Structural and Functional Characterisation of Hedgehog Ligand-Receptor Complexes

Benjamin F. Bishop

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

Wolfson College
Nuffield Department of Medicine
Division of Structural Biology
University of Oxford

Trinity 2012

As the creeper that girdles the tree-trunk the Law runneth forward and back —
For the strength of the Pack is the Wolf, and the strength of the Wolf is the Pack.

The Jungle Book, Rudyard Kipling

Structural and Functional Characterisation of Hedgehog Ligand-Receptor Complexes

Benjamin F. Bishop

A thesis submitted in partial fulfilment of the requirements for
the degree of *Doctor of Philosophy*
Wolfson College, Nuffield Department of Medicine and Division of Structural Biology
University of Oxford, Trinity Term 2012

Abstract

Members of the Hedgehog (HH) family of morphogenic signalling molecules are key mediators of many fundamental processes in embryonic development. A relatively small change in HH concentration results in the specification of distinct cell types. Such fine-tuning necessitates a range of regulatory cell-surface proteins to control the concentration of HH to which responding cells are exposed. This thesis focuses on the structural and functional characterisation of three human extracellular modulators of the HH pathway, namely the hedgehog-interacting protein HHIP, the glypican GPC3 and the Growth Arrest Specific 1 (GAS1). Cell culture robot protocols were developed to allow large-scale expression of these targets in mammalian cells (Chapter 3). In Chapter 4, the structure of the HH antagonist HHIP alone and in complex with HH ligands is described. These studies combined with functional experiments reveal a binding mode distinct from previously defined ligand interaction sites, with the HH metal-binding sites playing key roles. The structural and functional analysis of GPC3, another negative HH regulator, is described in Chapter 5. Binding studies suggest that the GPC3 core domain does not bind directly to SHH and that the GPC3-attched heparan sulphate chains play an important role in HH regulation. Chapter 6 reveals crystal structures of the GAS1 ectodomain, a HH agonist, allowing comparison to glial cell line-derived neurotrophic factor receptors, the identification of an unexpected ligand and mapping of the HH binding site. In summary, this work provides insights into the extracellular modulation of HH signalling and extends our current knowledge of this fundamental signalling pathway. It also offers a model to explain how both agonists and antagonists can adopt similar mechanisms of HH binding.

Acknowledgements

First, I would like to thank my supervisor Christian Siebold for his encouragement, support and unfailing enthusiasm and patience during the four years of my DPhil studies. STRUBI is a wonderful place to work and I am grateful to him for giving me the opportunity to study here.

I would also like to thank everyone in STRUBI for helping me in the wetlab and making my time in Oxford so enjoyable. I would like to acknowledge Amber Clayton, Margaret Jones, Weixian Lu, Tomas Malinauskas, Ross Robinson and Yuguang Zhao at STRUBI for their help in learning new techniques, great discussions and weekend support. Special thanks go to Radu Aricescu, Yuguang Zhao and Tomas Malinauskis for their practical help and the many useful discussions and Rebecca Qiezi Ting for reading through my thesis and keeping me sane. I would also like to acknowledge Karl Harlos and Tom Walter for their invaluable help in setting up crystals and preparing them for data collection and Geoff Sutton for his help with MALS. I would also like to thank Fleur and Kelly for supporting me throughout my Thesis.

Finally I want to thank my parents for their unwavering love and support. Without them, none of this would have been possible.

Abbreviations

BCC Basal Cell Carcinoma

BOC Brother of CDO

BOI Brother of IHOG

CD Cluster of Differentiation

CDO Cell adhesion molecule-related/down-regulated by oncogenes

CI Cubitus Interuptus

CI-A Active form of CI

CI-FL Full length CI

CKI casein kinase I

COS2 Costal 2

CRD cysteine-rich domain

DHH Desert Hedgehog

DISP Dispatched

DLP Dally-like protein

dNTP deoxyribonucleotide

EGF Epidermal Growth Factor

EVC Elli-van Creveld syndrome

FBS Foetal Bovine Serum

FTM fantom

FU Fused

FZ Frizzled

GAG Glycosamino-glycans

GAS1 Growth Arrest Specific 1

GDC-0449 Vismodegib, HH Antagonist 691

GLI Glioma

GPC Glypican

GPI Glycosylphosphatidylinositol

GSK3 Glycogen synthase kinase-3

HH Hedgehog

HHAT Hedgehog acyltransferase

HHIP Hedgehog Interactig Protein

HHN N-terminal domain of Hedgehog

HPE holoprosencephaly

HSPG Heperan Sulphate Proteoglycans

IFT Intraflagellar transport

IHH Indian Hedgehog

IHOG Interference hedgehog

Kif Kifunensine

KIF7 Kinesin-like protein KIF7

MIHF Middle Interhemispheric fusion

Mks1 Meckel syndrome type 1

OFD1 oral-facial-digital syndrome 1

PCR Polymerase Chain Reaction

PEI Polyethylenimine

PKA Protein Kinase A

PTCH Patched

rmsd root mean square deviation

RND resistance-nodulation division

SHF Shifted

SHH Sonic Hedgehog

SKI Skinny Hedgehog

SMO Smoothened

SSD Sterol sensing domain

SUFU Supressor of Fused

ZPA zone of polarizing activity

Table of Contents

Chapter 1

Introduction

Chapter 1 Introduction

1.1 Multicellularity

The emergence of multicellular organisms became possible with the advent and refinement of mechanisms by which cells could communicate with each other. Relatively few signalling pathways across the metazoa have been harnessed for developmental pathways. The central canon of developmental pathways includes: transforming growth factor-β (TGFβ), bone morphogenetic protein (BMP), the JAK-signal transducer and activator of transcription (STAT) protein kinases, epidermal growth factor receptor (EGFR), NOTCH, Fibroblast growth factor (FGF), WNT and Hedgehog (HH) (Ingham et al. 2011). The HH signalling pathway has been particularly difficult to study, due in part to the fact that HH signalling has been lost in one of the main tools used by geneticists along with *Drosophila melanogaster*; the nematode *Caenorhabditis elegans*, which has no HH orthologue (Aspöck et al. 1999), but does possess several genes encoding proteins homologous to the HH receptor PTCH (Kuwabara et al. 2000; Ingham & McMahon 2001), and in part due to its byzantine mode of action (Ingham et al. 2011).

1.2 Early development

The maternal genome controls virtually all aspects of early animal development. Maternal mRNAs and proteins, which are loaded into the egg during oogenesis, implement basic biosynthetic processes in the early embryo, direct the first mitotic divisions, and specify initial cell fate and patterning. As development proceeds, two processes are triggered that together form the maternal-to-zygotic transition (MZT):

first, a subset of the maternal mRNAs is eliminated; second, the transcription of the zygotic genome begins (Tadros & Lipshitz 2009).

Initially, maternally-encoded products accomplish the destruction of maternal mRNAs. However, zygotic transcription leads to the production of proteins and microRNAs (miRNAs) that provide feedback to enhance the efficiency of maternal mRNA degradation. In addition, among the earliest mRNAs synthesized *de novo* in the embryo are transcriptional activators that enhance the efficiency of zygotic transcription. The net result is that the control of development is transferred from the maternal to the zygotic genome (Tadros & Lipshitz 2009). From this point on, cells require positional information in order to develop correctly. This is achieved by use of signalling with proteins known as morphogens.

1.3 Morphogens

The term 'morphogen' was first coined by Alan Turing in his seminal paper, "The Chemical Basis of Morphogenesis" (Turing 1952). In this paper, he stated:

"…a system of chemical substances, called morphogens, reacting together and diffusing through a tissue, is adequate to account for the main phenomena of morphogenesis. Such a system, although it may originally be quite homogeneous, may later develop a pattern or structure due to an instability of the homogeneous equilibrium, which is triggered off by random disturbances."

Although more complex models have superseded Turing's model, it still offers the first insight as to how complex organisms could form from a ball of cells. Patterning

of the body plan takes place during development of the embryo, whereby cells acquire positional information specifying their place within tissues, organs and the body in general. At the tissue level, cells must receive cues relating to their relative location so that they acquire an appropriate cell fate. The instructive signals that deliver positional information to the cells are garnered by morphogens. The next big advance in the understanding of morphogens and morphogenesis was developed by Lewis Wolpert with his 'French Flag Model' (Wolpert 1968; Wolpert 1969). As Wolpert saw it:

"Spatial differentiation is the process by which the individual cells within a population are specified to undergo a particular molecular differentiation, which results in a characteristic spatial pattern." (Wolpert 1969)

To deal with this problem, he devised the tricolore system to represent the effect of a morphogen on cell differentiation. In this model, the concentration of the morphogen affects the cell state; the states being represented by the three colours of the French flag. High concentration of a morphogen activate a 'blue' gene, lower concentrations activate a 'white' gene while cells in an environment below the concentration threshold of the morphogen are in their default 'red' state (Wolpert 1968; Wolpert 1969) (**Fig. 1.1**).

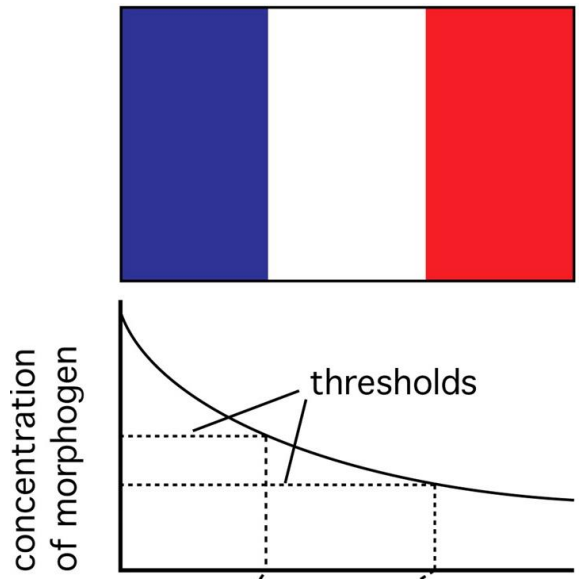


Fig. 1.1. The French flag model based on Wolpert(Wolpert 1968). A morphogen gradient provides positional information encoded by the local concentration, which can be read by cells. At distinct threshold concentrations cells respond with differential gene activity. The tricolore exemplifies the cell fate pattern resulting from the morphogen read-out. The figure is adapted from (Sohka et al. 2009).

This basic model still holds as a simple representation of how morphogens act. Morphogens are now seen as locally secreted molecules and potent ligands which spread forming concentration gradients that instruct receiving cells of their relative position and function within the tissue (Gallet 2011). For the purposes of this thesis the following definition of a morphogen will be used:

"Morphogens are molecules that form gradients of concentration from a source and activate different target genes at different concentration thresholds." (Lecuit & Le Goff 2007).

1.4 Hedgehog signalling

Hedgehog (HH) signalling is an ancient signalling pathway found across the metazoa (Ingham et al. 2011). The main ligand in the signalling pathway was discovered by

(Nüsslein-Volhard & Wieschaus 1980) who performed a genetic screen for mutations that disrupt the *Drosophila* larval body. Indeed, the HH ligand derives its name from the spiked 'hedgehog-like' appearance of *Drosophila* denticles projecting from imaginal discs when mutant HH was present (Nüsslein-Volhard & Wieschaus 1980). HH was subsequently discovered to be present in mammals following a multi-species collaboration between three groups studying mouse, chick and fish (Echelard et al. 1993; Riddle et al. 1993; Krauss et al. 1993). Unlike the fly, which possesses only one HH gene (fHH), vertebrates have several related genes. Three HH genes were identified in the mouse: Desert hedgehog (DHH), Indian hedgehog (IHH), and Sonic hedgehog (SHH) (Echelard et al. 1993). DHH is the most closely related to *Drosophila* hedgehog; IHH and SHH are more closely related to one another, representing a more recent gene duplication event (**Fig. 1.2**).

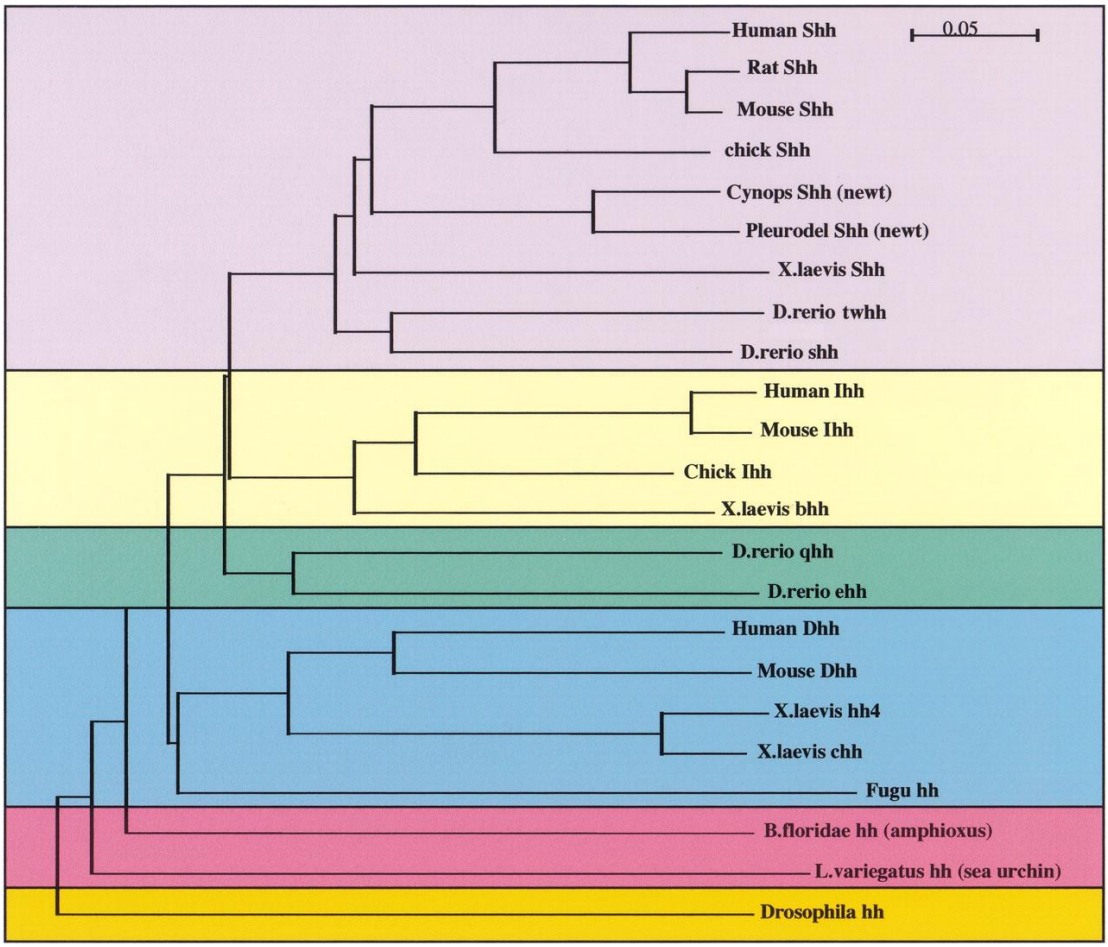


Fig 1.2. Phylogenetic relationship of members of the HH protein family from different vertebrate species, *Cephalochordates* and *Drosophila*. Amniote HH falls into three distinct subgroups: Sonic (lilac), Indian (yellow), and Desert (blue). The zebrafish TWHH (now known as SHHb) is a divergence of the Sonic subgroup. Zebrafish EHH(Currie & Ingham 1996) and QHH(Ingham & McMahon 2001) are divergent members of the Indian subgroup (now known as IHHa and IHHb respectively) (green). Figure adapted from (Ingham & McMahon 2001).

The HH protein fold is highly conserved in all kingdoms (**Fig 1.3**), although there are some differences. In vertebrates, the HH N-terminal signalling domain harbours a zinc binding site (identified in human DHH, SHH and IHH (Echelard et al. 1993)), whereas this site is absent in *Drosophila* HH (**Fig. 1.3A, 1.4 and 1.5A+C**). The lack of the zinc ion in *Drosophila* will be discussed further in one results chapter (Chapter 4) and the discussion (Chapter 7).

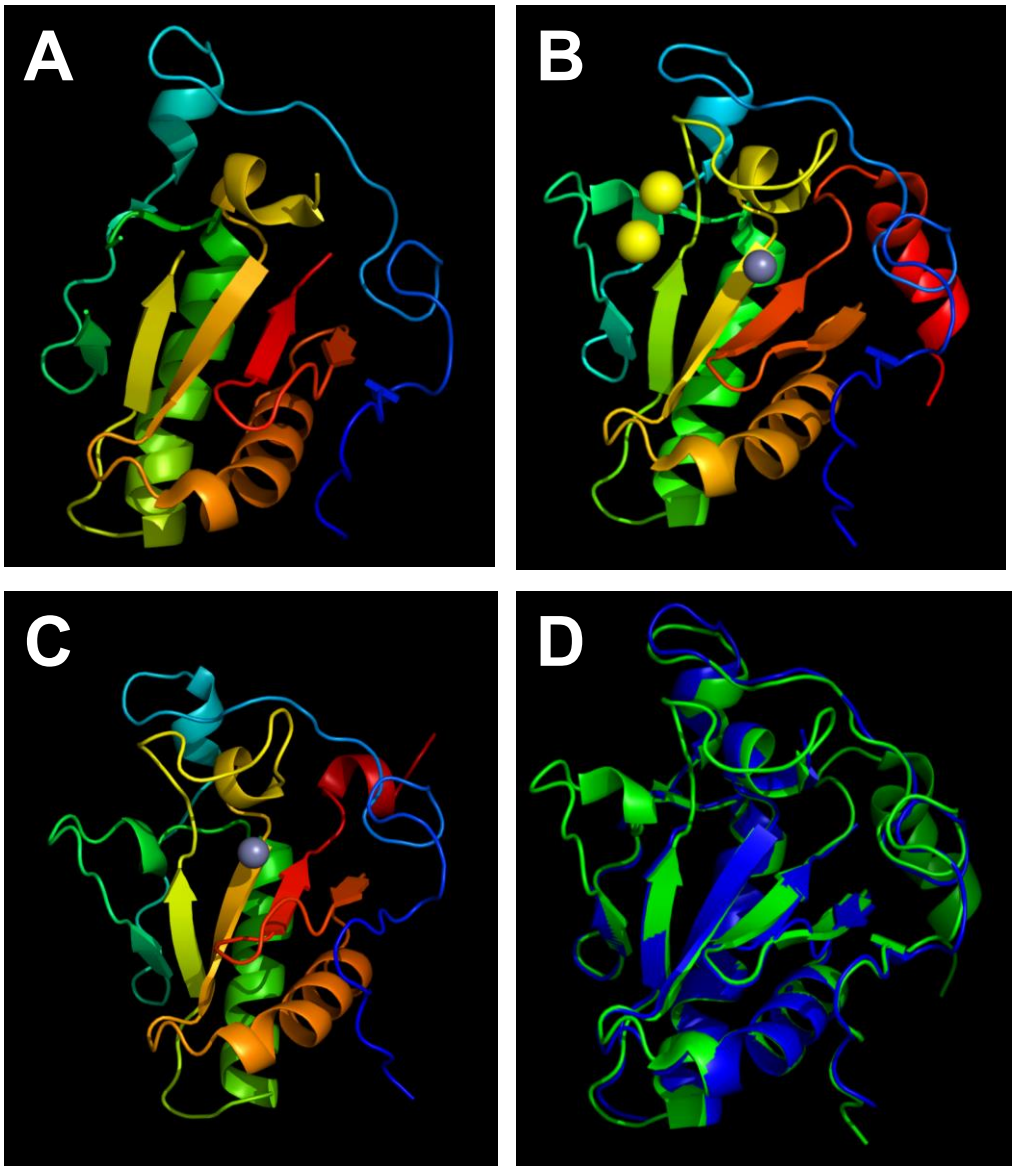


Fig 1.3. Crystal structure of the N-terminal signalling domain of HH. Cartoon representation of **(A)** *Drosophila* (fly) HH (fHH) (PDB ID: 2IBG) (McLellan et al. 2008), **(B)** human DHH (PDB ID: 2WFR) (Bishop et al. 2009) and **(C)** murine SHH (PDB ID: 1VHH) (Hall et al. 1995). Colouring is rainbow from N-terminal (blue) to C-terminal (red). Zinc (grey) and calcium (yellow) ions are depicted as spheres. (**D**) Superposition of human DHH (green) and *Drosophila* HH (blue). Figure was made with PyMOL (www.pymol.org).

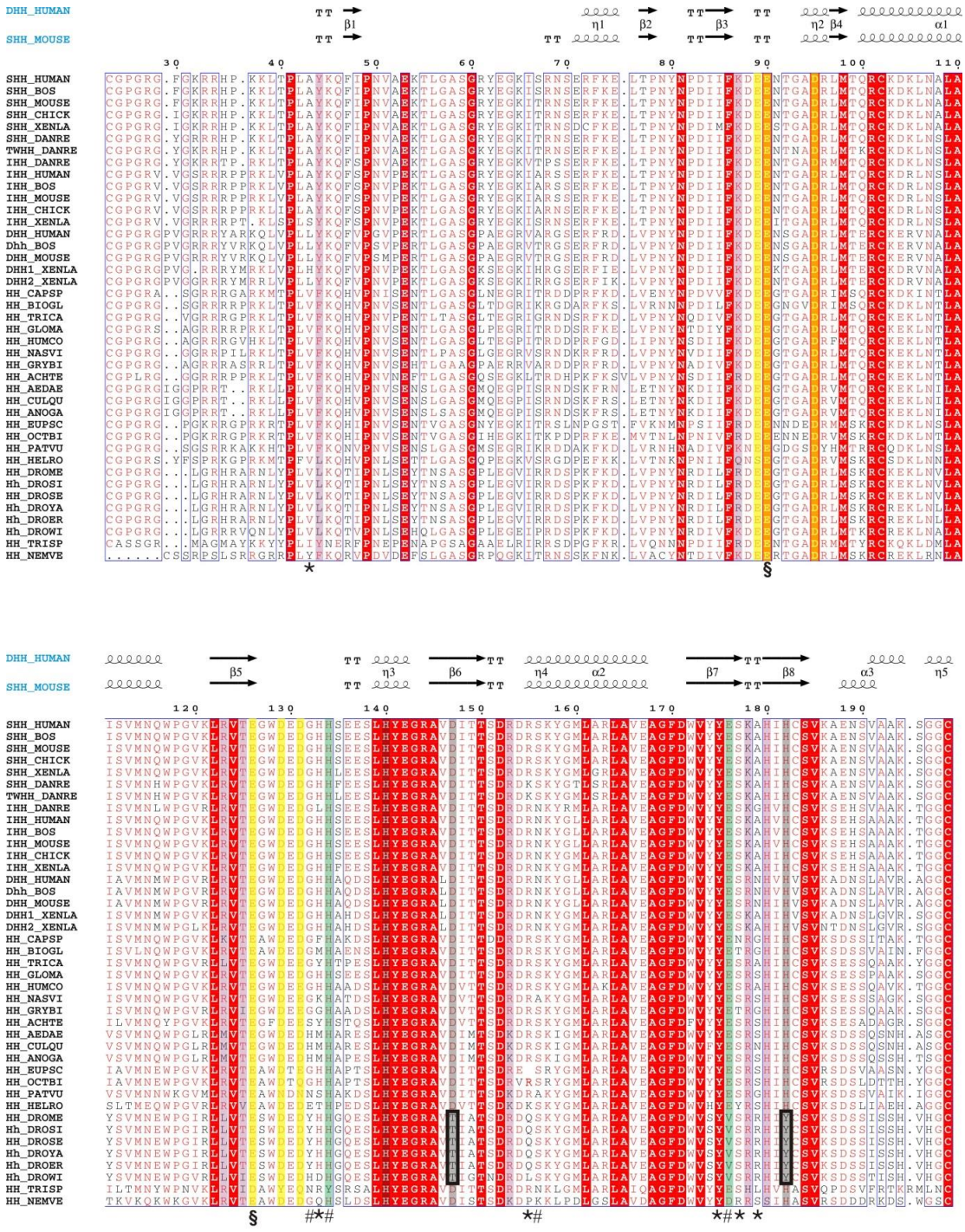


Fig 1.4. Sequence analysis of the N-terminal signalling domains of Hedgehog family members. The HH sequences were aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPript/ESPript/). Numbering is that of the full length human Sonic hedgehog. Secondary structure assignments for mouse SHHN (pdb accession code 1VHH) and human DHHN are displayed above the alignment. Residues forming hydrophilic interactions with HHIP are highlighted in violet and exclusively formed by SHHN in green. Zinc coordinating residues are in grey. Calcium-coordinating residues are in yellow. Residues forming hydophobic interactions with HHIP are marked below the sequences with an asterisk (*), exclusively formed by DHHN with a hash (#) and exclusively formed by SHHN with a section sign (§). The black boxes highlight sequence variations of the Drosophila family in the otherwise highly conserved zinc-binding site residues. Hh proteins only appear in cnidarians, sporadically in other invertebrate phyla (annelida, mollusca and nematoda) and in all arthropod species for which we could access sequence data. Vertebrate Hh

proteins are typically encoded by three genes: Sonic HH, Indian HH and Desert HH (plus Tiggy-winkle, a fourth gene in zebrafish). The abbreviations used are: BOS: Bos tauris; XENLA: Xenopus laevis; DANRE: Danio rerio; CAPSP: Capitella sp.; BIOGL: Biomphalaria glabrata; TRICA: Tribolium castaneum; GLOMA: Glomeris marginata; PEDHU: Pediculus humanus corporis; NASVI: Nasonia vitripennis; GRYBI: Gryllus bimaculatus; ACHTE: Achaearanea tepidariorum; AEDAE: Aedes aegypti; CULQU: Culex quinquefasciatus; ANOGA: Anopheles gambiae; EUPSC: Euprymna scolopes; OCTBI: Octopus bimaculoides; PATVU: Patella vulgata; HELRO: Helobdella robusta; DROME: Drosophila melanogaster; DROSI: Drosophila simulans; DROSE: Drosophila sechellia; DROYA: Drosophila yakuba; DROER: Drosophila erecta; DROWI: Drosophila willistoni; TRISP: Trichinella spiralis; NEMVE: Nematostella vectensis.

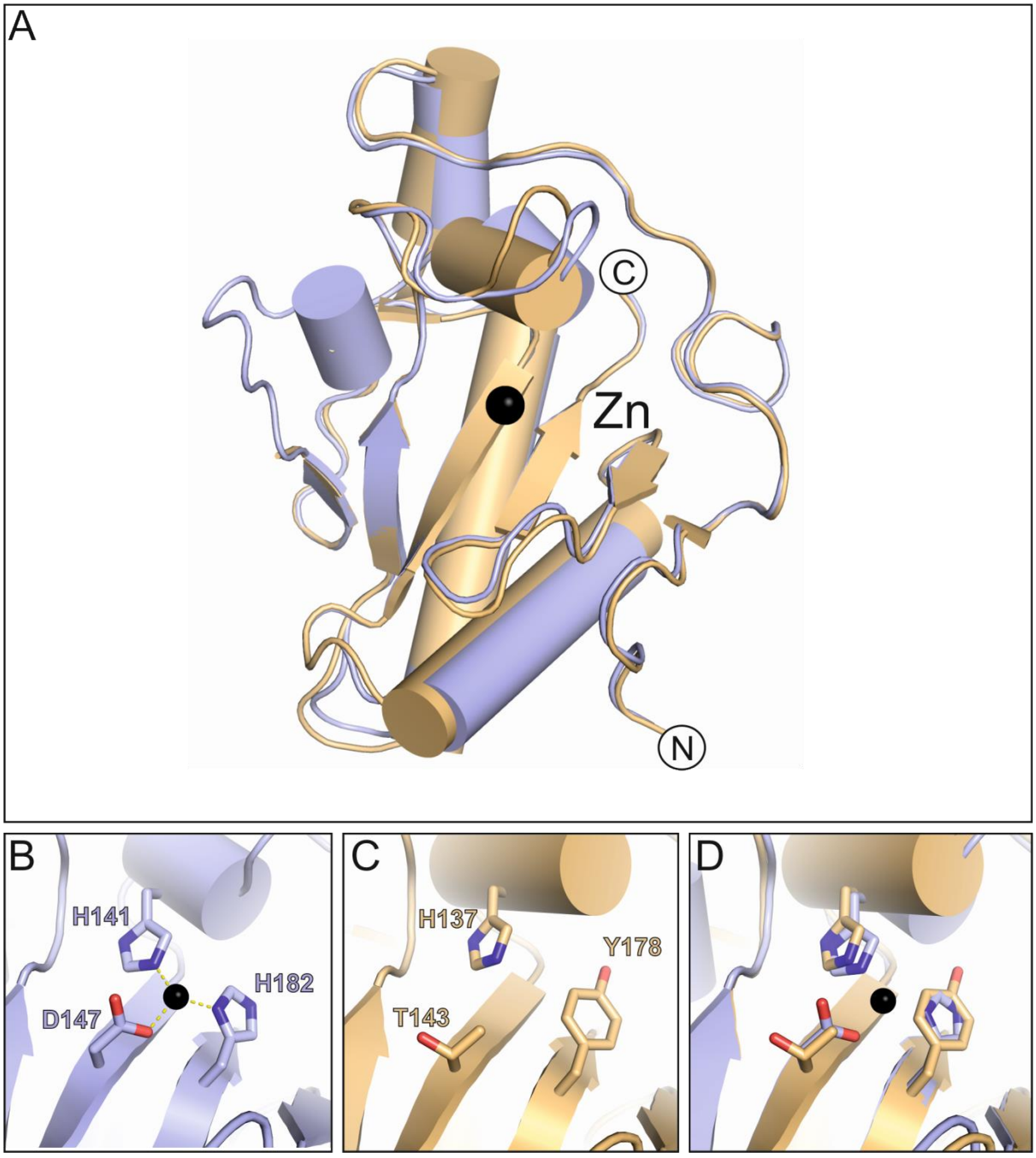


Fig 1.5. Structure of mammalian and *Drosophila* HH. Cartoon representation of (A) murine SHH (purple) and *Drosophila* HH (yellow), Zn^{2+} is represented by a black sphere. (B) mSHH with residues coordinating the Zn^{2+} ion marked. (C) *Drosophila* HH with residues corresponding to those in mSHH that coordinate the Zn^{2+}. (D) mSHH and *Drosophila* HH overlaid.

1.4.1 Hedgehog processing and release

HH proteins are synthesised as 45 kDa precursor proteins following signal sequence cleavage and entry into the secretory pathway. This precursor is palmitoylated at the

N-terminus (Pepinsky et al. 1998), which is promoted by an acyl transferase encoded in *Drosophila* by the skinny hedgehog (*SKI*) gene (Chamoun et al. 2001) and in vertebrates by its orthologue, *HHAT* (M.-H. Chen et al. 2004). In the next step, the C-terminal HH "Hog" domain is responsible for the autocatalytic cleavage of the HH ligand resulting in an attachment of a cholesterol molecule at the Carboxy-terminal end of the N-terminal domain. This dually lipid-modified N-terminal signalling domain of HH (HHN) is responsible for all known signalling functions (J. J. Lee et al. 1994; Porter, Young, et al. 1996b; Porter, Ekker, et al. 1996a) (**Fig. 1.3**).

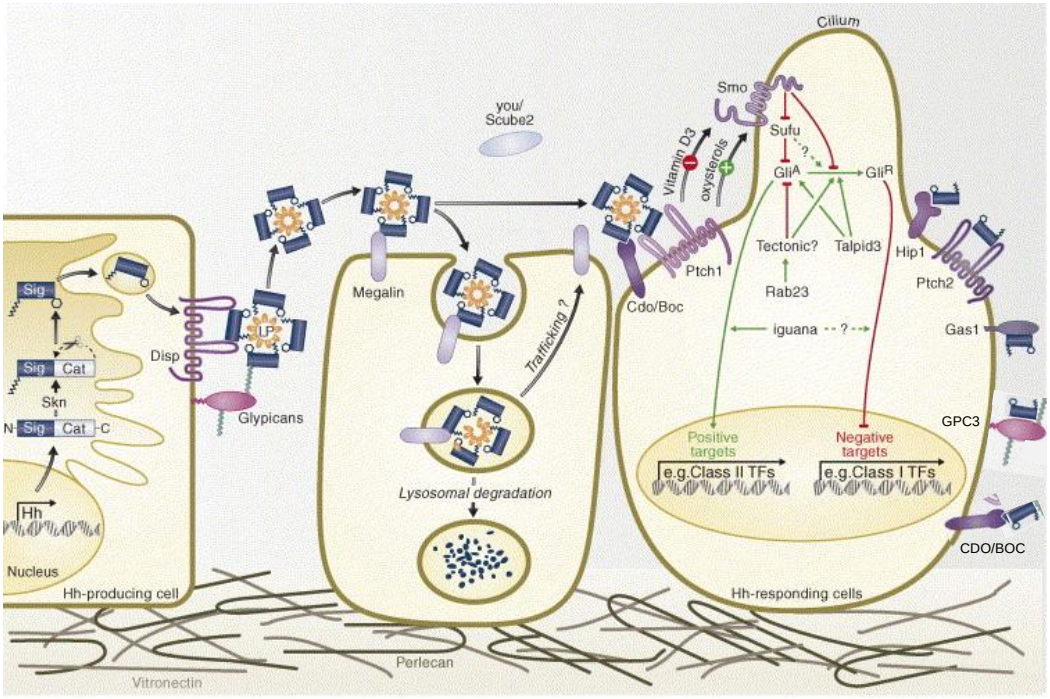


Fig 1.6. Overview of Hedgehog signalling. The 19 kDa, dually lipid-modified, N-terminal HH signalling protein is produced in the HH-producing cell. HH proteins are then trafficked to the cell surface and released from cells as lipoprotein (LP)-associated oligomers in a process mediated by the 12-pass-transmembrane protein Dispatched (DISP). These HH oligomers travel from HH-producing cells to HH-responding cells via interactions with both glypicans and MEGALIN, the latter of which also regulates HH protein turnover. For simplicity, in this model HH is depicted as trafficking only across a single cell via the actions of MEGALIN, while, *in vivo*, HH is capable of travelling over many cell diameters by a process that is probably regulated by multiple molecules and mechanisms. On responding cells, HH interacts with a complex of PTCH and CDO/BOC, resulting in de-repression of SMO and activation of the HH signalling cascade. SMO activates downstream pathways that culminate in activation of GLI activators and repression of GLI repressors. SMO, GLI activators and SUFU all localise to the tips of cilia. SUFU plays an essential role as a negative regulator between SMO and GLI. Adapted from (Y. Wang et al. 2007).

Despite these lipid modifications and the resulting membrane association, HHN can directly act on distant cells. This secretion requires the involvement of a number of different membrane proteins, for example Dispatched and Heparan sulphate proteoglycans (HSPGs) for release from secreting cells (Burke et al. 1999; Yan & X. Lin 2009). The released HHN morphogen appears to be multivalent and part of a bigger lipoprotein complex (M.-H. Chen et al. 2004; Panáková et al. 2005). Once released, the HHN morphogen is then able to signal over a range of cell distances. For long-range signalling, the protein SCUBE-2 is essential (J.-L. F. A. Johnson et al.

2012). MEGALIN (LRP2) is also involved in trafficking of HHN, binding to and endocytosing the HH multimer (McCarthy et al. 2002; Morales et al. 2006; Christ et al. 2012), thus regulating its availability for its canonical receptor, Patched (PTCH) (**Fig. 1.6**).

Interestingly, once HH has left the producing cell, it is found primarily in vesicular structures in the target cells rather than at the cell surface or in the extracellular space (Hooper & Scott 2005). Although this has led to speculation that HH traffics through tissues via transcytosis, the endocytic motor dynamin is not required for movement of HH through the tissue (C. Han et al. 2004). It is therefore more likely that HH traffics via the extracellular space and is rapidly internalised on reception by the HH responsive cells.

The extracellular distribution of oligomeric lipid-modified HHN is dependent on the glypican Dally-Like Protein (DLP) in *Drosophila* and is positively regulated by its mammalian homologue, Glypican-5 (GPC5) (F. Li et al. 2011a). Distribution of the lipid-modified HHN is, in particular, dependent upon the heparan sulphate side chains of these glypicans as shown by the importance of the glycosyltransferase *tout velu* (the EXT protein family in vertebrates), which adds heparan sulphate chains to DLP. In *Drosophila*, Shifted (SHF), which displays sequence similarity to WNT inhibitory factor (WIF), associates with both HH and HSPGs (Thérond 2012). Its loss decreases HH stability and movement as it binds both HH and DLP, providing a higher affinity docking site for the lipid-modified HHN oligomer (Hooper & Scott 2005). In the absence of Shifted (SHF), HH becomes dispersed or degraded and is no longer available for signalling (Hooper & Scott 2005).

1.4.2 PTCH and SMO, the core components of HH signal transduction

The HH receptor on HH responsive cells is the twelve-pass integral membrane protein, PTCH. PTCH constitutively suppresses the seven-pass transmembrane protein Smoothened (SMO), which is responsible for all known downstream HH signalling events (Ingham & McMahon 2001) (**Fig. 1.6**). On binding of HH to PTCH, the suppression of SMO by PTCH is released and signalling occurs. A quantitative study of PTCH activation indicates that PTCH functions as a trimer, in which all three subunits require HH binding before signalling is initiated (Casali & Struhl 2004). PTCH can repress SMO at a stoichiometry of up to 1:50(Taipale et al. 2002), suggesting that suppression of SMO by PTCH is not a direct interaction.

Although many aspects of *Drosophila* HH signalling are conserved in vertebrates, vertebrate HH signal transduction differs in its requirement for the primary cilium. Vertebrates have two types of cilia; motile and non-motile. Inside cilia and flagella is a microtubule-based cytoskeleton called the axoneme. The axoneme of primary cilia typically has a ring of nine outer microtubule doublets (called a 9+0 axoneme), while the axoneme of a motile cilium has two central microtubule singlets in addition to the nine outer doublets (called a 9+2 axoneme). Primary cilia are slim, non-motile structures that project from the surface of nearly all vertebrate cells but are strikingly absent in most *Drosophila* (Goetz & Anderson 2010; Ryan & Chiang 2012). Based on genetic studies that associated HH signalling with cilia, several groups have shown that all key components of the HH pathway are enriched in primary cilia (Goetz & Anderson 2010). PTCH and SMO show dynamic, HH-dependent trafficking in primary cilia (Rohatgi et al. 2007; Corbit et al. 2005). In the absence of HH ligand,

PTCH is localised to the base of the primary cilium and SMO is not associated with primary cilia, while on exposure to HH, PTCH exits the cilium and SMO moves in (Rohatgi et al. 2007; Corbit et al. 2005). Activating mutations in SMO, as well as pathway agonists, cause SMO to localise constitutively to the cilium (Corbit et al. 2005). Despite the enrichment of SMO in the cilium in response to SHH, in the absence of ligand, SMO also accumulates in cilia in cells that lack the dynein retrograde motor. This suggests that SMO traffics through the cilium in the absence of ligand, and that SHH increases ciliary accumulation of SMO (Goetz & Anderson 2010).

PTCH can bind all mammalian HH ligands (SHH, DHH and IHH) with nano molar affinity (Marigo et al. 1996). Once PTCH binds HH, the PTCH-HH complex localises in the lipid rafts away from the primary cilium and no longer suppresses SMO. The PTCH-HH complex is internalised and targeted to lysosomes where degradation of both PTCH and HH takes place (Y. Chen & Struhl 1996). Neither endocytosis nor degradation of PTCH is required for signalling to be initiated (Torroja et al. 2004), suggesting some alteration in PTCH activity in addition to localisation and concentration. PTCH is a multipass transmembrane protein with predicted structural similarities to the resistance-nodulation division (RND) family of prokaryotic permeases (Taipale et al. 2002). These proteins confer multiple drug resistance on bacteria by pumping out toxins. PTCH also has a sterol-sensing domain (SSD), which can directly bind cholesterol (Ohgami et al. 2004; Radhakrishnan et al. 2004) and alters protein trafficking in response to lipid binding. Both the RND-like structure and the SSD are important for PTCH function, as mutations of conserved residues in either render PTCH incapable of inhibiting SMO (Taipale et al. 2002; Martín et al.

2001; Strutt et al. 2001; R. L. Johnson et al. 2002). PTCH induces the secretion of pro-vitamin D3 and vitamin D3 from cells, both of which can inhibit SMO directly (Bijlsma et al. 2006). Cholesterols and oxysterols, which lie downstream of pro-vitamin D3 in the cholesterol biosynthetic pathway, activate SMO (Corcoran & Scott 2006). It has been speculated that PTCH acts to transport pro-vitamin D3 to inhibit SMO activity while removing oxysterols which have been shown to activate SMO (Ingham 2008); a theory that is supported by the susceptibility of SMO to modulation by small molecules such as the steroidal alkaloid cyclopamine (Cooper et al. 1998; Taipale et al. 2000; J. K. Chen et al. 2002; Ryan & Chiang 2012).

SMO is a member of the seven-pass transmembrane protein superfamily, consisting of a cysteine-rich N-terminal extracellular domain (CRD), and seven transmembrane domains which are highly conserved across phyla and show substantial similarity to the Frizzled (FZ) family of WNT receptors. In FZ proteins, the CRD binds WNT proteins, but the CRD of SMO binds neither WNT nor HH proteins (Nusse 2003). SMO responds to small molecules, such as oxysterols (Nachtergaele et al. 2012). PTCH had been implicated in pumping these into cells which activates SMO (Corcoran & Scott 2006). Although this gives a potential model for SMO activation whereby the concentration of pro-vitamin D3 and oxysterols are finely balanced, in the absence, or removal of PTCH, SMO is still able to localise at the cell surface even in the absence of HH. Conversely, SMO is inactive at the surface when both PTCH and SMO are retained at the surface by blocking endocytosis (Zhu et al. 2003). This suggests indirect modulation of SMO rather than a direct interaction with PTCH.

1.4.3 Extracellular HH co-receptors and modulators

A relatively small change in HH concentration (approximately twofold) results in the specification of distinct cell types (Ericson et al. 1997). Such fine-tuning of the HH signal would necessitate multiple levels of regulation to ensure a cell's position and fate were correctly determined. To this end, a range of regulatory cell-surface proteins are utilised to help control the concentration of HH to which responding cells are exposed.

In contrast to mammalian HH signalling, direct binary interaction of *Drosophila* HH to PTCH could not be detected (Zheng et al. 2010). Moreover, *Drosophila* HH seems to form a multimeric complex with PTC and the IHOG/BOI family of single transmembrane spanning adhesion-like receptors. IHOG/BOI is essential for HH biological responses in *Drosophila* and both IHOG/BOI and PTC are required for high-affinity HH binding (Yao et al. 2006). CDO and BOC, the closest vertebrate homologues of the IHOG/BOI family, have also been identified as positive regulators of HH signalling (Yao et al. 2006; W. Zhang et al. 2006; Tenzen et al. 2006)(**Fig. 1.6**). Both proteins can directly bind to all three human HH ligands, but, in contrast to their fly orthologues, were suggested to compete with PTCH for binding to SHH, rather than forming a three-component complex (McLellan et al. 2006; McLellan et al. 2008; Kavran et al. 2010). Thus the mechanism for their agonistic effect remains obscure.

HSPGs form an additional group of extracellular modulators of HH signalling characterised in both vertebrates and invertebrates (Yan & X. Lin 2009). This is not

surprising given the fact that HH ligands can bind to different types of glycosamino-glycans (GAGs) (e.g. heparan sulphate) and that these interactions affect their morphogen signalling range (F. Zhang et al. 2007; J. J. Lee et al. 1994; Ohlig et al. 2011) Depending on their tissue distribution, HSPGs can act as both positive and negative regulators (Desbordes & Sanson 2003; Lum et al. 2003; C. Han et al. 2004; Beckett et al. 2008; Capurro et al. 2008). In addition to the HH-GAG binding properties, some HSPGs, the glypican (GPC) family, have also been implicated in direct protein-protein interaction with the HH ligand (Capurro et al. 2008; Williams et al. 2010; M.-S. Kim et al. 2011). Further details of the glypicans will be given in Chapter 5.

Vertebrates, unlike invertebrates, express two extra high-affinity, membrane-attached HH receptors, the hedgehog-interacting protein (HHIP) and the Growth Arrest Specific 1 (GAS1) (**Fig. 1.6**). GAS1 is a GPI-anchored protein of unknown structure and has been shown to be an essential positive regulator of vertebrate HH signalling (Izzi et al. 2011; Allen et al. 2011; Martinelli & Fan 2007a; Kang et al. 2007), potentially synergizing with PTCH and the CDO/BOC receptors (Bae et al. 2011). In contrast, HHIP is, similar to PTCH, an inhibitor of HH signalling and cells exposed to HH up-regulate HHIP expression, suggesting a mechanism of feed-back regulation (Chuang & McMahon 1999) (**Fig. 1.6**). Decreased expression of human HHIP has been noted in several tumour types, suggesting a potential role for HHIP in tumour suppression (Olsen et al. 2004). HHIP is a multi-domain type I membrane protein with a predicted N-terminal Frizzled-like domain, followed by a β-propeller domain and two EGF repeats, and a C-terminal hydrophobic helix, required for membrane

attachment. More details of the vertebrate specific modulators HHIP and GAS1 can be found in Chapters 4 and 6 respectively.

1.4.4 Intracellular HH signal transduction

Once HH has bound PTCH, and repression of SMO has been released, intracellular pathways are activated. Although there are similarities between *Drosophila* and vertebrates, there are some major differences, mainly due to the requirement of the primary cilium in vertebrates (**Fig. 1.7**).

In *Drosophila,* HH reception by PTCH in conjunction with IHOG, BOI and DLP prevents PTCH repression of SMO, which is found mainly in intracellular vesicles (Ingham et al. 2011). SMO translocates to the plasma membrane where it is phosphorylated by PKA, GSK3 and CKI. This phosphorylation causes a conformational change in the C-terminal domain of SMO, enhancing its interaction with COS2 (**Fig. 1.6**). Phosphorylation of COS2 by Fused (FU) causes dimerisation of FU, leading to its activation, resulting in the release of CI from the HH signalling complex, abrogating its phosphorylation and so promoting the accumulation of its full-length form. FU-dependent phosphorylation of Suppressor of fused (SUFU) promotes its dissociation from CI-FL, allowing CI-FL to translocate to the nucleus where it undergoes further modification to its activated form (CI-A) and thus promotes the transcriptional activation of HH target genes (**Fig. 1.7**).

In mammalian cells, exposure to HH causes SMO to localise to the axoneme of the primary cilium (Rohatgi et al. 2007), quickly followed by the accumulation of the

SUFU-GLI complex at the cilium tip (Haycraft et al. 2005; Endoh-Yamagami et al. 2009) (**Fig. 1.7**). This translocation appears to facilitate the dissociation of the complex in response to SMO activity (Tukachinsky et al. 2010). Cleavage of GLI proteins is prevented and their full-length forms translocate into the nucleus where they are converted into transcriptional activators (Humke et al. 2010) (**Fig. 1.7**). How this dissociation is effected by SMO and how the GLI proteins move from the cilium to the nucleus is not fully understood (Ingham et al. 2011).

