13.57	2.27 42 14.35	2.24 41 14.36	1.84 51 23.13	1.99 61 19.7	1.53 71 34.69	1.53 66 34.33	1.54 63 32.47	1.55 66 34.83	2.06 52 16.44
Fz8-PAM (4F0A)		1.00 83 43.64	0.16 117 102.54	0.40 106 79.01	1.30 78 38.33	2.17 61 21.36	2.54 40 13.35	2.76 44 12.16	2.51 44 13.31	2.43 43 14.14	1.22 81 39.80
Smo (5L7D)			1.09 84 43.18	1.04 80 44.60	1.26 78 40.84	2.32 54 19.01	2.68 46 11.24	2.62 46 13.73	2.64 44 12.47	2.49 48 15.65	1.29 78 35.37
Fz8 (1IJY)				0.45 108 78.07	1.35 76 36.13	2.14 62 22.27	2.50 45 15.00	2.66 43 13.41	2.38 45 15.40	2.53 43 14.34	1.28 82 36.11
sFRP3 (1IJX)					1.36 79 36.65	2.41 60 17.28	2.31 45 17.60	2.59 47 14.13	2.49 39 13.87	2.47 46 15.62	1.24 83 36.46
MuSK (3HKL)						2.14 67 22.48	2.25 55 20.03	2.30 55 15.14	2.16 53 19.21	2.13 53 23.24	0.87 111 63.31
NPC1 (3GKI)							1.73 82 35.46	1.95 89 31.57	1.74 81 36.37	1.74 90 30.58	1.96 73 27.23
RFBP (not in PDB)								0.68 160 104.06	0.85 138 80.30	0.66 157 105.39	2.48 53 15.01
FRα (4LRH)									0.42 161 125.21	0.16 194 176.05	2.77 51 12.88
JUNO (5EJN)										0.46 157 120.83	2.36 48 15.46
FRβ (4KMZ)											2.42 51 15.98

Supplementary Table 5.2 – RMSD values calculated for equivalent Cα positions by SHP for the ROR2 CRD. See Supplementary Table 3.1 for a description of the data displayed.

Protein	Apolipoprotein(A) Kr IV (1I71)	Angiostatin (1KI0)	Plasminogen Kr IV (1PML)	UPA (2FD6)	HGF NK-1 (2QJ2)	Pre-Thrombin (3NXP)	Kremen (5WFS)	Plasminogen Kr V (5HPG)
ROR2 Kr	0.38 65 53.08	0.46 64 49.85	0.63 58 42.86	0.66 60 40.65	0.39 66 51.97	0.65 53 41.28	0.71 57 39.13	0.47 62 48.98
Apolipoprotein(A) Kr IV (1I71)		0.24 81 66.68	0.44 72 55.02	0.50 70 51.10	0.30 75 61.27	0.61 63 45.73	0.68 63 43.02	0.23 82 66.56
Angiostatin (1KI0)			0.53 71 50.55	0.58 71 47.91	0.32 75 60.32	0.57 61 46.90	0.64 67 44.99	0.16 83 72.03
Plasminogen Kr IV (1PML)				0.27 82 64.40	0.68 69 43.34	0.61 66 45.14	0.58 68 48.25	0.50 74 52.20
UPA (2FD6)					0.72 66 40.62	0.70 66 40.50	0.57 67 47.94	0.57 71 47.42
HGF NK-1 (2QJ2)						0.61 63 43.15	0.72 60 41.02	0.30 75 61.92
Pre-Thrombin (3NXP)							0.75 61 40.67	0.56 64 48.68

Supplementary Table 5.3 – RMSD values calculated for equivalent Cα positions by SHP for the ROR2 Kringle domain (Kr). See Supplementary Table 3.1 for a description of the data displayed. UPA - Urokinase Plasminogen Activator; HGF NK-1 – Hepatocyte Growth Factor N-terminal and Kringle-1 fragment.

Supplementary Bibliography (for the construction of Fig. 3.14B)

Akey, D.L., Brown, W.C., Konwerski, J.R., Ogata, C.M., and Smith, J.L. (2014). Use of massively multiple merged data for low-resolution S-SAD phasing and refinement of flavivirus NS1. Acta Crystallogr. D Biol. Crystallogr. *70*, 2719–2729.

Bian, C., Yuan, C., Chen, L., Meehan, E.J., Jiang, L., Huang, Z., Lin, L., and Huang, M. (2010). Crystal structure of a triacylglycerol lipase from *Penicillium expansum* at 1.3 Å determined by sulfur SAD. Proteins Struct. Funct. Bioinforma. *78*, 1601-1605

Bond, C.S., Shaw, M.P., Alphey, M.S., and Hunter, W.N. (2001). Structure of the macrocycle thiostrepton solved using the anomalous dispersion contribution of sulfur. Acta Crystallogr. D Biol. Crystallogr. *57*, 755–758.

Debreczeni, J.E., Bunkóczi, G., Girmann, B., and Sheldrick, G.M. (2003). In-house phase determination of the lima bean trypsin inhibitor: a low-resolution sulfur-SAD case. Acta Crystallogr. D Biol. Crystallogr. *59*, 393–395.

Doan, D.N.P., and Dokland, T. (2003). Structure of the nucleocapsid protein of porcine reproductive and respiratory syndrome virus. Struct. Lond. Engl. 1993 *11*, 1445–1451.

