Targeting the Hedgehog Pathway in Disease: a Structural Biology Approach

Lea Sefer

A thesis submitted in partial fulfillment of the degree
Doctor of Philosophy



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

Trinity 2017

'Chance is the only source of true novelty.'
Francis Crick

Acknowledgements

I would first like to thank my supervisors Professor Christian Siebold and Professor Philip Biggin for their supervision and constant encouragements throughout my DPhil studies. It has been a great privilege to be part of their groups, learn from them and their group members. I am thankful to the Wellcome Trust for their generous support that has enabled me to explore multiple research project and learn various aspects of structural biology.

Both Strubi and SBCB have been exceptionally welcoming environments with extremely knowledgeable students and post-docs. I would like to thank all Siebees (and ex-Siebees) for their help in and out of the lab. I would especially like to thank Ben and Christian for kindly reading and correcting my thesis. I would also like to thank Ben for his help with tissue culture and SPR, Ellie for first introducing me to the lab and teaching me all the basics, and Rebekka for introducing me to cell assays. I would also like to thank Eamon, Christiane, Joy, Amalie and Sam for all the prosecco Fridays and everyday support. Thanks to Laura, Theresa and Maria from SBCB for always answering my computing-related questions. Special thanks to Alice Douangamath for all her help and efforts with the fragment screening.

I would like to thank my dearest friends Sonja, Claudia, Sarah and Sabine for all the talks, dinners, travels, races and proseccos, and for always being there for me. Thanks to Bani for his support and motivating me to always do better. Finally, I would like to thank my parents Kornelija and Robert, and my grandparents, Katarina, Aleksandar and late Klara, for helping me get this far. My journey would not have been possible without their support and encouragements.

Technical acknowledgements

I would like to acknowledge the following people for their specific contributions to the work reported in this thesis:

- Christiane Kowatsch and Dr Amalie Rudolf from STRUBI, for crystallising the ShhN-SOS complex
- Dr Alice Douangamath and Dr Patrick Collins from XChem DLS, for their assistance with the fragment screening and data analysis
- Prof Angela Russell and Dr Graham Wynne from the Department of Chemistry, University of Oxford, for their assistance with the lead generation
- Dr Anthony Bradly and Dr Anthony Aimon from OxXChem, University of Oxford, for assistance with chemistry and lead generation
- Dr Theresa Paramo from SBCB, for writing the consensus docking script
- Dr Tom Lanyon-Hogg from Imperial College London, for carrying out the palmitoylation assays for Skinny Hh constructs