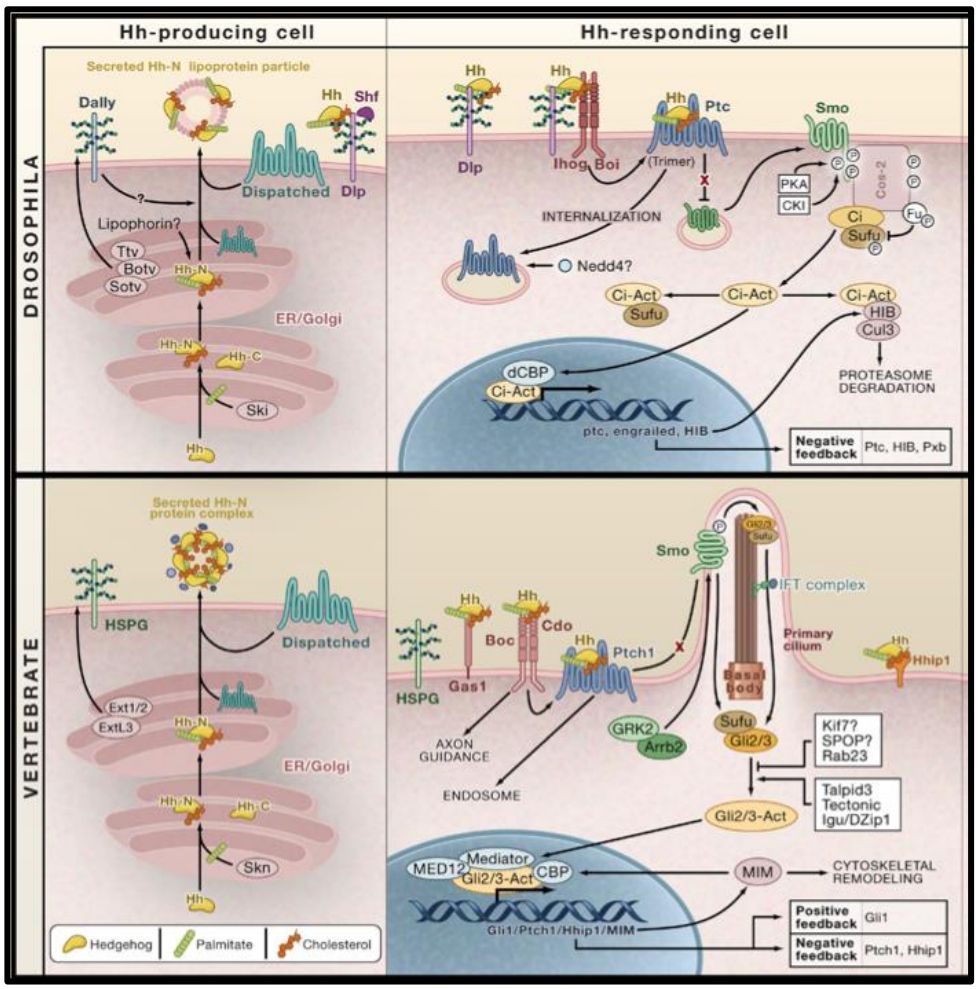


Fig. 1.7 Comparison of invertebrate and vertebrate HH signalling. Schematic of the processes involved in HH release by producing cells (left-hand panels), reception by HH responding cells (right-hand panels) in invertebrates (top panels) and vertebrates (bottom panels). Adapted from (M.-H. Chen et al. 2007a).

1.5 Vertebrate HH signalling in development

While DHH is involved in the development of male germ cells, IHH is an important regulator in long-bone growth and cartilage development (St-Jacques et al. 1999). SHH, however, performs the largest known range of biological actions including establishing left-right body asymmetry, central nervous system development, somite patterning, eye development and limb patterning (Ingham & McMahon 2001).

Ectopic expression studies in frog (Roelink et al. 1994), fish (Krauss et al. 1993), mouse (Echelard et al. 1993), and chick (Echelard et al. 1993) bolstered the initial evidence that SHH regulates dorsal-ventral polarity in the neural tube and anterior-posterior polarity in the limb. Dorsal expression of SHH in the neural tube resulted in the induction of ventral marker gene expression (Ingham & McMahon 2001). Anterior expression of SHH in the limb mesenchyme produced a mirror image duplication of limb pattern resembling those produced by anterior grafts of ZPA cells (Riddle et al. 1993; D. T. Chang et al. 1994). The role of SHH in patterning the ventral neural tube extends into regions of the brain where several cell identities have been shown to depend on ventrally-derived SHH for their induction (Ingham & McMahon 2001).

There are few parts of the vertebrate body plan that are not, in some way, influenced by HH signalling (Ingham & McMahon 2001). HH signals regulate cell fate specification, cell proliferation and cell survival in different target cells. This signalling can be either short or long-range, direct or indirect, and importantly, concentration-dependent, evoking distinct molecular responses at discrete concentration threshold (Ingham & McMahon 2001).

1.6 HH and disease

The HH signalling pathway has been implicated in various pathologies, ranging from developmental defects to cancer. These diseases can be caused by defects to HH ligand (SHH or IHH), to HH receptors (e.g. PTCH, HHIP or GAS1), to HH downstream signalling molecules (e.g. SMO, SUFU or GLIs) or indeed to the development and maintenance of the cilia and intracellular signalling components.

1.6.1 Developmental defects

Holoprosencephaly (HPE, OMIM 236100) is the most common developmental forebrain anomaly in humans, with an incidence in live-borns of approximately 1 in 8,000 (Leoncini et al. 2008; Muenke & Beachy 2000), and in embryos as high as 1 in 250 (Matsunaga & Shiota 1977). HPE results from incomplete cleavage of the prosencephalon into hemispheres. HPE varies in severity: in alobar HPE, the hemispheres do not divide and a single median ventricle is present; in semilobar HPE, the hemispheres and ventricles partially separate posteriorly; in lobar HPE, the ventricles are separated and there are well-formed lobes that may be of normal size, with some continuity across the frontal cortex; and in middle inter-hemispheric fusion (MIHF), the frontal and parietal lobes fail to separate posteriorly and the corpus callosum is absent (DEMYER & ZEMAN 1963; Dubourg et al. 2007; Muenke & Beachy 2000). Facial features associated with HPE include cyclopia or anophthalmia

(congenital absence of one or both eyes) with proboscis, ethmocephaly (extreme hypotelorism - closeness of eyes with proboscis), a single-nostril nose, mid-face hypoplasia with cleft lip and/or palate, hypotelorism, and a single central incisor (Pineda-Alvarez et al. 2012). *SHH* is the major gene implicated in HPE (12.7% of HPE cases: 50% of overall point mutations and 38% of overall large deletions) (Dubourg et al. 2004; Bendavid et al. 2006). Several human HPE mutations have been identified recently in the *DISP* gene (Ma et al. 2002), while mouse embryos deficient in MEGALIN exhibit fused cerebral hemispheres (Willnow et al. 1996). Mice lacking CDO display HPE (W. Zhang et al. 2006) as do GAS1-/- mice, which exhibit microform HPE (Seppala et al. 2007). Animal models showing enhanced PTCH activity have an HPE phenotype, which is further supported by the identification of *PTCH* mutations in patients presenting HPE (Ming et al. 2002). Mutations in the *GLI2* gene are also involved in a distinctive phenotype within the HPE spectrum (Roessler et al. 2003).

Mutations disrupting basal body proteins of the primary cilia, such as oral-facial-digital syndrome 1 (OFD1), fantom (FTM), Meckel syndrome type 1 (Mks1) and Ellis–van Creveld syndrome protein (EVC), cause human ciliopathies and affect mammalian HH signalling (Goetz & Anderson 2010). Mice mutant for OFD1 (Ferrante et al. 2006), FTM (Vierkotten et al. 2007) or Mks1 (Weatherbee et al. 2009) have abnormal or absent cilia and display HH signalling defects corresponding to the severity of the cilia disruption. EVC also localises to the basal body and is mutated in the human skeletal disorder, Ellis–van Creveld syndrome (Ruiz-Perez et al. 2000). Expression of EVC in mice is limited to developing skeletal structures, and EVC does not seem to be required for the formation of cilia in chondrocytes. Nevertheless,

EVC-/- mice show reduced IHH signalling specifically within skeletal structures. Therefore EVC does not apparently affect ciliogenesis but is required for HH signalling in a specific cell type (Goetz & Anderson 2010).

1.6.2 Cancer

HH signalling is inextricably linked to cancer. Indeed, the intracellular transcription factor, GLI, was named after its discovery in gliomas. An aberrant HH signalling pathway can arbitrate initiation, proliferation and propagation of many different types of cancers through constitutive activation of the pathway driven either in a ligand-dependent or a ligand-independent manner. This irregular expression of the pathway may be attributed either to the up-regulation of HH ligands (SHH, IHH or DHH) and pathway components such as PTCH, SMO, GLI, SUFU, or to genetic and epigenetic modifications in the pathway (Teglund & Toftgård 2010). Inappropriate activation of SHH signalling can cause medulloblastomas and rhabdomyosarcomas (paediatric tumours of the cerebellum and muscle, respectively), and is found in all cases of basal cell carcinoma (Goetz & Anderson 2010). In addition, growing evidence indicates that SHH signals promote the growth of other types of tumours (Teglund & Toftgård 2010). Recent studies show that the primary cilium regulates HH signalling in tumours. The role of these cilia in tumours depends on how the pathway is activated. Expression of activated SMO in the postnatal mouse brain can cause medulloblastomas, but removal of cilia prevents tumour formation (Y.-G. Han et al. 2009), consistent with earlier genetic experiments indicating that cilia are required for the activity of the pathway at a step downstream of SMO (Goetz & Anderson 2010).

Tumours with ligand-independent HH pathway activation have mutations in HH pathway components that confer cell-intrinsic growth promoting and/or survival properties to the tumours. Most commonly found are inactivating mutations. These include deletions, mRNA splice-site, and nonsense mutations in PTCH or other negative regulators of HH signalling, or activating missense mutations in SMO that result in ligand-independent constitutive HH pathway activation. Other loss-of-function mutations likely to result in cell-autonomous HH pathway activity have been described in *SUFU* and *REN*, a gene involved in proliferation and differentiation of neural progenitor cells (Teglund & Toftgård 2010). Ligand-independent mechanisms bestow cell-intrinsic growth promoting and survival properties to the tumour cells as is seen in the case of Basal Cell Carcinoma and Medulloblastoma (Kar et al. 2012). Ligand-dependent cancers require the action of HH to develop or be maintained. In normal HH signalling, producing cells release HH, which then acts on responding cells. In terms of cancer, ligand-dependent cancers can be either autocrine, where the HH ligand produced by tumour cells act on the tumour cells themselves to stimulate their propagation and survival, or they may involve a paracrine mechanism where HH ligand secreted from the epithelium signals the underlying stromal compartment for continued proliferation into malignancy. Autocrine signalling is seen in cases of lung cancer, pancreatic cancer and prostate cancer, while in the cases of ovarian and colorectal cancer, the paracrine mechanism is prevalent (Teglund & Toftgård 2010).

1.7 HH signalling and stem cells

There is emerging evidence that the HH signalling pathway may play an important role in the continuous self-renewal of tissues from stem cells which persists into post-natal and adult life (Onishi & Katano 2011). Activation of the SHH pathway in

developing thymocytes influences T-cell receptor repertoire selection and the differentiation from progenitor to effector cells (Rowbotham et al. 2007). SHH is produced by follicular dendritic cells and protects germinal-centre B-cells from apoptosis (Crompton et al. 2007). Addition of recombinant SHH to partially purified mouse and human T-cell populations *in vitro* has been shown to enhance the induction of T-cell activation and proliferation by suboptimal concentrations of CD3-specific antibody and CD28-specific antibody, whereas addition of hedgehog-specific antibody to these cultures reduced T-cell activation and proliferation (Crompton et al. 2007). It has also recently been suggested that apart from developmental diseases and cancer, HH signalling may also be involved in maintenance of adult stem cells and play a role in senescence (Dashti et al. 2012).

1.8 Hedgehog pathway-based therapies

GDC-0449 was the first systemic SMO inhibitor entering clinical trials with Genentech (Robarge et al. 2009). It was successfully tested in a Phase I clinical trial demonstrating good pharmacodynamic and pharmacokinetic properties and showing objective response and clinical benefit in several patients with BCC (Hoff et al. 2009; Low & de Sauvage 2010). In a Phase II study of 41 people with BCC, a team led by Ervin Epstein from the Children's Hospital Oakland Research Institute, California found that participants taking GDC-0449 developed only four new tumours on average over the course of a year, compared with 24 in subjects on placebo(Redmond et al. 2011). Recently, GDC-0449 was approved by the FDA (food and drug administration, USA) granting Genentech a patent for the SMO inhibitor (SAUVAGE 2012).

Another compound recently identified as an inhibitor of SHH signalling is Robotnikinin. Robotnikinin binds SHH and blocks its ability to induce pathway activity at the level of PTCH (Stanton et al. 2009). The SHH antibody 5E1 (Ericson et al. 1996; Maun et al. 2010) is also being developed as a HH signalling inhibitor (Nakamura et al. 2007). The promising initial results of HH pathway inhibitors indicate that HH signalling is a suitable target for cancer therapy (Onishi & Katano 2011).

1.9 Aims

The HH signalling pathway is an important developmental pathway, which also plays a role in stem cell maintenance and disease. Molecular information is needed to decipher these complex mechanisms, and understanding the molecular mechanisms underlying HH signalling could potentially enable the development of better and more specific cancer treatments and even prevent the ravages of aging. In the proceeding chapters, results pertaining to three mammal-specific protein regulators of the HH signalling pathway will be presented. In the first results chapter, a new, automated method of protein production, which was of vital importance to the completion of this thesis, will be described. The automated transient transfection protocols were of particular importance to target difficult proteins and protein complexes. Thereafter, the antagonists of HH signalling, HHIP and GPC3, will be dealt with. In the final results chapter, GAS1, a HH pathway agonist, will be covered. This thesis will attempt to answer the question of how these proteins interact directly with the HH ligand and how they are able to orchestrate such a fundamental pathway. Particular emphasis will be placed on how cell-surface receptors act as either agonists or antagonists of the HH signalling pathway.

To achieve these goals, a strategy utilising a range of techniques has been employed. X-ray crystallography was used to determine the high-resolution molecular structures of single components of the HH signalling pathway, and where applicable, their complexes with HH in conjunction with biophysical binding studies, mutagenesis and functional studies.

Chapter 2

Materials and Methods

Chapter 2 Materials and Methods

2.1 Molecular Biology

2.1.1 cDNA and Expression Vectors

For expression in bacteria, the human desert hedgehog N-terminal signalling domain (termed DHHN, residues 39-194 of DHH, GenBank ID: NP_066382) was cloned into the bacterial expression vector pET22b (Invitrogen), harbouring a C-terminal His6-tag. The mouse sonic hedgehog N-terminal signalling domain (termed SHHN, residues 40-194, GenBank ID: NP_033196), human eHHIP (residues 56-670 of HHIP, Gene Bank ID: NP_071920), human eHHIPS (residues 56-670 with Arg189, Arg210 and Lys211 mutated to alanine), human eHHIPΔN (residues 214-670), GPC3-WT (residues 31-580, GenBank ID: NP_001158089), GPC3-ΔGPI (residues 31-514), GPC3-ΔGAG (residues 31-493), GPC3-Δ1/2GAG (residues 31-506) and human GAS1-WT (residues 40-345, GenBank ID: NP_032112), GAS1-ΔGPI (residues 40-314), GAS1-ΔL1 (residues 40-246 minus residues 80-96), GAS1-ΔL2 (residues 40-246 minus residues 153-162), GAS1-ΔL1L2 (residues 40-246 minus residues 80-96 and 153-162), GAS1-ΔC (residues 40-148), GAS1-ΔCL1 (residues 40-148 minus residues 80-96) and GAS1-ΔN (residues 163-314) were cloned into the mammalian expression vectors pHLsec (Aricescu et al. 2006) for large expression. These vectors have a Kozak sequence, a secretion signal and a C-terminal His6-tag. Proteins used in Surface Plasmon Resonance (SPR) were cloned into vector pHLAvitag3 (Aricescu et al. 2006) introducing a C-terminal biotinylation tag. All constructs were cloned using primers purchased from MWG Operon.

2.1.2 Designing Constructs and PCR

Expression constructs were designed using the tertiary prediction program HHPred (http://hhpred.tuebingen.mpg.de/hhpred) and the disorder prediction server RONN (http://www.oppf.ox.ac.uk/RONN/). PCR reactions were carried out in a final volume of 50 µL including (in the order reagents were added): MilliQ H_2O to give a final volume of 50 µL, 5 µL 10X Pyrobest™ (Takara) Buffer II, 4 µL 2.5 mM dNTP Mixture, <500 ng Template DNA, 0.25 µL Pyrobest™ DNA polymerase (Takara), and a final concentration of 2mM for each primer. PCR cycles were performed using a Tetrad Thermocycler (BIORAD). For a 1 kbp DNA fragment, the following cycle was used:

1. 98 °C for 10 s
2. 98 °C for 10s ⎫
3. 60 °C for 30s ⎬ 30 cycles
4. 72 °C for 60s ⎭
5. 72 °C for 600s
6. 04 °C for ∞

For larger constructs, extension time 4 was increased adding 1 minute for every extra kbp. PCR products were then purified using the QIAquick PCR Purification Kit (Qiagen).

2.1.3 Restriction Digests, Ligations and Transformations

Purified PCR products and 2 µg of the required vector were digested for two hours with the appropriate restriction enzymes using the recommended buffer conditions, concentrations and temperatures (New England Biolabs). Digested PCR products and vector were then loaded on a 1% agarose gel. Relevant bands were excised and

purified using QIAquick Gel Extraction Kit (Qiagen). Ligations were conducted using Quick Ligation™ Kit (New England Biolabs) following the supplier instructions. 1 µL of the ligation mixture was then used to transform 40 µL DH5α library efficiency® cells (Life Technologies™). Transformations were carried out as specified by the supplier. After transfection, cells were spread on LB-Agar plates containing 100 µg/ml ampicillin and incubated overnight at 37 °C. Two colonies for each construct were picked and used to inoculate overnight cultures in order to prepare the plasmid DNA using the QIAprep Spin Miniprep Kit (Qiagen). 1 µg purified plasmid DNA was digested using the appropriate restriction enzymes and analysed by agarose gel electrophoresis. Clones containing an insert of the correct size were then sequence verified (Geneservice, Oxford) and stored at -20 °C.

2.1.4 Site directed mutagenesis

Site-directed mutagenesis for constructs used in eukaryotic expression was carried out following a two-step, overlap-extension PCR. In the first step, two PCR products were synthesized: one consisting of sequences spanning from the mutation primer to the vector-derived 5' terminus primer and one consisting of sequences from the mutation primer to the vector-derived 3' terminus primer. Following the first PCR, the products were separated and purified by gel electrophoresis followed by gel extraction using the Gel Extraction Kit (Qiagen). The reaction mixtures for the second PCR were composed of the combined PCR products from the first PCR step (~200 ng), terminal vector-derived primers and standard Pyrobest (Takara) PCR reagents. This reaction created PCR inserts with the desired mutation that could then be cloned into the appropriate vector.

2.2 Protein Expression and Purification

2.2.1 Vectors and Cell Lines

The vector chosen for development of an automated transfection protocol was pHLsec. This vector was built on the pLEXm backbone (Aricescu et al. 2006) and contains: pBR322 origin of replication giving the high copy number in *E. coli* needed to obtain the high amounts of plasmid DNA required for transient transfection; ampicillin resistance; a cytomegalovirus enhancer; chick β-actin promoter to give high levels of expression; the rabbit β-globin intron to increase RNA production; and a poly-A signal to increase RNA stability. Two cell lines were chosen for the development of automated transient transfection protocols for glycosylated mammalian proteins: HEK 293T (ATCC CRL-11268), grown in the presence of the N-glycosylation inhibitor, kifunensine (V. T. Chang et al. 2007) added immediately post-transfection at 1 mg L^{-1} final concentration, and the N-acetylglucosaminyltransferase I-negative HEK 293S GnTI- cells (Reeves et al. 2002), which are unable to synthesize complex glycans. These cell lines were chosen due to ease of handling, robust growth rates, excellent transfection efficiency, high capacity for recombinant protein expression and low cost media requirements. Both cell lines are also routinely used for manual transient transfections, thus allowing a direct comparison between automated and manual protocols.

2.2.2 DNA Purification

The Endotoxin-Free Plasmid High Speed Giga Kit (Qiagen) was used to purify plasmid DNA. High-quality DNA is essential for successful transfections and only samples with an OD260/280 ratio of 1.8 or higher are suitable. The pLEXm-derived

plasmids are very high copy number, therefore one can expect between 12 and 15 mg of pure DNA from a 2 L overnight bacterial culture (*E. coli* DH5α strain). It is essential that DNA samples are sterile, so care was taken to wash the DNA precipitates with 70% ethanol before dissolving them in sterile 10 mM Tris pH 8.

2.2.3 Transfection Reagent

Polyethylenimine (PEI) is an affordable and highly efficient transfection reagent. The PEI used in both manual and automated protocols is '25 kDa branched' (Sigma-Aldrich), which was reported to be effective in transfecting various HEK293 cell lines (Aricescu et al. 2006; Durocher et al. 2002). Stock solutions were made in water at 100 mg mL-1, the pH adjusted to 7 with HCl, filter sterilized and aliquoted. Aliquots and working solutions of 1 mg mL-1 can be stored frozen at -20 °C for long periods of time (months) without a reduction in transfection efficiency.

2.2.4 Cell Maintenance Protocol

Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM high glucose, Sigma) supplemented with L-glutamine, non-essential amino-acids (Invitrogen) and 10% Foetal Bovine Serum (FBS, Invitrogen). Cells were manually cultured from stocks stored at -196 °C, using a T25 flask (Corning). On reaching 90% confluence, cells were detached using trypsin-EDTA (Sigma), seeded in a T75 flask (Corning or Greiner) and finally transferred to a barcoded T175 flask (Corning or BD-Falcon) and imported into the CompacT SelecT Cellbase robot (The Automation Partnership, Royston, UK), where all subsequent cell maintenance was automated.

2.2.5 Manual Transfection in Roller Bottles

This was typically performed as described in (Aricescu et al. 2006) Half a milligram of plasmid DNA was used for transfection in each 2125 cm^2 expanded surface roller bottle (Greiner). Briefly, the DNA solution was added to 50 mL serum-free DMEM, mixed, supplemented with 1 mL PEI stock (1 mg mL^{-1}) and vortexed for 10 s. This solution was incubated for 10 min at room temperature to allow DNA-PEI complex formation. During complex formation, media from the roller bottles was changed, lowering the serum concentration (200 mL of DMEM, 2% FBS, were added per roller bottle). Finally, the DNA-PEI complex was added to each bottle, briefly rotated to allow mixing, after which the cells were placed in the incubator. Approximately 4-7 days later, conditioned medium was ready for collection and protein purification.

2.2.6 Automated Cell Maintenance and Seeding

Cells were harvested from 90% confluent T175 flasks, analysed by the Cedex automated cell counter and diluted to a concentration of 1.6 x 10^6 cells mL^{-1} in DMEM, 10% FBS. New bar-coded T175 flasks were then called from the CompacT SelecT Cellbase incubator hotel and seeded by pipetting 10^7 cells per flask, supplemented with DMEM, 10% FBS to a final volume of 30 mL. The robotic Stäubli arm then swirled the flask to mix the cells before placing it in the hotel incubator. Cells were passaged twice a week, using a trypsin-EDTA mix (Sigma) for detachment, to a maximum of 20 passages before new aliquots were retrieved from liquid nitrogen storage. Triple flasks (Nunc) were seeded with 2 x 10^7 cells, topped up to 110 mL with DMEM/10% FBS. Ten layer HYPERFlasks (High Yield PERformance Flask, Corning) were seeded with 6 x 10^7 cells, and topped up to 555

mL with DMEM, 10% FBS. Both flask types were returned to the robot incubator for four days, irrespective of the cell line used, prior to transfection.

2.2.7 Automated Transient Transfection Protocol

The transfection cocktail was prepared manually prior to the automated transient transfection of Triple flasks or HYPERFlasks. To transfect 12 Triple flasks, 1.8 mg plasmid DNA was chloroform extracted (to ensure sample sterility) prior to addition of 360 mL serum free DMEM. Mixing was followed by the addition of 3.6 mL PEI stock (1 mg mL^{-1}) with brief swirling followed by 10 min incubation at room temperature to allow DNA-PEI complex formation. The robot can follow only linear protocols, as it has but one arm. The automated protocol for Triple flasks differed from that used manually as the robot cannot work quickly enough to transfect 12 Triple flasks (a typical batch size, providing 1.2 L of conditioned media) in less than the hour before transfection efficiency begins to drop (see Chapter 6). To prevent this problem, it was necessary to stack two programs: the first program adds the transfection cocktail to each of the 12 flasks (30 mL per Triple flask), and when this procedure is completed for all the flasks, a second program is run to top up the flasks to the final volume of 100 mL with DMEM, containing a final FBS concentration of 1.6%. To transfect one HYPERFlask, chloroform extracted DNA was added manually to 100 mL serum-free DMEM, mixed, supplemented with 2 mL PEI stock (1 mg mL^{-1}) and briefly mixed by swirling followed by a 10 min incubation at room temperature to allow DNA-PEI complex formation. Afterwards, the automated transfection protocol was started. Flasks seeded four days prior to transfection were taken from the incubator and media removed by pouring into the waste receptacle. The DNA-PEI complexes were then added to the flasks (100 mL per HYPERFlask), followed by low

serum DMEM (455 mL per HYPERFlask), resulting in a final FBS concentration of 1.6%.

2.2.8 Large-scale purification of secreted protein constructs from 293T cells.

Secreted, His-tagged proteins constructs of HHIP, GAS1 and SHH were purified using an immobilised metal affinity chromatography (IMAC) batch procedure. Conditioned media were filtered through a 0.2μm membrane (Express filter, Millipore) and dialysed against 25 mM phosphate buffered saline (PBS, pH8.0) at 4 °C to prevent metal stripping from the affinity chromatography column. IMAC purification was performed using cobalt-coated Talon beads (Clontech), which were washed with 5.0-7.5 mM imidazole in PBS (pH 8.0), and proteins eluted by addition of 250 mM Imidazole in 10 mM Tris, pH 8.0, 250 mM NaCl. The proteins were further purified by size exclusion chromatography (SEC) on a Superdex200 16/60 (GE Healthcare) column using a buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl.

For GPC3 constructs an additional purification was implemented as the sample was not pure enough after the first purification: after the first IMAC purification, proteins were then dialysed using 10k MWCO 12 mL capacity Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific) to reduce the concentration of Imidazole to 0.25 mM. Dialysis was followed by a second purification step using cobalt-coated Talon beads (Clontech), which were washed with 10 mM imidazole in PBS (pH 8.0). Proteins were eluted in 3 x 1 mL 300mM imidazole in PBS (pH 8.0) and further purified via SEC as described above.

2.2.9 Protein Purification of DHHN from *E. coli*.

DHHN was expressed in *E. Coli* Rosetta(DE3)pLysS cells. Cultures were grown in the presence of 100 μg mL^{-1} ampicillin at 37 ºC to an A_{595} of 0.8, cooled to 20 ºC, induced with 0.25 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) and then allowed to grow for another ~15h before harvesting by centrifugation (6000 x g, 5 ºC, 10 min). Bacterial pellets were resuspended in 25 mM sodium phosphate, pH 8.0, 500 mM NaCl, 1 mM β-mercaptoethanol, supplemented with one EDTA-free protease inhibitor cocktail tablet (Roche) per 50 mL buffer. Cells were broken by two passages in a Basic Z model cell disruptor (Constant Systems) and fractionated by centrifugation (30,000 x g, 5 ºC, 30 min). DHHN was purified from the supernatant by IMAC purification using cobalt-coated Talon beads (Clontech) and followed by SEC (Superdex 200 10/30 column, GE Healthcare) as described above.

2.2.10 Western Blot Analysis

Small aliquots of conditioned media (10 μL) were analysed by Western blotting using the Penta-His monoclonal antibody (1:1000 dilution, Qiagen) and goat anti-mouse IgG peroxidase- conjugated secondary antibody (diluted 1:2000, Sigma). The signal was visualised by chemiluminescence using the ECL kit and hyper-film (GE Healthcare).

2.2.11 Fluorescence Microscopy

Imaging of transfection efficiency using EGFP-pLEXm constructs was carried out with a Nikon Eclipse TE2000U fluorescence inverted microscope equipped with a

Hamamatsu C4742-95 Orca low-noise B&W CCD and using IPLab imaging software to capture and process images.

2.3 Luciferase reporter assays

NIH 3T3 cells (ATCC) were cultured in DMEM supplemented with 10% FBS. On day 0, cells were seeded in 24 well plates at a density of $5x10^4$ cells/well. On day 1, when between 70% and 90% confluent, cells were transfected for 24 hours with Lipofectamine2000 (2 μL/well) according to the manufacturers instructions (without antibiotics). The 8x3'Gli-BSδ51LucII(Sasaki et al. 1997) reporter plasmids were transfected at 0.4 μg/well, while construct plasmids (full length GPC3 and full length HHIP) and *Renilla* internal control plasmid was co-transfected at 0.6 μg/well and 50 ng/well respectively. A total of 1.05 μg plasmid DNA was used per well (made up with empty pHLsec vector for the control wells). 24 hours post-transfection, wells were washed with 500 μL PBS and replaced with 500 μL serum-free DMEM (this step was necessary to remove SHH present in the serum which gave a high background signal if not removed). On day 2, purified murine SHH was added at 50 ng/well and the cells incubated for a further 48 hours. All experiments were performed in quadruplicate. Luciferase assays were performed using the Dual Luciferase Assay Reporter (Promega) system according to the manufacturer's instructions in a Veritas luminometer (Turner Designs). Results were expressed as changes in luciferase activity relative to *Renilla* activity.

2.4 Surface Plasmon Resonance (SPR)

SPR experiments were performed using a Biacore T100 machine (GE Healthcare) at 25 °C in 10 mM HEPES, pH 7.5, 150 mM NaCl, 0.05% (v/v) polysorbate 20, supplemented with 2 mM $CaCl_2$ where indicated. All proteins were homogeneous with full biological activity and underwent gel filtration in running buffer immediately before use. Protein concentrations were determined from the absorbance at 280 nm using calculated molar extinction coefficients. Proteins for surface attachment were enzymatically biotinylated within an engineered C-terminal Avitag (Schatz 1993). The native membrane insertion topology of the receptor molecules GPC3, GAS1 and HHIP was recapitulated by attachment of 200-2,000 response units (RU) of biotinylated protein to surfaces on which 3,000 RU of streptavidin were coupled via primary amines (O'Callaghan C et al. 1999). Experiments with wild-type proteins were performed in both orientations and with mutant proteins in one orientation. The signal from experimental flow cells was corrected by subtraction of a blank and reference signal from a mock or irrelevant protein coupled flow cell. In all experiments analysed, the experimental trace returned to baseline after each injection and the data fitted to a simple 1:1 Langmuir model of binding. K_d values were obtained by nonlinear curve fitting of the Langmuir binding isotherm (bound = C*max/ (K_d + C), where C is analyte concentration and max is the maximum analyte binding) to the data using the Levenberg-Marquardt algorithm implemented in the program Origin (Microcal Software) (Bishop et al. 2009).

2.5 Multi-angle light scattering (MALS)

MALS experiments were carried out using an analytical Superdex S200 10/30 column (GE Heathcare) with online static light-scattering (DAWN HELEOS II, Wyatt Technology), differential refractive index (Optilab rEX, Wyatt Technology) and Agilent 1200 UV (Agilent Techologies) detectors. Proteins used for MALS were fully glycosylated and had previously been purified by size-exclusion chromatography and concentrated to approximately 2 mg mL^{-1}. Data were analysed using the ASTRA software package (Wyatt Technology).

2.6 Protein crystallisation and crystallographic analysis

Crystallisation experiments were set-up using nanoliter crystallisation trials using a Cartesian Technologies robot (100 nL protein solution plus 100 nL reservoir solution) in 96-well Greiner plates(Walter et al. 2005), placed them in a tandem affinity purification (TAP) Homebase storage vault maintained at 295 K and imaged them via a Veeco visualization system (Mayo et al. 2005)

2.6.1 Crystallisation and X-ray analysis of HHIP, DHHN and the HHIP-HH complexes

Prior to crystallisation, proteins purified via SEC (see 2.2.7) were concentrated by ultrafiltration (eHHIPΔN, 4 mg mL^{-1}; eHHIPΔN-SHHN, 8 mg mL^{-1}; eHHIPΔN-DHHN, 10 mg mL^{-1}; DHHN, 7 mg mL^{-1}). Native and SeMet-labeled eHHIPΔN crystallized in 0.1 M Tris-HCl, pH 8.5, 3.5 M potassium formate, the eHHIPΔN-SHHN complex in 0.1 M sodium acetate, pH 4.6, 0.5 M potassium thiocyanate, the eHHIPΔN-DHHN in 1.4 M sodium or potassium phosphate, pH 5.0, and DHHN in 0.1 M Tris-HCl, pH 8.5, 0.2 M lithium sulphate, 0.2 M sodium thiocyanate, 30%

(w/v) PEG4000. For the eHHIPΔN-SHHN- and DHHN- calcium complexes, crystals were grown as above and subsequently soaked for 5 h with 20 mM $CaCl_2$ dissolved in reservoir solution.

All diffraction data were collected at 100 K, crystals being flash-cooled in a cryo N_2 gas stream prior to data collection. Prior to flash-freezing, crystals were treated with the appropriate cryo protectant solutions (eHIPΔN: 0.1 M Tris-HCl, 4.5 M potassium formate; eHIPΔN-SHHN: perfluoropolyether oil PFO-X125/03 (Lancaster Synthesis); eHIPΔN-DHHN and DHHN: 25% (v/v) glycerol in mother liquor). Data were collected at beamline ID29 (native and selenomethionine-labeled eHIPΔN and eHIPΔN-DHHN) at the European Synchrotron Radiation Facility (ESRF), France, and at beamlines I02 (DHHN) and I03 (eHIPΔN-SHHN) at the Diamond Light Source (DLS), UK. All X-ray data were processed and scaled with the HKL suite(Otwinowski & Minor 1997). The program XPREP (http://www.bruker-nonius.com) was used to calculate quality indicators and to merge data. Data collection statistics and structure solution strategies applied are given in Chapter 4, Table 4.1 and paragraph 4.2.2

2.6.2 Crystallisation and X-ray analysis of the human GPC3 ectodomain

Purified GPC3-ΔGPI was concentrated to a concentration of 4 mg mL^{-1} by ultrafiltration. Crystals were obtained in mother liquor containing 0.1M Tris Hydrochloride (pH 8.2), 10% v/v Jeffamine ED2001 (O,O'-Bis(2-aminopropyl) polypropylene glycol-block-polyethylene glycol-block-polypropylene glycol 1,900), pH 7.0. Prior to X-ray data collection, crystals were cryo-protected using 20% (V/V) glycerol supplemented to mother liquor and flash-frozen in liquid nitrogen.

Diffraction data were collected at 100 K at DIAMOND beamline I03. X-ray data were processed and scaled with Xia2 (Winter 2009). Data collection and structure solution approaches using the molecular replacement method combined with homology modelling and normal mode analysis are describe in details in Chapter 5 paragraph 5.2.3 and Table 5.1

2.6.3 Crystallisation and X-ray analysis of the human GAS1 ectodomain

Attempts to crystallise GAS1 complexes with SHHN and DHHN failed, but GAS1-ΔL2 alone yielded well-ordered crystals. Prior to crystallisation, GAS1-ΔL2 was concentrated to 6 mg mL^{-1} by ultrafiltration. Crystallisation resulted in three different, well-diffracting crystal forms grown in the mother liquor containing 0.1 M Tris pH 8.5, 2.5 M Ammonium Nitrate for GAS1-crystal form I, 5% v/v 2-Propanol 0.01 M Magnesium Acetate, 0.05 M Tris-HCl pH 7.5 for GAS1-crystal form II, and 5% w/v Polyethylene Glycol 6000, 0.1 M HEPES pH 7.0 for GAS1-crystal form III. All crystals were cryo-protected using 20% glycerol supplemented to the respective mother liquor. X-ray data were collected at ESRF beamline ID14-4 and DIAMOND beamline I03 (see Chapter 5, Table 5.1). All X-ray data were processed with Xia2 (Winter 2009). The structure for GAS1-crystal form I was solved using a Single Anomalous Dispersion (SAD) experiment of a crystal soaked with gold cyanide. Subsequent application of the molecular replacement method resulted in the solution of GAS1-crystal forms II and III. Details of data collection, structural solution and refinement are documented in Chapter 6, Table 6.1 and paragraph 6.2.5.

Chapter 3

Automated Cell Culture

'Automation of large scale transient protein expression in mammalian cells' (Zhao et al. 2011).

Chapter 3 Automated Cell Culture

Abstract

Traditional mammalian expression systems rely on the time consuming generation of stable cell lines, difficult to accommodate within a modern structural biology pipeline. Transient transfections are a fast, cost-effective solution, but require skilled cell culture scientists, making manpower a limiting factor in a setting where numerous samples are processed in parallel. Here a strategy employing a customised CompacT SelecT cell culture robot allowing the large-scale expression of multiple protein constructs in a transient format is reported. Successful protocols for automated transient transfection of HEK 293T and 293S GnTI⁻ cells in various flask formats have been designed. Protein yields obtained by this method were similar to those produced manually, with the added benefit of reproducibility, regardless of user. Automating cell maintenance and transient transfection allows the expression of high quality recombinant protein in a completely sterile environment with limited support from a cell culture scientist. Reducing human input has the added benefit of enabling continuous cell maintenance and protein production, of particular importance to structural biology laboratories, which typically use large quantities of pure recombinant proteins, and often require construct modifications at short notice.

3.1 Introduction

Fewer than 3% of the protein structures currently deposited in the Protein Database (PDB) were solved using samples produced in mammalian cells (Nettleship et al. 2010), despite the recent emergence of rapid and cost-effective transient expression methods (Aricescu et al. 2006; Durocher et al. 2002; Meissner et al. 2002). The

increasing focus on medically relevant proteins of eukaryotic, especially human, origin would seem, inevitably, to favour the use of mammalian expression systems where the protein is expressed in its native cell type, under physiological conditions, with numerous molecular systems working together for efficient production and quality control. However, the expression of eukaryotic proteins in a stable and soluble form is still a major bottleneck. This is due, in part, to the time consuming process required to maintain and expand cells and perform transient transfections. Automated mammalian cell culture using stably transfected cell lines is becoming increasingly prevalent in many stages of drug discovery. In its current form however, time considerations mean that it is of little use to modern structural biology laboratories, where many different constructs are typically processed in parallel and rapid feedback loops, supported by automated protein crystallization for example (Walter et al. 2003; Walter et al. 2005), may necessitate construct re-design within a two week cycle (Aricescu et al. 2006). Manual cell culture, typically performed by experienced staff, is highly repetitive and requires considerable time commitment and is therefore an ideal candidate for automation. Robotic transient transfection poses a significant challenge for existing cell culture systems, largely due to the exchange of media and delivery of various transfection mixtures to flasks in a linear sequence and within a strict time frame. Automated culture using the SelecT robot has already been reported for stable cell lines used in drug discovery programmes, demonstrating advantages such as consistency between operators and batches, ease of use and reduced risks of contamination (SZYMANSKI et al. 2008).

Simple and reproducible protocols for automated large-scale transient transfection of constructs cloned into the pHLsec vector (Aricescu et al. 2006) in two human cell

lines: HEK 293T and the N-glycosylation deficient HEK 293S GnTI$^-$ (Reeves et al. 2002) are presented here. A survey of 75 different protein batches, expressed using this robot over a four month period, allowed examination of the distribution of pure protein yields and the assessment of the usefulness of such a machine in the demanding environment of an academic crystallography laboratory.

3.2 Results

3.2.1 Automated transient transfection: protocol design, implementation, optimization

The CompacT SelecT robot (**Fig. 3.1A-G**), recently developed by The Automation Partnership (Royston, Cambridge, UK), is a smaller version of the SelecT automated cell culture platform. This system retained more than 70% of the SelecT capacity and, importantly, is able to manipulate 10 layer HYPERFlasks. Typically, it is used for cell-based drug screening or protein production using stable cell lines (SZYMANSKI et al. 2008). The CompacT Cellbase, a CompacT SelecT without a plating module, was chosen for our mammalian protein expression. The system consists of two compartments. The first is a humidified 5% CO_2, 37 °C carousel incubator/flask hotel unit (**Fig. 3.1B**) with space for 40 new input flasks and 90 production flasks for cell maintenance and transfection. The second is a laminar flow compartment containing a Stäubli robotic arm (Stäubli Robotics, Faverges, France) (**Fig. 3.1C**), a pipette head (**Fig. 3.1D**), a waste receptacle and "cocktail bar" (**Fig. 3.1E**). The laminar flow compartment also has a barcode reader to track flasks and a flask de-capper (**Fig. 3.1F**). The unit also incorporates a Cedex automated cell counting module (**Fig. 3.1G**). Cells are automatically maintained by regular harvesting and seeding of T175

flasks, while production is performed in Triple flasks and HYPERFlasks, following the flowchart described in **Fig. 3.1H**.

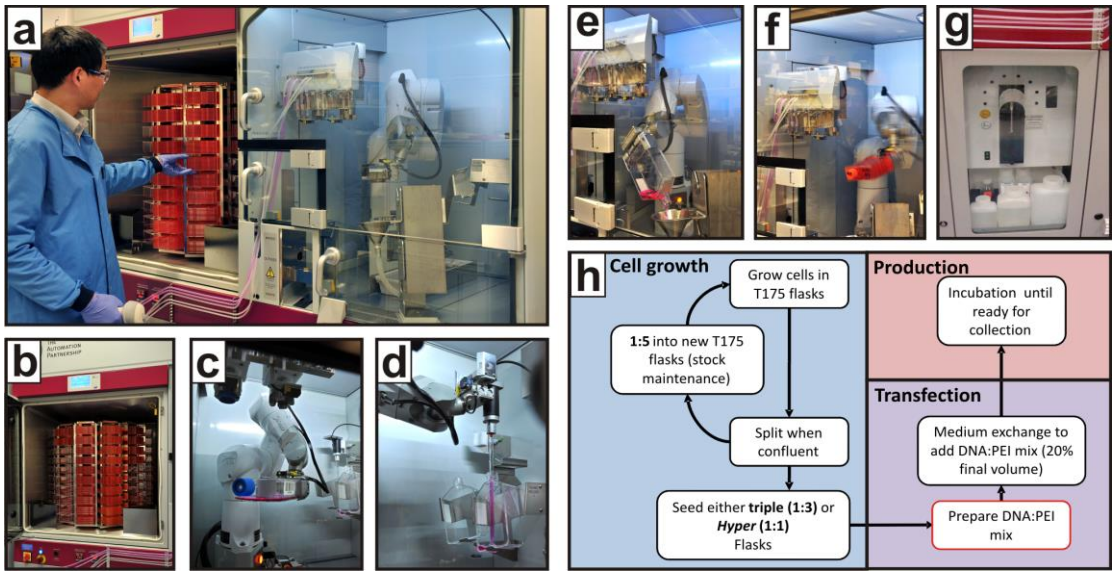


Figure 3.1. **The different CompacT Cellbase robot modules, employed for cell maintenance, seeding and transfection. (A)** The two robot units: incubator/hotel on the left, accessible to the operator; laminar flow cabinet for flask processing, on the right. The robot arm can access the cell flasks within the incubator carousel through an automated internal door that is synchronised with carousel turning. All liquid cell culture reagents, such as media, PBS and trypsin are connected to the system through silicon tubing and driven via peristaltic pumps. There are 10 pumps for up to 10 lines of different cell culture liquids. The robot system also coordinates the capping or decapping of flasks and presents the flask to the relevant media line, delivering precise amounts of liquid. The system uses 10 ml pipettes for transferring cell suspensions or adding solutions from the static liquid holder. **(B)** Carousel consisting of nine vertical hotels, each with a ten flask capacity, for cell maintenance and four hotels with ten positions for new flasks. **(C)** The Stäubli robot arm gripping a single flask, ready to be returned to the carousel incubator. **(D)** Splitting flasks using the pipette head to transfer cells. **(E)** Pouring media from a flask into the waste receptacle. **(F)** The robot arm performing the "hyperswirl" command after transfection; the "cocktail bar" can also be seen on the left of the photo. **(G)** The Cedex cell-counting module. **(H)** Workflow of automated cell culture, including cell maintenance, transfection and protein production steps.

A series of protocols was designed and extensively tested for routine cell maintenance (this work was carried out by both Yuguang Zhao and Benjamin Bishop) as well as transfection in various flask formats. To determine the optimal number of flasks which could be efficiently transfected with one batch of DNA:PEI mix, a range of incubation times was investigated (this work was carried out by Yuguang Zhao). Transfection efficiency, determined by monitoring live-cell EGFP fluorescence, remained stable for one hour after mixing PEI with DNA, but subsequently started to

decrease (**Fig. 3.2A**). Each HYPERFlask takes eight minutes to complete the transfection protocol, therefore no more than six HYPERFlasks can be transfected from one batch of DNA-PEI mixture if efficiency is to be maintained. Reliability of the whole procedure was tested by using this system to seed and independently transfect 24 different T175 flasks with the same pHLsec construct, encoding a 50 kDa glycoprotein. A sample of media from each flask was collected and analysed by western blot with an anti-His tag antibody, showing consistent protein expression (**Fig. 3.2B**) (Western blot was performed by Jordan Clay).