El Omari, K., Iourin, O., Kadlec, J., Fearn, R., Hall, D.R., Harlos, K., Grimes, J.M., and Stuart, D.I. (2014). Pushing the limits of sulfur SAD phasing: de novo structure solution of the N-terminal domain of the ectodomain of HCV E1. Acta Crystallogr. D Biol. Crystallogr. *70*, 2197–2203.

Fukakusa, S., Kawahara, K., Nakamura, S., Iwashita, T., Baba, S., Nishimura, M., Kobayashi, Y., Honda, T., Iida, T., Taniguchi, T., et al. (2012). Structure of the CFA/III major pilin subunit CofA from human enterotoxigenic *Escherichia coli* determined at 0.90 Å resolution by sulfur-SAD phasing. Acta Crystallogr. D Biol. Crystallogr. *68*, 1418–1429.

Gordon, E.J., Leonard, G.A., McSweeney, S., and Zagalsky, P.F. (2001). The C 1 subunit of α-crustacyanin: the *de novo* phasing of the crystal structure of a 40 kDa homodimeric protein using the anomalous scattering from S atoms combined with direct methods. Acta Crystallogr. D Biol. Crystallogr. *57*, 1230–1237.

Gorgel, M., Bøggild, A., Ulstrup, J.J., Weiss, M.S., Müller, U., Nissen, P., and Boesen, T. (2015). Against the odds? *De novo* structure determination of a pilin with two cysteine residues by sulfur SAD. Acta Crystallogr. D Biol. Crystallogr. *71*, 1095–1101.

Lakomek, K., Dickmanns, A., Mueller, U., Kollmann, K., Deuschl, F., Berndt, A., Lübke, T., and Ficner, R. (2009). *De novo* sulfur SAD phasing of the lysosomal 66.3 kDa protein from mouse. Acta Crystallogr. D Biol. Crystallogr. *65*, 220–228.

Lemke, C.T., Smith, G.D., and Howell, P.L. (2002). S-SAD, Se-SAD and S/Se-SIRAS using Cu *K* α radiation: why wait for synchrotron time? Acta Crystallogr. D Biol. Crystallogr. *58*, 2096–2101.

Li, Y., Bahti, P., Shaw, N., Song, G., Chen, S., Zhang, X., Zhang, M., Cheng, C., Yin, J., Zhu, J.-Y., et al. (2008). Crystal structure of a novel non-Pfam protein AF1514 from Archeoglobus fulgidus DSM 4304 solved by S-SAD using a Cr X-ray source. Proteins Struct. Funct. Bioinforma. *71*, 2109–2113.

Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancia, F., and Hendrickson, W.A. (2012). Structures from anomalous diffraction of native biological macromolecules. Science *336*, 1033–1037.

Pakharukova, N., Tuittila, M., and Zavialov, A. (2013). Crystallization and sulfur SAD phasing of AggA, the major subunit of aggregative adherence fimbriae type I from the *Escherichia coli* strain that caused an outbreak of haemolytic-uraemic syndrome in Germany. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. *69*, 1389–1392.

Pal, A., Debreczeni, J.É., Sevvana, M., Gruene, T., Kahle, B., Zeeck, A., and Sheldrick, G.M. (2008). Structures of viscotoxins A1 and B2 from European mistletoe solved using native data alone. Acta Crystallogr. D Biol. Crystallogr. *64*, 985–992.

Ramagopal, U.A., Dauter, M., and Dauter, Z. (2003). Phasing on anomalous signal of sulfurs: what is the limit? Acta Crystallogr. D Biol. Crystallogr. *59*, 1020–1027.

Vasur, J., Kawai, R., Larsson, A.M., Igarashi, K., Sandgren, M., Samejima, M., and Ståhlberg, J. (2006). X-ray crystallographic native sulfur SAD structure determination of laminarinase Lam16A from *Phanerochaete chrysosporium*. Acta Crystallogr. D Biol. Crystallogr. *62*, 1422–1429.

Wagner, A., Pieren, M., Schulze-Briese, C., Ballmer-Hofer, K., and Prota, A.E. (2006). Structure determination of VEGF-E by sulfur SAD. Acta Crystallogr. D Biol. Crystallogr. *62*, 1430–1434.

Yang, C., and Pflugrath, J.W. (2001). Applications of anomalous scattering from S atoms for improved phasing of protein diffraction data collected at Cu *K* α wavelength. Acta Crystallogr. D Biol. Crystallogr. *57*, 1480–1490.

Yogavel, M., Tripathi, T., Gupta, A., Banday, M.M., Rahlfs, S., Becker, K., Belrhali, H., and Sharma, A. (2014). Atomic resolution crystal structure of glutaredoxin 1 from *Plasmodium falciparum* and comparison with other glutaredoxins. Acta Crystallogr. D Biol. Crystallogr. *70*, 91–100.

Zhu, J.-Y., Fu, Z.-Q., Chen, L., Xu, H., Chrzas, J., Rose, J., and Wang, B.-C. (2012). Structure of the *Archaeoglobus fulgidus* orphan ORF AF1382 determined by sulfur SAD from a moderately diffracting crystal. Acta Crystallogr. D Biol. Crystallogr. *68*, 1242–1252.