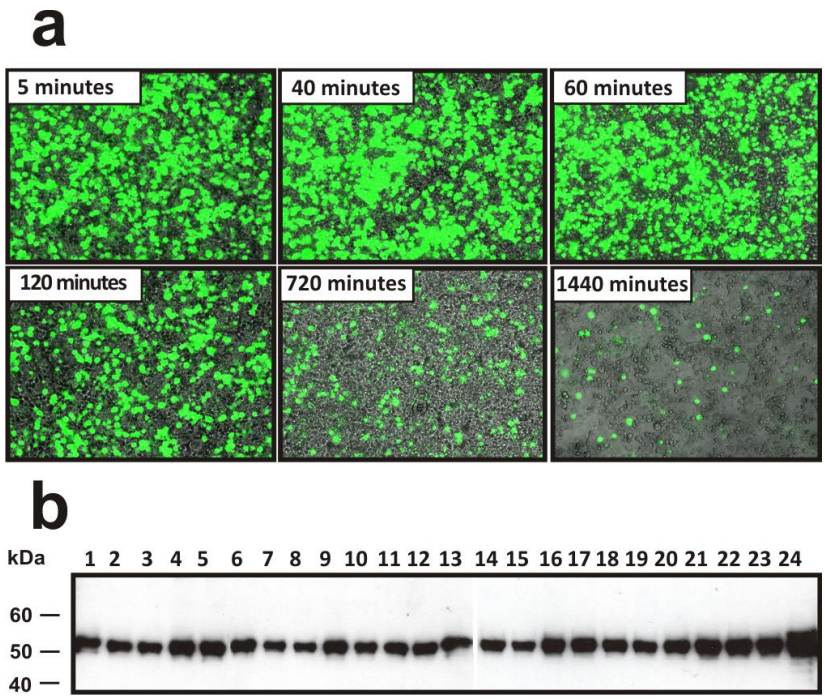
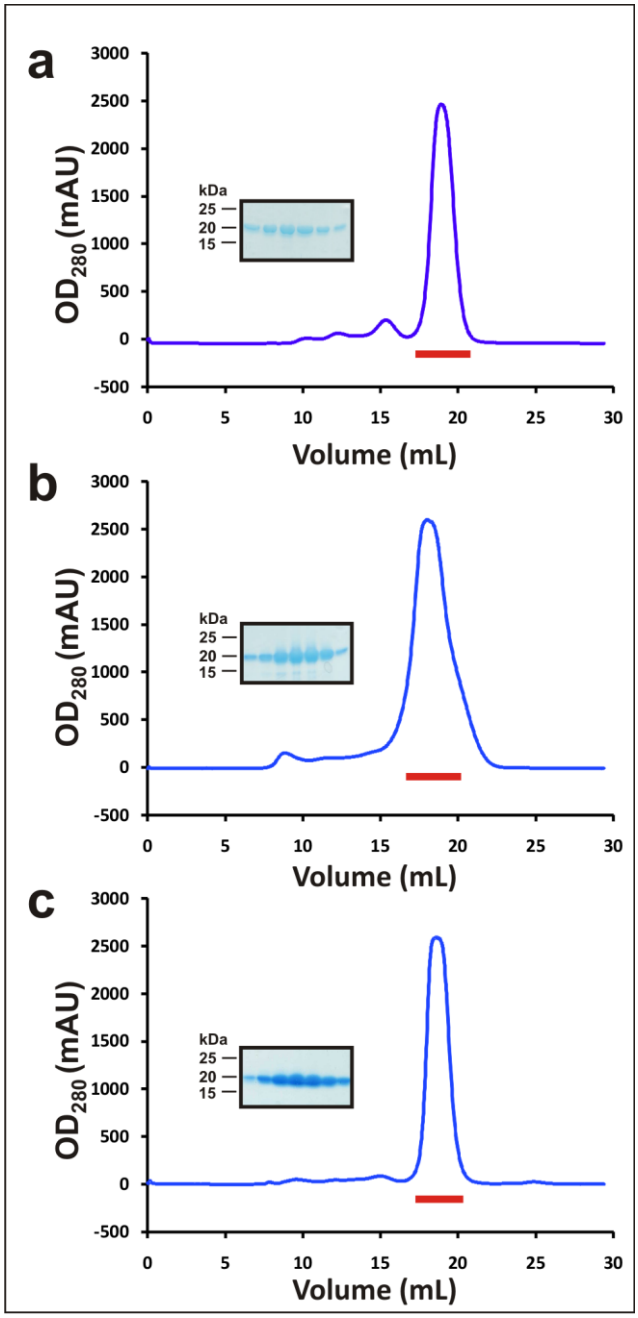


Figure 3.2. **Optimisation of the transient transfection protocol and reliability tests. (A)** Cells transfected by the robot at different time points after PEI:DNA mixing. Expression of EGFP allowed visualisation of transfection efficiency following different incubation periods. Transfection protocols were designed to cater for this decrease in efficiency over time (work carried out by Yuguang Zhao). **(B)** Western blot analysis of samples from 24 separate T175 flasks after automated transfection with the pHLsec vector encoding the same 50 kDa glycoprotein construct (Western blot performed by Jordan Clay).

3.2.2 Comparison between large-scale automated and manual transfection procedures

Ensuring that the efficiency of the automated transfection protocol matched that of the manual system was important in verifying that automated transient transfection was viable. To this end, test expression experiments using the N-terminal signalling domain of mouse SHH, solubly secreted in the conditioned medium(Bishop et al. 2009), were performed in three configurations: 4 roller bottles manually transfected with 2 mg DNA; automated transfection of 12 Triple flasks with 1.8 mg DNA; and automated transfection of 2 HYPERFlasks with 2 mg DNA. Four days post-transfection, media were harvested and purified before analysing the protein samples by SEC and SDS-PAGE (**Fig. 3.3A-C**). The construct was well expressed in all procedures (**Fig. 3.3D**), demonstrating that the robot is a viable alternative to manual tissue culture. The smaller footprint of HYPERFlasks provides an advantage over the Triple variant, occupying fewer slots in the incubator hotel for a comparable output. However the protein produced is roughly half the concentration of that observed in roller bottles. In terms of real "hands-on" time, the benefit of automation is directly proportional to the scale of the experiment. A typical experiment of manual transfection in 12 roller bottles takes approximately 30 minutes of operator input, similar to automated transfection of 6 HYPERFlasks (giving an equivalent volume of media). Doubling the output on the robot adds 10 further minutes of operator time, significantly less than doubling the number of roller bottles handled manually, which typically takes a lot longer due to operator fatigue. The greatest benefit of automation results from cell maintenance and seeding procedures, which require minimal operator input. Media and flasks must be supplied by an operator, but all subsequent procedures can then be fully automated, running overnight if necessary.



d

Flask	Surface (cm^2)	Volume (L)	Yield/Surface (µg/100cm^2)	Yield/Vol (mg/L)
(a) Roller	8500	1.0	49.41	4.20
(b) Triple	6000	1.2	60.6	3.03
(c) Hyper	3440	1.1	63.95	2.00

Figure 3.3. **Comparison of three protein expression strategies.** The secreted form of the mouse SHH N-terminal domain was purified from conditioned media by IMAC and SEC (blue traces) and analysed by SDS-PAGE (inset) of peak fractions (shown by a red bar on SEC traces). The protein was produced using: **(A)** four Roller bottles seeded and transfected manually, **(B)** twelve Triple flasks seeded and transfected by the robot and **(C)** two HYPERFlasks, seeded and transfected by the robot. **(D)** Table summarizing the experimental output.

3.2.3 Analysis of a 4-month production run

Data from HYPERFlask transient transfections performed on the CompacT SelecT Cellbase robot over a period of four months, using both HEK 293T and HEK 293S GnTI$^-$ cells and using only constructs confirmed to be secreted in small-scale experiments, revealed a broad range of pure, crystallization grade, protein yields (**Fig. 3.4**). It should be noted that constructs showing low expression in small-scale tests were often put forward for automated transfection, as manual protocols were typically more variable. These data represent the efforts of many users with different skill levels, handling 75 different samples of varied molecular weight, number of N-linked glycosylation sites, fold and number of domains and oligomeric state. An overall trend of higher expression levels in HEK 293T (**Fig. 3.4A**) versus HEK 293S GnTI$^-$ cells (**Fig. 3.4B**) was observed, but in all cases the protein quantities obtained were sufficient to perform extensive crystallization trials using nanolitre scale technologies(Walter et al. 2003; Walter et al. 2005). The inherent abilities of the two cell lines to produce recombinant proteins may be responsible for the observed differences, however other factors such as the addition of kifunensine, an α-mannosidase inhibitor used to reduce the complexity of N-linked glycosylation(V. T. Chang et al. 2007), which inhibits ER-associated degradation(Tokunaga et al. 2003) cannot be excluded.

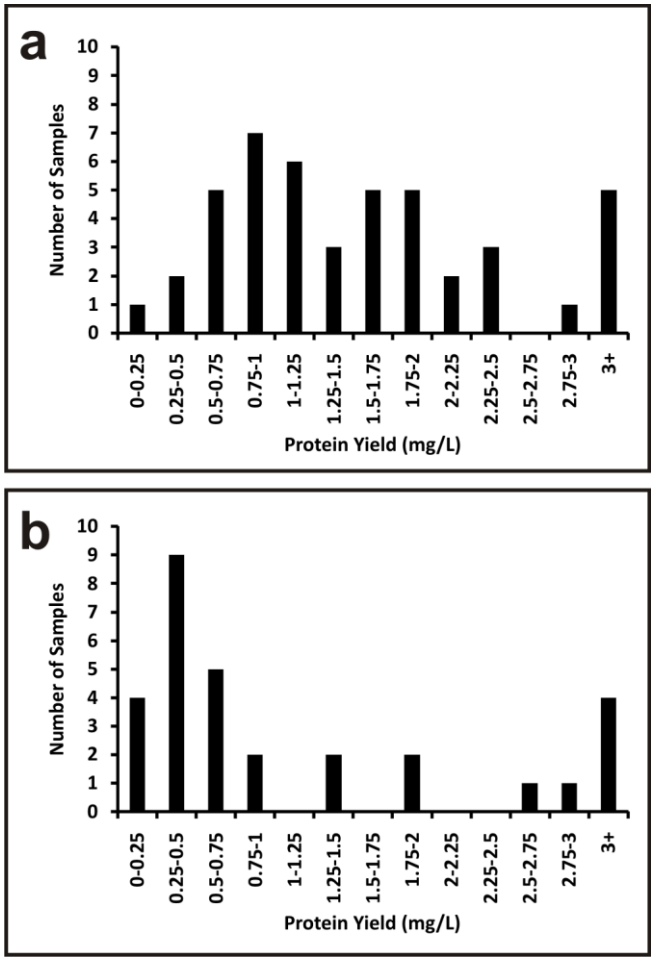


Figure 3.4. **Pure protein yields obtained from HYPERFlasks during a four-month automated production run in STRUBI. (A)** 42 samples produced in HEK 293T cells and **(B)** 33 samples produced in HEK 293S GnTI$^-$ cells. Each sample represents an individual construct which was transfected in a minimum of two and maximum of twelve HYPERFlasks.

3.3 Discussion

The automated protocols described here are equivalent to their manual counterparts in terms of quality and quantity of secreted mammalian proteins. From October 2009 to November 2010, automated transfection protocols were responsible for the production of ~300 protein batches in our laboratory. Proteins produced ranged in size from 10-400 kDa, including secreted intracellular and membrane proteins (Coles et al. 2011; Janssen et al. 2010; Malinauskas et al. 2011; Seiradake et al. 2010). The most obvious benefits of the automated protocols over the equivalent manual ones is the increased capacity for cell culture and protein production that becomes available, whilst human

effort is released for other tasks. The manual system is entirely dependent on the skills and presence of an experienced cell culture scientist. In the automated system, there is no such constraint. As long as media reservoirs are filled as needed and flasks loaded into the robot, automated cell culture can continue 24 hours a day 7 days a week without a break. Transient transfections still require the presence of an operator to prepare the cocktail of reagents and thus can only be completed during normal working hours. With efficient timetabling, all fully automated procedures can be completed out of normal working hours and the machine monitored and operated remotely via an Internet based video link and control software if required. A further advantage over the manual roller bottle system is that due to the low volume of media required in the roller bottles, HEK 293T cells in the presence of kifunensine typically survive for only 4 days post transfection while in HYPERFlasks, efficient gas exchange, allowed by the permeable membrane on which the cells grow, results in cells producing proteins for up to 10 days post-transfection. For HEK 293S GnTI$^-$ efficient expression time is increased from a maximum of 7 days post-transfection in roller bottles, to 14 days post-transfection in HYPERFlasks. This increased production time may result in increased yields, although this was not utilised for most of the protein samples discussed here as users allowed the same maximum times for the expression of their protein for HYPERFlasks as they did for roller bottles.

Automation of the manual mammalian expression protocol, based on the pLEXm vectors (Aricescu et al. 2006), required some modification to allow for the limitations of the robot being used. Flasks of suitable geometry include T175 (for maintenance) and either Triple flasks or HYPERFlasks for transient production. Both offer the same ease of handling for the robot, but the Triple flasks occupy three times more volume

in the incubator for the equivalent surface area of cell growth. Both types of flask use a lot more media than the roller bottles used in manual transfections. Roller bottles require 250 mL culture media for a surface area of 2125 cm^2 while Triple flasks minimally require 100 mL each for a surface area of 500 cm^2 and HYPERFlasks require 555 mL for a surface area of 1720 cm^2. This then requires a modification of the amount of DNA and PEI used in the transfection cocktail to ensure the correct concentration of both in larger volumes. For roller bottles 0.5 mg DNA with 1 mL PEI was required per bottle. For Triple flasks, 150 μg DNA is required per flask with 250 μL PEI, while for HYPERFlasks, 1 mg DNA per flask with 2 mL PEI. Transfections are carried out in the automated system using peristaltic pump delivery rather than pipettes, as the robot is only able to handle 10 mL pipettes, not a viable option for large volume transfers required by multi-layered flasks. A major drawback of the HYPERFlasks is that they must have 555 ± 1 mL in order to function properly. This requires daily calibration of the pumps, which cannot currently be performed remotely. The T175 and Triple flasks do not require such rigorous calibration; a weekly procedure offers sufficient accuracy.

Having used the CompacT SelecT Cellbase for more than a year, many opportunities for improvements in both the hardware and software have come to light. In terms of hardware, a more consistent delivery system for media would be useful, which would relax the requirement for daily calibration. It is currently time consuming to wash the peristaltic pump lines between transfection constructs in order to prevent cross-contamination, a significant problem when the number of flasks transfected with a given construct is small (one or two). Allowing volumes greater than 40 mL into the static liquid holder, located inside the laminar flow unit, would offer greater

flexibility in protocol design. Another issue during full time production is space limitation in the robot incubator. All transfected flasks are routinely removed to free spaces in the nine incubator hotels, but despite this, space remains an issue and is now the limiting factor in protein production. A current drawback of the software used to control the robot is its inability to make use of incubation periods in which the robot is dormant. This greatly increases the operating time required for stacked commands.

The protocols described here are the first implementation of an automated tissue culture system for large-scale transient expression of secreted recombinant proteins in mammalian cells in a quantity and of a quality suitable for X-ray crystallography. Although more expensive than manual systems in terms of consumables and initial set-up cost, the time saved and the increased productivity makes up for this initial investment, removing the protein production bottleneck. A major advantage of the automated system is the reproducibility of expression between users and at different times.

Chapter 4

Hedgehog Interacting Protein

'Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HIP' (Bishop et al. 2009)

Chapter 4 Hedgehog Interacting Protein

Abstract

HH signalling molecules are involved in many developmental processes, including tissue growth, patterning and morphogenesis. HHIP binds members of the vertebrate HH family proteins inhibiting their function. This inhibition regulates HH function, but dysregulation has been linked to congenital defects, neurodegenerative diseases and cancer. The structure of human HHIP in complex with HH family members reveals a binding mode distinct from previously defined ligand interaction sites, with metal ions playing key roles in the interaction. Zinc is essential for the binding of HH to HHIP, while calcium prevents electrostatic repulsion between the two proteins, thus playing a major modulatory role. This interplay of several metal-binding sites suggests a tuneable mechanism for regulation of HH signalling.

4.1 Introduction

The HH signalling pathway is highly conserved in vertebrates and invertebrates. The number of ligands, has however, increased during evolution. There is one HH ligand in Drosophila, and three family members in mammals: SHH, DHH and IHH (Ingham & McMahon 2001). Although their spatial and temporal expression patterns vary significantly in different tissues, mechanistically these ligands are thought to work in a similar fashion (McMahon et al. 2003; Varjosalo & Taipale 2008).

4.1.1 Discovery and biological function

Diverse strategies are employed for the extracellular regulation of HH signalling. The interference hedgehog protein family (IHOG in fly and CDO and BOC in human,

respectively) (Yao et al. 2006; W. Zhang et al. 2006; Tenzen et al. 2006) and HSPGs (Capurro et al. 2008) were identified as crucial modulators in both vertebrates and invertebrates, while HHIP (Chuang & McMahon 1999) and GAS1 (Allen et al. 2007; Martinelli & Fan 2007a) are vertebrate-specific. All of these membrane proteins have been shown to directly bind the HH ligand.

HHIP is a type I membrane protein attached to the cell surface by a hydrophobic C-terminal helix(Chuang & McMahon 1999) and binds all three HH ligands (SHH, IHH and DHH) with equally high affinity to that of PTCH, functioning as an inhibitor of HH signalling (Chuang & McMahon 1999; Pathi et al. 2001). Like PTCH, expression of HHIP is up-regulated in response to HH signalling (Chuang & McMahon 1999), but unlike PTCH, there is no evidence that it acts by directly regulating SMO. HHIP therefore provides an additional layer of control to HH signalling. A secreted version of HHIP is still able to bind and sequester HH, thus interfering with HH signalling (Coulombe et al. 2004).

4.1.2 HHIP in normal development

At the cell membrane, HHIP binds SHH with high affinity and acts as a negative regulator of HH signalling in embryos (Coulombe et al. 2004). HHIP is important in cartilage development in mouse embryos (Chuang & McMahon 1999), normal nerve development via Schwann cell derived DHH (Parmantier et al. 1999), and axon guidance, where it mediates the repulsive effect of SHH on post-commissural axons (Bourikas et al. 2005). HHIP is particularly important in the development of the nervous system where SHH signalling is required for the correct generation of various classes of neurons including motor neurons and gabaergic interneurons in the ventral

neural tube, dopaminergic neurons in the midbrain, serotoninergic neurons in the hindbrain, and participates in oligodendroglia commitment and HHIP is found in both membrane bound and secreted form (Ingham & McMahon 2001; Coulombe et al. 2004). HHIP is expressed next to HH expressing regions and transgenic mice which overexpress HHIP in the cartilage display a phenotype reminiscent of IHH knockout mice (Chuang & McMahon 1999; Coulombe et al. 2004). HHIP-/- mutant mice do not exhibit pronounced defects in tissues related to HH signalling during embryonic development, except in the presumptive lung (Chuang et al. 2003).

HHIP expression in transgenic animals inhibits HH function (Chuang & McMahon 1999; Treier et al. 2001), while loss of HHIP function results in increased HH signalling (Chuang et al. 2003). HHIP and PTCH expression has been reported during lung development (Chuang et al. 2003), demonstrating that the function of these regulatory proteins during normal embryogenesis is to restrict overt HH signalling by sequestration of HH ligands (Kawahira et al. 2003). HHIP can be expressed in both soluble and membrane-associated forms in the adult brain (Coulombe et al. 2004), suggesting that HHIP participates in SHH signalling in the mature brain pointing to a possible role for soluble HHIP in regulating the SHH pathway in the adult brain (Coulombe et al. 2004). As negative regulators, HHIP and PTCH function jointly to restrict HH signalling activity during pancreas organogenesis, and, as has been shown for other tissues, pancreatic tissue responds to HH in a dose dependent manner (Kawahira et al. 2003).

4.1.3 HHIP and disease

HHIP has been implicated either directly or by association with a number of different diseases. As HHIP is a HH signalling antagonist, it can result in aberrant HH signalling in a number of ways. It can lead to increased HH signalling if absent or to a HH-/- phenotype if overexpressed. However, HHIP overexpression should not be confused with a phenotype that simply mimics that of a HH-/- phenotype. Mice that ectopically express HHIP in the intestinal epithelium have expanded smooth muscle, but there is also a paracrine mechanism at play, whereby the surrounding mesenchyme is affected by loss of HH signalling and in turn it signals back and stimulates increased epithelial proliferation so altering the pattern of crypt-villus; an affect not observed in the IHH or SHH null mice (Madison et al. 2005; Parkin & Ingham 2008). Mice lacking HHIP, on the other hand, display a range of morphogenetic abnormalities of the pancreas. In severe cases the ventral and dorsal pancreas fuse to form a small mass of tissue, although this retains pancreatic character. This is consistent with an increase in proliferation, but not differentiation, of pancreatic endocrine cells (Kawahira et al. 2003) which is consistent with increased SHH signalling. Additionally, ectopic pancreatic tissue can be found in the stomach and duodenum whereas the spleen is also reduced in size (Kawahira et al. 2003). Functional redundancy between PTCH and HHIP in limiting HH diffusion is indicated by the increase in severity of pancreatic defects in HHIP mutants that also lack one PTCH allele (Kawahira et al. 2003; Parkin & Ingham 2008). As HHIP has been shown to have importance in the development of mouse embryonic lungs, it is little surprise, therefore, that HHIP is associated with the risk of developing COPD (Pillai et al. 2009; X. Li et al. 2011b).

In vitro and *in vivo* studies have suggested that SHH may have a therapeutic value in neurodegenerative diseases such as Alzheimer or Parkinson (Coulombe et al. 2004) and long-acting forms of SHH have been shown to be efficacious in a sciatic nerve crush model (Coulombe et al. 2004). A SHH antagonist could abrogate these beneficial effects of long range HH signalling, therefore reducing the effects of HHIP could be a potential therapy for neurodegenerative diseases. In addition, HHIP expression is altered in a variety of cancers (Olsen et al. 2004; Tada et al. 2008; Taniguchi et al. 2007; Tojo et al. 2002). Indeed, loss of HHIP activity may be a contributing factor to pancreatic cancer as expression is reduced or absent in many cell lines because of hyper-methylation of the HHIP promoter(Martin et al. 2005; Parkin & Ingham 2008). In addition, HH neutralizing antibodies or recombinant HHIP protein are able to reduce pancreatic cancer cell growth (Kayed et al. 2005; Thayer et al. 2003).

Despite this wealth of functional data, the molecular mechanisms underlying HHIP function are still unknown. We therefore undertook an analysis of the structural and binding characteristics of human HHIP-HH interactions.

4.2 Results

4.2.1 Production and characterisation of the HHIP ectodomain

All HHIP constructs were produced in mammalian 293T cells using transient expression and the glycosylation inhibitor kifunensine, which allowed the cleavage of N-linked sugars with endoglycosidase F1 (see methods chapter and(Aricescu et al. 2006)). When expressing the full-length ectodomain of human HHIP (eHIP, 56-670), lacking the last 30 residues (which are crucial for membrane attachment (Chuang &

McMahon 1999)), we observed fragmentation of the protein and identified two proteolytic cleavage sites (at Arg189 and Arg209) using N-terminal amino acid sequencing (Fig. 4.1A). Mutations of the two proteolytic recognition sequences (Arg189, Arg209 and Lys210 to alanine) or an N-terminal truncation of the first 210 amino acid residues (eHIPΔN construct) resulted in proteolytically resistant and homogeneous proteins. Size exclusion chromatography and multiangle laser light scattering (MALS) revealed that the stabilised full-length ectodomain is dimeric in solution whereas the eHIPΔN is a monomer (**Fig. 4.1B,C**). These experiments suggest a function of the N-terminal cysteine-rich domain in mediating dimerisation.

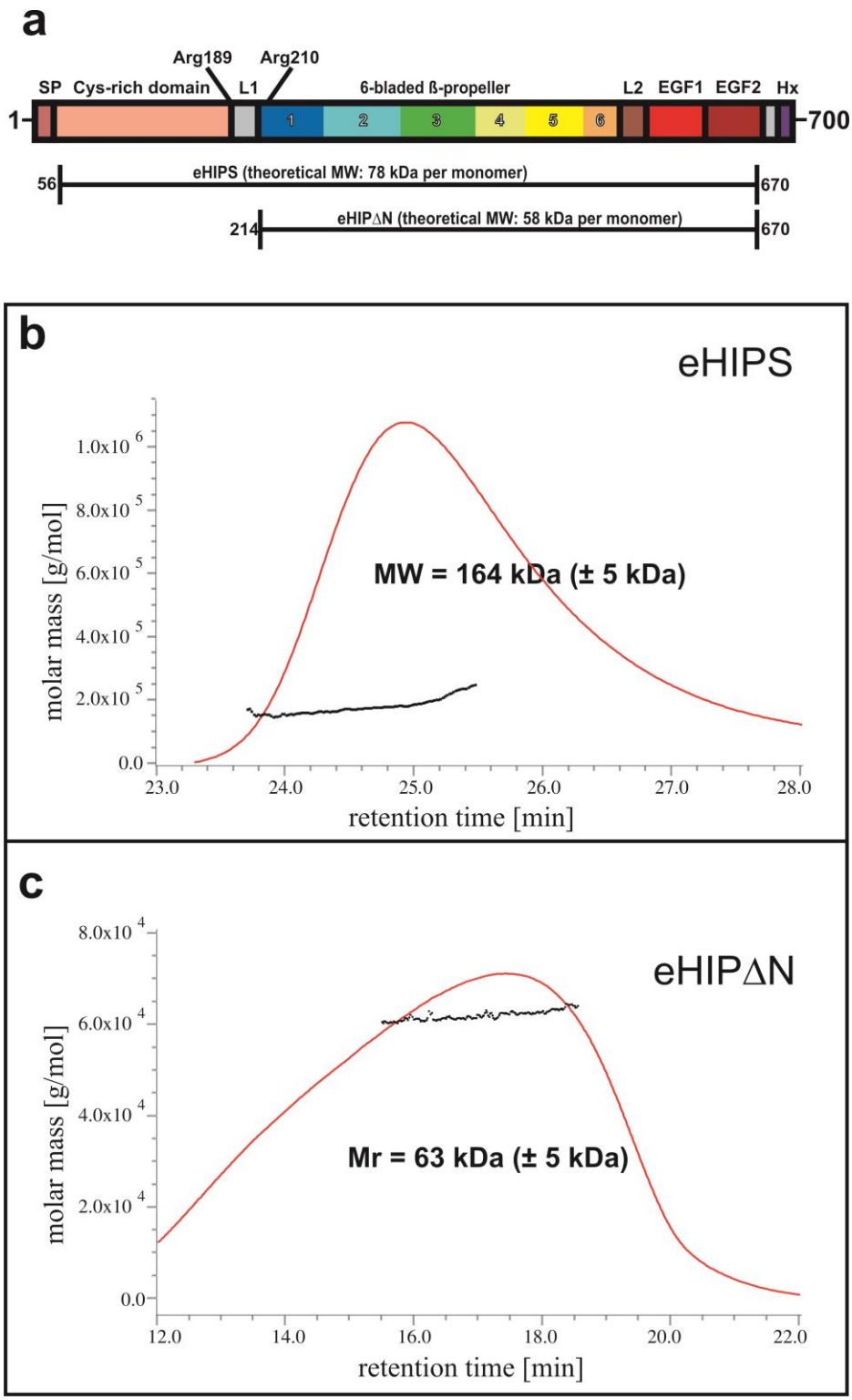


Fig. 4.1 Measurements of the molecular weight of HHIP ectodomain constructs by multiangle laser light scattering (MALS). (A) Schematic representation of the HHIP domain organisation. SP: signal peptide; L1, L2: interdomain linker regions; EGF1, EGF2: epidermal growth factor repeat domains; Hx: membrane attachment helix. β-propeller blades are color-coded and numbered. Proteolytic cleavage site residues Arg189 and Arg210, identified by N-terminal sequencing, are highlighted. Theoretical molecular weights of glycosylated proteins were calculated using the PROTPARAM program at the EXPASY server (www.expasy.ch). (B,C). The measured 164 kDa molecular weight for the glycosylated, proteolytically stabilised full-length HHIP ectodomain (with Arg189, Arg210 and Lys211 mutated to alanine, eHIPS) is consistent with a dimer, whereas the observed mass for the N-

terminal truncated eHIPΔN (63 kDa) corresponds to a monomer. These results are consistent with size exclusion chromatography data (not shown).

4.2.2 Structure of the HHIP ectodomain

Attempts to crystallise the stabilised full ectodomain of HHIP failed, but crystals of eHIPΔN were obtained in space group $P3_121$, with two molecules in the asymmetric unit and a solvent content of 71% (Table 4.1). For phasing, selenomethionine-labeled protein was also produced in HEK 293T cells (Aricescu et al. 2006) and the structure was solved by a single anomalous dispersion experiment resulting in a high quality, experimental electron density map (**Fig. 4.2B**). The structure refinement resulted in an R-factor of 20.1% (R-free: 23.9%) to 2.8 Å resolution (Table 4.1) with two eHIPΔN molecules in the crystallographic asymmetric unit (**Fig. 4.2C**). The eHIPΔN monomer is composed of an N-terminal 6-bladed β-propeller domain (residues 214-583) linked to two epidermal growth factor (EGF) repeat domains (residues 607-637and 638-670, respectively) (**Fig. 4.2D**). The β-propeller domain shows the classical fold observed in a variety of proteins from all kingdoms with functions as diverse as enzymes, transcriptional regulators or cell-surface glycoproteins, as for example the low-density lipoprotein receptors(Rudenko et al. 2002). The major atypical characteristic of the HHIP β-propeller domain is a large protrusion of two loops from blade 3 of the propeller forming a highly negative charged patch on the surface (highlighted in **Fig. 4.2E**).

The β-propeller domain is connected with the two EGF repeats via a 28 residues linker region (coloured orange in **Fig. 4.2B,D**). This linker stretches over blades 6, 1 and 2, is well ordered and locks the EGF repeats on the opposite side of the β-propeller C-terminus. The conformation of the linker and the orientation of the first

EGF repeat are further stabilised by several disulphides (disulphides II, IV and VI, see **Fig. 4.2D**) conserved in all known HHIP receptors (**Fig. 4.3**). This rigid arrangement forms a structural unit; attempts to produce soluble forms of HHIP without the first EGF repeat failed. The EGF repeats have a linear, elongated conformation with the C-terminus of the second EGF repeat pointing away from the β-propeller domain. Although the EGF repeats form rigid domains, each stabilised by 3 disulphide bonds, there seems to be flexibility in the inter-domain interactions (the major difference between the two HHIP copies in the asymmetric unit is the relative orientation of the second EGF repeat) (**Fig. 4.2C**). This hinge is likely to introduce some flexibility in the membrane-proximal region of the ectodomain. The hydrophobic helix, essential for the membrane attachment of the HHIP receptor (Chuang & McMahon 1999), is located 9 residues downstream of the second EGF repeat.

Table 4.1. Crystallographic data collection and refinement statistics

	SeMet-eHIPΔN	**Native eHIPΔN**	**eHIPΔN-SHHN**	**eHIPΔN-SHHN without calcium**	**eHIPΔN-DHHN without calcium**	**DHHN-without calcium**	**DHHN with calcium**
DATA							
X-ray source	ESRF-ID29	ESRF-ID29	ESRF-ID29	Diamond-I03	ESRF-ID29	Diamond-I02	ESRF-ID29
Resolution [Å]	20.0-3.5 (3.6-3.5)	20.0-2.8 (2.9-2.8)	20.0-3.2 (3.3-3.2)	20.0-3.15 (3.25-3.15)	20.0-2.6 (2.7-2.6)	30.0-1.85 (1.95-1.85)	40.0-1.85 (2.05-1.95)
Space group	$P3_121$	$P3_121$	$P3_221$	$P3_221$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions [Å]	a= 101.1 Å b = 101.1 Å c = 308.8 Å	a = 101.0 Å b =101.0 Å c = 305.9 Å	a = 89.2 Å b = 89.2 Å c = 171.4 Å	a = 88.0 Å b = 88.0 Å c = 171.9 Å	a = 102.8 Å b = 110.7 Å c = 144.4 Å	a = 40.6 Å b = 42.1 Å c = 84.3 Å	a = 40.1 Å b = 41.9 Å c = 83.8 Å
Solvent content [%] (mols per asymmetric unit)	71 (2 mols)	71 (2 mols)	67 (1 complex, 1:1 stoichiometry)	67 (1 complex, 1:1 stoichiometry)	57 (2 complexes, 1:1 stoichiometry)	41 (1 mol)	41 (1 mol)
Wavelength [Å]	0.9792	0.9793	0. 97623	0.8900	0.9793	0.9507	0.97623
Unique reflections	56853 (5019)[d]	45408 (4436)	13529 (1331)	13878 (1265)	49775 (4920)	12853 (1849)	10765 (1003)
Completeness [%]	99.5 (99.7)	99.5 (99.9)	100 (100)	99.2 (99.8)	100.0 (100.0)	99.6 (99.7)	93.6 (91.8)
R_{merge} [%][a]	12.9 (62.3)	11.5 (73.7)	18.7 (72.8)	20.9 (88.5)	18.4 (81.0)	12.9 (68.0)	11.8 (74.9)
I/σI	15.7 (3.4)	14.7 (3.4))	10.6 (2.5)	9.9 (2.8)	15.3 (2.7)	14.6 (4.5)	10.5 (2.6)
Redundancy	11.3 (8.1)	9.1 (8.9)	6.5 (6.7)	7.5 (7.8)	10.0 (10.1)	12.7 (12.0)	6.8 (6.5)
PHASING							
No. of Heavy atom sites	12						
SHELXD CC/CCweak for substructure solution	32.0/12.5						
FOM (acen/cen) (SHARP)	0.10/0.03						
Anomalous Cullis R-factor (SHARP)	0.932						
Anomalous phasing power (SHARP)	0.578						
REFINEMENT							
Resolution range [Å]		20.0-2.8 (2.86-2.80)	20.0-3.2 (3.35-3.20)	20.0-3.15 (3.35-3.15)	20.0-2.6 (2.67-2.60)	30.0-1.85 (1.90-1.85)	40.0-1.95 (2.00-1.95)
Number of reflections		45399 (2611)	16254 (1640)	13191	48634 (3477)	12205 (876)	10413 (718)
No. of atoms (protein/Zn/Ca/ water)		6578/0/0/38	4465/1/3	4455/1/0	17348/2/135	2599/1/104	2609/1/2/96
B factors [Å^2] (protein/Zn/Ca/ water)		81/0/0/60	75/60/73/0	63/44/0/0	38/28/0/34	32/30/0/25	32/30/46/22
R_{factor} [%)][b]		20.1 (29.7)	24.1 (32.7)	23.4 (28.9)	22.7 (33.0)	18.9 (23.8)	18.2 (25.8)
R_{free} [%] [c]		23.9 (34.2)	29.9 (38.5)	29.2 (34.1)	28.1 (40.2)	22.9 (27.9)	22.5 (30.0)
r.m.s.d. bonds [Å]		0.014	0.004	0.004	0.011	0.013	0.014
r.m.s.d. angles [deg]		1.794	0.524	0.557	1.329	1.382	1.455
Ramachandran statistics							
Favoured [%]		93.0	92.4	92.7	95.7	96.3	96.3
Disallowed [%]		0	0	0	0.1	0	0

r.m.s.d. stands for root mean square deviation from ideal geometry. Numbers in parentheses refer to the appropriate outer shell. CC stands for correlation coefficient, FOM for figure of merit.

[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I(hkl;i) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I(hkl;i)$, where $I(hkl;i)$ is the intensity of an individual measurement and $\langle I(hkl) \rangle$ is the average intensity from multiple observations.

[b] $R_{factor} = \Sigma_{hkl} ||F_{obs}| - k|F_{calc}|| / \Sigma_{hkl} |F_{obs}|$.

[c] R_{free} equals the *R*-factor against 5% of the data removed prior to refinement. [d] Friedel partners are accounted as different reflections.

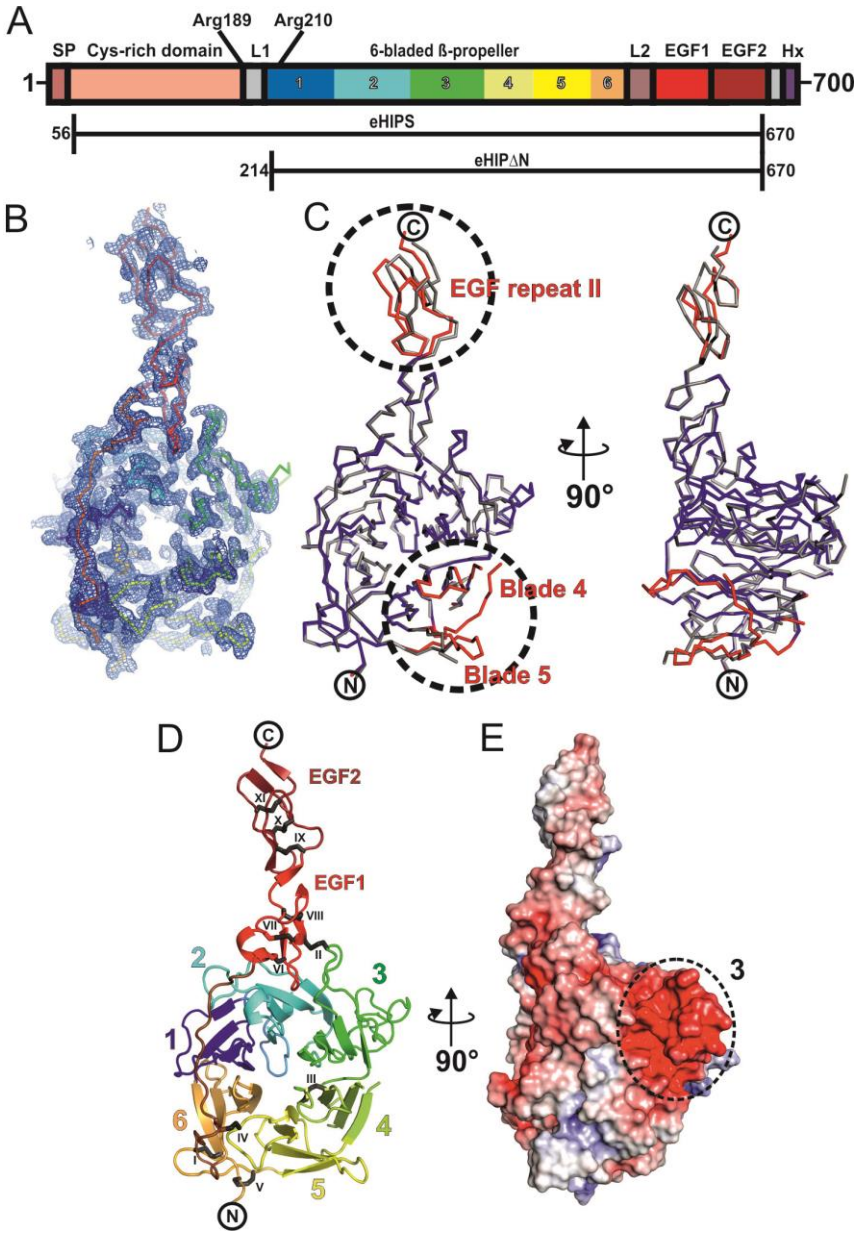


Fig. 4.2 Structure of the HHIP ectodomain. (A) Schematic domain organization of the HHIP receptor. SP: signal peptide; L1, L2: interdomain linker regions; EGF1, EGF2: epidermal growth factor repeat domains; Hx: membrane attachment helix. β-propeller blades are colour-coded and numbered. Proteolytic cleavage site residues Arg189 and Arg210, identified by N-terminal sequencing, are highlighted. The crystallisation construct (eHIPΔN) and the stabilised full-length ectodomain construct (eHIPS) are shown. (B) Selenomethionine SAD-phased and phase-extended electron density map (calculated to 2.8 Å, contoured at 1 σ) with rainbow-coloured Cα trace of eHIPΔN. (C) Superposition of the two copies of eHIPΔN in the asymmetric unit. One copy is coloured in grey and the other one is colour-coded (red where there are significant differences between the structures and blue where parts are superimposable) using the significance level defined by default criteria of program ESCET (http://shelx.uni-ac.gwdg.de/~trs/escet/). The rigid body movement of the second EGF domain and the differences of blade 4 and 5 of the β-propeller domain are indicated with dotted circles. The right panel is rotated by 90º anti-clockwise around the x axis. (D) Ribbon diagram of eHIPΔN with colour coding as in (A). The six blades of the β-propeller domain (each consisting of a 4-stranded β-sheet) are numbered as in a. The eleven disulphide bridges are shown in black stick representation and marked with Roman numerals. (E) Electrostatic properties. eHIPΔN is shown as solvent accessible surface coloured by electrostatic potential contoured at +/- 10 kT (red, acidic; blue, basic). The prominent negatively charged patch which interacts with HH ligands is marked with a dotted circle.

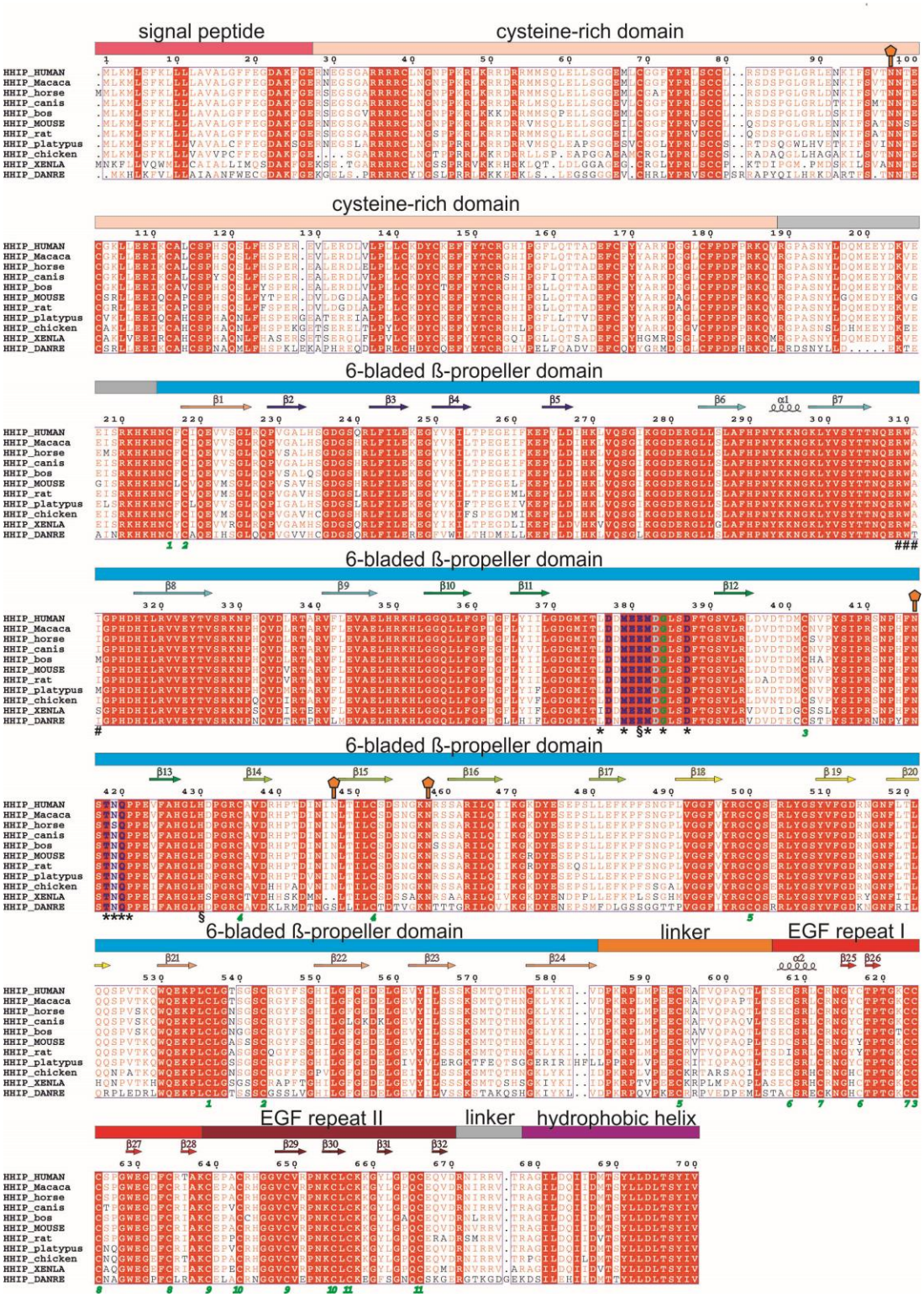


Fig. 4.3. Sequence alignment of the HHIP family members. The HHIP sequences were aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPript/ESPript/). Numbering corresponds to the full length human HHIP (including the secretion signal). Domain organisation and secondary structure assignments for human HHIP are displayed above the alignment. Following refinement, clear electron density corresponding to the N-linked N-acetylglucosamine (GlcNAc) could be seen for only one out of the 4 glycosylation sites predicted by the NetNGlyc server (http://www.cbs.dtu.dk/services/NetNGlyc/): Asn99, Asn416, Asn447 and Asn459 (marked by filled orange pentagons). Residues forming hydrophilic interactions with DHHN and SHHN are highlighted in blue, residues exclusively forming hydrogen bonds with

DHHN are coloured in green. Asp383 which participates in the zinc-binding site is coloured cyan. Residues forming hydrophobic interactions with DHHN and SHHN are marked below the sequences with an asterisk (*), with exclusively DHHN with a hash (#) and with exclusively SHHN with a section sign (§). Disulphide bridges are numbered in green.

4.2.3 Structure of the HHIP-SHH complex

We next produced (**Fig. 4.4A** and methods, Chapter 2, paragraph 2.2.8) the complex between eHIPΔN and the N-terminal signalling domain of SHH (SHHN, a monomeric construct that lacks the lipid-modification sites) and determined its crystal structure to 3.2 Å resolution (Table 4.1). The HH ligand uses a surface oriented away from its termini (which are lipid-modified in vivo) to interact with the two distinctive loops (BL1 and BL2) protruding from blade 3 of the HHIP β-propeller (**Fig. 4.5A,B**). Two disulphide bridges at the beginning and the end of blade 3 (disulphide bridges II, III in **Fig. 4.5A**) stabilise the blade for HH ligand reception whilst the C-terminal EGF repeats and membrane proximal region of HHIP are oriented away from the ligand-binding site. The complex interface buries 840 $Å^2$ of eHIPΔN and 759 $Å^2$ of SHHN with a surface complementarity score (sc) of 0.66, and consists of 13 potential hydrogen bonds, 157 non-bonded van-der-Waals contacts and one zinc ion, involving a total of 14 residues of the HHIP receptor and 20 residues of SHHN (**Fig. 4.5C-F**). Detailed interactions are depicted in **Fig. 4.5C-F**. In eHIPΔN, BL1 (residues 370-391, harbouring a small α-helical segment) forms the major interaction interface inserting into the SHH zinc-binding cavity (**Fig. 4.5B,C**). The carboxylate group of HHIP-Asp383, located at the tip of BL1, completes the coordination of the zinc-binding site, which is formed by SHH residues His140, Asp148 and His183 (**Fig. 4.5D**). The presence of the zinc was verified by coordination geometry (Harding 2001), X-ray absorption and fluorescence scans and anomalous scattering (**Fig. 4.6**). Two calcium ions, coordinated by SHH residues Glu90, Glu91, Asp96, Thr126, Glu127 and Asp130 (**Fig. 4.5F**), are also clearly visible in the electron density maps (**Fig. 4.6**) and appear to play an important role in the stabilisation of a SHH loop (Lys88-Gly94; discussed below). The second HHIP loop contributing to binding (BL2) interacts with

a SHH surface patch adjacent to the zinc-binding cavity forming three potential hydrogen bonds and several van-der-Waals contacts (**Fig. 4.5E**).

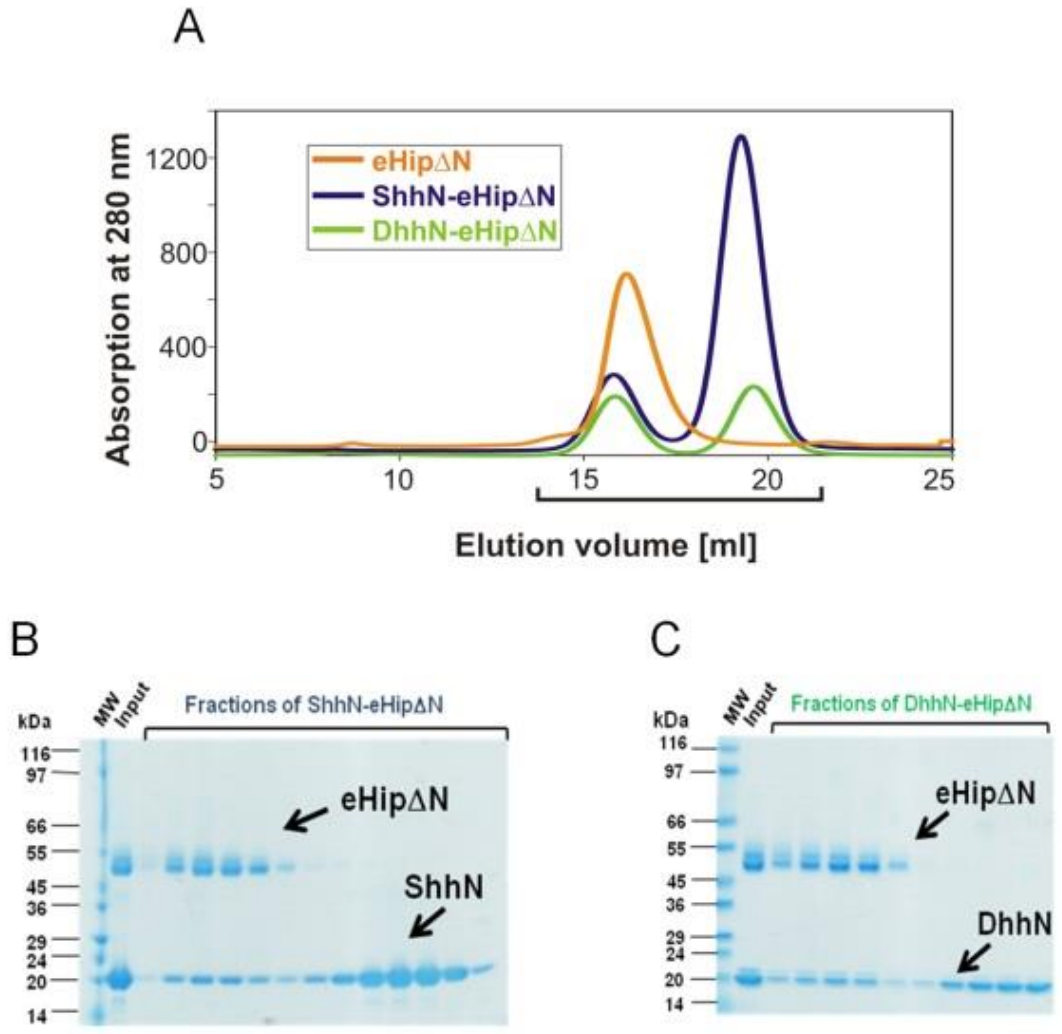


Fig. 4.4 Size-exclusion purification of HH-HHIP complexes. (A) The elution profiles of eHIPΔN (orange), SHHN-eHIPΔN (blue) and DHHN-eHIPΔN (green) are overlaid. Prior to complex formation, eHIPΔN was deglycosylated using endoglycosidase F1 (see methods for details). Fractions analysed by SDS-PAGE are marked below the x axis. (B,C), SDS-PAGE analysis of the fractions of the SHHN-eHIPΔN (B) and DHHN-eHIPΔN (C) complexes. The first peak corresponds to the HH-eHIPΔN complexes, the second to the uncomplexed HH ligands, which were added in a 1.5 molar excess prior to size exclusion chromatography. The single peak (orange) is eHIPΔN alone. MW: Sigma Molecular Weight Marker S8445.

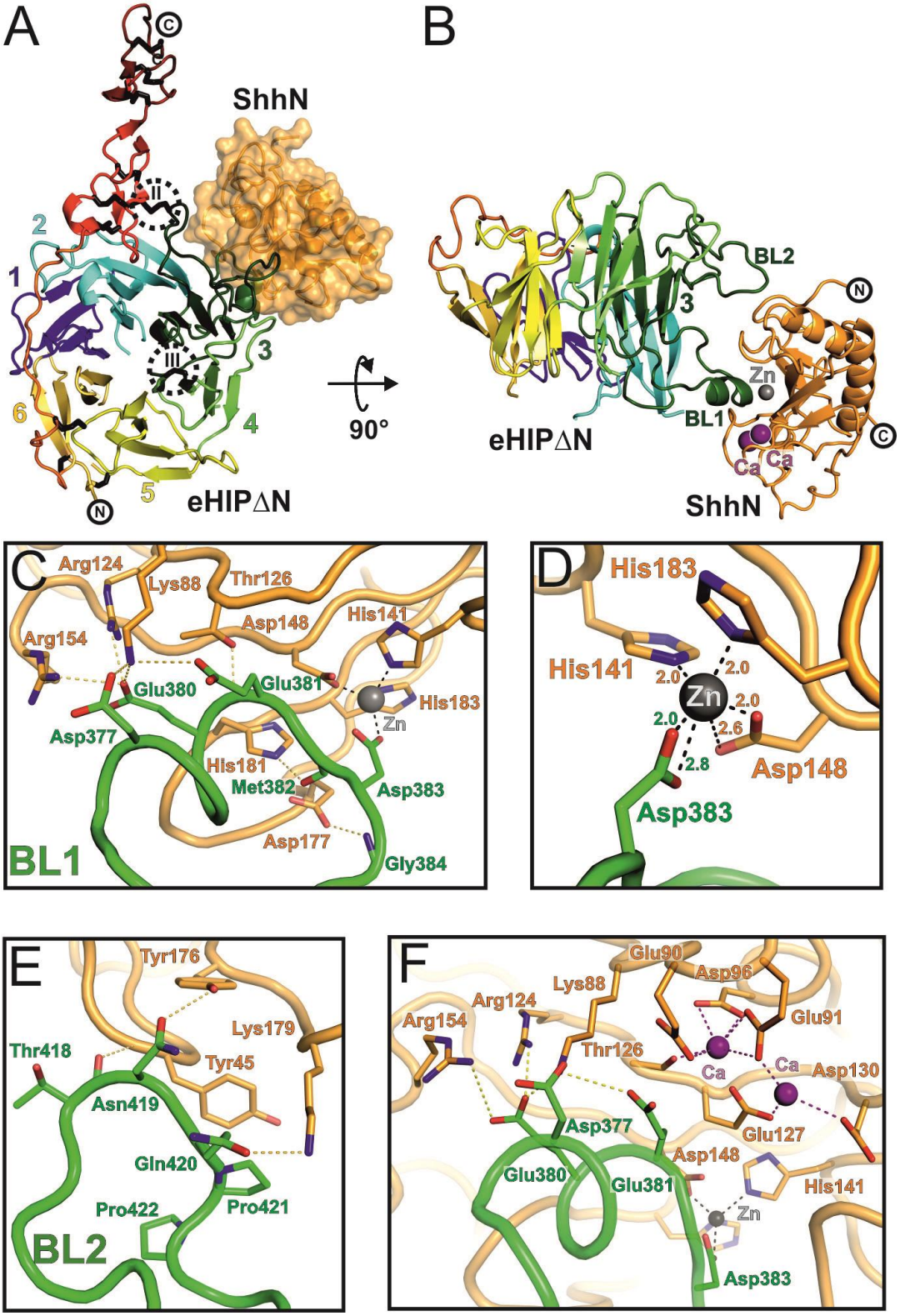


Fig. 4.5 Structure of the SHH-HHIP complex. (A) Cartoon representation of eHIPΔN (orientation, β-propeller blade numbering, and colour coding as in **Fig. 4.2**) in complex with SHHN depicted as solvent accessible surface (in orange). Disulphide bridges (II and III), stabilizing HHIP β-propeller blade 3, are highlighted (dotted circles). (B) Cartoon representation of the eHIPΔN-SHHN complex. The two binding loops (BL1 and BL2) of HHIP are labelled. The zinc ion is shown in grey and the two calcium ions in violet. (C-F) Close up views of the SHH-HHIP interactions. eHIPΔN (green) and SHHN (orange) main chains are shown as coils. Residues involved in complex interactions are drawn

in stick representation (oxygen: red, nitrogen: blue). The zinc ion is shown in grey and the calcium ions in violet. Potential hydrogen bonds are marked as yellow dotted lines, interactions with the zinc ion as grey dotted lines and with calcium as violet dotted lines. Contacts of HHIP-BL1 with SHHN are shown in C, the zinc-binding site is detailed in D, HHIP-BL2 interactions are shown in E and the properties of the calcium-binding site in F.

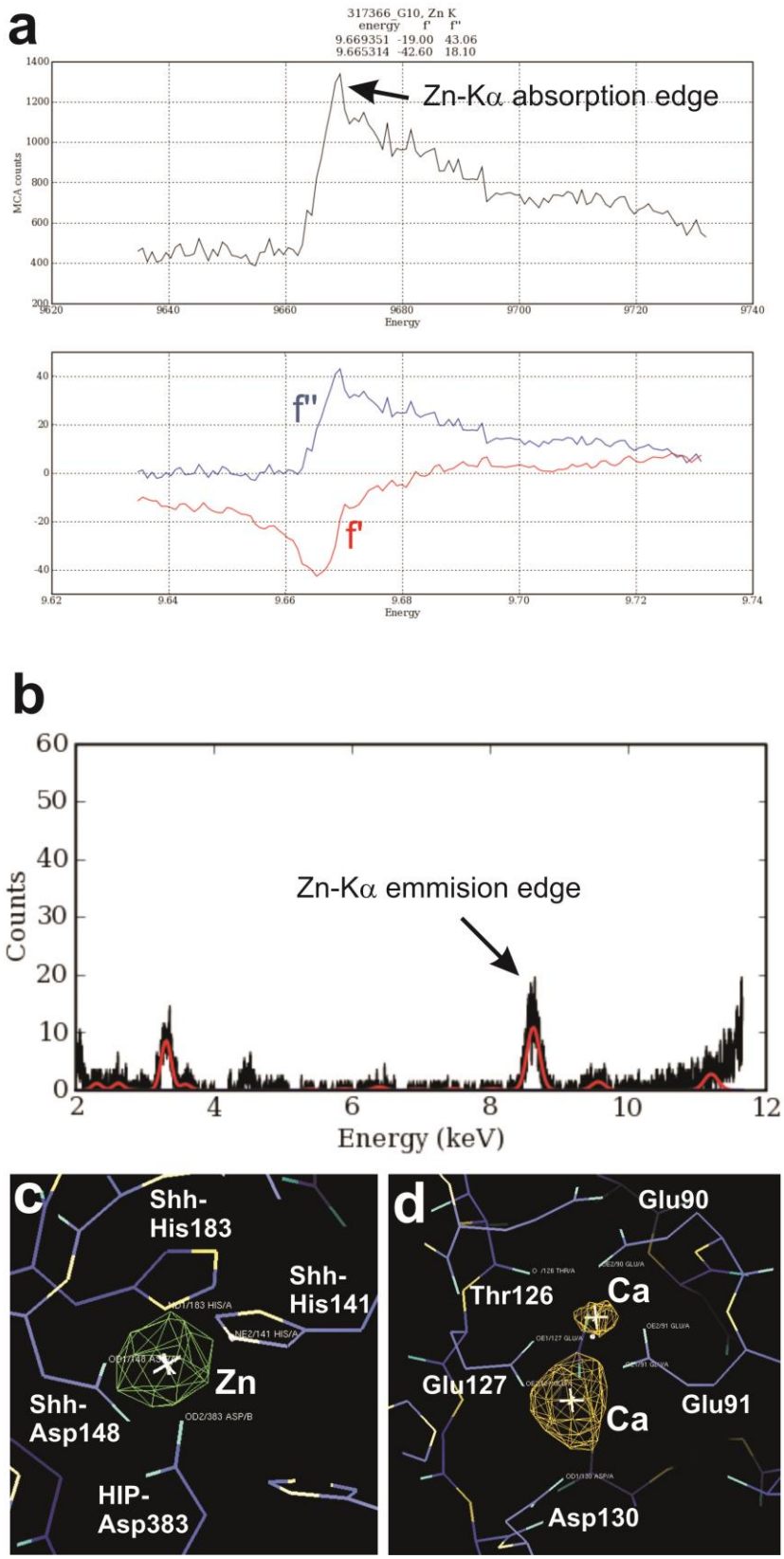


Fig. 4.6. Biophysical validation of the presence of zinc in HH ligand-HHIP complex crystals. (A) Fluorescence scan of a SHHN-eHIPΔN crystal around the Zn K-edge performed at ESRF beamline BM14. Upper panel is the X-ray absorption spectrum, lower panel shows the anomalous scattering factors f' and f''. The spectra clearly demonstrate the presence of zinc in the crystal. (B) X-ray fluorescence emission analysis of a DHHN-eHIPΔN complex crystal collected at ESRF beamline

ID29. The peak at the Zn-Kα egde clearly indicates the presence of zinc. In both cases, no zinc was added at any stage of protein purification or crystallization. (C) Close-up view of the HHIP-complexed SHHN zinc-binding site. The anomalous difference Fourier map based on the SHHN-eHIPΔN complex calculated to 3.2 Å and contoured at 6 σ is shown in green. (D) Close-up view of the HHIP-complexed SHHN calcium-binding site. The 3.2 Å Fo-Fc simulated annealed omit electron density map calculated without the two calcium ions at 4 σ is shown in yellow.

4.2.4 Importance of metal-binding sites for HHIP-HH interactions

Although the eHIPΔN-SHHN interface is relatively small and the surface complementarity score is at the lower end of the scale for physiological protein-protein interactions, binding constants in the nanomolar range were observed using both cell-based (Chuang & McMahon 1999) and surface plasmon resonance (SPR) experiments (**Fig. 4.7**). HHIP ectodomain (including the N-terminal cysteine-rich dimerisation domain; **Fig. 4.2A**) and eHIPΔN show similar HH-binding behaviour, although the full-length form has somewhat enhanced binding which may reflect its dimeric state (**Fig. 4.7**). For both HHIP constructs the presence of zinc in the HH metal-binding site is crucial (**Fig. 4.7**). Addition of 10 mM EDTA (a chelator for divalent cations) completely abolished HH-binding (**Fig. 4.7D**) or, in the presence of calcium, mutation of the zinc-coordinating HHIP-Asp383 (**Fig. 4.5C,D**) to arginine or alanine abolished or reduced HH-binding by greater than 25-fold, respectively (**Fig. 4.7A**). Thus the tight HH-HHIP interaction can be explained by the contribution of HHIP to the coordination of zinc in the HH ligand metal-binding site (**Fig. 4.5D**) (Hall et al. 1995) first identified zinc in the crystal structure of SHHN, and pointed out a similarity to the zinc-containing active sites of some metalloproteases. Zinc appears to play a crucial structural role, binding SHHN with very high affinity ($K_d \leq$ 100 pM) (Day et al. 1999), however, the HHIP-SHH structure reveals a second role. Zinc-binding is the key mediator of complex formation; indeed, the zinc-binding site is not conserved in *Drosophila*, which correlates with the absence of HHIP in flies.

The low affinity ($K_d \geq 100$ μM (McLellan et al. 2008)) calcium-binding site (**Fig. 4.5F**) adopts the same conformation as described by (McLellan et al. 2008) and has been found to be important for SHH binding to various cell-surface receptors. We confirmed this finding in SPR experiments where addition of 2 mM calcium caused a marked increase of the HH-HHIP binding affinity (K_d from 100 nM to 6 nM; **Fig. 4.7A-C,E**). However, our structure reveals that neither the two calcium ions nor the loops coordinating them contribute specific interactions to the complex interface suggesting a secondary role for the calcium ions in HHIP-binding. We therefore investigated the effect of calcium absence on the structure of HH ligands and, subsequently, on HH-HHIP structures. In the previously published uncomplexed SHHN structure (Hall et al. 1995), in the absence of calcium, the side chains of Glu90, Glu91 are oriented towards the solvent (**Fig. 4.8A**). To verify whether this applies to other HH ligands, we solved crystal structures of the N-terminal signalling domain of human DHH (DHHN, equivalent to the SHHN construct) with and without bound calcium. The overall architecture of DHHN, including the zinc-binding site, closely matches SHHN (r.m.s.d 0.5 Å for 150 Cα atoms; **Fig. 4.8A**) as expected from the high degree of sequence identity (77%). The conformations of the Glu90 and Glu91 side chains in the calcium-free forms of SHHN (Hall et al. 1995) and DHHN match, while for DHHN with calcium bound, these residues contribute to metal-coordination and adopt the conformation observed in the SHHN-eHIPΔN complex (**Fig. 4.8A, Fig. 4.9A-D**).

I produced, crystallized and solved the structures of SHHN-eHIPΔN and DHHN-eHIPΔN complexes in the absence of calcium (**Fig. 4.4**). The structures show that the

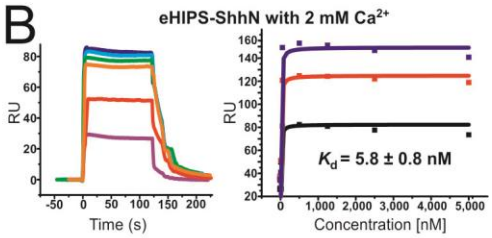
HH residues contributing to the HH-HHIP interface are highly similar across family members and vertebrate species (**Fig. 4.8B**), conserving the architecture of the ligand-receptor interfaces (**Fig. 4.8C, Fig. 4.9E-H**). In the absence of calcium the HH loop Lys88-Gly94 becomes disordered, suggesting that calcium-binding is important to prevent electrostatic repulsion between the negatively charged residues of HH loop Lys88-Gly94 and a negatively charged patch (formed by HHIP residues Asp377, Glu380, Glu381, Asp383) on the apposing HHIP surface (**Fig. 4.5F**). In the presence of calcium, stabilisation of the loop allows the side chains of HHIP-Glu381 and SHH-Glu90 to contribute to the complex by forming additional van-der-Waals interactions (**Fig. 4.8D**), consistent with the observed 20- to 60-fold increase in affinity between HH ligands and HHIP under these conditions (**Fig. 4.7A**).

A previously reported structure of SHHN in complex with the third fibronectin type III domain of CDO showed calcium to be essential for binding(McLellan et al. 2008) and the conformation of the calcium-binding site is essentially identical to the ones we observed in the DHHN and SHHN-eHIPΔN structures (**Fig. 4.9A-D**). A comparison of the SHH-HHIP and SHH-CDO complexes shows a partial overlap of the receptor-binding interfaces which includes residues of the SHH loop Lys88-Gly94; however, HHIP and CDO have very different orientations relative to the HH ligand (**Fig. 4.8E**). The SHH-CDO interface does not include direct interactions with the zinc-binding site and is less extensive, in agreement with solution binding assays: $K_d^{SHH\text{-}CDO}$ = 1300 nM (McLellan et al. 2008) and $K_d^{SHH\text{-}HHIP}$ = 6 nM (**Fig. 4.7A**). In our SHH-HHIP and DHH-HHIP complexes the zinc-binding site is crucial for complex formation and even in the absence of calcium the HH-HHIP affinities are 10-times higher than that reported for SHH-CDO (**Fig. 4.7A**).

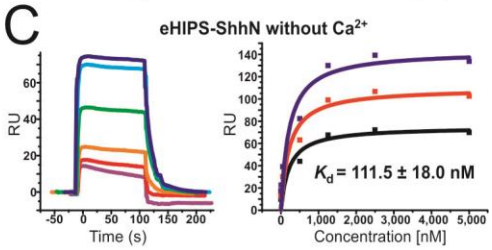
A

Interaction	K_d [nM]
eHIPS/ShhN with Ca^{2+}	5.8 ± 0.8
eHIPS/ShhN without Ca^{2+}	111.4 ± 18.0
eHIPS/ShhN with EDTA	ND
eHIPS/DhhN with Ca^{2+}	3.0 ± 0.5
eHIPS/DhhN without Ca^{2+}	202.9 ± 33.5
eHIPΔN/ShhN with Ca^{2+}	14.0 ± 1.6
eHIPΔN/ShhN without Ca^{2+}	73.9 ± 12.3
eHIPΔN/DhhN with Ca^{2+}	8.7 ± 1.1
eHIPΔN/DhhN without Ca^{2+}	73.6 ± 12.9
eHIPΔN-D383A/ShhN with Ca^{2+}	384 ± 11.6
eHIPΔN-D383A/ShhN without Ca^{2+}	ND
eHIPΔN-D383R/ShhN with Ca^{2+}	ND
eHIPΔN-D383R/ShhN without Ca^{2+}	ND
eHIPΔN-W310A/ShhN without Ca^{2+}	79.3 ± 3.8

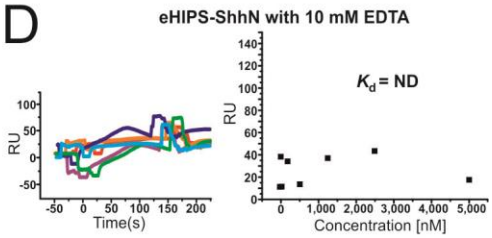
B



C



D



E

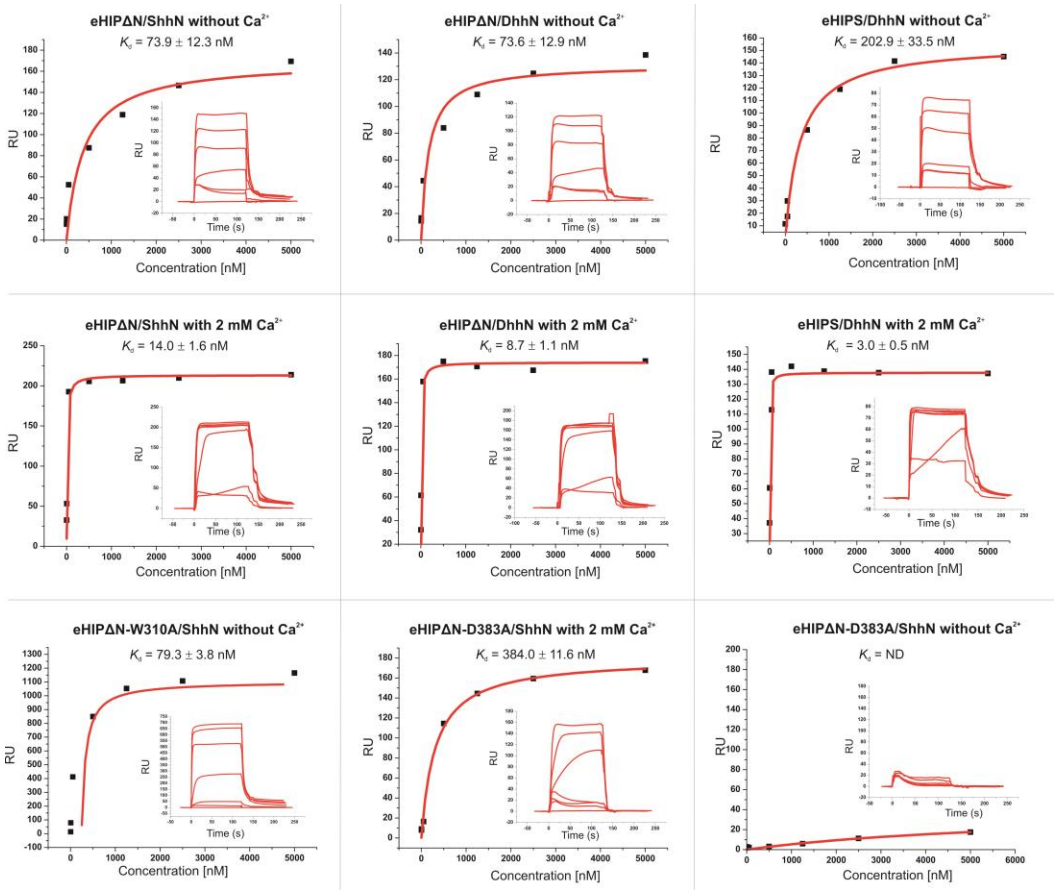


Fig. 4.7 Binding properties of HH-HHIP interactions. (A) Table of binding constants (K_d) measured by SPR between different HHIP constructs and SHHN or DHHN, respectively. Data are expressed as mean ± standard error. ND: not determinable. (B-D) Binding of the stabilized HHIP ectodomain (eHIPS) to SHHN. The left panels are representative sets of experimental sensorgrams from typical equilibrium-based binding experiments, with reference subtraction. Different concentrations of SHHN were injected over surfaces coupled with eHIP in the presence (B) or the absence (C) of calcium and in the presence of 10 mM EDTA (D), respectively. For all injections the primary experimental traces reached equilibrium and returned to baseline after the injection. The right panels are plots of the equilibrium binding response (RU) from the sensorgrams as a function of SHHN concentration (0.5 to 5000 nM). Each plot is derived from a different independent experiment, but within each plot, the three curves are derived from a single series of injections of SHHN over a biosensor chip in which three experimental surfaces have been coated with different amounts of eHIPS. This allowed the use of global fitting to improve data averaging. Best-fit binding curves with a global fit of the experimental data to a uniform Kd value are shown as coloured lines. (E) SPR equilibrium binding data of different HHIP constructs with SHHN or DHHN, respectively. Sensorgrams (insets) and plots of the equilibrium binding response (RU) from the sensorgrams as a function of SHHN or DHHN concentration (0.5 to 5000 nM) are displayed.

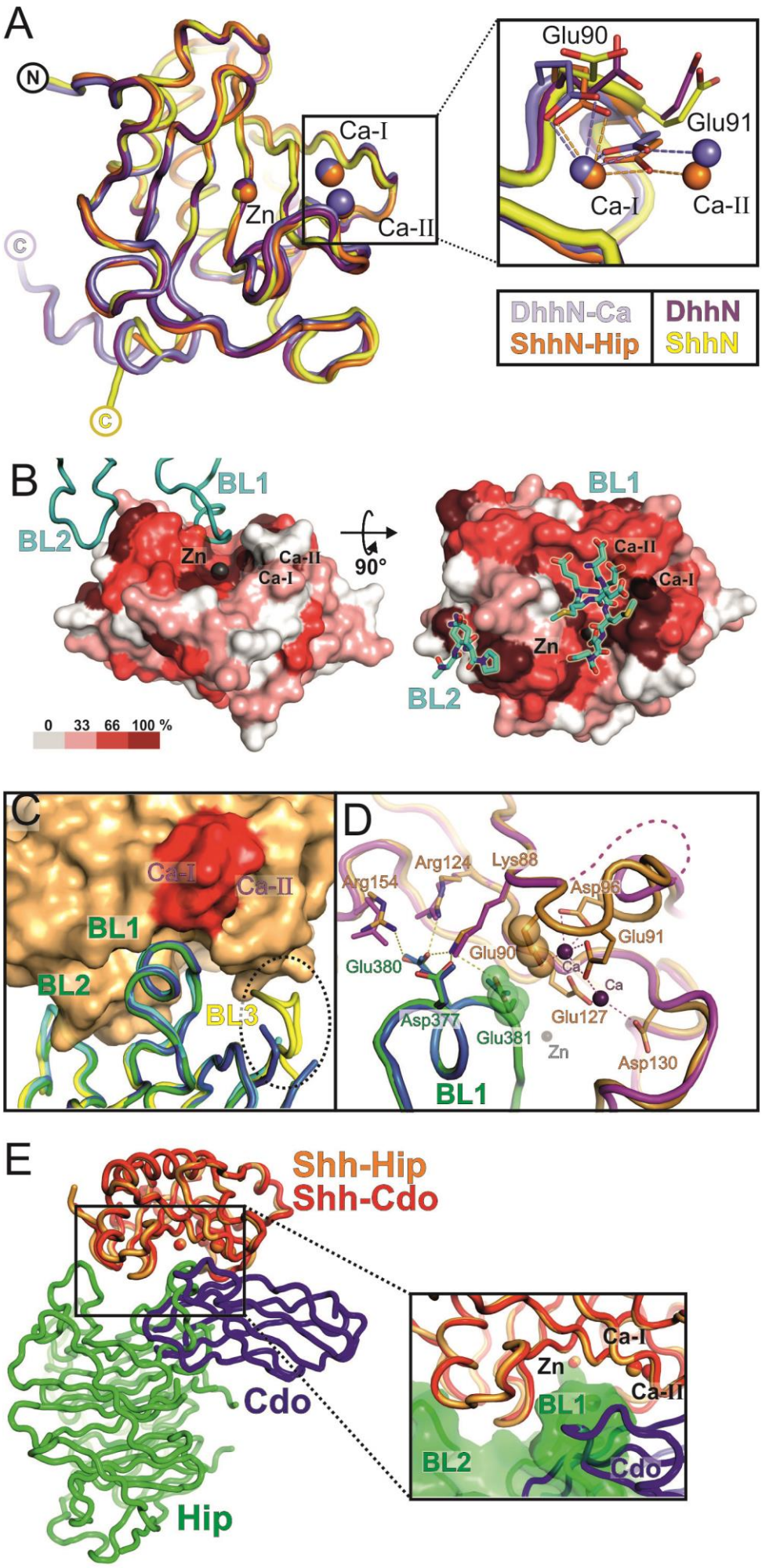
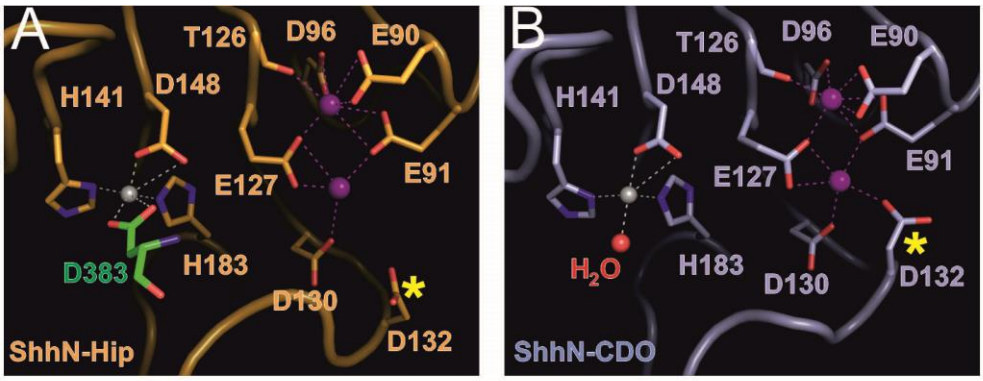
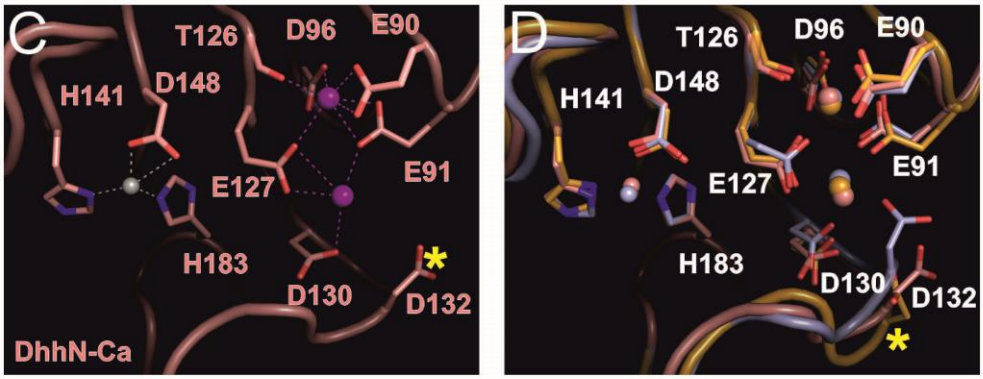
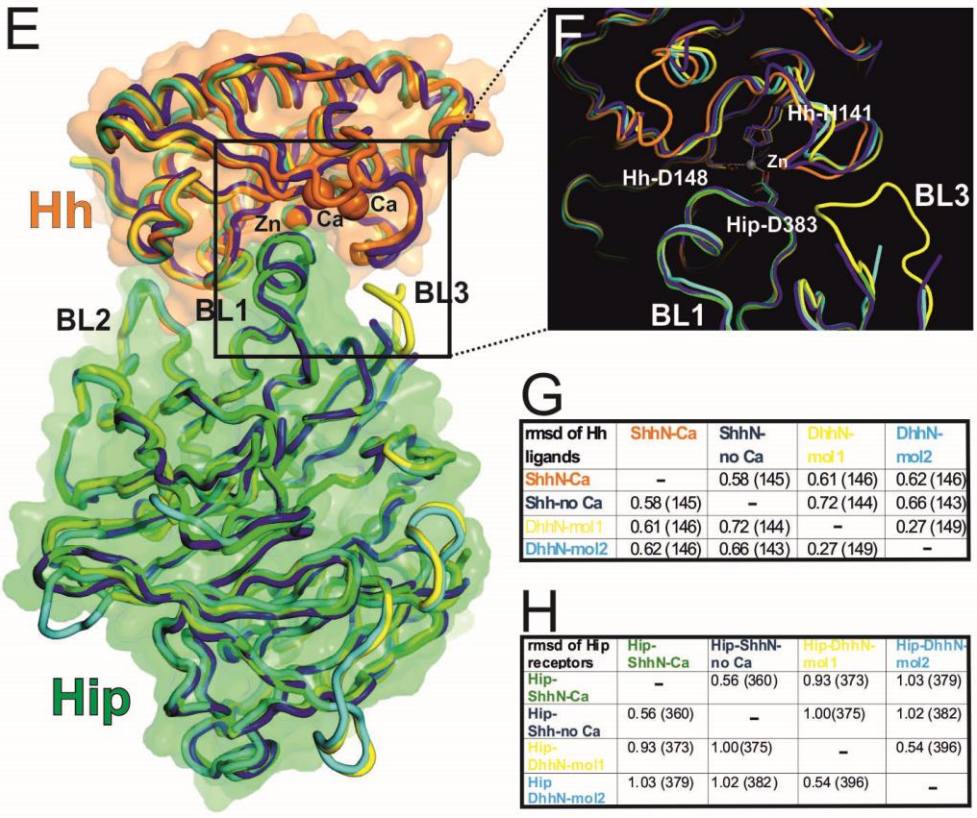

A
N
C
C
Zn
Ca-I
Ca-II
Glu90
Glu91
Ca-I
Ca-II
DhhN-Ca
ShhN-Hip
DhhN
ShhN
B
BL1
BL2
Zn
Ca-II
Ca-I
90°
BL1
Ca-II
Ca-I
Zn
BL2
0 33 66 100 %
C
Ca-I
Ca-II
BL1
BL2
BL3
D
Arg124
Lys88
Arg154
Asp95
Glu91
Glu90
Glu380
Ca
Ca
Glu127
Asp377
Glu381
Zn
Asp130
BL1
E
Shh-Hip
Shh-Cdo
Cdo
Hip
Ca-I
Zn
Ca-II
BL1
BL2
Cdo

Fig. 4.8 Molecular determinants of HH-receptor interactions. (A) Structural superpositions. DHHN with calcium (slate) and without (magenta), and SHHN with calcium (orange) and without (yellow; PDB 1VHH) are shown as coils. Zinc and calcium ions are depicted as spheres. The close-up highlights loop Lys88-Gly94, Glu90 and Glu91 (stick representation) change conformation upon calcium-binding. (B) SHHN solvent-accessible surface coloured by residue conservation. HH-binding loops of HHIP are shown in cyan. (C) Comparison of SHHN- and DHHN-HHIP complexes. SHHN surface is coloured orange. SHH loop Lys88-Gly94, ordered only in the calcium-bound SHHN-HHIP, is highlighted in red. HH-binding loops of HHIP are shown as coils (SHHN-eHIPΔN with calcium: green; SHHN-eHIPΔN without calcium: blue; DHHN-eHIPΔN without calcium - molecule 1: yellow, molecule 2: cyan). A third ligand-binding loop of HHIP (BL3) is observed in only one copy of the DHHN-eHIPΔN asymmetric unit (dotted ellipse). (D) Effects of calcium-binding. SHHN-eHIPΔN complexes are shown as coils (with calcium: orange/green, without calcium: violet/blue). Spheres highlight HHIP-Glu381 and SHH-Glu90. SHH-loop Lys88-Gly94 is depicted as a dotted line. (E) Comparison of the SHH-HHIP and SHH-CDO (PDB 3D1M) complexes. SHH ligands are superimposed (HHIP: green, CDO: blue, HHIP-complexed SHHN: orange, CDO-complexed SHHN: red). Metal ions are depicted as spheres.







G

rmsd of Hh ligands	ShhN-Ca	ShhN-no Ca	DhhN-mol1	DhhN-mol2
ShhN-Ca	-	0.58 (145)	0.61 (146)	0.62 (146)
Shh-no Ca	0.58 (145)	-	0.72 (144)	0.66 (143)
DhhN-mol1	0.61 (146)	0.72 (144)	-	0.27 (149)
DhhN-mol2	0.62 (146)	0.66 (143)	0.27 (149)	-

H

rmsd of Hip receptors	Hip-ShhN-Ca	Hip-ShhN-no Ca	Hip-DhhN-mol1	Hip-DhhN-mol2
Hip-ShhN-Ca	-	0.56 (360)	0.93 (373)	1.03 (379)
Hip-Shh-no Ca	0.56 (360)	-	1.00(375)	1.02 (382)
Hip-DhhN-mol1	0.93 (373)	1.00(375)	-	0.54 (396)
Hip DhhN-mol2	1.03 (379)	1.02 (382)	0.54 (396)	-

Fig. 4.9 Comparison of SHHN-HHIP and DHHN-HHIP complexes and metal-binding properties of HH ligands. (A-D) Close up views of (A) HHIP-complexed SHHN (orange), (B) CDO-complexed SHHN (slate) and (C) DHHN in the calcium bound form (salmon). Residues involved in metal interactions are drawn in stick representation (oxygen: red, nitrogen: blue). Asp383 from the HHIP-BL1 loop is shown in green. The zinc ion is shown as a grey sphere, the calcium ions as violet spheres. Interactions with the zinc ion is depicted by grey dotted lines and with calcium by violet dotted lines. Asp132, which is coordinating calcium only in the CDO-complexed SHHN, is marked with a yellow asterisk. (D) Superposition of the three binding sites depicted in A-C. (E) Superposition of SHHN-HHIP with calcium (orange/green) and without calcium (blue), and DHHN-HHIP asymmetric unit molecule 1 (yellow) and asymmetric unit molecule 2 (cyan), respectively, are calculated using the HHIP receptor as template. SHHN-HHIP with calcium is highlighted as solvent accessible surface (probe radius 1.4 Å). HHIP loops BL1, BL2 and BL3 interacting with the HH ligand are depicted. Zinc and calcium ions are shown as spheres. In the DHHN-eHIPΔN crystal there are two copies of the complex in the asymmetric unit, one of which reveals additional interactions that involve a third ligand binding loop (BL3) of HHIP. BL3 is disordered in all other complexes. Since in SPR assays both SHHN and DHHN bind HHIP with similar affinities and mutation of the main interface residue (Trp310 to alanine) has no significant effect (**Fig. 4A**), it appears that the BL3 interaction is caused by crystal packing. (F) Close-up view of the HH-HHIP interface. Residues coordinating the zinc are shown in stick representation in atomic colouring (oxygen: red, nitrogen: blue). The zinc is depicted as a grey sphere. (G,H) Tables of root mean square deviations (rmsd) of the different HH ligands (G) and HHIP receptors (H) of the HH-HHIP complexes. Brackets indicate the number of Cα atoms used for rmsd calculations. Colour coding is as in (A).

4.3 Discussion

The structural similarity between the HH zinc-binding site and the active site of zinc-hydrolases strongly suggests that these classes of proteins are evolutionarily related. However, alteration of SHH residues, which by analogy should be essential for hydrolytic activity, did not impair signalling via PTCH, thus ruling out a functional requirement for enzymatic catalysis (Fuse et al. 1999). The fact that this site is not conserved in any *Drosophila* species sequenced to date, although it is maintained in all other closely related HH orthologues from insects and other invertebrates, further suggests that it is not required for the canonical signalling pathways such as PTCH. Nevertheless, vertebrates appear to have exploited the presence of this zinc-binding site to diversify the repertoire of HH interacting proteins to include the HHIP receptor, thus introducing an additional level of regulation in this signalling pathway.

Mutations of SHH-His140 and -Cys183 (residues directly or indirectly involved in zinc-chelation) cause holoprosencephaly (HPE) and further facial anomalies as well as cancer in human patients (Orioli et al. 2001; Ribeiro & Richieri-Costa 2005). The SHHN-CDO interaction appears to be zinc independent (McLellan et al. 2008) and, as discussed above, *Drosophila* HH lacks the zinc binding site. Although site-directed mutagenesis of residues forming the zinc-binding site led to a marked decrease of SHH potency in a cellular differentiation assay, it was not clear whether this effect was due solely to a decrease in protein stability (Day et al. 1999). However, since functional activity was not completely abolished, interactions with other vertebrate HH receptors (e.g. PTCH) may not be entirely dependent on the presence of zinc. In contrast, zinc clearly plays a central role in HH ligand-binding to HHIP, thus the zinc-

and calcium-binding sites have distinct functions in defining the specificities of HH interactions with cell-surface receptors.

For the HH-HHIP interaction our results reveal a secondary role for the calcium-binding sites. The local concentration of extracellular calcium is highly variable (Hofer & Brown 2003), thus the occupancy of the low affinity-binding site in HH may vary. The juxtaposition of one metal-binding site, which plays the lead role in protein-protein interaction, with additional sites, which modulate the binding affinity, is strongly reminiscent of ligand-binding by the integrin family of cell adhesion molecules. For example, the structure of the integrin αIIbβ3 – fibrinogen γC peptide interface is centred on the integrin metal ion-dependent adhesion site (MIDAS; a high affinity magnesium-binding site) but has direct contributions from the adjacent to MIDAS (ADMIDAS) calcium-binding site (Springer et al. 2008). In general, for the integrin system, the MIDAS metal ion has been shown to have a direct role in ligand-binding while the ADMIDAS metal ion has a regulatory function. Similarly, our results suggest that the interplay between the zinc- and calcium-binding sites in vertebrate HH ligands may allow extracellular calcium concentration to play an important modulatory role.

Chapter 5

Glypican 3

Chapter 5 Glypican 3

ABSTRACT

GPC3 plays an important functional role in (A) development and HH signalling when in normal circumstances, and (B) in disease when mutated; a possible function of aberrant HH signalling. This chapter reveals the crystal structure solution and description of the GPC3 core domain. Solving this structure required homology modelling and normal mode analysis. The crystal structure of the GPC3 core domain allows a comparison between the structures of other glypican family members. This analysis reveals the rearrangement of GPC3 lobes resulting in structural plasticity, one reason for difficulties in solving the structure of the GPC3 core domain. Structural analysis of GPC3 also allowed classification and mapping of disease-related mutations and provides a rationale for disease mechanism. GPC3 biophysical binding studies show that the GPC3 core domain does not bind directly to SHH. Moreover the HS chains seem to play an important role in GPC3-mediated HH regulation.

5.1 Introduction

5.1.1 Heparan Sulphate Proteoglycans

Heparan sulphate proteoglycans (HSPGs) are secreted or membrane-bound cell-surface macromolecules composed of a protein core to which heparan sulphate (HS) glycosaminoglycan (GAG) chains are attached (Bernfield et al. 1999). On the basis of their core protein structure, HSPGs can be classified into several families (**Fig. 5.1**). Glypicans and syndecans are cell surface HSPGs, linked to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor or a transmembrane domain, respectively. Perlecans are secreted HSPGs that are distributed predominantly in the

extracellular matrix (ECM). Glypicans and perlecans bear exclusively HS GAG chains, while syndecans are modified with both HS and chondroitin sulphate (CS). All three families of HSPGs are evolutionarily conserved from vertebrates to *Drosophila* (Yan & X. Lin 2009).

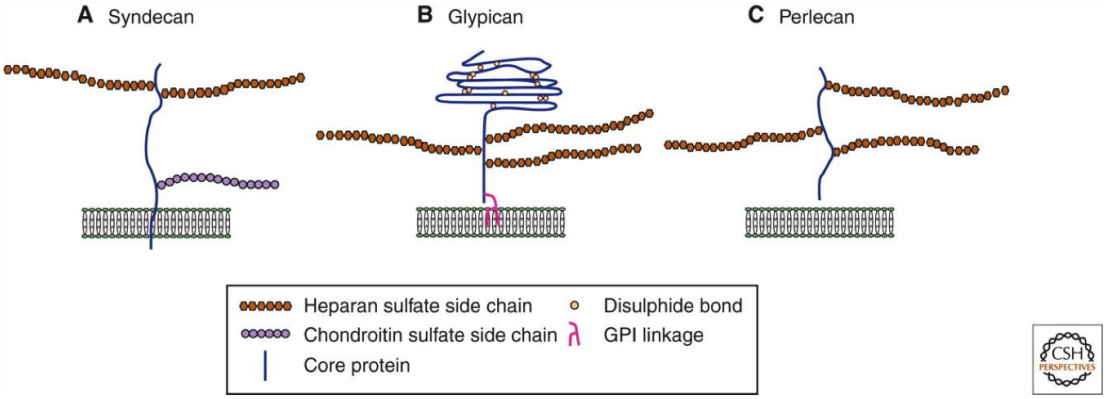


Figure 5.1 The three main classes of HPSGs. (A) Syndecan core proteins are transmembrane proteins that contain a highly conserved carboxy-terminal cytoplasmic domain. Heparan sulphate (HS) chains attach to serine residues distal from the plasma membrane. Some syndecans also contain chondroitin sulphate (CS) chain(s) that attaches to serine residue(s) near the membrane. (B) Glypicans are composed of a globular, disulphide-stabilized core domain that is bound to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor. HS chains link to serine residues adjacent to the plasma membrane. (C) Perlecans are secreted HSPGs that carry HS chains(Yan & X. Lin 2009). Figure is adapted from (Yan & X. Lin 2009).

HSPGs can carry one or more HS chains (Esko & Selleck 2002). These chains comprise long, linear polymers of glucuronic acid/N-aceltylglucosamine repeating disaccharide units. The length of the chains is highly variable, and a significant proportion of the disaccharides are modified by N-deacetylation/N-sulphation, epimerization and O-sulphation (Kreuger et al. 2006). This sulphation gives a highly negative charge to the HS chains that enables HSPGs to interact with a large range of proteins that display positively charged domains. These interacting proteins include growth factors such as: fibroblast growth factors (FGFs), WNTs, Hedgehogs (HHs)

and bone morphogenetic proteins (BMPs) (Kirkpatrick & Selleck 2007; Filmus & Capurro 2008).

Most of the *in vivo* evidence published so far indicates that the main function of membrane-attached glypicans is to regulate the signalling of WNTs, HHs, FGFs, and BMPs (Filmus et al. 2008). For example, mice that carry a homozygous deletion of the glypican GPC3 display alterations in WNT and HH signalling (Song et al. 2005; Capurro et al. 2008), and *Drosophila* glypican mutants have defective HH, WNT, BMP and FGF signalling in specific tissues (X. Lin & Perrimon 1999; Ohkawara et al. 2003; Desbordes & Sanson 2003; Jackson et al. 1997). Furthermore, GPC3 promotes the growth of hepatocellular carcinoma cells by stimulating WNT signalling (Capurro, Xiang, et al. 2005b). The function of glypicans is not limited to the regulation of growth factor activity. For example Dally-like protein (DLP), a *Drosophila* glypican, has been shown to play a role in synapse morphogenesis and function by binding and inhibiting the receptor phosphatase LAR (K. G. Johnson et al. 2006) In addition, it has been proposed that glypicans are involved in the uptake of polyamines (Fransson 2003; Filmus et al. 2008).

5.1.2 Glypicans

Glypicans can be found throughout the metazoa. Six glypican family members are present in *Homo sapiens* (GPC1 to GPC6) and *Mus musculus*, while there are no clear GPC homologues outside of the metazoa (Filmus et al. 2008). These six glypicans can be placed into two broad sub-families based on their sequence: glypicans 1, 2, 4 and 6 and glypicans 3 and 5. A single representative of each group is found in *Drosophila*;

DLP, an orthologue of mammalian glypicans 1, 2, 4 and 6, and DALLY, an orthologue of mammalian glypicans 3 and 5 (Filmus et al. 2008).

5.1.3 Invertebrate Glypicans and Hedgehog

In 2003 and 2004 three groups reported that the *Drosophila* glypican DLP regulates HH signalling by stimulating signal reception, and by playing a critical role in the transport of the HH morphogen in the *Drosophila* wing (C. Han et al. 2004; Desbordes & Sanson 2003). One of these reports also implicated DALLY in HH morphogen gradient formation (C. Han et al. 2004). More recently, binding of SHH to a DLP construct lacking the GAG attachment was undetectable, despite being able to rescue the HH response in the presence of DLP RNAi (Williams et al. 2010), suggesting the HS attachment sites are not essential for DLP function. DALLY has also been implicated in long range signalling, positively regulating apical HH levels in producing cells, promoting long range HH spreading when DALLY is cleaved form the cell surface (Ayers et al. 2010).

5.1.4 Vertebrate Glypicans and Hedgehog – GPC3 interactions

GPC3 involvement in HH signalling is by far the best characterised in terms of both function and disease association. There is a report that also implicates GPC5 as a positive regulator of HH signalling, by localisation of GPC5 to the primary cilium where its HS chains bind both SHH and PTCH (F. Li et al. 2011a).

GPC3 has been shown to regulate embryonic growth by inhibiting the Hedgehog (HH) signalling pathway (Capurro et al. 2008). GPC3 can bind to HH at the cell membrane preventing HH from binding with PTCH. Additionally, binding of HH to

GPC3 triggers endocytosis and degradation of the GPC3-HH complex, thus reducing the amount of HH available for binding to PTCH(Capurro et al. 2008). Furthermore, GPC3-null mice display increased SHH and IHH protein expression levels, while having comparable amounts of the corresponding transcripts to that seen in normal litter mates (Capurro et al. 2008; Capurro et al. 2009). Studies performed in GPC3-expressing mouse embryonic fibroblasts showed that soon after internalization, the GPC3-HH complex can be localised in early endosomes and, subsequently, in the late endosome/lysosome compartment, which is in agreement that incubation of mouse embryonic fibroblasts with SHH-containing medium significantly reduces the cellular levels of GPC3. This further supports the model of GPC3-HH complex endocytosis as a way of reducing the HH signal (Capurro et al. 2008).

5.1.5 Mutations in glypican 3 cause human disorders

GPC3 loss-of-function mutations cause the X-linked Simpson-Golabi-Behmel overgrowth syndrome (SGBS), a disorder characterised by both pre- and post-natal overgrowth a wide range of visceral and skeletal abnormalities, resulting in an increased risk of developing embryonic tumours (Pilia et al. 1996; Neri et al. 1998; Filmus & Capurro 2008). This was further supported by the developmental overgrowth displayed by GPC3-deficient mice (Cano-Gauci et al. 1999; Chiao et al. 2002). In addition, GPC3 polymorphisms have a significant impact on the body size of mice (Oliver et al. 2005; Capurro et al. 2012). GPC3 also has implications in cancer: data from several groups suggest that GPC3 expression is frequently down regulated in breast, ovarian and lung cancer (Xiang et al. 2001; H. Lin et al. 1999; H. Kim et al. 2003). Of particular interest to this thesis, is the reported increase in HH signalling in many cases of these cancers (Yuan et al. 2007; Kubo et al. 2004; X.

Chen et al. 2007b). It is, therefore, possible that the loss of GPC3 may contribute to cancer progression by activating HH signalling (Filmus & Capurro 2008).

Knowledge of the structural architecture of GPC3, can therefore help decipher the disease mechanisms underlying GPC3 mutation. Furthermore, the molecular determinants of the GPC3-HH interaction are evidently crucial in both normal development and disease. This chapter aims to address three questions: (1) What is the structure of GPC3, (2) how do disease mutants affect the structure of GPC3, and (3) what is the nature of the HH-GPC3 interaction?

5.2 Results

5.2.1 Design and production of human GPC3 constructs

Three human GPC3 constructs for secreted expression in HEK 293T cells (see methods; Chapter 2, paragraph 2.2.8) were designed, based on the predicted sites for two HS attachment sites (at Ser495 and Ser509) and the predicted GPI anchor site at Ser560 (using http://mendel.imp.ac.at/gpi/cgi-bin/gpi_pred.cgi) (Eisenhaber et al. 2000) (**Fig. 5.2A**). The decision to make constructs lacking the GPI anchor was to facilitate secretion rather than retention at the cell membrane (which would have made purification much more difficult), while HS attachment sites were removed from other constructs to improve the crystallizability of protein targets. All three constructs could be successfully expressed and secreted in HEK 293T cells (**Fig. 5.2B**). The larger molecular bands observed for GPC3-Δ1/2GAG and GPC3-ΔGPI indicate the presence of HS attachment sites, while GPC3-ΔGAG shows a band corresponding to the globular core domain alone indicating a complete lack of HS attachment.

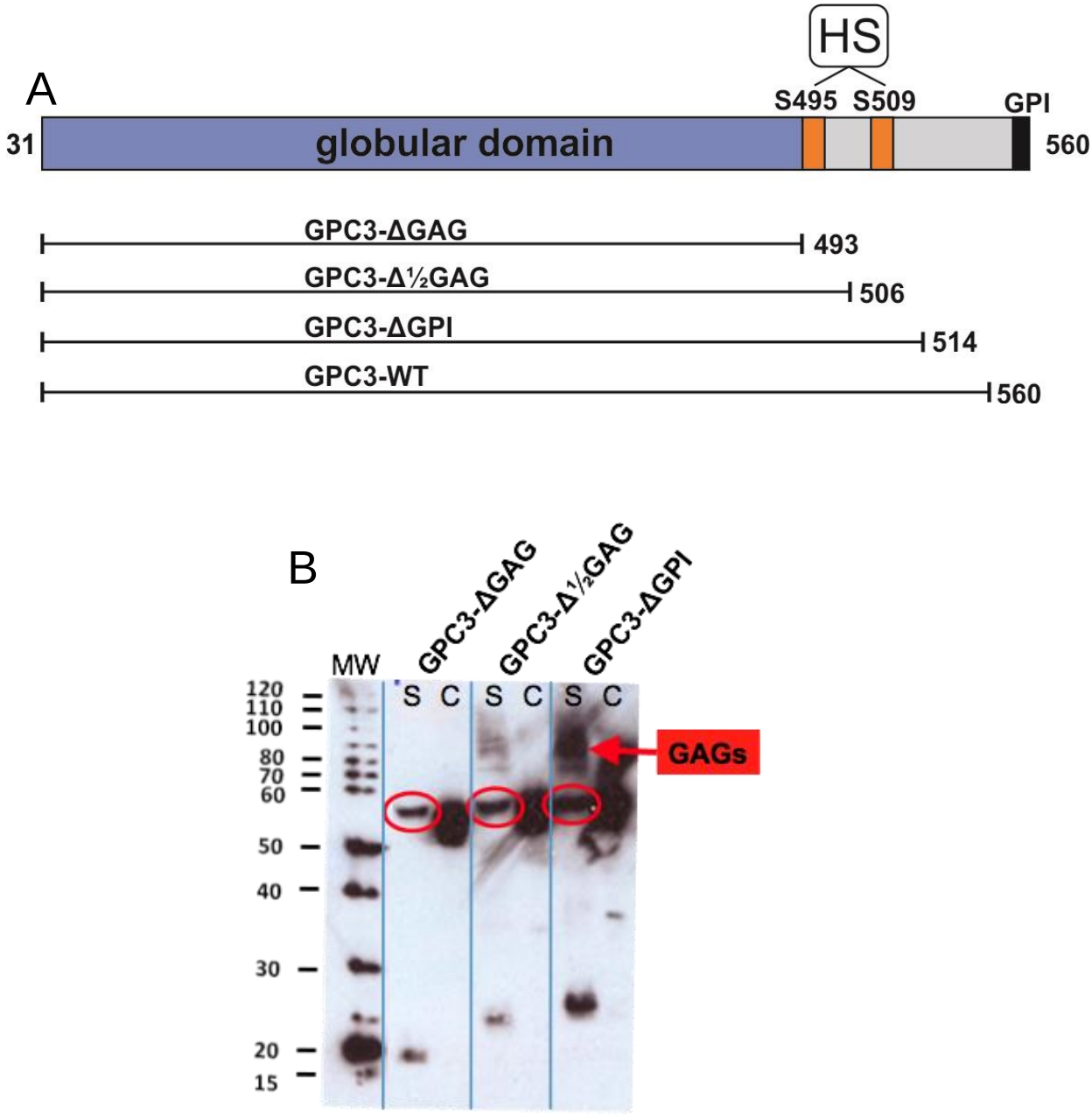


Figure 5.2 Construct design and expression of GPC3 constructs. (A) Domain organisation of human GPC3. The globular protein core domain is in blue. HS attachment sites are shown in orange. The constructs used in this study are shown below the schematic. **(B)** Western blot analysis of GPC3 constructs under reducing conditions expressed in HEK293T cells (S: conditioned medium; C: cell pellet), probed with an anti-penta-His antibody (see methods; Chapter 2, paragraph 2.2.10). The three constructs contain a C-terminal hexa-histidine tag. GPC3-Δ1/2GAG and GPC3-ΔGPI showed also a higher molecular weight band, likely attributed to the presence of GAGs (marked with a red arrow). Red circles indicate the protein bands corresponding to the expected protein contribution only. Molecular weight (kDa) is shown on the left-hand side of the blot.

5.2.2 Purification and crystallisation of GPC3-ΔGAG

Large-scale expression in HEK 293T cells and subsequent purification via IMAC and SEC (see methods; Chapter 2, paragraph 2.2.8) resulted in the observation of three bands on a reduced SDS-PAGE (one band showing the expected molecular weight and two smaller molecular weight bands, which combined equal the mass of the former band) (**Fig. 5.3**), presumably due to cleavage at a potential furin-like cleavage site. This is in agreement with (1) SEC analysis resulting in a single monodisperse species (**Fig. 5.3A**) and, (2) non-reduced SDS-PAGE analysis showing a band at the expected molecular weight while the smaller band of the cleavage product is no longer present (**Fig. 5.3C**). The GPC3 core protein domain harbours 14 highly conserved cysteine residues, likely forming disulphide bridges, which potentially maintain the structural integrity of the core domain under non-reducing conditions, while under reducing conditions and potential furin-cleavage, result in the observation of multiple bands on SDS-PAGE analysis (**Figure 5.3).**

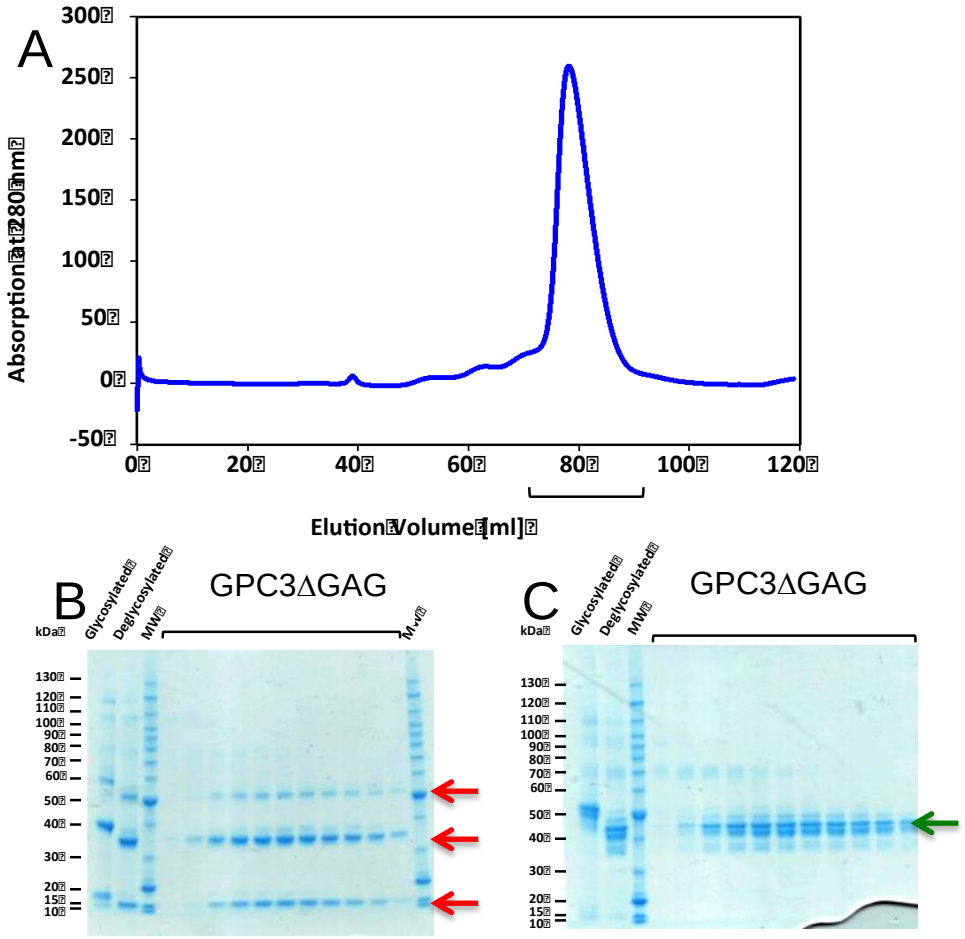


Figure 5.3 Large scale purification of GPC3-ΔGAG from HEK293T cells. (A) SEC profile of purified GPC3-ΔGAG. SEC resulted in a single monodisperse species. **(B)** Reduced and **(C)** non-reduced SDS-PAGE analysis revealed multiple bands in the reduced sample (indicated by red arrows), but only one band in the non-reduced sample (indicated by a green arrow) suggesting proteolytic cleavage.

5.2.3 Crystallization and structure determination of the human GPC3 core domain

Native GPC3-ΔGAG crystallized form mother liquor 0.1M Tris Hydrochloride (pH 8.2), 10% v/v Jeffamine ED2001 (O,O'-Bis(2-aminopropyl) polypropylene glycol-block-polyethylene glycol-block-polypropylene glycol 1,900), pH 7.0, resulting in small cube-shaped crystals (**Fig 5.4A,** see methods, Chapter 2, paragraph 2.6.2 for details). These crystals allowed the collection of a native data set to 3.2 Å resolution

(**Fig. 5.4B** and **Table 5.1**, see methods; Chapter 2, paragraph 2.6.2). At this time, there were no structural homologues of GPC3 available in the PDB database (http://www.pdb.org), so strategies using experimental phasing techniques were explored. Attempts to produce SeMet-labelled protein for crystallisation failed because the yields were too low. Furthermore, heavy atom (HA) soaking for structure determination using single isomorphous replacement with anomalous scattering (SIRAS) failed because of the fragility of the GPC3 crystals – addition of any HA solution tested resulted in the breaking of the crystal and subsequent loss of X-ray diffraction capability.

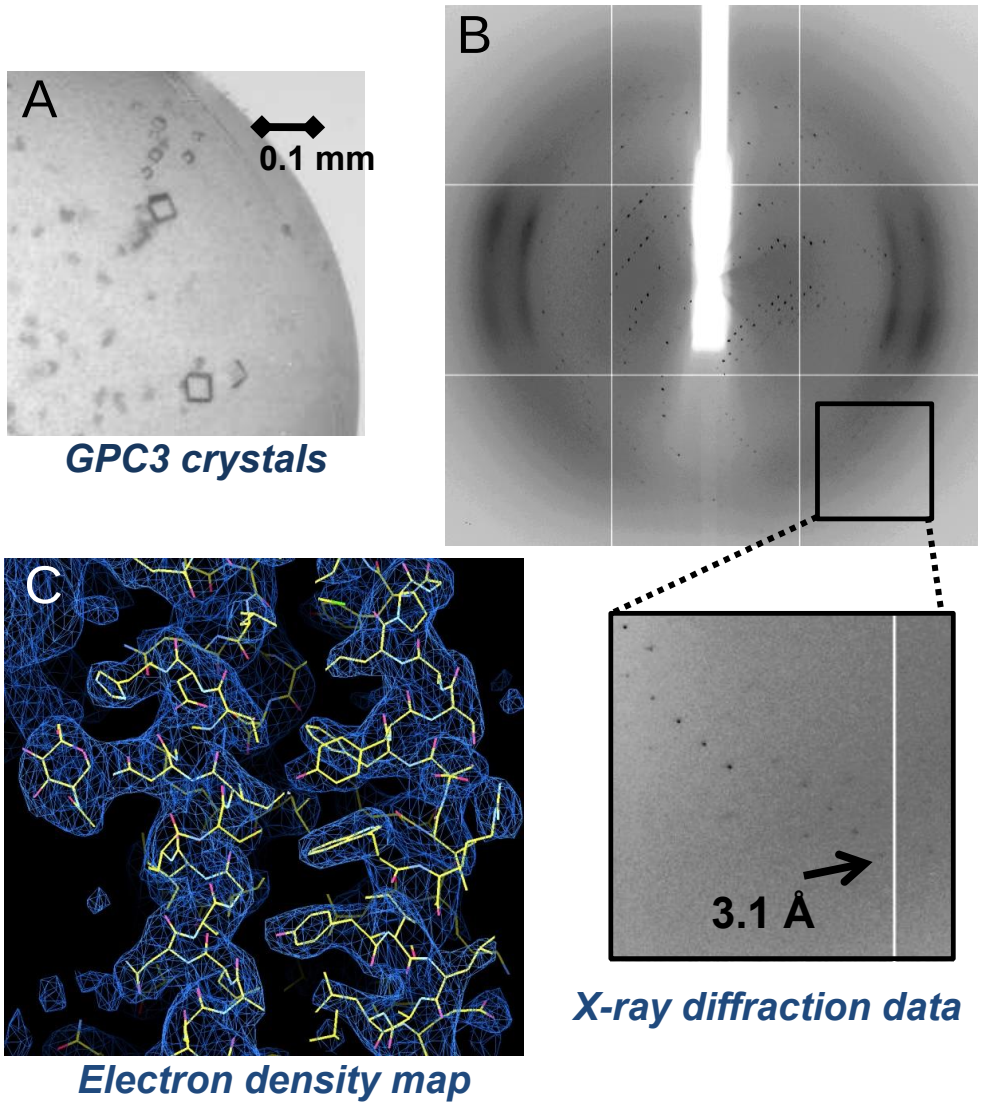


Fig. 5.4 Crystallisation and data collection of GPC3-ΔGAG (A) GPC3 crystals with scale bar (B) Diffraction image of a GPC3 crystal collected at DIAMOND-I03. (C) SigmaA-weighted $2F_O$-F_C map of the final GPC3 model from autoBUSTER(G et al. 2011) contoured at 1.2σ.

Table 5.1. Data collection and refinement statistics of GPC3-ΔGAG

DATA COLLECTION	
X-ray source	DIAMOND-I03
Resolution [Å]	85-3.2 (3.3-3.2)
Space group	$P3_121$
Cell Dimension [Å]	a = b = 99.1 c = 88.6
Wavelength [Å]	0.9763
Unique reflections	8549 (638)
Completeness [%]	99.9 (99.9)
Rmerge [%]	12.9 (94.9)
I/σI	15.5 (2.7)
Redundancy	9.0 (9.1)
REFINEMENT	
Resolution range [Å]	50.0-3.2 (3.6-3.2)
Number of reflections	8597 (2288)
No. atoms (Protein/NAG)	2917/42
B factors [$Å^2$] (Protein/ NAG)	80/101
Rfactor [%]	18.5 (21.0)
Rfree [%]	24.6 (28.5)
r.m.s.d. bonds [Å]	0.010
r.m.s.d. angles [deg]	1.2
Ramachandran statistics	
Favoured [%]	96.4
Disallowed [%]	0

r.m.s.d. stands for root mean square deviation from ideal geometry. Numbers in parentheses refer to the appropriate outer shell.

[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I(hkl;i) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I(hkl;i)$, where $I(hkl;i)$ is the intensity of an individual measurement and $\langle I(hkl) \rangle$ is the average intensity from multiple observations.

[b] $R_{factor} = \Sigma_{hkl} ||F_{obs}| - k|F_{calc}|| / \Sigma_{hkl} |F_{obs}|$.

[c] R_{free} equals the *R*-factor against 5% of the data removed prior to refinement.

[d] from MolProbity (V. B. Chen et al. 2010)

During the process of data collection and attempts to experimentally phase the GPC3 structure, the group of Professor Daniel Leahy (Johns Hopkins University School of Medicine, Baltimore, MA) published the crystal structure of the *Drosophila* glypican DLP (M.-S. Kim et al. 2011). Initial molecular replacement (MR) attempts using the DLP structure as a search model failed, because of the low sequence identity (some 18% sequence identity between the core domains of DLP and GPC3), thus a more convoluted method was necessary: a homology model of GPC3 using program MODELLER (Söding et al. 2005) was prepared using the DLP structure (PDB Id. 3ODN) as a template. Side chains of the surface exposed residues of the homology model were changed to alanine and loop regions were omitted. The truncated model was used to generate a set of 55 normal mode perturbed models for MR using elNémo web server (Suhre & Sanejouand 2004) with default settings (**Fig. 5.5**) (with the help of T. Malinauskas, STRUBI Oxford). 5 out 55 models were correctly placed into the unit cell using Phaser (McCoy et al. 2007) with translation function Z scores (TFZ) in the range of 6.5-8.4. The MR solution with the highest score (TFZ=8.4) was further used for refinement. The RMSD between the best homology model and the final structure is 1.86 Å for 294 Cα equivalents, whereas comparison of the original homology model from MODELLER (Söding et al. 2005) without normal mode calculation with the final structure resulted in a RMSD of 3.09 Å for the same Cα range. This clearly shows the importance of the normal mode analysis for MR applied here. An initial model was then built using iterative cycles of automatic building in Buccaneer(Cowtan 2006), manual adjustments in Coot (Emsley & Cowtan 2004) and refinement in BUSTER (Blanc et al. 2004) (see **Table 5.1** for statistics).

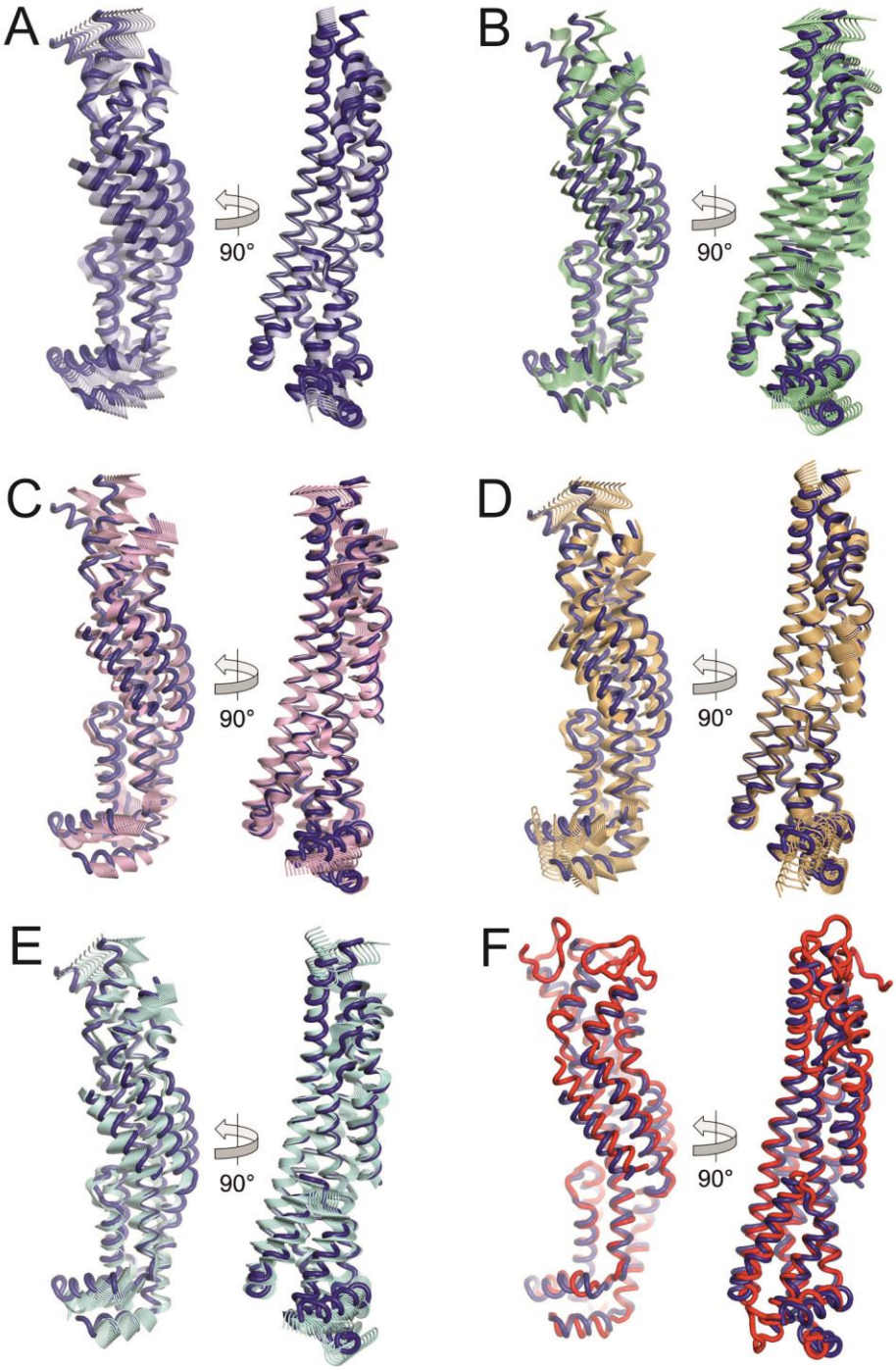


Fig. 5.5. Normal mode perturbed models used as templates for GPC3 diffraction data phasing by MR. (A-E) Set of 55 normal mode models using the elNémo web server (Suhre & Sanejouand 2004) distributed in 5 sets of 11 models perturbed along directions of normal modes 7, 8, 9, 10 and 11, respectively. The normal mode 7 perturbed model, which resulted in the correct MR solution is shown as a thick dark blue tube in panels A-F. Additional four normal mode 7 perturbed models which also gave the right MR solution are shown as thin dark blue tubes in A. **(F)** Superposition of the normal mode model used for MR phasing and the final GPC3 model. R.M.S.D.: 1.855 Å for 294 Cα equivalents. Each panel shows GPC3 in two views that differ by a 90° rotation around the vertical axis. This figure was produced with the help of T. Malinauskas, STRUBI, Oxford.

5.2.4 Structure of the GPC3 core domain

GPC3-ΔGAG adopts a cylindrical, all α-helical structure measuring approximately 104 Å in length and 25 Å in diameter. The N-terminal helix α1 traverses the long axis, while breaks or kinks in the additional 12 helices define three distinct lobes within the GPC3 structure. These lobes have been termed the N-lobe (N-terminal segment of α1, α6, α7 and α12), M-lobe (middle segment of α1, α5, α8, α11 and α13), and C-lobe (C-terminal segment of α1, α2, α4, α5, α9 and α10). Six of the seven disulphide bonds conserved in glypicans map to the N-lobe (disulphides I, II, IV, V, VI) and one to the C-lobe (disulphide III) (**Fig. 5.6 and 5.7A**). One of the conserved disulphide bridges predicted to be located in the N-lobe is not visible in electron density maps and remained unmodelled. In WNT signalling, the GPC3 core protein is processed by a furin-like convertase which is essential for modulation of WNT signalling and cell survival (De Cat et al. 2003). Cleavage of GPC3-ΔGAG was evident in the structure, supporting the identification of the cleavage site R355Q/YR358 (De Cat et al. 2003) (**Fig. 5.6**), which explains not only the triple band on SDS-PAGE analysis of GPC3-ΔGAG SEC fractions under reduced conditions (**Fig. 5.3B**), but also the large distance between helices α10 and α11 (**Fig. 5.7**). Mutations of this cleavage site resulted in unprocessed forms of GPC3 (De Cat et al. 2003), although a mutated form of GPC3 (R355A/R358A) retained its ability to interact with WNT in co-immunoprecipitation experiments and had comparable activity levels to that of cleaved GPC3 in luciferase assays (Capurro, Shi, et al. 2005a). This cleavage between α10 and α11 (**Fig. 5.6 and 5.7A**) results in the C-terminus of GPC3-ΔGAG emerging not from the end of the molecule, but rather form the boundary between M- and C-lobes (**Fig. 5.7A**).

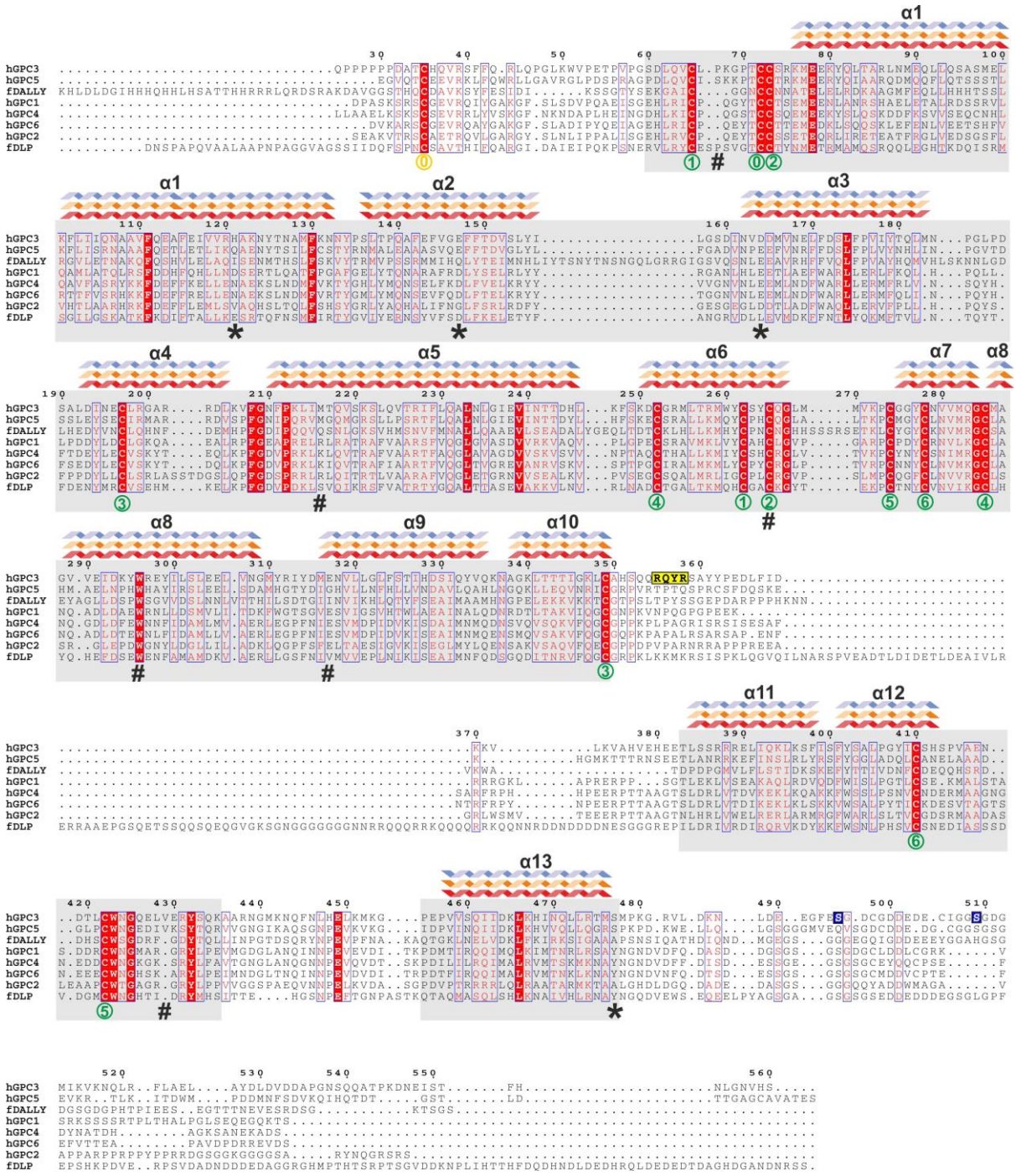


Fig. 5.6 . Sequence alignment of GPC family members. Sequences were aligned using MultAlin (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPript (espript.ibcp.fr/ESPript/ESPript/). Numbering corresponds to human GPC3 (including the secretion signal). Secondary structure assignments are displayed above the alignment (blue: humanGPC3, orange: human GPC1 (PDB ID. 4ACR), red: fly DLP (PDB ID. 3ODN). The identified furin cleavage site in GPC3(De Cat et al. 2003) is highlighted as yellow box. Blue boxed residues indicate the predicted heparan sulphate attachment sites for GPC3. Disease-related residues identified in human GPC3 are marked below the sequences (hashtag: buried, asterisk: solvent-exposed). Disulphide bridges are numbered.

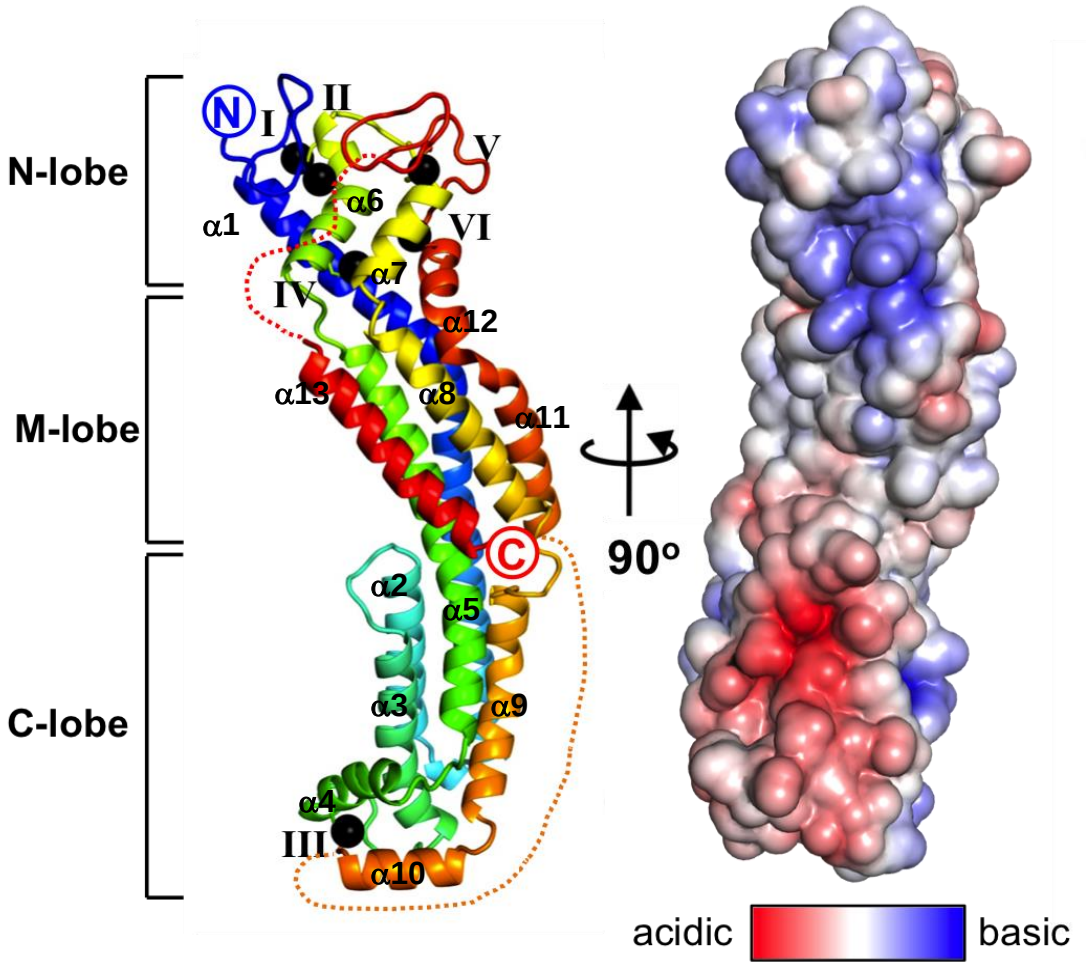
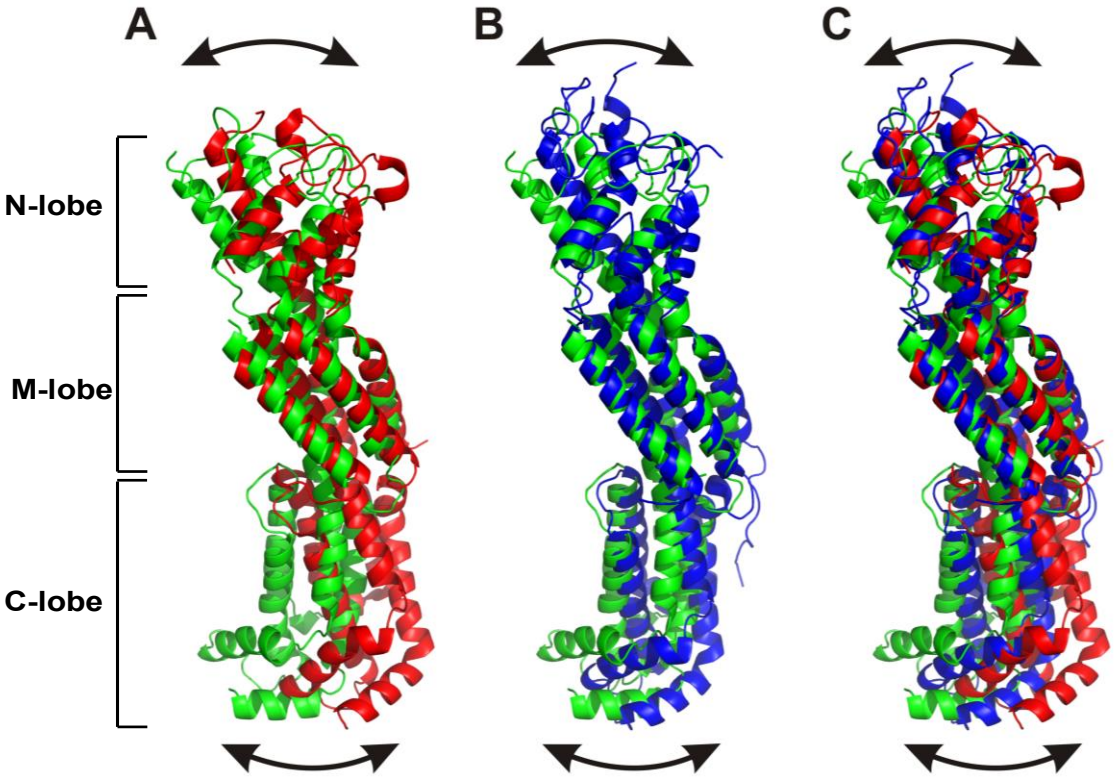


Fig. 5.7. Structure of the human GPC3 ectodomain. (A) Cartoon representation of the GPC3 ectodomain. Colour coding is in rainbow colouring from blue (N-terminus) to red (C-terminus). Disulphide bridges are depicted in roman numerals. N- and C-termini are highlighted. The three lobes are labelled (N-lobe, M-lobe and C-lobe). (**B**) Electrostatic potential from red (-8 k_bT/e_c) to blue (+8 k_bT/e_c) calculated with APBS as implemented in PYMOL (www.pymol.org). (B) is rotated by 90° around the vertical axis compared to (A). Helices are numbered α1-α13

5.2.5 Structural comparison of GPC3 to other Glypicans

The closest structural homologues of GPC3 identified using structural similarity searches implemented in PDBeFold (http://www.ebi.ac.uk/msd-srv/ssm/) are the fly glypican DLP (PDB Id. 3ODN (M.-S. Kim et al. 2011)) and human GPC1 (PDB Id. 4ACR (Svensson et al. 2012)). Structural superposition of GPC3 with DLP gives a RMSD of 3.27 Å for 269 Cα (**Fig. 5.8A**) with a sequence identity of 19%, while that

of GPC3 with GPC1 gives a RMSD of 2.85 Å for 322 Cα with a sequence identity of 19% (**Fig. 5.8B**). Such a high RMSD could preclude MR as a method for solving the structure of GPC3, but when the RMSD for the individual lobes is investigated, a different story emerges. The RMSD for the superposition of GPC3 and DLP N- and M- and C-lobes are 1.12 Å for 71 Cα, 2.71 Å for 51 Cα and 2.94 Å for 112 Cα respectively (**Fig. 5.8D**). The elNémo (Suhre & Sanejouand 2004) server allowed movement of individual lobes when calculating normal mode perturbed models finding a much closer match than would otherwise have been possible. Superposition of GPC1 with DLP gives a RMSD of 2.3 Å for 305 Cα; N-, M- and C-lobes having an RMSD of 0.99 Å for 96 Cα, 1.29 Å for 65 Cα and 1.08 Å for 105 Cα respectively, with a sequence identity of 32%, indicating their much closer evolutionary heritage than GPC3 shares with DLP. There is flexibility between the different lobes in relation to each other (**Fig. 5.8A-C**). Superposition of GPC3-ΔGAG, GPC1 and DLP (**Fig. 5.8**) using the M-lobe as the region of rigid template demonstrates the relative flexibility of the N-lobe and C-lobe compared to the M-lobe. This structural plasticity potentially made the structure solution via MR difficult, but without this plasticity, the structural solution using this method would have proved impossible.



D

		DLP			GPC1	
	Lobe	rmsd	Cα	Lobe	rmsd	Cα
GPC3	All	3.27	269	All	2.85	322
	N	1.12	71	N	1.03	85
	M	2.71	51	M	3.625	55
	C	2.94	112	C	2.46	106
GPC1	All	2.3	305			
	N	0.99	96			
	M	1.29	65			
	C	1.08	105			

Figure 5.8 Comparison of the GPC3 structure to DLP and GPC1 structures. (A) Superposition of GPC3 (green) with DLP (red). (B) Superposition of GPC3 (green) with GPC1 (blue). (C) Superposition of GPC3 (green), DLP (red) and GPC1 (blue). All superpositions were calculated with PYMOL (www.pymol.org) with GPC3ΔGAG (residue range 98-112;α1, 228-241;α5, 243-310;α8, 385-402;α11 and 455-477α13) as template. Colour-coding is as follows: GPC3: green, DLP (PDB Id. 3ODN(M.-S. Kim et al. 2011)): red, GPC1 (PDB Id. 4ACR(Svensson et al. 2012)): blue. Black arrows show the degree of flexibility of N- and C- lobes in relation to the M-lobe.

5.2.6 Mutational analysis based on the GPC3 structure

Mutations in GPC3 have been implicated in a range of diseases (**Table 5.2**) including:

Beckwith-Wiedemann syndrome (BWS) (White et al. 2002), Simpson-Golabi Behmel

Syndrome (SGBS) (Shi & Filmus 2009; Veugelers et al. 2000; Pénisson-Besnier et al. 2008) and a range of cancers from adenocarcinoma (Seshagiri et al. 2012) to carcinomas, malignant melanomas and non-small cell carcinoma (from the Catalogue Of Somatic Mutations In Cancer **(**COSMIC) database http://www.sanger.ac.uk/genetics/CGP/cosmic/). GPC3 mutations appear in a number of forms. Most disease mutations are point mutations, which lead to a loss of function, rather than frame-shifts, which would lead to a lack of expression. To obtain a better understanding of the disease mechanisms, we analysed previously published GPC3 mutations on the basis of our crystal structure.

Of the point mutations identified from both the literature and the COSMIC database, ten are present in our structure and have been mapped (**Fig. 5.9A**). This allowed classification of two distinct groups: (1) surface-exposed mutations: H121Y, E147D, S447F and (2) buried mutations: P67T, M216V, C265R, W296R, E317K and V429M. For example, the buried mutation M216V, could potentially destabilize the interaction between helices α3 and α5, while C265R (**Fig. 5.9A,B**), would disrupt the disulphide bridge with C73 breaking the link between α1 and α6. W296R (**Fig. 5.9A,C**), is found in sufferers of SGBS and leads to loss of function (Veugelers et al. 2000). The exchange of a tryptophan (W296) to an arginine in the centre of the five-helix core of the M-lobe, formed by helices α1, α5, α8, α11, and α13 (**Fig 5.9C**), is likely to destabilize the hydrophobic core of the M-lobe. E317K (**Fig. 5.9A+D**) has the potential to rearrange the relative orientation of helices α9 and α11. E317, located on helix α9, forms a salt bridge with R388 on α11 and mutation of E317 to a lysine may abolish this hydrophilic interaction. Another mutation of note, not visible in the structure, is the mutation of Glycine 556 to Argenine, which potentially disrupts

addition of a GPI anchor (Shi & Filmus 2009; Pénisson-Besnier et al. 2008); this mutation would render GPC3 incapable of inhibiting HH signalling as it would no longer be able to internalize upon HH binding (Capurro et al. 2008).

Table 5.2 Disease mutations identified for human GPC3

Mutation	Disease	Tissue	Reference	Location of Mutation	Potential Function
P67T*	Adenocarcinoma	Colon	COSMIC Database	Buried	Potentially important for stabilizing the hydrophobic core of the N-lobe
H121Y	BWS	Multiple tissues	(White et al. 2002)	Surface-exposed	Unknown
E147D	Adenocarcinoma	Colon	COSMIC Database	Surface-exposed	Unknown
D164H	Non small cell carcinoma	Lung	COSMIC Database	Surface-exposed	Unknown
M216V	Adenocarcinoma	Colon	COSMIC Database	Partially buried	Potentially destabilizes the interaction between helices α3 and α5
C265R*	Adenocarcinoma	Colon	(Seshagiri et al. 2012).	Buried	Very likely to destabilize the N-lobe. Leads to disruption of a disulphide bridge with C73
W296R*	SGBS	Multiple tissues	(Veugelers et al. 2000)	Buried	Very likely destabilizes the hydrophobic core of the M-lobe
E317K*	Adenocarcinoma	Lung	COSMIC Database	Partially buried	Potentially destabilizes the interaction of helices α9 and α11 leads to disruption of a salt bridge with R388
H377R	Adenocarcinoma	Large Intestine	COSMIC Database	Not in structure	Unknown
V429M	BWS	Multiple tissues	(White et al. 2002).	Partially buried	Potentially destabilizes the hydrophobic core of the N-lobe
S477F	Malignant melanoma	Skin	COSMIC Database	Surface-exposed	Potentially destabilizes interaction of helices α13 and α8
E548K	Adenocarcinoma	Large Intestine	COSMIC Database	Not in structure	Unknown
G556R	SGBS	Multiple tissues	(Shi & Filmus 2009). (Pénisson-Besnier et al. 2008)	Not in structure	Potentially inhibits GPI anchor attachment(Shi & Filmus 2009)
A569T	BWS	Multiple tissues	(White et al. 2002)	Not in structure	Unknown
H580L	Carcinoma	Prostate	COSMIC Database	Not in structure	Removed in mature form

SGBS: Simpson-Golabi-Behmel syndrome
HS: Heparan sulphate
BWS: Beckwith-Wiedemann syndrome
Asterisks (*) indicate residues shown as close-ups in Fig. 4.7.

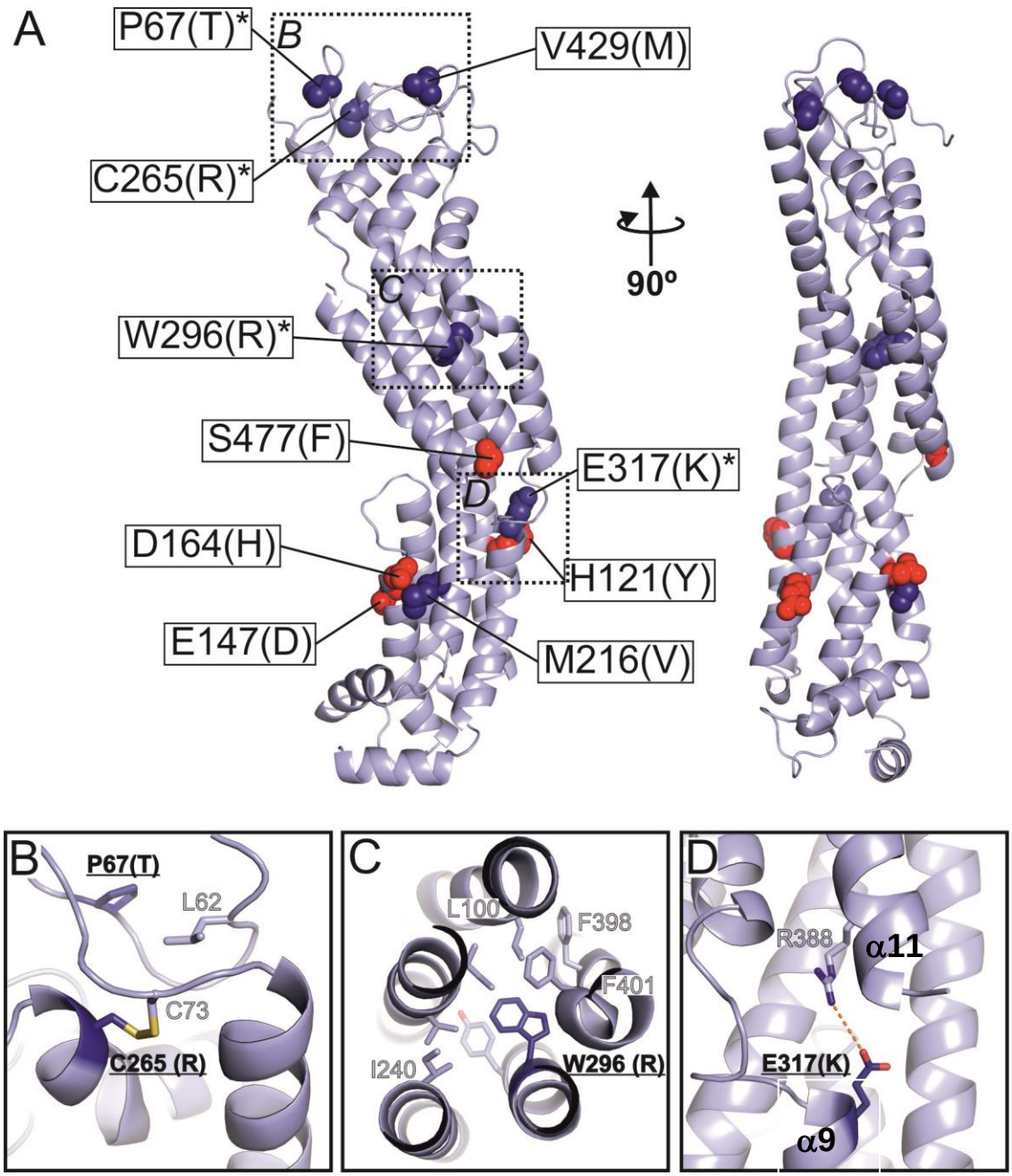


Fig. 5.9. Disease mutations mapped onto the structure of GPC3. (**A**) Potential disease-related mutations from Table 5.2 mapped onto a cartoon representation of GPC3 (orientation as in Fig. 5.5). Mutations are colour-coded as either buried (blue) or surface-exposed (red). (**B**-**D**) Close-up views of selected mutation hotspots indicated in (A). The mutation C265R results in disruption of a conserved disulphide bridge. (B) Mutations of W296R is likely to disrupt the hydrophobic core of the five-helix bundle in the M-lobe (C). The mutation E317K will inhibit the formation of the salt bridge with R388, potentially leading to a rearrangement of helices 9 and 11.

5.2.7 GPC3 is a Negative Regulator of Hedgehog Signalling

To investigate the role of GPC3 in HH signalling, it was first necessary to assess the role of full length, membrane-attached GPC3. This was achieved with a HH luciferase response assay in 3T3 fibroblasts using a reporter plasmid containing eight directly repeated copies of 3′Gli-binding sites, followed by the firefly luciferase gene (Sasaki et al. 1997) (see methods; Chapter 2, paragraph 2.3). In agreement with previous reports, GPC3 showed significant inhibition of HH signalling that was comparable to that of the HH antagonist HHIP (**Fig. 5.10**). GPC3 regulates HH signalling by binding HH and then internalizing, and thus removing HH from the ECM (Capurro et al. 2008).

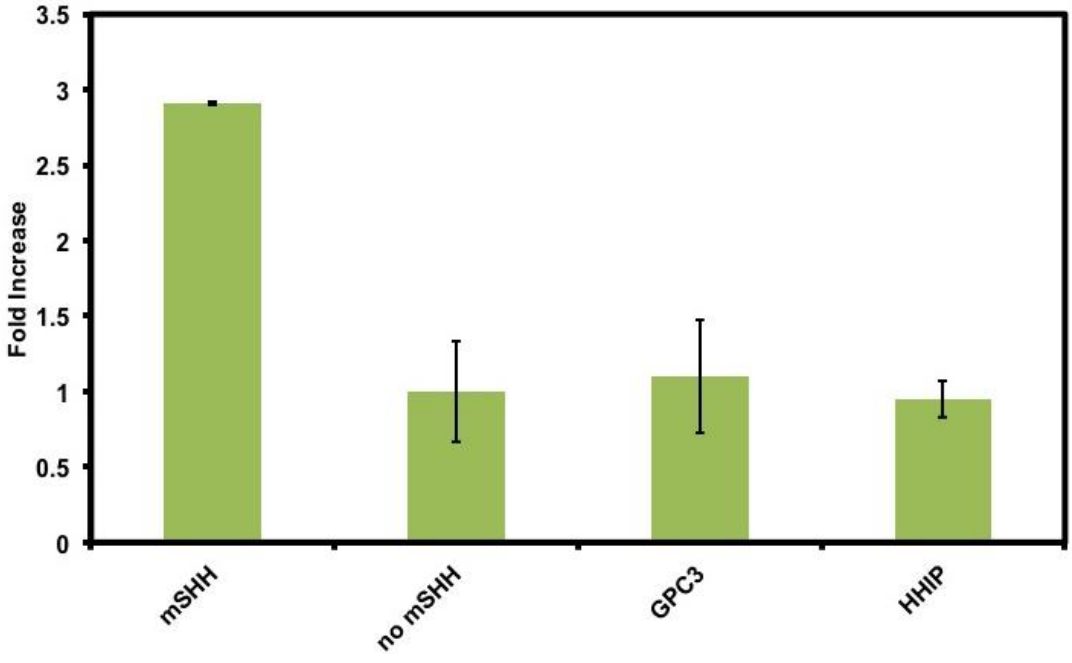


Figure 5.10 GPC3 inhibits HH signalling. A HH-Luciferase response assay was used to measure the relative inhibition of GPC3 and HHIP. NIH 3T3 mouse fibroblasts were transfected with either: HH-luciferase response plasmid (Sasaki et al. 1997) alone (mSHH and no mSHH), full length, membrane-bound GPC3, including GAG attachment sites and GPI anchor co-transfected with HH-luciferase response plasmid (GPC3), and full length, HHIP membrane protein co-transfected with the HH-luciferase response plasmid (HHIP). Results are the product of a single experiment done in quadruplicate.

5.2.8 Characterisation of the GPC3-SHH interactions.

The ability of the protein core of GPC3 to bind HH remains contentious, with conflicting data having been published. In order to find whether HS chain modification of GPC3 was necessary for HH binding, or whether the protein core alone was sufficient, two constructs were used in a series of SPR experiments. These two constructs were bound to a CM5 biacore chip (see methods; Chapter 2, paragraph 2.4) in order to carry out SPR and answer this question. As calcium has been found to be important in both CDO binding (McLellan et al. 2008), HHIP (Bishop et al. 2009) (see results chapter 4) and GAS1 (see results chapter 6), 5 mM calcium was used in all GPC3-HH SPR experiments. The two constructs used were: GPC3-ΔGAG, containing the protein core alone, and GPC3-ΔGPI, consisting of the protein core and both HS attachment sites. No binding was observed for GPC3-ΔGAG (**Fig. 5.11A**), while a dissociation constant (K_d = 81 μM) was calculated for GPC3-ΔGPI (**Fig. 5.11B**). The protein core of GPC3 is, therefore, unable to bind HH directly and, although the role of the protein stalk cannot be discounted, it appears that the HS chains are required for HH binding. Two prior studies investigated whether GPC3 can bind HH via the protein core alone. The first, using SPR, found that HH could bind the protein core of GPC3 alone (Capurro et al. 2008), while the second, using a pull-down assay, found that the protein core of GPC3 alone could not bind HH (M.-S. Kim et al. 2011). The authors of the second study attributed this discrepancy to their use of a shorter GPC3 construct. Interestingly, DLP does not bind HH via the protein core alone, although this, again relied on a pull-down assay (Williams et al. 2010).

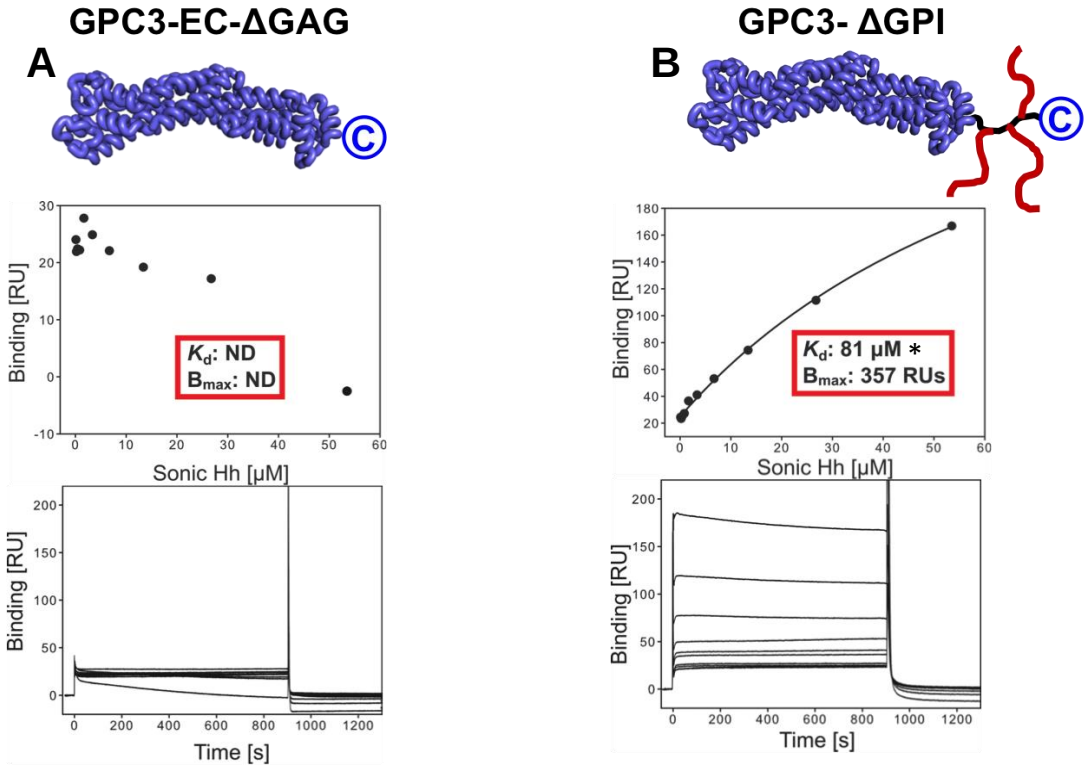


Figure 5.11 SPR equilibrium binding experiments of different GPC3 constructs with SHH. Ribbon representation of the construct used in each experiment is shown above the sensorgram and plotted equilibrium binding for (A) GPC3-ΔGAG and (B) GPC3-ΔGPI. Each SHH injection is shown in the lower graph, with all injections reaching equilibrium. RU measured on reaching equilibrium is plotted in the upper graph against time (s). * saturation was never reached, so the B_{max} and K_d are purely theoretical.

5.3 Discussion

Structural comparison of GPC3 to DLP and GPC1 revealed the apparent flexibility between the different domains of the core protein unit. This plasticity made solving the structure of GPC3 by MR very difficult, an obstacle, which could be resolved by applying normal mode analysis (taking this inherent plasticity into account). Analysis of GPC3 disease mutations on the basis of the GPC3 crystal structure demonstrates the importance of the tightly packed helices and hydrophobic core of the protein. Mutations buried in the structure all have the potential to disturb the structural integrity of GPC3 either by destroying disulphide bonds, or by disrupting interactions between helices in either the N- or M-lobe. Such structural changes would render GPC3 incapable of abrogating the HH signal as it would be unable to internalise on

binding HH (Capurro et al. 2008). Mutations which map onto the surface of GPC3 could potentially impact on protein-protein interactions with other interacting proteins. For example, BMP ligands have been identified as a direct binding partner for glypicans in *Drosophila* and that vertebrate BMP4 could directly bind DALLY (Kirkpatrick et al. 2006).

There is controversy about whether the protein core of GPC3 is able to bind HH directly, or whether HS modification is required. Two groups have presented conflicting data; one group finding that there was a direct interaction between the protein core of GPC3 and HH (Capurro et al. 2008), the other that no binding of HH to the protein core alone was observed (M.-S. Kim et al. 2011). Our own experiments support the latter findings, with no observable binding of HH to GPC3-ΔGAG, but observable binding (K_d = 81 μM) with GPC3-ΔGPI, a construct that has both HS insertion sites, but no GPI anchor attachment site. The construct used in our experiments and those by Kim et al (2011) was truncated to exclude the HS attachment sites, while the construct used by Capurro et al. (2008) contained the full-length sequence carrying only mutations of the HS attachment site residues.

There are two possibilities to explain this discrepancy. First, SHH may not interact with the protein core domain, but rather with the protein stalk of GPC3, present in neither our, nor the construct used by Kim et al (2011). This argument is further complicated as no binding was found in a DLP construct which did include the protein stalk (Williams et al. 2010). Second, there could be an as yet unidentified co-factor required to facilitate binding of HH to GPC3.

The work presented here will enable the design of further experiments to advance our understanding of the role GPC3 plays in development and disease. Mutant constructs with the disease mutations outlined above will be made to test for expression in small scale expression tests. Those that do express will be used for binding studies utilising both the HH-luciferase response assay to test their ability to reduce HH signalling in comparison to the wild type and SPR to obtain a binding constant. Mutations of the HS attachment sites, rather than truncating the C-terminus as is the case with GPC3ΔGAG, would also give important data which may help unravel the mystery of whether the protein core alone is sufficient to bind HH. It would also be interesting to test whether GPC3 binds any of the BMP ligands. This would also be tested using SPR and a BMP-luciferase response assay. Once the requirements for binding of HH to GPC3 have been elucidated, it would be imperative to set up a full crystallization screen of GPC3 in complex with HH.

Chapter 6

Growth Arrest Specific 1

Chapter 6 Growth Arrest Specific 1

ABSTRACT

Functional studies have identified GAS1 as an important agonist of the HH signalling pathway. GAS1 forms a complex with PTCH and works redundantly with CDO/BOC as a co-receptor for the HH ligand. This chapter discloses the crystal structure solution and description of GAS1 in three different space groups. Solving this structure required the prediction of flexible disordered loop regions in the GAS1 structure, which prevented crystallization. To achieve this, a combination of bioinformatics and homology modelling was used. Constructs based on this model were successfully crystallized and the structure of GAS1 solved. Analysis of the structure revealed the importance of the disordered loop 2 region in the crystallization of GAS1. The GAS1 structure enabled a comparison with GFRs, giving the first hint of how GAS1 could interact with GDNF. An unexpected ligand buried within domain 2 of the GAS1 structure was observed in the electron density map, which was analysed by mass spectrometry. Biophysical binding studies were conducted to determine: (1) The binding specificity of GAS1 domains to HH; (2) the role of clinically relevant disease mutants mapped onto the structure of GAS1 in HH binding, and; (3) the role of the unknown small molecule identified within domain 2 of GAS1.

6.1 Introduction

6.1.1 Discovery and biological function of GAS1

GAS1, first isolated from NIH 3T3 cells (C. Schneider et al. 1988), belongs to the family of cell-derived neurotrophic factor receptor α (GFRα) GPI-anchored cell surface proteins, based on sequence homology alignments (Cabrera et al. 2006). GAS1 consists of 279 amino acid residues and is predicted to have two domains

(Cabrera et al. 2006). It was originally identified as a gene whose expression is linked to arrested proliferation of cultured fibroblasts (C. Schneider et al. 1988). Subsequently, GAS1 was shown to directly bind vertebrate HH ligands and to positively regulate the HH signalling output (C. S. Lee & Fan 2001).

6.1.1.1 Biological function

The expression pattern of GAS1 shows increased expression at regions that are distant from sources of HH secretion; a reverse gradient to that of the HH signalling gradient. This counter-gradient in GAS1 expression is due to GAS1 being down-regulated by HH signalling (Martinelli & Fan 2007a; Allen et al. 2007). Further, GAS1 is a positive regulator of the SHH signalling cascade acting to promote signalling in a SHH dose-dependent manner (Allen et al. 2007).

Chick electroporation experiments directly established that GAS1 is capable of promoting SHH signalling in a cell-autonomous manner; only cells with the GAS1 genotype have increased signalling. These results are consistent with previous reports examining GAS1 function in other tissues where removal of GAS1 also results in phenotypes observed in cases of reduced SHH signalling (C. S. Lee & Fan 2001; Liu et al. 2001).

Embryos mutant for GAS1, CDO and BOC completely abolish SHH signalling in the neural tube, demonstrating that in vertebrates these receptors do not function merely as auxiliary SHH receptors, but rather are an absolute requirement for SHH signalling *in vivo* (Izzi et al. 2011; Allen et al. 2011). Current data argue strongly that GAS1 is a novel positive component of the SHH signalling pathway (Allen et al. 2007).

6.1.1.2 GAS1 activity and knockout phenotype

Analyses of GAS1 mouse knock-outs reveal several defects. These include craniofacial, limb, and axon guidance deficiencies that are the hallmark of reduced SHH signalling. Detailed examination of ventral neural tube patterning, a process that depends critically on graded SHH signalling, also reveals deficiencies in both floor plate and vp3 neuro-progenitor cell specification in Gas1-/- embryos. Craniofacial, limb, and neural tube defects seen in GAS1 mutants are all significantly exacerbated by reducing the SHH dosage (Allen et al. 2011).

Proliferation of cerebellar granule neuron progenitors (CGNPs) in response to SHH in the developing cerebellum revealed that CGNPs express GAS1 and BOC, but not CDO. Interestingly, the cerebellum is smaller in BOC-/- mice, and BOC-/- CGNPs have lower proliferation than wild-type CGNPs in response to SHH. Similarly, GAS1-/- CGNPs are also less responsive to SHH. Most striking, however, is that GAS1-/-;BOC-/- CGNPs are completely unable to proliferate in response to SHH (Izzi et al. 2011).

Compound mutants between GAS1 and SHH showed that GAS1 is a genetic modifier of SHH in multiple developmental systems, including craniofacial structure, heart, spinal cord, somite and limb (Martinelli & Fan 2007a; Seppala et al. 2007). Interestingly, despite the strong cooperation seen between GAS1 and CDO in the promotion of SHH signalling during craniofacial and neural tube development, there appears to be no such cooperative interactions in the limb. This is despite the expression of GAS1 and CDO in overlapping domains of the limb(Allen et al. 2007). Genotypes of GAS1-/-, GAS1- /- ;SHH+/- , SHH- /- , to GAS1- /- ;SHH- /-

reveal a progressive reduction in SHH signalling (relative to wild type and GAS1+/-;SHH+/- controls). Together, these data indicates that GAS1 acts positively rather than negatively in HH signalling (Martinelli & Fan 2007b).

6.1.2 GAS1-SHH interactions

In *Drosophila*, IHOG and BOI (the orthologues of CDO and BOC) are an absolute requirement for HH signalling (Zheng et al. 2010). If HH receptor functions were entirely conserved, it would have been expected that vertebrate cells lacking BOC and CDO would not be able to respond to SHH. However, the inactivation of BOC in CGNPs (which do not express CDO) leads to a partial decrease in their proliferative response to SHH. Moreover, inactivation of CDO in BOC-/- cerebella did not further decrease CGNP proliferation compared to inactivation of BOC alone. Therefore, unlike *Drosophila* which absolutely requires either IHOG or BOI for HH signalling, vertebrate cells are able to respond to SHH in the absence of CDO and BOC. These results highlight a fundamental difference between *Drosophila* and vertebrate HH signalling and raised two possible mechanisms for HH reception in vertebrates: Either PTCH is sufficient for SHH signalling or, unlike *Drosophila*, additional SHH binding molecules enable vertebrate cells to respond to SHH. Recent data, supports a model where GAS1 acts with BOC and CDO as essential SHH receptors (Allen et al. 2011; Izzi et al. 2011).

GAS1, as a positive regulator of HH signalling, has two possible modes of action; it either facilitates diffusion of HH, or it increases its activity. When GAS1 expression was reduced by siRNA in HH- responsive NIH-3T3 cells, the cells became less responsive to exogenously supplied SHH (Martinelli & Fan 2007a). Over-expression

of GAS1 in the chick neural tube extended the range of HH signalling by activating HH downstream genes ectopically at farther distances (Martinelli & Fan 2007a; Allen et al. 2007). Clonal analysis of these activated cells revealed that GAS1 acts cell-autonomously to enhance HH signalling; only cells overexpressing GAS1 responded to HH at greater distances from the source of HH. Therefore, GAS1 is directly responsible for increased sensitivity to a lower HH gradient, demonstrating the function of GAS1 in receiving cells to enhance HH responsiveness (Martinelli & Fan 2007b). Immunohistochemistry to detect SHH in the developing limb bud revealed no reduction in either GAS1- /- or GAS1- /- ;SHH+/- genotypes in its range of diffusion despite a reduction in SHH downstream gene expression in both genotypes(Martinelli & Fan 2007b), suggesting that GAS1 facilitates an increase in HH activity rather than diffusion. Mutations of SHH amino acids important for binding to GAS1, BOC and CDO have been identified in HPE and in brachydactyly (McLellan et al. 2008) both in patients (Ribeiro et al. 2010) and in mouse, where inactivation of GAS1 and CDO leads to HPE (Seppala et al. 2007).

In contrast to the pronounced defects in SHH signalling, there are no apparent abnormalities in IHH-dependent long bone growth detected in GAS1 mutant embryos (Allen et al. 2011). This is a surprising result because GAS1 binds IHH with similar affinity to SHH (C. S. Lee et al. 2001a), and there seems to be significant overlap between GAS1 and IHH expression in developing bone (St-Jacques et al. 1999; K. K. Lee et al. 2001b).

6.1.3 GAS1-PTCH interactions

Upon binding to PTCH, HH relieves PTCH's inhibitory effect on SMO, allowing SMO to translocate to the primary cilia and trigger downstream signalling events (Singla & Reiter 2006). In NIH3T3 cells, a constitutively active form of SMO activates the HH pathway regardless of GAS1 overexpression or GAS1 RNAi (Martinelli & Fan 2007a). Conversely, a constitutively active form of PTCH that cannot bind HH represses HH pathway activation but such repression can be overcome by GAS1 overexpression plus SHH application. This suggests that GAS1 acts at a level above SMO and PTCH (Martinelli & Fan 2007b).

Cells co-expressing GAS1 and PTCH display a SHH binding capacity that is greater than the combined binding capacity provided by cells expressing each protein individually. Thus, the HH enhancing activity of GAS1 may rest on its ability to facilitate loading of HH onto PTCH thereby enhancing HH repression of PTCH activity (Martinelli & Fan 2007b). To further support this, a mutated form of SHH (SHHN E90A) incapable of binding GAS1, CDO and BOC, but able to bind PTCH is not sufficient to induce SHH-dependent signalling (Izzi et al. 2011), suggesting that binding to GAS1, CDO or BOC is a prerequisite for SHH-dependent signal transduction to occur (Izzi et al. 2011). GAS1, CDO and BOC interact with PTCH to form GAS1/PTCH, CDO/PTCH or BOC/PTCH complexes (Izzi et al. 2011). Although these complexes form, there was no detectable BOC in the GAS1/PTCH complexes or vice versa which makes it unlikely that a trimer of GAS1/BOC/PTCH forms (Izzi et al. 2011). Although clear that the presence of GAS1, CDO and BOC are essential for SHH-mediated signalling, the nature of the molecular interaction between HH, PTCH, GAS1, CDO and BOC remains poorly understood.

6.2 Results

6.2.1 Expression of the full-length ectodomain

The *GAS1* gene has a high GC content (>80%), which made it an extremely difficult target for cloning. As a result, it was necessary to design a synthetic, codon-optimised gene, which had a reduced GC content (64%). Initial expression experiments in HEK 293T cells using a full-length wild-type construct of GAS1 (GAS1-WT), lacking only the GPI anchor, resulted in multiple species in SEC (**Fig. 6.1A**), and multiple bands on SDS-PAGE (**Fig. 6.1B**). GAS1-WT was predicted to have no O-linked glycosylation sites using NetOGlyc 3.1 Server (http://www.cbs.dtu.dk/services/NetOGlyc/) and one N-linked glycosylation site using NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/). The problem of multiple species could be only partially resolved by the addition of Endoglycosidase H (EndoH), which cleaves asparagine-linked mannose rich oligosaccharides, thus deglycosylating the protein. Nevertheless, a full crystallization screen was performed, but yielded no crystals.

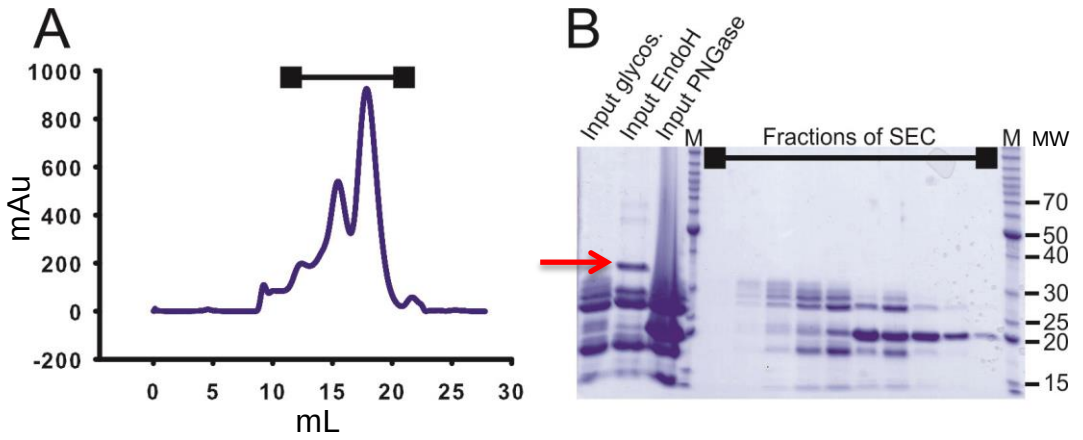


Fig. 6.1. SEC purification of the ectodomain of human GAS1. (A) Secreted GAS1 expressed in large scale yielded mg quantities of protein, but gave a heterogeneous profile in SEC (on a Superdex S200 10/30 column). **(B)** Analysis of SEC fractions by SDS-PAGE revealed multiple species. Input is the protein sample injected onto SEC, EndoH is a deglycosylation enzyme, which leaves only one *N*-

Acetyl-glucosamine attached to the Asparagine residue. PNGase is a deglycosylation enzyme which cleaves all sugars from the Asparagine residue. EndoH is marked by a red arrow.

6.2.2 Redesign of constructs based on a homology model

To remedy the heterogeneity of GAS1-WT, it was necessary to redesign the construct. Homology modelling implemented in HHPRED (Söding et al. 2005) using the closest structural relative to GAS1, GDNF receptor-α (GFR-α), with a sequence identity of less than 19% was carried out. Despite this low similarity, a homology model was created. The resulting model, combined with prediction of disordered secondary structure-lacking regions of the protein as implemented in RONN (Yang et al. 2005), was used for the design of new constructs with less flexibility and potentially better crystallizability (**Fig. 6.2**). The homology model from HHPRED suggested a structural architecture with two distinct domains (termed domain 1 and 2, **Fig. 6.2A, B**). Disorder prediction identified a potentially disordered loop region in domain 1 (called loop 1, residues G80-A96) and a flexible linker region between the two domains (called loop 2, G153-G162) (**Fig 6.2A, B**). Additionally, the C-terminal 60 residues (E247-G314) also showed a high degree of disorder. Based on this analysis, constructs for secretion in HEK 293T cells were designed removing one or both of the predicted disordered loop regions, as well as the disordered C-terminal region (**Fig 6.2** and **6.4**)

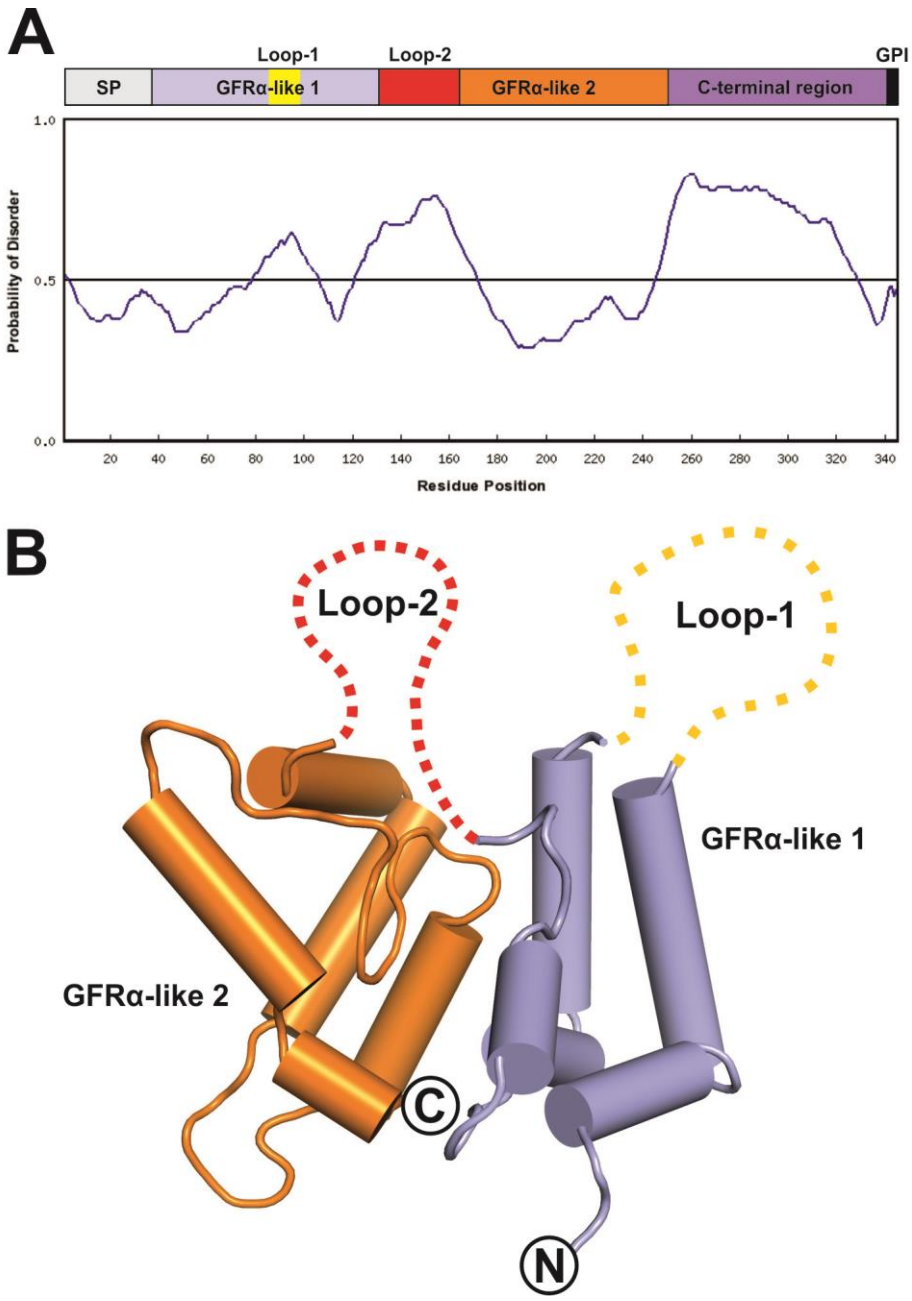


Fig. 6.2 Bioinformatic analysis of human GAS1. (A) Disorder prediction plot from RONN(Yang et al. 2005) showing the probability of disorder (y-axis) against the GAS1 residue position (x-axis). The top bar shows the predicted GAS1 domain organisation. SP: secretion signal. GPI: Glycosylphosphatidylinositol. **(B)** Homology model of human GAS1 calculated with MODELLER(Sali & Blundell 1993) as implemented in the HHPRED(Söding et al. 2005) server (http://toolkit.tuebingen.mpg.de/hhpred). Loops 1 (yellow) and 2 (red), predicted to be disordered, are highlighted. N- and C-termini are marked.

6.2.3 A construct harbouring deletions of loop 2 and C-terminal region gives the highest protein yield

Of the new constructs, the one with highest secreted expression lacked loop 2 and also had a C-terminal truncation (lacking residues G153-G162 and E247-G314, hereafter called GAS1-ΔL2). This construct yielded milligram quantities of secreted protein from 3 L of harvested conditioned media (**Fig 6.3A**) and could be purified (see methods, Chapter 2, paragraph 2.2.8) to homogeneity (**Fig. 6.3B**).

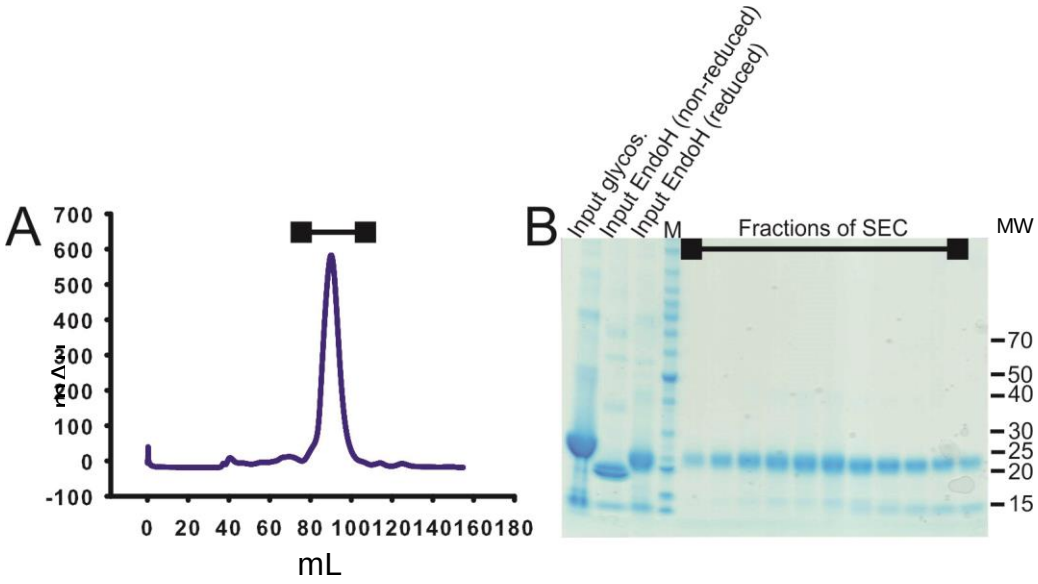


Fig. 6.3 SEC purification of GAS1-ΔL2. (**A**) SEC elution profile in SEC (on a superdex S200 16/60 column) shows a monodisperse peak. The black bar shows the range of fractions taken for analysis by SDS-PAGE (**B**) Fractions from SEC revealed a single species of the GAS1-ΔL2 construct. The black bar shows the region taken from SEC fractions for analysis.

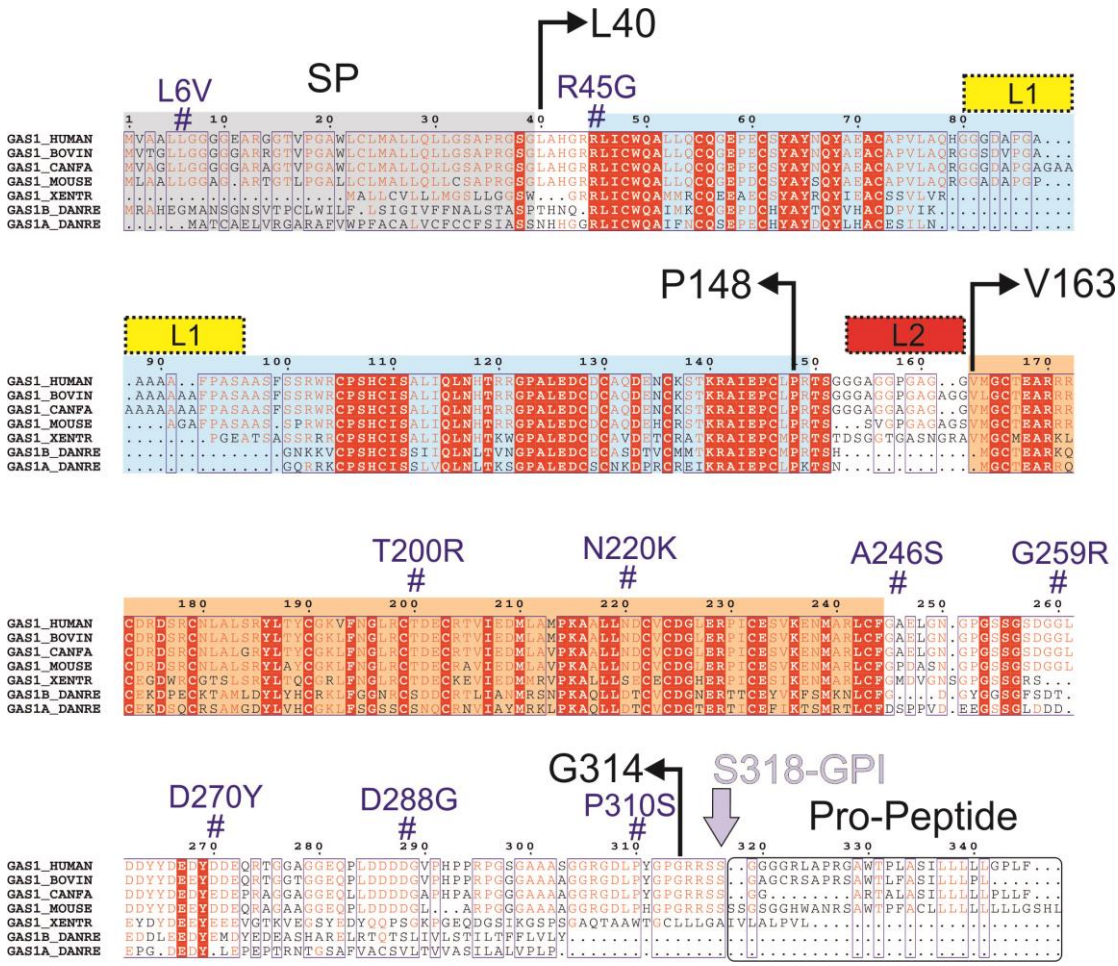


Fig. 6.4 Sequence alignments of the GAS1 family members. Sequences were aligned using MULTALIN (bioinfo.genotoul.fr/multalin/multalin.html) and formatted with ESPRIPT (espript.ibcp.fr/ESPript/ESPript/). Numbering corresponds to full length human GAS1 (including the secretion signal). Domains are coloured as in **Fig 6.2**. The disordered loops 1 and 2 are marked. Holoprosencephaly-related residues identified in human GAS1 are marked above the sequences with hashtags. (#). The GPI anchor residue S318 is depicted with an arrow. SP: secretion signal.

6.2.4 Crystallization and structural determination of the GAS1 ectodomain

Attempts to crystallize the HH-GAS1 complex failed, but GAS1-ΔL2 yielded well-ordered crystals using previously reported robotic technologies and protocols(Walter et al. 2005). GAS1 crystals frozen in reservoir solution (see Table 6.1) plus 20% glycerol diffracted to 2.1 Å resolution at ESRF-ID14-4 and DIAMOND-I03. This resulted in the collection of three datasets all in different space groups (see Table 6.1). Additionally, a heavy atom derivative dataset of crystal form 1 was collected at DIAMOND-I03 from a crystal soaked in a gold cyanide-saturated crystallization

solution for 1 hour. This allowed the structure determination using SAD analysis. The positions of three gold atoms were determined using SHELXD (T. R. Schneider & Sheldrick 2002). This solution was input into PHENIX (McCoy et al. 2007) for phase calculation and resultant improvement in the map. The obtained map was of high quality (**Fig. 6.5A,B**) and allowed automatic chain-tracing implemented in BUCCANEER (Cowtan 2006). The model was completed by manual rebuilding using COOT (Emsley & Cowtan 2004) (**Table 6.1**). The crystal form 2 and 3 structures were determined by molecular replacement using the experimentally-phased crystal form 1 structure. All structures contain one molecule of GAS1-ΔL2 in the crystallographic asymmetric unit and show excellent refinement and Ramachandran statistics (**Table 6.1**).

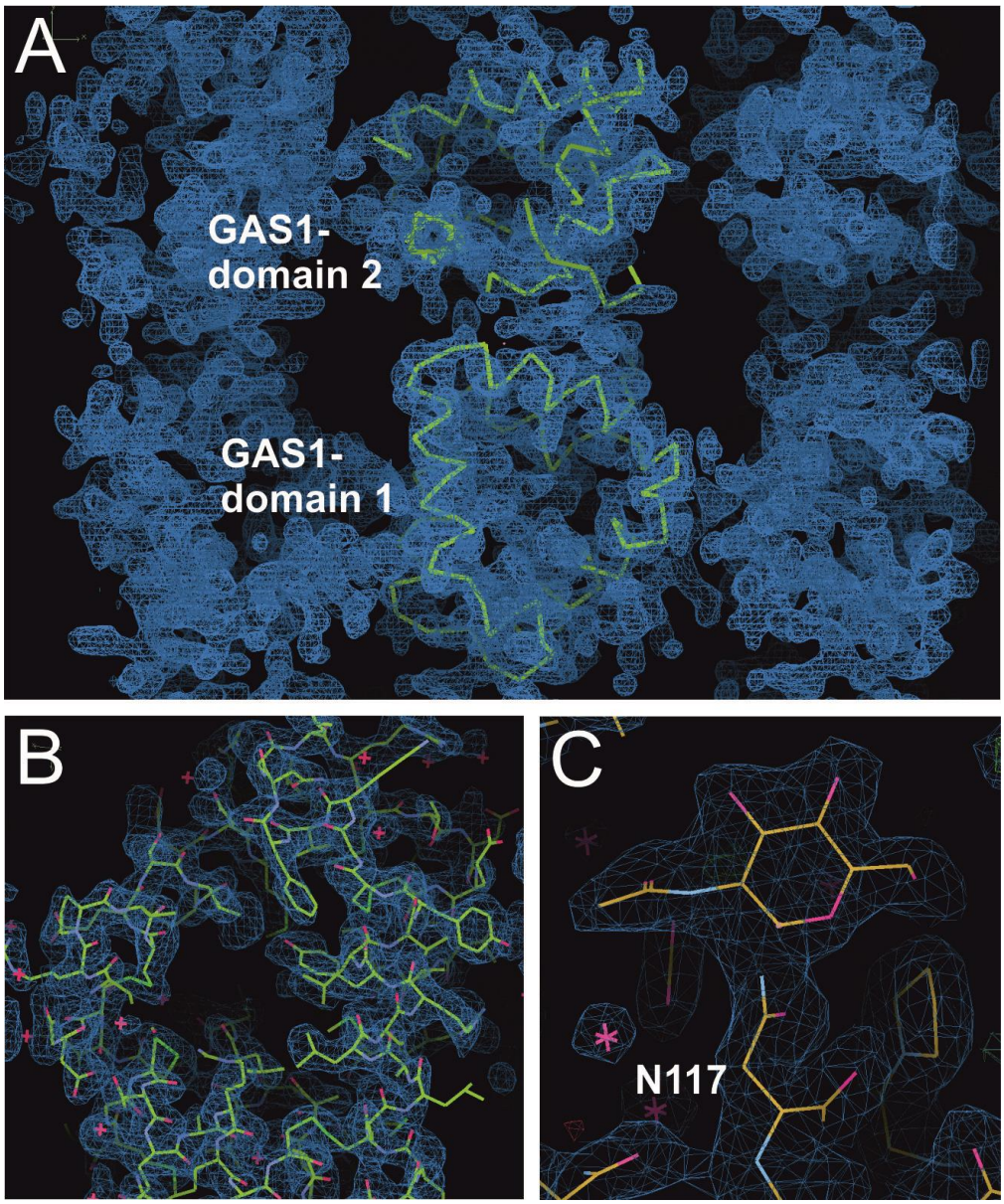


Fig.6.5 Electron density map of the GAS1 structure. (**A-B**) Initial electron density map of GAS1 after SAD phasing and density modification using PHENIX(McCoy et al. 2007) contoured at 1.0σ. The obtained map was of high quality and allowed automatic chain-tracing implemented in BUCCANEER(Cowtan 2006). The final model of the GAS1 structure is shown (A: ribbon, B: sticks). (**C**) SigmaA-weighted $2F_O$-F_C map of the final crystal form 1 model from autoBUSTER(G et al. 2011) contoured at 1.2σ of the GAS1 N-linked glycosylation site N117.

Table 6.1. Crystallographic data collection and refinement statistics

	hGAS1 crystal form 1 $KAu(CN)_2$-soak	hGAS1 Crystal form 1	hGAS1 Crystal form 2	hGAS1 Crystal form 3
DATA COLLECTION				
X-ray source	DIAMOND-I03	ESRF-ID14-EH4	DIAMOND-I03	ESRF-ID14-EH4
Resolution [Å]	30.0-2.40 (2.47-2.40)	30.0-2.11 (2.16-2.11)	50.0-2.25 (2.33-2.25)	50.0-2.08 (2.14-2.08)
Space group	$P2_1$	$P2_1$	$P6_322$	$P2_12_12_1$
Cell dimensions [Å]	a = 32.7 Å b = 59.7 Å c = 43.2 Å β = 105.5°	a = 32.6 Å b = 59.6 Å c = 43.2 Å β = 105.6°	a = b = 89.7 Å c = 119.0 Å	a= 46.5 Å b = 57.9 Å c = 58.4 Å
Wavelength [Å]	1.0300	0.9393	0.9763	0.9393
Unique reflections	6021 (313)	9287 (683)	13937 (1341)	9051 (430)
Completeness [%]	95.7 (68.9)	99.8 (99.6)	100 (100)	91.7 (61.1)
R_{merge} [%][a]	3.4 (9.3)	10.4 (58.4)	8.3 (87.7)	9.2 (52.4)
I/σI	19.8 (6.8)	12.2 (2.6)	31.4 (3.0)	15.2 (2.4)
Redundancy	3.1 (2.3)	6.4 (4.7)	12.2 (12.0)	6.3 (3.9)
REFINEMENT				
Resolution range [Å]		26.19-2.11 (2.36-2.11)	50.0-2.25 (2.43-2.25)	41.0-2.08 (2.33-2.08)
Number of reflections		9270 (2472)	13914 (2766)	9018 (1900)
No. of atoms (Protein/Water/NAG/nitrate)		1297/59/14/0	1266/51/14/40	1331/75/14/0
B factors [$Å^2$] (Protein/Water)		33/34/50/0	66/65/84/78	28/33/31/0
R_{factor} [%)][c]		19.0 (19.6)	19.9 (20.7)	19.5 (19.7)
R_{free} [%] [d]		22.7 (25.8)	23.4 (23.4)	24.8 (27.4)
r.m.s.d. bonds [Å]		0.010	0.010	0.010
r.m.s.d. angles [deg]		1.04	1.05	1.03
Ramachandran statistics[d]				
Favoured [%]		98.1	98.7	98.8
Disallowed [%]		0	0	0

r.m.s.d. stands for root mean square deviation from ideal geometry. Numbers in parentheses refer to the appropriate outer shell.

[a] $R_{merge} = \Sigma_{hkl} \Sigma_i|I(hkl;i) - \langle I(hkl)\rangle|/\Sigma_{hkl} \Sigma_i I(hkl;i)$, where I(hkl;i) is the intensity of an individual measurement and $\langle I(hkl)\rangle$ is the average intensity from multiple observations.

[b] $R_{factor} = \Sigma_{hkl}||F_{obs}| - k|F_{calc}||/\Sigma_{hkl} |F_{obs}|$.

[c] R_{free} equals the R-factor against 5% of the data removed prior to refinement.

[d] *from MolProbity* (V. B. Chen et al. 2010).

6.2.5 Structure of the GAS1 ectodomain

The ectodomain of GAS1 is composed of two GFR-like repeat domains connected by a flexible linker region (**Fig. 6.6A**). Each GFR-like repeat domain consists of 5 α-helices connected by five disulphide bridges, resulting in a rigid triangular helix spiral arrangement with an opening at the base of the triangle. There is no interpretable electron density present for the loop region in domain 1 (Loop-1 residues 153-162), which we predicted to be disordered (**Fig. 6.2A,B**, and **6.6A**). Structural analysis revealed a surface with mixed hydrophilic and hydrophobic patches (**Fig. 6.6B**) and a mixed pattern of surface conservation (**Fig. 6.6C**). The two GFR-like repeats share a common architecture, which stack in two roughly triangular spirals (**Fig. 6.6D**). The 5 helices of each GFR-like repeat are held in rigid confirmation by 5 disulphide bonds. Domain 1 has disulphide bonds between helices α1 and α2 (C55-C61), α2 and α5 (C72-C146), α3 and α5 (C105-C146), α4 and α1 (C128-C48), and α4 and α5 (C130-C136), while domain 2 has disulphide bonds between helices α4 and α5 (C224-C232), α1 and α4 (C166-C222), α1 and α2 (C173-C179), α2 and α3 (C190-C203), and α3 and α5 (C199-C243) (**Fig. 6.4** and **6.6A,D**). This disulphide bridge pattern and the relative orientation of the helices are highly conserved with a RMSD of 1.17 Å for 71 equivalent Cα positions and a sequence identity of 23%.

X-ray data was collected for three different crystal forms, which crystallized in different space groups and packing arrangements (Table 6.1). These three crystal forms reveal the flexible nature of the loop 2 region, which acts as a linker between the two GFR-like domains (**Fig. 6.6A,D**) as predicted (**Fig. 6.2A,B**). This flexibility allows ~50° rotation of domain 2 relative to domain 1 (**Fig. 6.6E**), with crystal

domain 2 of crystal form 1 rotated 180° parallel to the long axis of this page relative to crystal forms 2 and 3. Crystal forms 2 and 3 are rotated 90° around this same axis and rotated 90° relative to domain 1 around an axis perpendicular to the page (**Fig. 6.6E**). Surprisingly, crystal form 2 (space group $P6_322$) showed no contacts at all between domains 1 and 2. This observation suggests that there is no physiological interaction between GAS1 domain 1 and 2 and that the domains potentially have functionalities independent of each other.

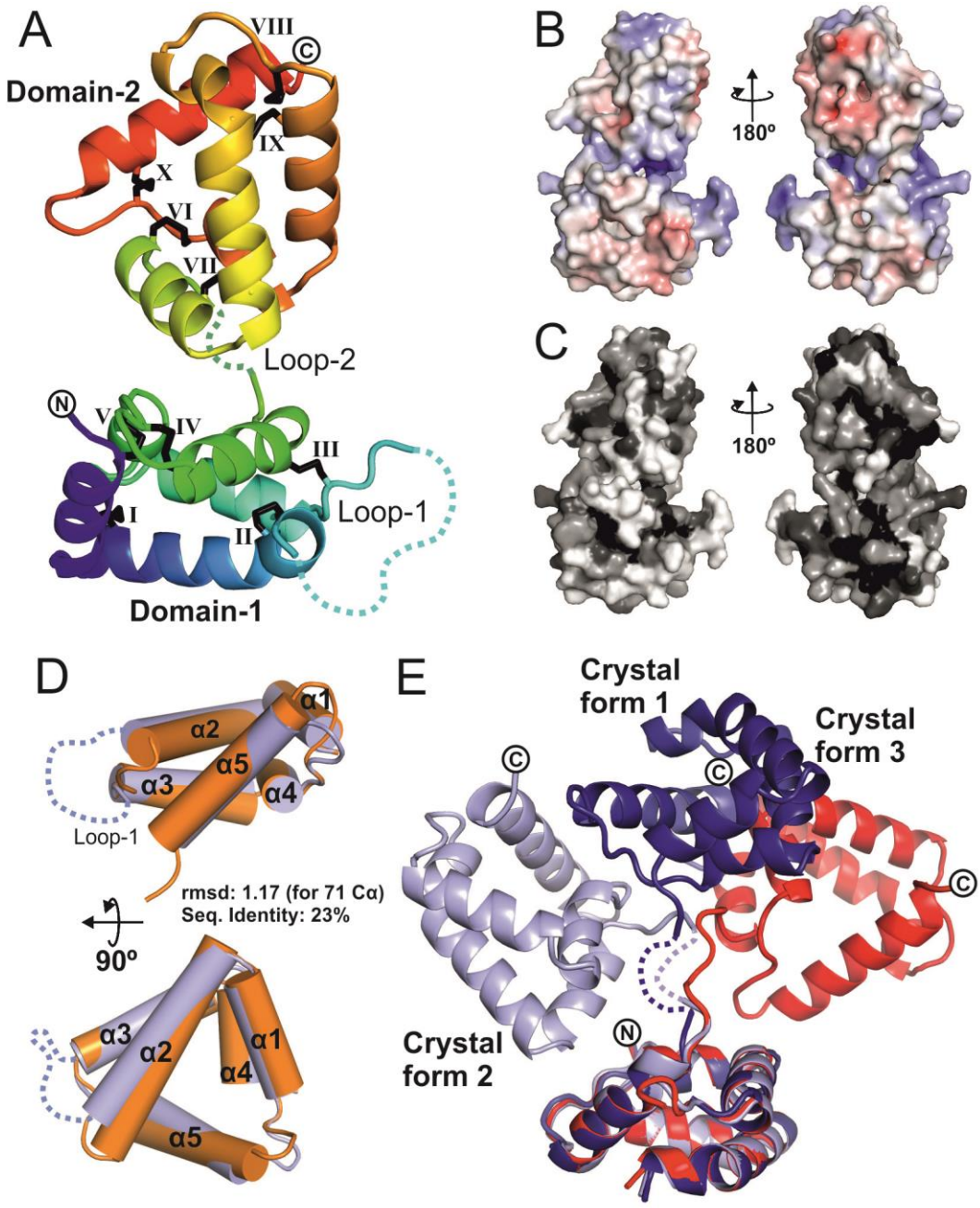


Fig. 6.6. Structure of the human GAS1 ectodomain. (A) Cartoon representation of the GAS1 ectodomain (crystal form 1). Colour coding is in rainbow colouring from blue (N-terminus) to red (C-terminus). Disulphide bridges are depicted in roman numerals. The disordered Loop 1 and Loop 2 regions are shown. (**B**) Electrostatic potential from red (-8 k_bT/e_c) to blue (+8 k_bT/e_c) calculated with APBS as implemented in PyMOL (www.pymol.org). (**C**) Residue conservation. Shown from non-conserved (white) to conserved (black)based on alignments containing sequences from all available vertebrate GAS1 family members. (**D**) Superposition of GAS1 domain 1 (light blue) and 2 (orange) reveals high level of structural homology with RMSD of 1.17 for 71 Cα and a sequence identity of 23%. Helices are numbered N- to C -terminally (α1-5). (**E**) Superposition of structures from all three crystal forms of GAS1ΔL2 using domain 1 as template showing the domain rearrangement.

6.2.6 An unknown ligand in the GAS1 structure

Surprisingly, extra electron density was visible within the cavity of the triangular helix spiral of domain 2 (**Fig. 6.7A**). This density is present in all three different crystal forms, independent of the crystal packing and crystallization condition which included a high salt condition (2.5M NH_4NO_3; crystal form 1), an alcohol condition (5% isopropanol; crystal from 2) and a PEG condition (5% PEG6000; crystal from 3) (**Fig. 6.7B-D**). This extra density is located in the pocket of the triangular helix spiral, and is roughly orientated with its long axis in line with the base and apex of the triangle. The pocket of domain 2 is formed by the side chains of Y186, F194, I207, M210, L211, L219 and M239 resulting in a hydrophobic environment.

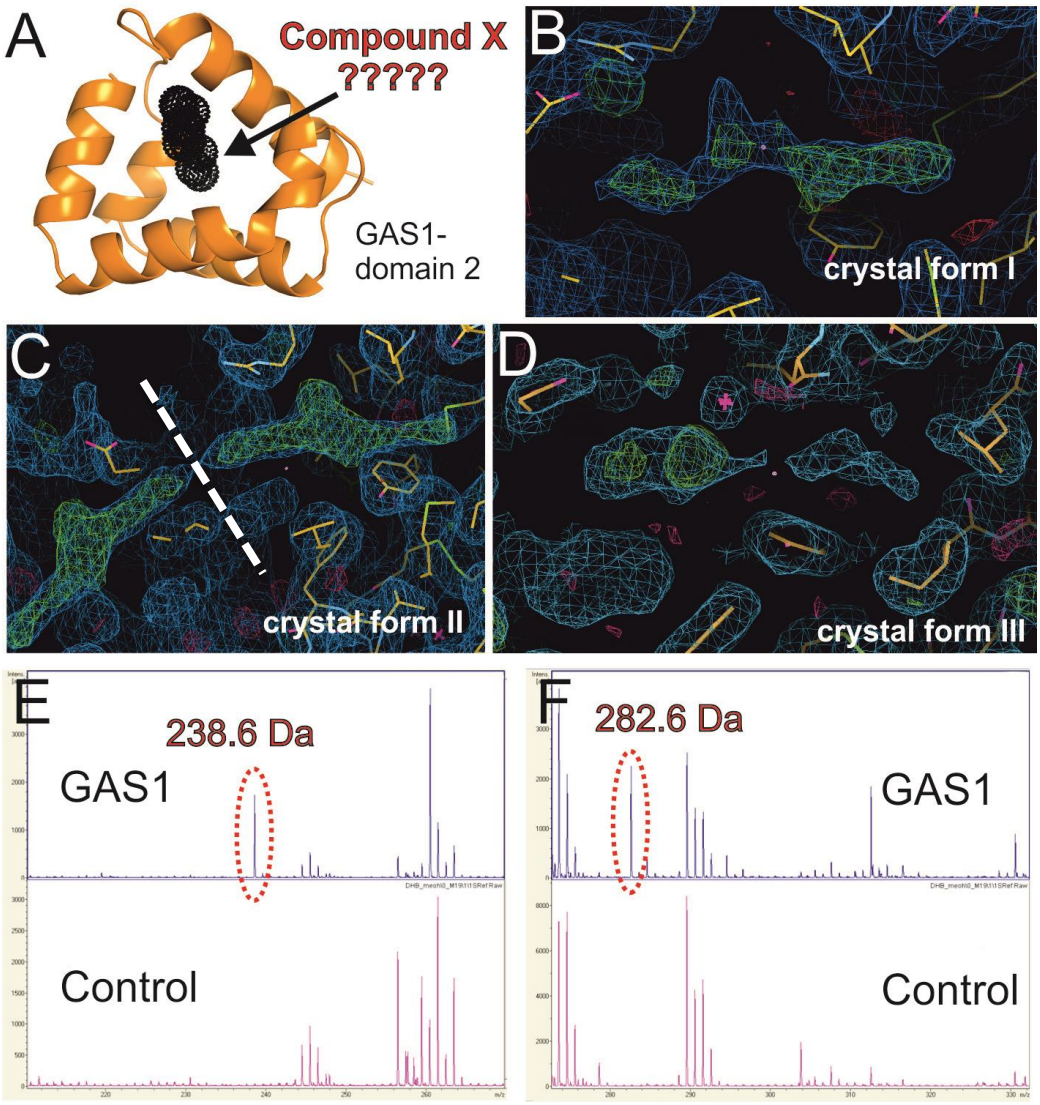


Fig. 6.7 Analysis of an unknown compound identified in the GAS1 structure. (A) Cartoon representation of GAS1 domain 2. The position of the unknown compound is depicted as black dots. **(B-D)** SigmaA-weighted $2F_O$-F_C (1.0σ) and $2F_O$-F_C (±3.0σ) map of the final model from autoBUSTER (G et al. 2011) for the three independent crystal forms of GAS1. The electron density suggests the presence of a ligand independent of the crystal packing and crystallization condition. The asterisk in panel (B) shows a salt bridge. The dashed white line in panel (C) indicates crystal symmetry **(E,F)** MALDI-TOF-MS analysis of purified GAS1 protein reveals the presence of two distinct peaks in the low molecular mass range that are unique to the GAS1 sample as compared to control.

In order to identify this potential ligand, we carried out mass spectrometry analyses using MALDI-TOF-MS (in collaboration with Dr. Benedikt Kessler, CCMP Oxford) with purified GAS1 proteins. These reveal the presence of two distinct peaks in the low molecular mass range that are unique to the GAS1 sample as compared to the

control alluding to the presence of compounds with a molecular mass of 238.6 and 282.6 Da, respectively. Analysis of this small molecule is on-going with further work required to identify the exact nature of the compound.

6.2.7 Structural comparison with GFRα

The closest structural homologues of GAS1 identified using structural similarity searches implemented in PDBeFold (http://www.ebi.ac.uk/msd-srv/ssm/) are the GFR alpha receptors GFRα1 (PDB code 3FUB (Parkash et al. 2008)) and GFRα3 (PDB code 2GH0(X. Wang et al. 2006)), which are receptors for the glial cell line-derived neurotrophic factor (GDNF) family ligands (GFLs), include GDNF (L. F. Lin et al. 1993), neurturin (NRTN) (Kotzbauer et al. 1996), persephin (PSPN) (Milbrandt et al. 1998), and artemin (ARTN) (Baloh et al. 1998). GFLs play critical roles in supporting the development and survival of distinct sets of central and peripheral neurons, in which GAS1 has also been implicated (Allen et al. 2011; Parkash & Goldman 2009). All GFLs signal through a two-receptor system. The first receptor, GDNF-family receptor (GFR), which is GPI anchored, is required for ligand binding. The GFL homodimer binds two molecules of GFR and the binding is specific: GDNF binds GFR1, NRTN binds GFR2, ARTN binds GFR3 and PSPN binds GFR4 (Airaksinen & Saarma 2002; Parkash & Goldman 2009). The second receptor, Rearranged during Transfection (RET), is a common signalling receptor for all the GFLs(Airaksinen et al. 1999). RET activation requires the formation of a heterohexameric complex GDNF(2):GFR1(2):RET(2) which leads to transphosphorylation and subsequent intracellular signalling (Airaksinen et al. 1999; Parkash & Goldman 2009). Intriguingly, it has been recently shown that GAS1 is able to inhibit the GDNF-

induced intracellular signalling pathway (López-Ramírez et al. 2008; Zarco et al. 2012).

Structural comparisons of GFRs and GAS1 reveal a similar overall arrangement of α-helices and disulphide bridges with the exception of disulphide bridge 4 (**Fig. 6.8A**). A major difference in the domain structure between the GFRs and GAS1 is the helix α5. In GFRs, helix α5 has a break caused by a rearrangement of disulphide bridge 4 (**Fig. 6.8A,D,E**). Structural superposition of GAS1 domain 1 with GFRα3 gives a RMSD of 1.73 Å for 61 Cα with a sequence identity of 16%, while that of GAS1 domain 2 with GFRα3 gives a RMSD of 1.52 Å for 55 Cα with a sequence identity of 12% (**Fig. 6.8D,E**). This data reveals a high degree of structural conservation between GAS1 and GFRs despite the low sequence identity. The structural superpositions also reveal that GAS1 domains 1 and 2 map onto the GFR region important for ligand interaction with GDNF and ARTN (**Fig. 6.8F,G**). Two major interfaces between the GFRs and their ligands exist. The first involves helices α1 and α2. These interface residues are broadly conserved between GAS1 domain 1 and 2 and the GFRs (**Fig 6.8A, F,G**). The second is helix α5, which is completely rearranged between GAS1 and GFRs. This makes it very difficult to predict whether GAS1 is able to bind either GDNF or ARTN, thus more functional experiments will need to be conducted before a definitive answer can be arrived at.

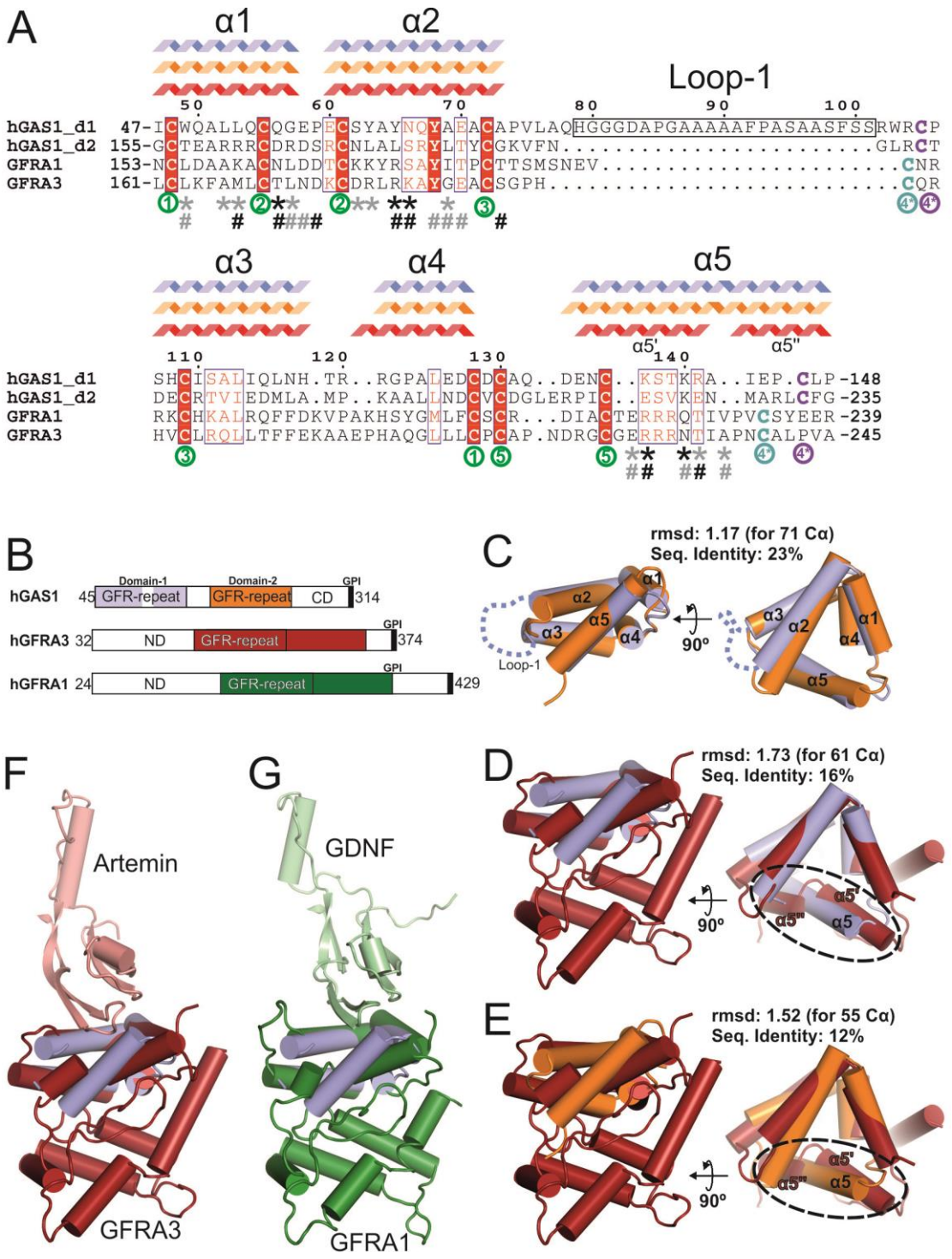


Fig. 6.8 Structural comparison of GAS1 and GDNF receptors GFR1 and GFRA3. (A) Structure based sequence alignment calculated with Protein3Dfit (http://protein3dfit.tu-bs.de/) and formatted with ESPript (espript.ibcp.fr/ESPript/ESPript/). Residue numbers of the domain boundaries used in the analysis below are shown. Secondary structure elements are shown above the sequence. Disulphide bridges are numbered. Note the rearrangement of disulphide bridge 4 (violet for GAS1 domains 1 and 2; cyan for GFRA1 and GFRA3). Interactions of GFRA1 with the GNDF ligand (PDB 3FUB (Parkash

et al. 2008)) are marked by asterisks (*), interactions of GFRA3 with ARTN (PDB 2GH0 (X. Wang et al. 2006)) by hashtags (#). The disordered GAS1 domain loop 1 is highlighted. (B) Schematic domain representations of human GAS1, GFRA1 and GFRA3. Same colour-coding is used in (C) to (G). **(C-E)** Structural superpositions of individual GFR repeat domains. C: GAS1 domain 1 and 2; D: GAS1 domain 1 and GFRA3; E: GAS1 domain 2 and GFRA3. RMSDs and sequence identities calculated with program SHP (Stuart et al. 1979) are shown. Helices of GAS1 domain 1 are marked. The rearrangement of helix α5 is highlighted in (D) and (E). (F-G) Superpostion of GAS1 domain 1 onto the GFRA3-ARTN (F) and the GFRA1-GNDF (G) complex.

6.2.8 GAS1 binds to SHH

To characterise HH binding to GAS1, a series of constructs based on the crystal structure of GAS1-ΔL2 were designed in order to determine which GAS1 domain binds to SHH. Three single domain constructs were designed using GAS1-ΔGPI as a template. A domain 1 construct extending from the N-terminal region after the signal peptide to the beginning of loop2 (L40-P148, termed GAS1-D1) and two domain 2 constructs were made. The first domain 2 construct started at the end of loop 2 to the end of the disordered C-terminal tail (V163-G314, termed GAS1-D2ΔGPI), while a shorter domain 2 construct started in the same position, ending before the disordered C-terminal region (V163-N250, termed GAS1-D2ΔC). All constructs were tested in small scale and shown to be secreted (**Fig 6.9B,C**). Previous studies have shown that the GAS1-SHH interaction is calcium dependent, although this was elucidated using pull-down experiments rather than SPR (McLellan et al. 2008). It was, therefore, important to test the role of calcium in the GAS1-HH interaction in SPR. Binding studies were carried out with wild type GAS1 and SHHN both in the presence and absence of calcium (**Fig. 6.10A,B**), the results of which agreed with the previous studies in pull-down experiments. Therefore, all subsequent SPR experiments were carried out in the presence of 5 mM $CaCl_2$.

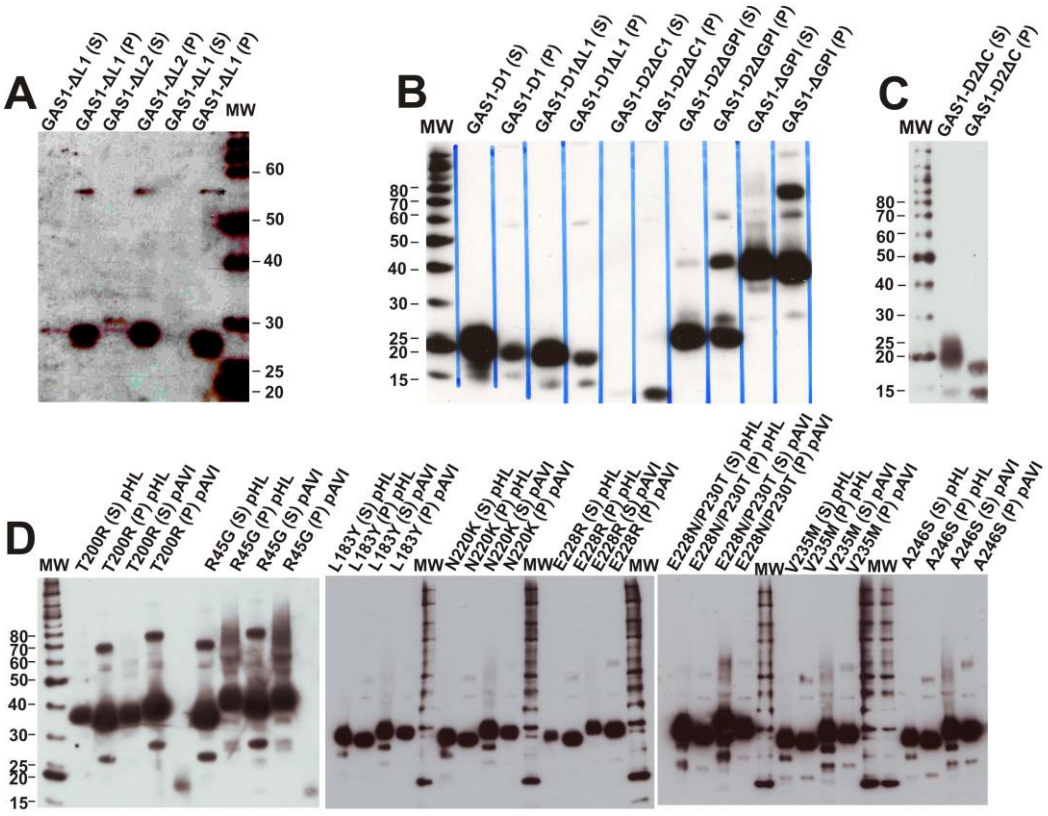


Fig. 6.9. Western blot analysis of GAS1 small-scale expression in HEK 293T cells. All constructs were tested for expression in small-scale HEK 293T. For all western blots, both secreted protein from conditioned media (S) and cell mass pellet (P) are shown. (**A**) Expression test for GAS1 loop deletion constructs. GAS1-ΔL2 was the best expresser, so was selected for large-scale production and crystallization. (**B**) Domain 1 (GAS1-D1 (L40-P148) and GAS1-D1ΔL1 (L40-P148 minus G80-A96)) and 2 (GAS1-D2ΔC1 (V163-A246) and GAS1-D2ΔGPI (V163-G314)) constructs and full GAS1 ectodomain (GAS1-ΔGPI (L40-G314)) of GAS1 including loop 1 and 2 in the biotinylation vector for SPR experiments. (**C**) Shorter domain 2 (GAS1-D2ΔC (V163-N250)) construct for SPR experiments. (**D**) GAS1 mutants based on the GAS1-ΔGPI template in both pHLsec and pHLAvitag3 (Aricescu et al. 2006).

Full length GAS1 (GAS1-ΔGPI) gave a dissociation constant (K_d) of 63±8 μM. Once this was established, the individual domains were then tested. The results clearly showed that domain 2 did bind to SHHN, although weaker binding was observed for GAS1-D2ΔGPI than for GAS1-ΔGPI (**Fig. 6.10A**), enabling the calculation of a dissociation constant (**Fig. 6.10B**), while GAS1-D1 did not bind (**Fig. 6.10D**). Of equal interest was the fact that GAS1-D2ΔC did not bind SHHN while GAS1-D2ΔGPI did. This suggests a role for the disordered C-terminal region (**Fig. 6.10C**).

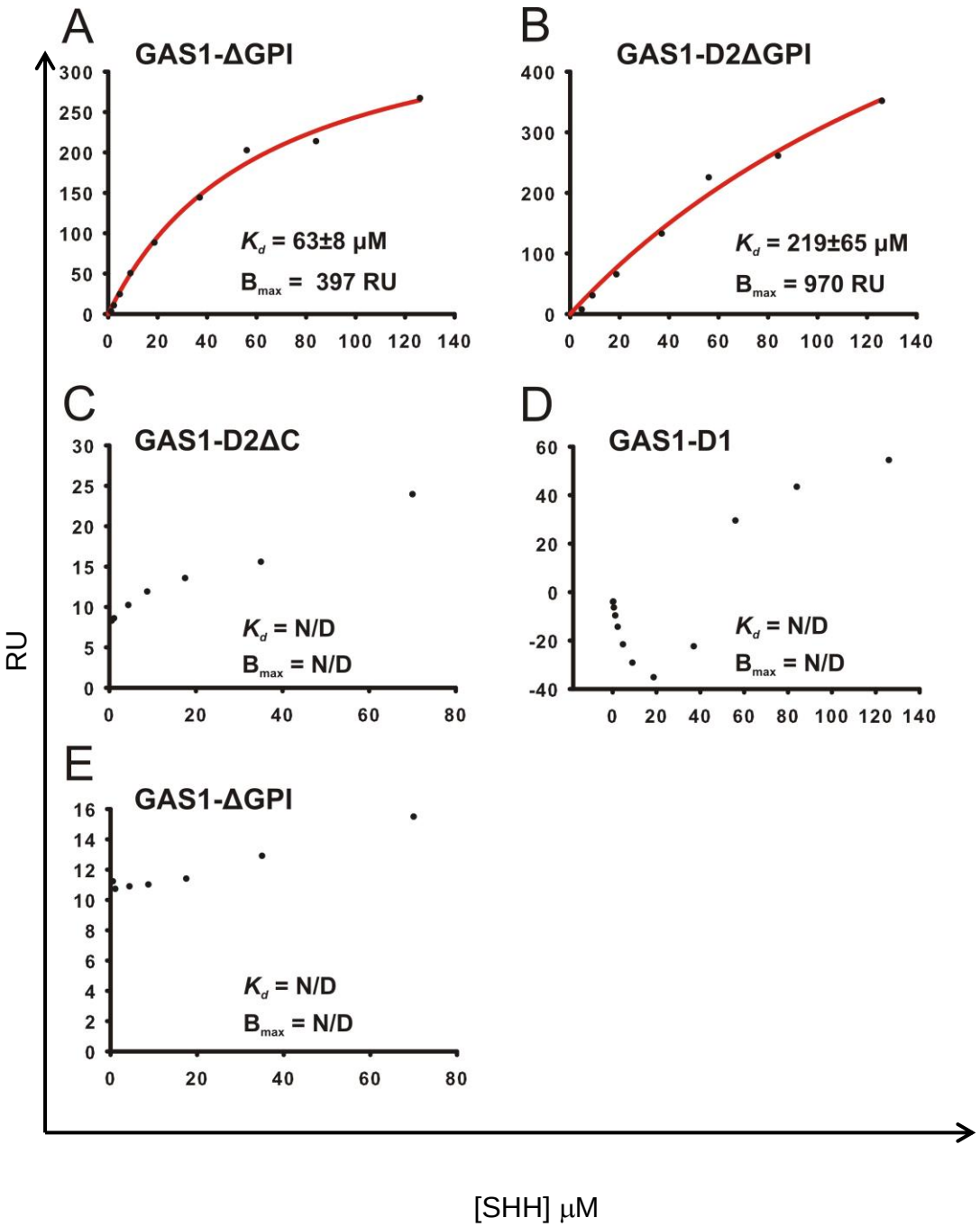


Fig. 6.10 SPR experiments of GAS1-SHH interactions. The role of the GAS1 domains was tested using SPR with GAS1 constructs as biotinylated ligand on the chip and murine SHH as analyte. **(A)** Wild-type construct lacking the GPI anchor, but containing both loop 1 and 2 (GAS1-ΔGPI). **(B)** Domain 2 construct with disordered C-terminal 60 residues present (GAS1-D2ΔGPI). **(C)** Domain 2 construct lacking the C-terminal 60 residues (GAS1-D2ΔC). **(D)** Domain 1 construct ending at the start of loop 2 and including loop 1 (GAS1-D1). **(E)** Wild-type construct lacking the GPI anchor, but containing both loop 1 and 2 in the absence of calcium. A binding constant could be calculated for GAS1-GPI and GAS1-D2ΔGPI, but there was no binding for either GAS1-D2ΔGPI or GAS1-D2ΔGPI or GAS1-ΔGPI in the absence of calcium.

6.2.9 Role of mutants linked to disease and the small molecule on HH binding

Several mutations linked to HPE have been identified (**Table 6.2** and **Fig. 6.10A**) (Pineda-Alvarez et al. 2012; Ribeiro et al. 2010). To understand the function of these mutations, binding studies were carried out using SPR. Two sets of mutants were designed on the GAS1-ΔGPI template to include both loop 1 and loop 2 in all proteins expressed for binding studies. The first set was based on clinical data (**Table 6.2**) to investigate the role of disease mutations. The second set was designed to determine the role of the small molecule seen in the electron density map. The first was a series of disease mutants base on clinical data (**Table 6.2**).

Small molecule mutants were designed to either disrupt a salt bridge located at the entrance of the binding pocket (**Fig. 6.7B**, marked by an asterisk) or to block a potential entry point to the interior of domain 2. Constructs incorporating disease mutants were based on mutations found in patients presenting with defective HH signalling phenotypes(Pineda-Alvarez et al. 2012). All mutants were sequence verified and tested in HEK 293T small-scale expression in which all were expressed (**Fig. 6.9D).**

Table 6.2 Disease mutations observed in GAS1. Patients presenting with various defective HH signalling phenotypes were tested for GAS1 mutations. Table adapted from Pineda-Alvarez et al. (2012)(Pineda-Alvarez et al. 2012).

GAS1 Disease Mutation	Clinical Data	Reference
R45G	Semilobar HPE, developmental delay, agenesis of the corpus callosum, microcephaly and spasticity	(Pineda-Alvarez et al. 2012)
T200R	Semilobar HPE, developmental delay, hypoplastic midface, upslanting palpebral fissures, proptosis, depressed nasal bridge, absent collumella, and bilateral cleft lip and palate	(Ribeiro et al. 2010)
N220K	Hypoplastic corpus callosum, diabetes insipidus, craniosynostosis of metopic suture, pyriform apeture stenosis, midline cleft lip with complete premaxillary agenesis	(Pineda-Alvarez et al. 2012)
A246S	No data available	(Ribeiro et al. 2010)

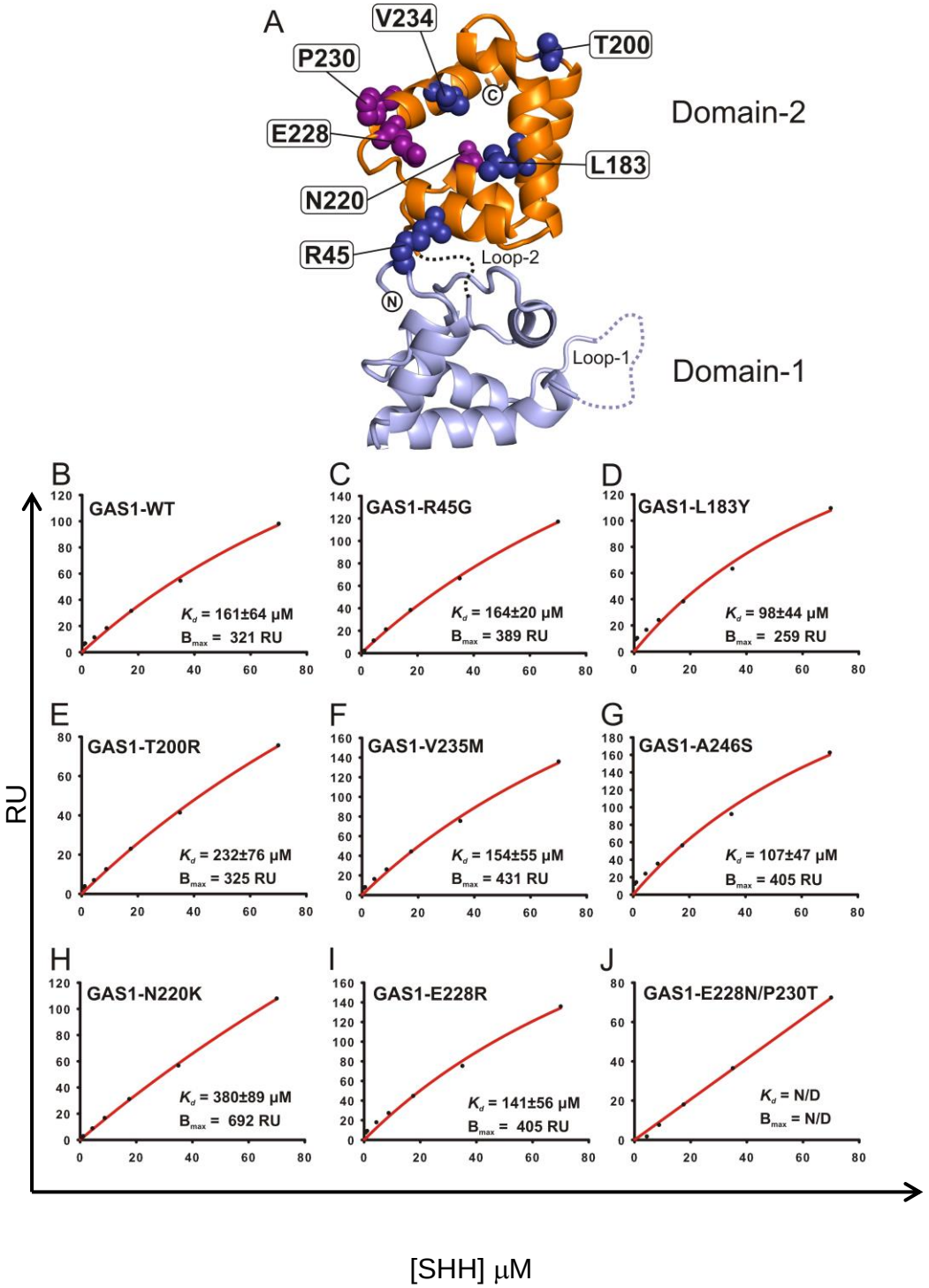


Fig. 6.11 SPR measurements of GAS1 disease and small molecule mutations. The role of disease mutations in GAS1 was tested using SPR with GAS1 constructs as biotinylated ligand on the chip and murine SHHN as analyte. (**A**) Structure of GAS1-ΔL2 with disease mutations (blue spheres) and small molecule mutations (lilac spheres) mapped onto the structure. (**B**) GAS1 construct lacking the GPI anchor, but containing both loop 1 and 2 (GAS1-WT). (**C**) GAS1-WT with the mutation R45G (GAS1-R45G). (**D**) GAS1-WT with the mutation L183Y (GAS1-L183Y). (**E**) GAS1-WT with the mutation T200R (GAS1-T200R). (**F**) GAS1-WT with the mutation V235M (GAS1-V235M). (**G**) GAS1-WT with the mutation A246S (GAS1-A246S). (**H**) GAS1-WT with the mutation N220K (GAS1-N220K). (**I**) GAS1-WT with the mutation E228R (GAS1-E228R). (**J**) GAS1-WT with the mutation E228N/P230T (GAS1-E228N/P230T).

Initial binding data shows that disease mutants (**Fig. 6.11D-G**) have no measurable effect on the binding constant. The small molecule mutations showed greater disruption to binding. GAS1-N220K showed a reduced dissociation constant from a K_d of 161 μM to 380 μM (**Fig. 6.11H**), while the GAS1-E228N/P230T double mutant gave data for which a binding constant could not be calculated (**Fig 6.10J**).

6.2.10 Discussion

Determination of the structure in three different crystal forms revealed that GAS1 has two independent domains that do not interact with each other. Due to the high structural similarity of GAS1 to GFRα1 and GFRα3, it will be interesting to investigate whether there is any interaction between GAS1 and GDNF ligands. Previous work (McLellan et al. 2008) revealed the importance of calcium in HH binding, but no SPR studies of GAS1 had previously been undertaken, all studies having been pull down studies (Martinelli & Fan 2007a). Calcium is clearly important for the binding of GAS1 to HH, without which, there is no binding. The K_d obtained from SPR (which has a dissociation constant in the 100-200 μM range) is much weaker than that reported previously in cell-based binding studies (with a dissociation constant of 6 nM) (C. S. Lee et al. 2001a). To determine why this is the case, SPR experiments using the full length, lipid modified, multimeric HH morphogen rather than the SHHN unmodified monomer with GAS1 will be required. It may be that this discrepancy is due to the presence of CDO and/or BOC being present in the cell-based study, which would have drastically altered the binding constant without the knowledge of the investigators. Domain analysis of GAS1 allowed the mapping of HH interactions with HH to domain 2. Therefore, GAS1-D2ΔGPI, which binds

murine SHHN will be produced in large scale and a full crystallization screen of GAS1-D2ΔGPI-SHHN complex will be set up.

The results obtained to date present a range of opportunities for further study. First, collaboration with Dr Frédéric Charron's Laboratory, (Institut de recherches cliniques de Montréal, Quebec Canada) is under way to test our constructs in animal models to assess the structure-based hypothesis *in vivo*. Additionally, experiments with GDNF will be conducted to (i) test whether GAS1 is able to bind GDNF and (ii) determine which domain is involved in GDNF binding. It is also imperative to identify what the small molecule seen in GAS1 domain 2 is, and what function it plays in either HH or GDNF signalling.

Chapter 7

Discussion

Chapter 7 DISCUSSION

This thesis set out to answer the question of how HH interacts with HSPGs, HHIP and GAS1, and what mechanisms were employed by agonists and antagonists in HH signalling. How do these proteins orchestrate such different outcomes? To answer this question, novel methods were required to increase the quantity and purity of protein being produced, which would allow enough time to both produce the protein required, and also to purify, crystallize, collect X-ray diffraction data, determine the three-dimensional structure of it, and carry out structure-guided biophysical and functional assays.

7.1 Automation of expression

An automated high-throughput mammalian cell-based expression system provides an important tool to obtain milligram quantities of high-quality proteins, especially for structural biology methods such as crystallization and NMR, which require large quantities of protein. An automated protein expression system should be time-efficient, allowing expression of multiple protein constructs in parallel. Manual tissue culture would allow time for at most two constructs to be expressed per week, if any other work is to be done. Automated methods allow one worker to express multiple large-scale constructs while still having time to carry out all the other work required of a structural biologist in a modern structural laboratory. Moreover, automated methods of cell culture and transfection result in a more reproducible system as the cell-culture robot executes every procedure identically, seeding the same number of cells in every flask, and in the same time-frame, allowing more precise planning to be implemented. Manual methods can give highly variable levels of protein for the same

construct, due in part to the time taken by the cell culture biologist as well as the relative experience of the worker. The automated system minimises handling by cell culture scientists, thereby reducing the risk of contamination. Minimising the risk of bacterial infection is particularly important in a structural biology laboratory as antibiotics are not added to culture media due to the reduction in yield associated with their addition. Media is always ready to harvest at the same time post-transfection for a given construct, again aiding planning. Automated methods were used extensively in the work presented in this thesis. GPC3, GAS1 and SHH were all made by both manual and automated methods, but it was protein from the automated method which yielded the crystals that gave the best diffraction data (Zhao et al. 2011).

7.2 HH receptors

In this thesis, the mammalian-specific HH receptors HHIP, GAS1, GPC3 were studied with the aim of answering three major questions: (1) Do these proteins interact directly with HH? (2) What is the nature of the interaction of these proteins with HH? (3) How do these antagonists and agonists orchestrate this fundamental signalling pathway?

7.2.1 SHH metal binding sites provide a hotspot for receptor interactions

HHIP, GPC3 and GAS1 provide an additional level of regulation for more complex organisms and are not found in invertebrates. What differences exist between vertebrates and invertebrates that allow this enhanced regulation of HH signalling? A major difference between invertebrate and vertebrate HH protein is the presence of three metal-binding sites in the vertebrate HH. One of these is lacking in the invertebrate HH protein and two have not been reported. The discovery of a zinc-

binding site in vertebrate HH proteins (Hall et al. 1995), which shares homology with the active sites of zinc hydrolases, caused great excitement as it suggested hydrolytic activity may be a feature of HH function. However, as mutagenesis studies (Day et al. 1999; Fuse et al. 1999) discounted this theory, the zinc site fell out of favour and was ignored (Beachy et al. 2010). The discovery that HHIP binds HH by occupying the zinc cleft and that the HH-bound zinc ion is directly coordinated by D383 from binding loop 1 of the third blade of the HHIP β-propeller (Chapter 4, **Fig. 4.5** and **Fig. 7.1A**) attributed a functionally important role for zinc in vertebrate HH (Bishop et al. 2009). The absence of the zinc ion binding site in *Drosophila* HH is surprising, but may suggest that it is the large substrate-binding cleft rather than the zinc ion itself that is most essential for mediating HH–PTCH interactions. Mutating binding loop 1 of HHIP (D383A) does not completely abolish binding between HHIP and SHH (Chapter 4, **Fig. 4.7**) (Bishop et al. 2009; Bosanac et al. 2009).

Two other metal binding sites, which bind calcium, were discovered recently(McLellan et al. 2008). They have proven to be very important in vertebrate regulation of the HH signalling pathway, while their role, if any, is unknown in *Drosophila*. The calcium binding sites were initially discovered to be essential for the interaction of SHH with its receptor CDO(McLellan et al. 2006) (**Fig. 7.1B**). Binding experiments in the presence of EGTA, a calcium chelator, and with SHH variants bearing mutations in their calcium-coordinating residues, along with the SPR data in this thesis (Chapter 4, **Fig. 4.7** and Chapter 6, **Fig. 6.10**) reveal that the calcium site contributes to the interaction between SHH and HHIP(Bishop et al. 2009) (Chapter 4 **Fig. 4.7**), GAS1 (Chapter 6, **Fig. 6.10**), PTCH (McLellan et al. 2008), and CDO/BOC (McLellan et al. 2008), although the reduction in binding to these partners is not as

significant as that observed for CDO/BOC (McLellan et al. 2008; Bishop et al. 2009; Kavran et al. 2010).

7.2.2 Multivalence of the active HH morphogen *in vivo* and implications in receptor binding

Lipid modification of HH does not seem to mediate HH-PTCH interactions. What then is the function of these HH-linked lipids? HH proteins were shown to be multimeric *in vivo* as part of a lipoprotein complex(Panáková et al. 2005; M.-H. Chen et al. 2004), resulting in high-density clustering of HH binding partners at localised patches on the surface of target cells (McLellan et al. 2008). The HH lipid modifications were shown to be essential for the formation of this multivalent HH complex (Panáková et al. 2005). The multimeric nature of the HH ligand can result in an avidity effect at sites of high local HH concentration, making the weak interactions of HH to binding partners, as measured by SPR using monomeric HH, functionally relevant (Kuriyan & Eisenberg 2007).

PTCH binds HH near the zinc cleft region on SHH (Fuse et al. 1999; Pepinsky et al. 1998). A region with great amino acid sequence similarity to the D383-containing loop of HHIP was identified in the second extracellular loop of PTCH in which 7 of 18 positions, including a D383 homologue, are conserved (Bosanac et al. 2009). A peptide encompassing this region from PTCH binds weakly to the zinc site in SHH and competes for binding with the HHIP D383-containing loop peptide, indicating that the binding sites for HHIP and PTCH on HH are overlapping (Bosanac et al. 2009).

Studies on the SHH-specific antibody 5E1, that blocks HH signalling(Ericson et al. 1996), further support this mode of binding of HH to PTCH. The recent structure of a complex of SHH and the 5E1 Fab revealed that 5E1 binds SHH across the zinc containing cleft that largely overlaps with the HHIP binding site (Maun et al. 2010) (**Fig 7.1C**). 5E1 also competes with a HHIP peptide designed on the SHH-binding D383 containing loop (binding loop 1) of HHIP, highlighting the functional dependence of both HHIP and 5E1 on the zinc cleft (Maun et al. 2010). Although 5E1 does not directly contact SHH-bound zinc or calcium ions, 5E1 requires both for optimal binding to SHH (Beachy et al. 2010). The 5E1-SHH structure further implicates the importance of the zinc cleft for HH signalling, suggesting that 5E1 inhibits HH signalling by competing directly with PTCH for SHH binding. Cell-based competition data involving pre-incubation of SHH with either 5E1 or soluble HHIP is consistent with this interpretation, as it prevents binding to PTCH-expressing cells (Beachy et al. 2010).

The known structures of HH in complex with both agonists (CDO) and antagonists (HHIP, PTCH, 5E1) show that HH binds in an overlapping manner with all four receptors (**Fig 7.1D**). Why then do some interactions confer a positive effect, while others are negative? It is possible that the 100-fold tighter binding of SHH to HHIP relative to CDO permits physiological levels of HHIP to fully occupy ligand sites in a multivalent complex. It may be that the lower affinity of positive regulators may result in occupancy of only a fraction of available ligand sites, and thus function to increase the concentration of ligands near the membrane that are available for interacting with PTCH. The determining factor for the effect of a particular protein would thus be whether its binding affinity and physiological concentration range

enable it to engage the HH proteins within multivalent complexes partially or fully. Given that the agonists CDO, BOC and GAS1 bind PTCH directly in the absence of HH (Izzi et al. 2011), it is not surprising that, with ~1000 times weaker binding of HH than PTCH, GAS1 and CDO/BOC have an agonistic effect on HH signalling when exposed to a multivalent HH complex. Conversely, HHIP, with its comparable binding of HH to PTCH, has an antagonistic effect.

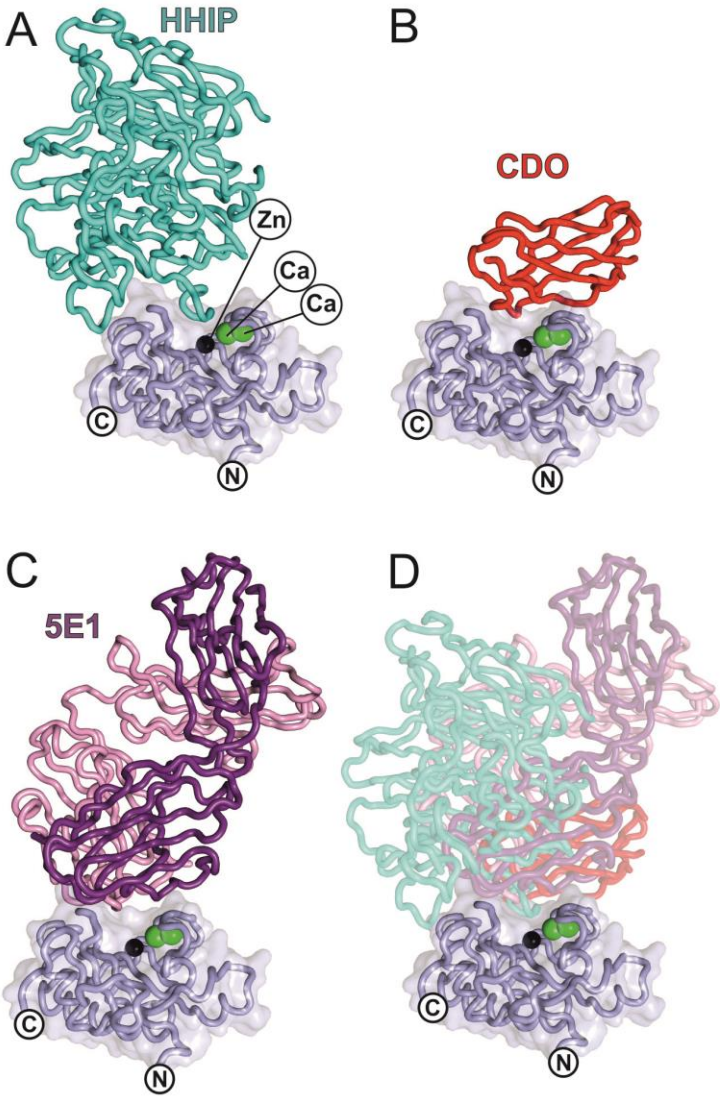


Fig. 7.1. Comparison of SHH-receptor complexes Ribbon representation of SHH (light-blue) complexes with (**A**) HHIP (PDB ID: 2WFX(Bishop et al. 2009)), (**B**) CDO (PDB ID: 3D1M(McLellan et al. 2008)) and (**C**) Antibody 5E1 (PDB ID: 3MXW(Maun et al. 2010)). (**D**) Superposition of the complexes from (**A-C**). The calcium (green) and zinc (black) sites are shown as spheres. SHH is in the same orientation in all panels. N- and C- termini are marked with a circled N or C respectively.

In summary, three extracellular modulators of the HH signalling pathway were studied in this thesis. In all cases, structural information from X-ray crystallography was combined with biophysical assays, suggesting specific signalling or activation mechanisms that had not been previously described. The work presented here clearly indicates a binding 'hot-spot' on HH where multiple ligands bind, offering a model to explain the mode by which HH signalling is modulated by extra-cellular proteins. Characterisation of this region is of paramount importance to fully understand HH's mode of action. The structures presented in this thesis have furthered our understanding of this fundamental developmental signalling pathway and could aid the design of structure specific drugs to combat developmental diseases and cancer.

References

Airaksinen, M.S. & Saarma, M., 2002. The GDNF family: signalling, biological functions and therapeutic value. *Nature reviews. Neuroscience*, 3(5), pp.383–394.

Airaksinen, M.S., Titievsky, A. & Saarma, M., 1999. GDNF family neurotrophic factor signaling: four masters, one servant? *Molecular and cellular neurosciences*, 13(5), pp.313–325.

Allen, B.L. et al., 2011. Overlapping roles and collective requirement for the coreceptors GAS1, CDO, and BOC in SHH pathway function. *Developmental cell*, 20(6), pp.775–787.

Allen, B.L., Tenzen, T. & McMahon, A.P., 2007. The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development. *Genes & development*, 21(10), pp.1244–1257.

Aricescu, A.R., Lu, W. & Jones, E.Y., 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 10), pp.1243–1250.

Aspöck, G. et al., 1999. Caenorhabditis elegans has scores of hedgehog-related genes: sequence and expression analysis. *Genome research*, 9(10), pp.909–923.

Ayers, K.L. et al., 2010. The long-range activity of Hedgehog is regulated in the apical extracellular space by the glypican Dally and the hydrolase Notum. *Developmental cell*, 18(4), pp.605–620.

Bae, G.-U. et al., 2011. Mutations in CDON, encoding a hedgehog receptor, result in holoprosencephaly and defective interactions with other hedgehog receptors. *American journal of human genetics*, 89(2), pp.231–240.

Baloh, R.H. et al., 1998. Artemin, a novel member of the GDNF ligand family, supports peripheral and central neurons and signals through the GFRalpha3-RET receptor complex. *Neuron*, 21(6), pp.1291–1302.

Beachy, P.A. et al., 2010. Interactions between Hedgehog proteins and their binding partners come into view. *Genes & development*, 24(18), pp.2001–2012.

Beckett, K., Franch-Marro, X. & Vincent, J.P., 2008. Glypican-mediated endocytosis of Hedgehog has opposite effects in flies and mice. *Trends in cell biology*, 18(8), pp.360–363.

Bendavid, C. et al., 2006. Molecular evaluation of foetuses with holoprosencephaly shows high incidence of microdeletions in the HPE genes. *Human genetics*, 119(1-2), pp.1–8.

Bernfield, M. et al., 1999. Functions of cell surface heparan sulfate proteoglycans. *Annual review of biochemistry*, 68, pp.729–777.

Bijlsma, M.F. et al., 2006. Repression of smoothened by patched-dependent (pro-

)vitamin D3 secretion. *PLoS biology*, 4(8), p.e232.

Bishop, B. et al., 2009. Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. *Nature structural & molecular biology*, 16(7), pp.698–703.

Blanc, E. et al., 2004. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 12 Pt 1), pp.2210–2221.

Bosanac, I. et al., 2009. The structure of SHH in complex with HHIP reveals a recognition role for the Shh pseudo active site in signaling. *Nature structural & molecular biology*, 16(7), pp.691–697.

Bourikas, D. et al., 2005. Sonic hedgehog guides commissural axons along the longitudinal axis of the spinal cord. *Nat Neurosci*, 8(3), pp.297–304.

Burke, R. et al., 1999. Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell*, 99(7), pp.803–815.

Cabrera, J.R. et al., 2006. Gas1 is related to the glial cell-derived neurotrophic factor family receptors alpha and regulates Ret signaling. *The Journal of biological chemistry*, 281(20), pp.14330–14339.

Cano-Gauci, D.F. et al., 1999. Glypican-3-deficient mice exhibit developmental overgrowth and some of the abnormalities typical of Simpson-Golabi-Behmel syndrome. *The Journal of cell biology*, 146(1), pp.255–264.

Capurro, M.I. et al., 2008. Glypican-3 inhibits Hedgehog signaling during development by competing with patched for Hedgehog binding. *Developmental cell*, 14(5), pp.700–711.

Capurro, M.I., Li, F. & Filmus, J., 2009. Overgrowth of a mouse model of Simpson-Golabi-Behmel syndrome is partly mediated by Indian hedgehog. *EMBO reports*, 10(8), pp.901–907.

Capurro, M.I., Shi, W. & Filmus, J., 2012. LRP1 mediates the Shh-induced endocytosis of the GPC3-Shh complex. *Journal of cell science*.

Capurro, M.I., Shi, W., et al., 2005a. Processing by convertases is not required for glypican-3-induced stimulation of hepatocellular carcinoma growth. *The Journal of biological chemistry*, 280(50), pp.41201–41206.

Capurro, M.I., Xiang, Y.-Y., et al., 2005b. Glypican-3 promotes the growth of hepatocellular carcinoma by stimulating canonical Wnt signaling. *Cancer research*, 65(14), pp.6245–6254.

Casali, A. & Struhl, G., 2004. Reading the Hedgehog morphogen gradient by measuring the ratio of bound to unbound Patched protein. *Nature*, 431(7004), pp.76–80.

Chamoun, Z. et al., 2001. Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. *Science (New York, N.Y.)*, 293(5537), pp.2080–2084.

Chang, D.T. et al., 1994. Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development (Cambridge, England)*, 120(11), pp.3339–3353.

Chang, V.T. et al., 2007. Glycoprotein structural genomics: solving the glycosylation problem. *Structure (London, England : 1993)*, 15(3), pp.267–273.

Chen, J.K. et al., 2002. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes & development*, 16(21), pp.2743–2748.

Chen, M.-H. et al., 2004. Palmitoylation is required for the production of a soluble multimeric Hedgehog protein complex and long-range signaling in vertebrates. *Genes & development*, 18(6), pp.641–659.

Chen, M.-H., Wilson, C.W. & Chuang, P.-T., 2007a. SnapShot: hedgehog signaling pathway. *Cell*, 130(2), p.386.

Chen, V.B. et al., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 1), pp.12–21.

Chen, X. et al., 2007b. Hedgehog signal pathway is activated in ovarian carcinomas, correlating with cell proliferation: it's inhibition leads to growth suppression and apoptosis. *Cancer science*, 98(1), pp.68–76.

Chen, Y. & Struhl, G., 1996. Dual roles for patched in sequestering and transducing Hedgehog. *Cell*, 87(3), pp.553–563.

Chiao, E. et al., 2002. Overgrowth of a mouse model of the Simpson-Golabi-Behmel syndrome is independent of IGF signaling. *Developmental biology*, 243(1), pp.185–206.

Christ, A. et al., 2012. LRP2 is an auxiliary SHH receptor required to condition the forebrain ventral midline for inductive signals. *Developmental cell*, 22(2), pp.268–278.

Chuang, P.-T., Kawcak, T. & McMahon, A.P., 2003. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes & development*, 17(3), pp.342–347.

Chuang, P.T. & McMahon, A.P., 1999. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature*, 397(6720), pp.617–621.

Coles, C.H. et al., 2011. Proteoglycan-specific molecular switch for RPTPσ clustering and neuronal extension. *Science (New York, N.Y.)*, 332(6028), pp.484–488.

Cooper, M.K. et al., 1998. Teratogen-mediated inhibition of target tissue response to Shh signaling. *Science (New York, N.Y.)*, 280(5369), pp.1603–1607.

Corbit, K.C. et al., 2005. Vertebrate Smoothened functions at the primary cilium. *Nature*, 437(7061), pp.1018–1021.

Corcoran, R.B. & Scott, M.P., 2006. Oxysterols stimulate Sonic hedgehog signal transduction and proliferation of medulloblastoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, 103(22), pp.8408–8413.

Coulombe, J. et al., 2004. Hedgehog interacting protein in the mature brain: membrane-associated and soluble forms. *Molecular and cellular neurosciences*, 25(2), pp.323–333.

Cowtan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 9), pp.1002–1011.

Crompton, T., Outram, S.V. & Hager-Theodorides, A.L., 2007. Sonic hedgehog signalling in T-cell development and activation. *Nature reviews. Immunology*, 7(9), pp.726–735.

Currie, P.D. & Ingham, P.W., 1996. Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish. *Nature*, 382(6590), pp.452–455.

Dashti, M., Peppelenbosch, M.P. & Rezaee, F., 2012. Hedgehog signalling as an antagonist of ageing and its associated diseases. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 34(10), pp.849–856.

Day, E.S. et al., 1999. Zinc-dependent structural stability of human Sonic hedgehog. *Biochemistry*, 38(45), pp.14868–14880.

De Cat, B. et al., 2003. Processing by proprotein convertases is required for glypican-3 modulation of cell survival, Wnt signaling, and gastrulation movements. *The Journal of cell …*.

DEMYER, W. & ZEMAN, W., 1963. Alobar holoprosencephaly (arhinencephaly) with median cleft lip and palate: clinical, electroencephalographic and nosologic considerations. *Confinia neurologica*, 23, pp.1–36.

Desbordes, S.C. & Sanson, B., 2003. The glypican Dally-like is required for Hedgehog signalling in the embryonic epidermis of Drosophila. *Development (Cambridge, England)*, 130(25), pp.6245–6255.

Dubourg, C. et al., 2007. Holoprosencephaly. *Orphanet journal of rare diseases*, 2, p.8.

Dubourg, C. et al., 2004. Molecular screening of SHH, ZIC2, SIX3, and TGIF genes in patients with features of holoprosencephaly spectrum: Mutation review and genotype-phenotype correlations. *Human mutation*, 24(1), pp.43–51.

Durocher, Y., Perret, S. & Kamen, A., 2002. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic acids research*, 30(2), p.E9.

Echelard, Y. et al., 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell*, 75(7), pp.1417–1430.

Eisenhaber, B. et al., 2000. Automated annotation of GPI anchor sites: case study C. elegans. *Trends in biochemical sciences*, 25(7), pp.340–341.

Emsley, P. & Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 12 Pt 1), pp.2126–2132.

Endoh-Yamagami, S. et al., 2009. The mammalian Cos2 homolog Kif7 plays an essential role in modulating Hh signal transduction during development. *Current biology : CB*, 19(15), pp.1320–1326.

Ericson, J. et al., 1997. Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell*, 90(1), pp.169–180.

Ericson, J. et al., 1996. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell*, 87(4), pp.661–673.

Esko, J.D. & Selleck, S.B., 2002. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annual review of biochemistry*, 71, pp.435–471.

Ferrante, M.I. et al., 2006. Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification. *Nature genetics*, 38(1), pp.112–117.

Filmus, J. & Capurro, M., 2008. The role of glypican-3 in the regulation of body size and cancer. *Cell cycle (Georgetown, Tex.)*, 7(18), pp.2787–2790.

Filmus, J., Capurro, M. & Rast, J., 2008. Glypicans. *Genome Biology*, 9(5), p.224.

Fransson, L.-A., 2003. Glypicans. *The international journal of biochemistry & cell biology*, 35(2), pp.125–129.

Fuse, N. et al., 1999. Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proceedings of the National Academy of Sciences of the United States of America*, 96(20), pp.10992–10999.

G, B. et al., 2011. autoBuster.

Gallet, A., 2011. Hedgehog morphogen: from secretion to reception. *Trends in cell biology*, 21(4), pp.238–246.

Goetz, S.C. & Anderson, K.V., 2010. The primary cilium: a signalling centre during vertebrate development. *Nature reviews. Genetics*, 11(5), pp.331–344.

Hall, T.M. et al., 1995. A potential catalytic site revealed by the 1.7-A crystal structure of the amino-terminal signalling domain of Sonic hedgehog. *Nature*, 378(6553), pp.212–216.

Han, C. et al., 2004. Drosophila glypicans control the cell-to-cell movement of Hedgehog by a dynamin-independent process. *Development (Cambridge, England)*, 131(3), pp.601–611.

Han, Y.-G. et al., 2009. Dual and opposing roles of primary cilia in medulloblastoma development. *Nat Med*, 15(9), pp.1062–1065.

Harding, M.M., 2001. Geometry of metal-ligand interactions in proteins. *Acta crystallographica. Section D, Biological crystallography*, 57(Pt 3), pp.401–411.

Haycraft, C.J. et al., 2005. Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. *PLoS genetics*, 1(4), p.e53.

Hofer, A.M. & Brown, E.M., 2003. Extracellular calcium sensing and signalling. *Nature reviews. Molecular cell biology*, 4(7), pp.530–538.

Hoff, Von, D.D. et al., 2009. Inhibition of the Hedgehog Pathway in Advanced Basal-Cell Carcinoma. *New England Journal of Medicine*, 361(12), pp.1164–1172.

Hooper, J.E. & Scott, M.P., 2005. Communicating with Hedgehogs. *Nature reviews. Molecular cell biology*, 6(4), pp.306–317.

Humke, E.W. et al., 2010. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes & development*, 24(7), pp.670–682.

Ingham, P.W., 2008. Hedgehog signalling. *Current biology : CB*, 18(6), pp.R238–41.

Ingham, P.W. & McMahon, A.P., 2001. Hedgehog signaling in animal development: paradigms and principles. *Genes & development*, 15(23), pp.3059–3087.

Ingham, P.W., Nakano, Y. & Seger, C., 2011. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nature reviews. Genetics*, 12(6), pp.393–406.

Izzi, L. et al., 2011. Boc and Gas1 each form distinct Shh receptor complexes with Ptch1 and are required for Shh-mediated cell proliferation. *Developmental cell*, 20(6), pp.788–801.

Jackson, S.M. et al., 1997. dally, a Drosophila glypican, controls cellular responses to the TGF-beta-related morphogen, Dpp. *Development (Cambridge, England)*, 124(20), pp.4113–4120.

Janssen, B.J.C. et al., 2010. Structural basis of semaphorin-plexin signalling. *Nature*, 467(7319), pp.1118–1122.

Johnson, J.-L.F.A. et al., 2012. Scube activity is necessary for Hedgehog signal

transduction in vivo. *Developmental biology*.

Johnson, K.G. et al., 2006. The HSPGs Syndecan and Dallylike bind the receptor phosphatase LAR and exert distinct effects on synaptic development. *Neuron*, 49(4), pp.517–531.

Johnson, R.L., Zhou, L. & Bailey, E.C., 2002. Distinct consequences of sterol sensor mutations in Drosophila and mouse patched homologs. *Developmental biology*, 242(2), pp.224–235.

Kang, J.-S., Zhang, W. & Krauss, R.S., 2007. Hedgehog signaling: cooking with Gas1. *Science's STKE : signal transduction knowledge environment*, 2007(403), p.pe50.

Kar, S. et al., 2012. Intricacies of hedgehog signaling pathways: A perspective in tumorigenesis. *Experimental cell research*, 318(16), pp.1959–1972.

Kavran, J.M. et al., 2010. All mammalian Hedgehog proteins interact with cell adhesion molecule, down-regulated by oncogenes (CDO) and brother of CDO (BOC) in a conserved manner. *The Journal of biological chemistry*, 285(32), pp.24584–24590.

Kawahira, H. et al., 2003. Combined activities of hedgehog signaling inhibitors regulate pancreas development. *Development (Cambridge, England)*, 130(20), pp.4871–4879.

Kayed, H. et al., 2005. Localization of the human hedgehog-interacting protein (Hip) in the normal and diseased pancreas. *Molecular carcinogenesis*, 42(4), pp.183–192.

Kim, H. et al., 2003. The heparan sulfate proteoglycan GPC3 is a potential lung tumor suppressor. *American journal of respiratory cell and molecular biology*, 29(6), pp.694–701.

Kim, M.-S. et al., 2011. Structure of the protein core of the glypican Dally-like and localization of a region important for hedgehog signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32), pp.13112–13117.

Kirkpatrick, C.A. & Selleck, S.B., 2007. Heparan sulfate proteoglycans at a glance. *Journal of cell science*, 120(Pt 11), pp.1829–1832.

Kirkpatrick, C.A. et al., 2006. The function of a Drosophila glypican does not depend entirely on heparan sulfate modification. *Developmental biology*, 300(2), pp.570–582.

Kotzbauer, P.T. et al., 1996. Neurturin, a relative of glial-cell-line-derived neurotrophic factor. *Nature*, 384(6608), pp.467–470.

Krauss, S., Concordet, J.P. & Ingham, P.W., 1993. A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing

activity in zebrafish embryos. *Cell*, 75(7), pp.1431–1444.

Kreuger, J. et al., 2006. Interactions between heparan sulfate and proteins: the concept of specificity. *The Journal of cell biology*, 174(3), pp.323–327.

Kubo, M. et al., 2004. Hedgehog signaling pathway is a new therapeutic target for patients with breast cancer. *Cancer research*, 64(17), pp.6071–6074.

Kuriyan, J. & Eisenberg, D., 2007. The origin of protein interactions and allostery in colocalization. *Nature*, 450(7172), pp.983–990.

Kuwabara, P.E. et al., 2000. A C. elegans patched gene, ptc-1, functions in germ-line cytokinesis. *Genes & development*, 14(15), pp.1933–1944.

Lecuit, T. & Le Goff, L., 2007. Orchestrating size and shape during morphogenesis. *Nature*, 450(7167), pp.189–192.

Lee, C.S. & Fan, C.M., 2001. Embryonic expression patterns of the mouse and chick Gas1 genes. *Mechanisms of development*, 101(1-2), pp.293–297.

Lee, C.S., Buttitta, L. & Fan, C.M., 2001a. Evidence that the WNT-inducible growth arrest-specific gene 1 encodes an antagonist of sonic hedgehog signaling in the somite. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), pp.11347–11352.

Lee, J.J. et al., 1994. Autoproteolysis in hedgehog protein biogenesis. *Science (New York, N.Y.)*, 266(5190), pp.1528–1537.

Lee, K.K. et al., 2001b. Functions of the growth arrest specific 1 gene in the development of the mouse embryo. *Developmental biology*, 234(1), pp.188–203.

Leoncini, E. et al., 2008. Frequency of holoprosencephaly in the International Clearinghouse Birth Defects Surveillance Systems: searching for population variations. *Birth defects research. Part A, Clinical and molecular teratology*, 82(8), pp.585–591.

Li, F. et al., 2011a. Glypican-5 stimulates rhabdomyosarcoma cell proliferation by activating Hedgehog signaling. *The Journal of cell biology*, 192(4), pp.691–704.

Li, X. et al., 2011b. Importance of hedgehog interacting protein and other lung function genes in asthma. *The Journal of allergy and clinical immunology*, 127(6), pp.1457–1465.

Lin, H. et al., 1999. Frequent silencing of the GPC3 gene in ovarian cancer cell lines. *Cancer research*, 59(4), pp.807–810.

Lin, L.F. et al., 1993. GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science (New York, N.Y.)*, 260(5111), pp.1130–1132.

Lin, X. & Perrimon, N., 1999. Dally cooperates with Drosophila Frizzled 2 to transduce Wingless signalling. *Nature*, 400(6741), pp.281–284.

Liu, Y., May, N.R. & Fan, C.M., 2001. Growth arrest specific gene 1 is a positive growth regulator for the cerebellum. *Developmental biology*, 236(1), pp.30–45.

Low, J.A. & de Sauvage, F.J., 2010. Clinical experience with Hedgehog pathway inhibitors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 28(36), pp.5321–5326.

López-Ramírez, M.A. et al., 2008. Gas1 reduces Ret tyrosine 1062 phosphorylation and alters GDNF-mediated intracellular signaling. *International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience*, 26(5), pp.497–503.

Lum, L. et al., 2003. Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. *Science (New York, N.Y.)*, 299(5615), pp.2039–2045.

Ma, Y. et al., 2002. Hedgehog-mediated patterning of the mammalian embryo requires transporter-like function of dispatched. *Cell*, 111(1), pp.63–75.

Madison, B.B. et al., 2005. Epithelial hedgehog signals pattern the intestinal crypt-villus axis. *Development (Cambridge, England)*, 132(2), pp.279–289.

Malinauskas, T. et al., 2011. Modular mechanism of Wnt signaling inhibition by Wnt inhibitory factor 1. *Nature structural & molecular biology*, 18(8), pp.886–893.

Marigo, V. et al., 1996. Biochemical evidence that patched is the Hedgehog receptor. *Nature*, 384(6605), pp.176–179.

Martin, S.T. et al., 2005. Aberrant methylation of the Human Hedgehog interacting protein (HHIP) gene in pancreatic neoplasms. *Cancer biology & therapy*, 4(7), pp.728–733.

Martinelli, D.C. & Fan, C.-M., 2007a. Gas1 extends the range of Hedgehog action by facilitating its signaling. *Genes & development*, 21(10), pp.1231–1243.

Martinelli, D.C. & Fan, C.-M., 2007b. The role of Gas1 in embryonic development and its implications for human disease. *Cell cycle (Georgetown, Tex.)*, 6(21), pp.2650–2655.

Martín, V. et al., 2001. The sterol-sensing domain of Patched protein seems to control Smoothened activity through Patched vesicular trafficking. *Current biology : CB*, 11(8), pp.601–607.

Matsunaga, E. & Shiota, K., 1977. Holoprosencephaly in human embryos: epidemiologic studies of 150 cases. *Teratology*, 16(3), pp.261–272.

Maun, H.R. et al., 2010. Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. *The Journal of biological chemistry*, 285(34), pp.26570–26580.

Mayo, C.J. et al., 2005. Benefits of Automated Crystallization Plate Tracking, Imaging, and Analysis. *Structure (London, England : 1993)*, 13(2), pp.175–182.

McCarthy, R.A. et al., 2002. Megalin functions as an endocytic sonic hedgehog receptor. *The Journal of biological chemistry*, 277(28), pp.25660–25667.

McCoy, A.J. et al., 2007. Phaser crystallographic software. *Journal of applied crystallography*, 40(Pt 4), pp.658–674.

McLellan, J.S. et al., 2006. Structure of a heparin-dependent complex of Hedgehog and Ihog. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), pp.17208–17213.

McLellan, J.S. et al., 2008. The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. *Nature*, 455(7215), pp.979–983.

McMahon, A.P., Ingham, P.W. & Tabin, C.J., 2003. Developmental roles and clinical significance of hedgehog signaling. *Current topics in developmental biology*, 53, pp.1–114.

Meissner, W., Thomae, R. & Seifart, K.H., 2002. The activity of transcription factor IIIC1 is impaired during differentiation of F9 cells. *The Journal of biological chemistry*, 277(9), pp.7148–7156.

Milbrandt, J. et al., 1998. Persephin, a novel neurotrophic factor related to GDNF and neurturin. *Neuron*, 20(2), pp.245–253.

Ming, J.E. et al., 2002. Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. *Human genetics*, 110(4), pp.297–301.

Morales, C.R. et al., 2006. Epithelial trafficking of Sonic hedgehog by megalin. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 54(10), pp.1115–1127.

Muenke, M. & Beachy, P.A., 2000. Genetics of ventral forebrain development and holoprosencephaly. *Current opinion in genetics & development*, 10(3), pp.262–269.

Nachtergaele, S. et al., 2012. Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nature chemical biology*, 8(2), pp.211–220.

Nakamura, M. et al., 2007. Anti-patched-1 antibodies suppress hedgehog signaling pathway and pancreatic cancer proliferation. *Anticancer research*, 27(6A), pp.3743–3747.

Neri, G. et al., 1998. Clinical and molecular aspects of the Simpson-Golabi-Behmel syndrome. *American journal of medical genetics*, 79(4), pp.279–283.

Nettleship, J.E. et al., 2010. Recent advances in the production of proteins in insect and mammalian cells for structural biology. *Journal of structural biology*, 172(1), pp.55–65.

Nusse, R., 2003. Wnts and Hedgehogs: lipid-modified proteins and similarities in

signaling mechanisms at the cell surface. *Development (Cambridge, England)*, 130(22), pp.5297–5305.

Nüsslein-Volhard, C. & Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. *Nature*, 287(5785), pp.795–801.

O'Callaghan C, A. et al., 1999. BirA enzyme: production and application in the study of membrane receptor-ligand interactions by site-specific biotinylation. *Analytical biochemistry*, 266(1), pp.9–15.

Ohgami, N. et al., 2004. Binding between the Niemann-Pick C1 protein and a photoactivatable cholesterol analog requires a functional sterol-sensing domain. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34), pp.12473–12478.

Ohkawara, B. et al., 2003. Role of glypican 4 in the regulation of convergent extension movements during gastrulation in Xenopus laevis. *Development (Cambridge, England)*, 130(10), pp.2129–2138.

Ohlig, S. et al., 2011. Sonic hedgehog shedding results in functional activation of the solubilized protein. *Developmental cell*, 20(6), pp.764–774.

Oliver, F. et al., 2005. Regulatory variation at glypican-3 underlies a major growth QTL in mice. *PLoS biology*, 3(5), p.e135.

Olsen, C.L. et al., 2004. Hedgehog-interacting protein is highly expressed in endothelial cells but down-regulated during angiogenesis and in several human tumors. *BMC Cancer*, 4, p.43.

Onishi, H. & Katano, M., 2011. Hedgehog signaling pathway as a therapeutic target in various types of cancer. *Cancer science*, 102(10), pp.1756–1760.

Orioli, I.M. et al., 2001. Identification of novel mutations in SHH and ZIC2 in a South American (ECLAMC) population with holoprosencephaly. *Human genetics*, 109(1), pp.1–6.

Otwinowski, Z. & Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, 276, pp.307–326.

Panáková, D. et al., 2005. Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature*, 435(7038), pp.58–65.

Parkash, V. & Goldman, A., 2009. Comparison of GFL-GFRalpha complexes: further evidence relating GFL bend angle to RET signalling. *Acta crystallographica. Section F, Structural biology and crystallization communications*, 65(Pt 6), pp.551–558.

Parkash, V. et al., 2008. The structure of the glial cell line-derived neurotrophic factor-coreceptor complex: insights into RET signaling and heparin binding. *The Journal of biological chemistry*, 283(50), pp.35164–35172.

Parkin, C.A. & Ingham, P.W., 2008. The adventures of Sonic Hedgehog in development and repair. I. Hedgehog signaling in gastrointestinal development and disease. *American journal of physiology. Gastrointestinal and liver physiology*, 294(2), pp.G363–7.

Parmantier, E. et al., 1999. Schwann cell-derived Desert hedgehog controls the development of peripheral nerve sheaths. *Neuron*, 23(4), pp.713–724.

Pathi, S. et al., 2001. Comparative biological responses to human Sonic, Indian, and Desert hedgehog. *Mechanisms of development*, 106(1-2), pp.107–117.

Pepinsky, R.B. et al., 1998. Identification of a palmitic acid-modified form of human Sonic hedgehog. *The Journal of biological chemistry*, 273(22), pp.14037–14045.

Pénisson-Besnier, I. et al., 2008. Carotid artery dissection in an adult with the Simpson-Golabi-Behmel syndrome. *American journal of medical genetics. Part A*, 146A(4), pp.464–467.

Pilia, G. et al., 1996. Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nature genetics*, 12(3), pp.241–247.

Pillai, S.G. et al., 2009. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS genetics*, 5(3), p.e1000421.

Pineda-Alvarez, D.E. et al., 2012. Missense substitutions in the GAS1 protein present in holoprosencephaly patients reduce the affinity for its ligand, SHH. *Human genetics*, 131(2), pp.301–310.

Porter, J.A., Ekker, S.C., et al., 1996a. Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell*, 86(1), pp.21–34.

Porter, J.A., Young, K.E. & Beachy, P.A., 1996b. Cholesterol modification of hedgehog signaling proteins in animal development. *Science (New York, N.Y.)*, 274(5285), pp.255–259.

Radhakrishnan, A. et al., 2004. Direct binding of cholesterol to the purified membrane region of SCAP: mechanism for a sterol-sensing domain. *Molecular cell*, 15(2), pp.259–268.

Redmond, E.M. et al., 2011. Investigational Notch and Hedgehog inhibitors--therapies for cardiovascular disease. *Expert opinion on investigational drugs*, 20(12), pp.1649–1664.

Reeves, P.J. et al., 2002. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proceedings of the National Academy of Sciences of the United States of America*, 99(21), pp.13419–13424.

Ribeiro, L.A. & Richieri-Costa, A., 2005. Single median maxillary central incisor, hypophyseal tumor, and SHH mutation. *American journal of medical genetics. Part A*, 136A(4), pp.346–347.

Ribeiro, L.A. et al., 2010. Holoprosencephaly and holoprosencephaly-like phenotype and GAS1 DNA sequence changes: Report of four Brazilian patients. *American journal of medical genetics. Part A*, 152A(7), pp.1688–1694.

Riddle, R.D. et al., 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell*, 75(7), pp.1401–1416.

Robarge, K.D. et al., 2009. GDC-0449-a potent inhibitor of the hedgehog pathway. *Bioorganic & medicinal chemistry letters*, 19(19), pp.5576–5581.

Roelink, H. et al., 1994. Floor plate and motor neuron induction by vhh-1, a vertebrate homolog of hedgehog expressed by the notochord. *Cell*, 76(4), pp.761–775.

Roessler, E. et al., 2003. Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23), pp.13424–13429.

Rohatgi, R., Milenkovic, L. & Scott, M.P., 2007. Patched1 regulates hedgehog signaling at the primary cilium. *Science (New York, N.Y.)*, 317(5836), pp.372–376.

Rowbotham, N.J. et al., 2007. A novel role for Hedgehog in T-cell receptor signaling: implications for development and immunity. *Cell cycle (Georgetown, Tex.)*, 6(17), pp.2138–2142.

Rudenko, G. et al., 2002. Structure of the LDL receptor extracellular domain at endosomal pH. *Science (New York, N.Y.)*, 298(5602), pp.2353–2358.

Ruiz-Perez, V.L. et al., 2000. Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodental dysostosis. *Nature genetics*, 24(3), pp.283–286.

Ryan, K.E. & Chiang, C., 2012. Hedgehog secretion and signal transduction in vertebrates. *The Journal of biological chemistry*, 287(22), pp.17905–17913.

Sali, A. & Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3), pp.779–815.

Sasaki, H. et al., 1997. A binding site for Gli proteins is essential for HNF-3beta floor plate enhancer activity in transgenics and can respond to Shh in vitro. *Development (Cambridge, England)*, 124(7), pp.1313–1322.

SAUVAGE, F.D., 2012. US2011054877 MUTANT SMOOTHENED AND METHODS OF USING THE SAME. *WO Patent WO/2012/047,968*.

Schatz, P.J., 1993. Use of Peptide Libraries to Map the Substrate Specificity of a

Peptide-Modifying Enzyme: A 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia coli. *Nature biotechnology*, 11(10), pp.1138–1143.

Schneider, C., King, R.M. & Philipson, L., 1988. Genes specifically expressed at growth arrest of mammalian cells. *Cell*, 54(6), pp.787–793.

Schneider, T.R. & Sheldrick, G.M., 2002. Substructure solution with SHELXD. *Acta crystallographica. Section D, Biological crystallography*, 58(Pt 10 Pt 2), pp.1772–1779.

Seiradake, E. et al., 2010. An extracellular steric seeding mechanism for Eph-ephrin signaling platform assembly. *Nature structural & molecular biology*, 17(4), pp.398–402.

Seppala, M. et al., 2007. Gas1 is a modifier for holoprosencephaly and genetically interacts with sonic hedgehog. *The Journal of clinical investigation*, 117(6), pp.1575–1584.

Seshagiri, S. et al., 2012. Recurrent R-spondin fusions in colon cancer. *Nature*, 488(7413), pp.660–664.

Shi, W. & Filmus, J., 2009. A patient with the Simpson-Golabi-Behmel syndrome displays a loss-of-function point mutation in GPC3 that inhibits the attachment of this proteoglycan to the cell surface. *American journal of medical genetics. Part A*, 149A(3), pp.552–554.

Singla, V. & Reiter, J.F., 2006. The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science (New York, N.Y.)*, 313(5787), pp.629–633.

Sohka, T. et al., 2009. An externally tunable bacterial band-pass filter. *Proceedings of the National Academy of Sciences of the United States of America*, 106(25), pp.10135–10140.

Song, H.H. et al., 2005. The loss of glypican-3 induces alterations in Wnt signaling. *The Journal of biological chemistry*, 280(3), pp.2116–2125.

Söding, J., Biegert, A. & Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research*, 33(Web Server issue), pp.W244–8.

Springer, T.A., Zhu, J. & Xiao, T., 2008. Structural basis for distinctive recognition of fibrinogen gammaC peptide by the platelet integrin alphaIIbbeta3. *The Journal of cell biology*, 182(4), pp.791–800.

St-Jacques, B., Hammerschmidt, M. & McMahon, A.P., 1999. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes & development*, 13(16), pp.2072–2086.

Stanton, B.Z. et al., 2009. A small molecule that binds Hedgehog and blocks its signaling in human cells. *Nature chemical biology*, 5(3), pp.154–156.

Strutt, H. et al., 2001. Mutations in the sterol-sensing domain of Patched suggest a role for vesicular trafficking in Smoothened regulation. *Current biology : CB*, 11(8), pp.608–613.

Stuart, D.I. et al., 1979. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *Journal of molecular biology*, 134(1), pp.109–142.

Suhre, K. & Sanejouand, Y.-H., 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic acids research*, 32(Web Server issue), pp.W610–4.

Svensson, G. et al., 2012. Crystal Structure of N-Glycosylated Human Glypican-1 Core Protein: STRUCTURE OF TWO LOOPS EVOLUTIONARILY CONSERVED IN VERTEBRATE GLYPICAN-1. *The Journal of biological chemistry*, 287(17), pp.14040–14051.

SZYMANSKI, S. et al., 2008. Automated Application of a Novel High Yield, High Performance Tissue Culture Flask. *Journal of the Association for Laboratory Automation*, 13(3), pp.136–144.

Tada, M. et al., 2008. Down-regulation of hedgehog-interacting protein through genetic and epigenetic alterations in human hepatocellular carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(12), pp.3768–3776.

Tadros, W. & Lipshitz, H.D., 2009. The maternal-to-zygotic transition: a play in two acts. *Development (Cambridge, England)*, 136(18), pp.3033–3042.

Taipale, J. et al., 2000. Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature*, 406(6799), pp.1005–1009.

Taipale, J. et al., 2002. Patched acts catalytically to suppress the activity of Smoothened. *Nature*, 418(6900), pp.892–897.

Taniguchi, H. et al., 2007. Intrahepatic mRNA levels of type I interferon receptor and interferon-stimulated genes in genotype 1b chronic hepatitis C. Association between IFNAR1 mRNA level and sustained response to interferon therapy. *Intervirology*, 50(1), pp.32–39.

Teglund, S. & Toftgård, R., 2010. Hedgehog beyond medulloblastoma and basal cell carcinoma. *Biochimica et biophysica acta*, 1805(2), pp.181–208.

Tenzen, T. et al., 2006. The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Developmental cell*, 10(5), pp.647–656.

Thayer, S.P. et al., 2003. Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature*, 425(6960), pp.851–856.

Thérond, P.P., 2012. Release and transportation of Hedgehog molecules. *Current opinion in cell biology*, 24(2), pp.173–180.

Tojo, M. et al., 2002. Expression of a sonic hedgehog signal transducer, hedgehog-interacting protein, by human basal cell carcinoma. *Br J Dermatol*, 146(1), pp.69–73.

Tokunaga, F., Hara, K. & Koide, T., 2003. N-linked oligosaccharide processing, but not association with calnexin/calreticulin is highly correlated with endoplasmic reticulum-associated degradation of antithrombin Glu313-deleted mutant. *Archives of biochemistry and biophysics*, 411(2), pp.235–242.

Torroja, C., Gorfinkiel, N. & Guerrero, I., 2004. Patched controls the Hedgehog gradient by endocytosis in a dynamin-dependent manner, but this internalization does not play a major role in signal transduction. *Development (Cambridge, England)*.

Treier, M. et al., 2001. Hedgehog signaling is required for pituitary gland development. *Development (Cambridge, England)*, 128(3), pp.377–386.

Tukachinsky, H., Lopez, L.V. & Salic, A., 2010. A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu-Gli protein complexes. *The Journal of cell biology*, 191(2), pp.415–428.

Turing, A.M., 1952. The Chemical Basis of Morphogenesis.

Varjosalo, M. & Taipale, J., 2008. Hedgehog: functions and mechanisms. *Genes & development*, 22(18), pp.2454–2472.

Veugelers, M. et al., 2000. Mutational analysis of the GPC3/GPC4 glypican gene cluster on Xq26 in patients with Simpson-Golabi-Behmel syndrome: identification of loss-of-function mutations in the GPC3 gene. *Human molecular genetics*, 9(9), pp.1321–1328.

Vierkotten, J. et al., 2007. Ftm is a novel basal body protein of cilia involved in Shh signalling. *Development (Cambridge, England)*, 134(14), pp.2569–2577.

Walter, T.S. et al., 2005. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta crystallographica. Section D, Biological crystallography*, 61(Pt 6), pp.651–657.

Walter, T.S. et al., 2003. A procedure for setting up high-throughput nanolitre crystallization experiments. I. Protocol design and validation. *Journal of applied crystallography*, 36(2), pp.308–314.

Wang, X. et al., 2006. Structure of artemin complexed with its receptor GFRalpha3: convergent recognition of glial cell line-derived neurotrophic factors. *Structure (London, England : 1993)*, 14(6), pp.1083–1092.

Wang, Y., McMahon, A.P. & Allen, B.L., 2007. Shifting paradigms in Hedgehog signaling. *Current opinion in cell biology*, 19(2), pp.159–165.

Weatherbee, S.D., Niswander, L.A. & Anderson, K.V., 2009. A mouse model for

Meckel syndrome reveals Mks1 is required for ciliogenesis and Hedgehog signaling. *Human molecular genetics*, 18(23), pp.4565–4575.

White, G.R.M. et al., 2002. Somatic glypican 3 (GPC3) mutations in Wilms' tumour. *British journal of cancer*, 86(12), pp.1920–1922.

Williams, E.H. et al., 2010. Dally-like core protein and its mammalian homologues mediate stimulatory and inhibitory effects on Hedgehog signal response. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13), pp.5869–5874.

Willnow, T.E. et al., 1996. Defective forebrain development in mice lacking gp330/megalin. *Proceedings of the National Academy of Sciences of the United States of America*, 93(16), pp.8460–8464.

Winter, G., 2009. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography*, 43(1), pp.186–190.

Wolpert, L., 1969. Positional information and the spatial pattern of cellular differentiation. *Journal of theoretical biology*, 25(1), pp.1–47.

Wolpert, L., 1968. Wolpert: Towards a Theoretical Biology, I, Prolegomena - Google Scholar.

Xiang, Y.Y., Ladeda, V. & Filmus, J., 2001. Glypican-3 expression is silenced in human breast cancer. *Oncogene*, 20(50), pp.7408–7412.

Yan, D. & Lin, X., 2009. Shaping morphogen gradients by proteoglycans. *Cold Spring Harbor perspectives in biology*, 1(3), p.a002493.

Yang, Z.R. et al., 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics (Oxford, England)*, 21(16), pp.3369–3376.

Yao, S., Lum, L. & Beachy, P., 2006. The ihog cell-surface proteins bind Hedgehog and mediate pathway activation. *Cell*, 125(2), pp.343–357.

Yuan, Z. et al., 2007. Frequent requirement of hedgehog signaling in non-small cell lung carcinoma. *Oncogene*, 26(7), pp.1046–1055.

Zarco, N. et al., 2012. GAS1 induces cell death through an intrinsic apoptotic pathway. *Apoptosis : an international journal on programmed cell death*, 17(6), pp.627–635.

Zhang, F. et al., 2007. Kinetic and structural studies on interactions between heparin or heparan sulfate and proteins of the hedgehog signaling pathway. *Biochemistry*, 46(13), pp.3933–3941.

Zhang, W. et al., 2006. Cdo functions at multiple points in the Sonic Hedgehog pathway, and Cdo-deficient mice accurately model human holoprosencephaly. *Developmental cell*, 10(5), pp.657–665.

Zhao, Y. et al., 2011. Automation of large scale transient protein expression in mammalian cells. *Journal of structural biology*, 175(2), pp.209–215.

Zheng, X. et al., 2010. Genetic and biochemical definition of the Hedgehog receptor. *Genes & development*, 24(1), pp.57–71.

Zhu, A.J. et al., 2003. Altered localization of Drosophila Smoothened protein activates Hedgehog signal transduction. *Genes & development*, 17(10), pp.1240–1252.