Abbreviations

ADMET Absorption, distribution, metabolism and secretion

AR library from Prof. Angela Russell's group

BCC Basal Cell Carcinoma

BDA1 Brachydactyly type a1

Boc Brother of CDO

BMP Bone morphogenetic protein

Boi Brother of IHOG

C6S Chondroitin 6 sulphate

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

CHS Cholesterol hemisuccinate

CNS Central nervous system

CoA Coenzyme A

CRD Cysteine-rich domain

CSPG Chondroitin sulphate proteoglycan

CV Column volume

CW Cardin-Weintraub heparin binding motif

DDM n-Dodecyl-β-D-Maltopyranoside

Dhh Desert Hedgehog

Disp Dispatched

DLS Diamond Light Source

DMSO Dimethyl Sulfoxide

DMNG Decyl Maltose Neopentyl Glycol dNTP deoxyribonucleotide trisphosphate

DSPL Diamond-SGC poised fragment library

EB Elution buffer

EG Ethylene glycol

EGF Epidermal Growth Factor

ER Endoplasmic reticulum

FBLD Fragment-based lead discovery

FBS Foetal Bovine Serum

FDA US Food and drug administration

FL Full-length

FRG Fragment

FSEC Fluorescence size-exclusion chromatography

GAGs Glycosaminoglycans

Gas1 Growth Arrest Specific 1

Gli Glioma-associated oncogene

GOAT ghrelin-O-acyltransferase

GPC Glypican

GPCR G-protein coupled receptor

GPI Glycosylphosphatidylinositol

HBA Hydrogen bond acceptor

HBD Hydrogen bond donor

HDAC Histone deacetylase

HEK Human embryonic kidney cells

Hh Hedgehog

Hhat Hedgehog acyltransferase, Skinny hedgehog

HhC Hedgehog C-terminal domain

Hhip Hedgehog Interactig Protein

HhN Hedgehog N-terminal domain

HPE Holoprosencephaly

HSPG Heparan Sulphate Proteoglycan

HTS High-throughput screening

Ig Immunoglobulin

Ihh Indian Hedgehog

Ihog Interference hedgehog

IPTG Isopropyl-D-1-thiogalactopyranoside

ITC Isothermal titration calorimetry

KCOT Keratocystic odontogenic tumour

Kif7 Kinesin-4 protein

LCP Lipidic-cubic phase

β-ME 2-Mercaptoethanol

MES 2-(N-morpholino)ethanesulfonic acid

MBOAT Membrane-bound O-acyltransferase

MB Medulloblastoma

MD Molecular dynamics

MSA Mobility shift assay

mV mono-Venus tag

NaBu Sodium Butyrate

NCS Newborn calf serum

NG Neopentyl Glycol NMR Nuclear magnetic resonance

NPS Sodium nitrate, Sodium phosphate dibasic dihydrate, Ammonium sulfate

β-OG n-Octyl–D-Glucopyranoside

OGNG Octyl Glucose Neopentyl Glycol

QPCR Quantitative polymerase chain reaction

PanDDa Pan-dataset analysis software

PCR Polymerase Chain Reaction

PCT Pre-crystallisation test

PDB Protein data bank

PEG Polyethylene glycol PEI Polyethylenimine

Ptch Patched

RMSD root mean square deviation

RND Resistance-nodulation-division transporter family

RUSKI small-molecule inhibitor of Skinny Hh

RT Room temperature

RT-PCR Reverse transriptase polymerase chain reaction

SAG Smoothened agonist

SAR Structure-activity relationship

SEC Size-exclusion chromatography

Shh Sonic Hedgehog

ShhN N-terminal domain of Sonic Hedgehog

Ski Skinny Hedgehog, Hhat

Smo Smoothened

SOS Sucrose octasulphate

SPR Surface plasmon resonance

SSD Sterol sensing domain

Sufu Suppressor of Fused

TLH small-molecule inhibitor of Skinny Hh

TM Transmembrane

VPA valproic acid

WT Wild-type

ZPA zone of polarising activity

1D4 Rho1D4 peptide tag

Targeting the Hedgehog Pathway in Disease: a Structural Biology Approach

Lea Sefer

A thesis submitted in partial fulfillment of the degree *Doctor of Philosophy*

Wolfson College, Nuffield Department of Medicine, Division of Structural Biology, University of Oxford, Trinity 2017

Abstract

The Hedgehog (Hh) pathway is a morphogen pathway essential for proper embryonic development of all bilaterians that remains active in adults where it contributes to tissue homoeostasis. Malfunction of the pathway results in developmental defect, while constitutive activation was correlated with multiple cancers. The Hh pathway is a validated cancer drug target with two FDA approved inhibitors targeting the Hh GPCR signal transducer Smoothened. This thesis focuses on two alternative Hh targets, the Sonic Hh (ShhN) morphogen and its acyltransferase Skinny Hh. In Chapter 2, the aim was to structurally characterise a complex between ShhN and its only small-molecule inhibitor Robotnikinin using *in silico* docking and crystallography. The results revealed that the most likely binding site of Robotnikinin is the pseudo-active site of ShhN and that the interactions involve the His134-Ser136 loop. In Chapter 3, the aim was to identify novel ShhN inhibitors using crystallographic fragment-based screening of 1,200 unique fragments. The screening resulted in 13 potential hits, 5 of which were found within the ion binding cleft. In Chapter 4, the aim was to validate the obtained hits and conduct preliminary structure activity relationship (SAR) assays. Based on the results, FRG 155 was identified as the preferred hit that bound to the the ShhN pseudo-active site with low micromolar affinity. Preliminary SAR identified positive and negative features of the fragment. In Chapter 5, the aim was to structurally and functionally characterise Skinny Hh. Protocols for successful expression and purification of milligram quantities of thermostable and active Skinny Hh were developed, providing a platform for crystallisation and drug screening. The first luminal loop of Skinny Hh was found to be essential for the activity of Skinny Hh. In summary, alternative Hh pathway targets, ShhN and Skinny Hh, were characterised using structural biology approaches and will contribute to the development of novel inhibitors with the application in Hh ligand-dependent tumours.

Contents

1 Introduction **16**

1.1 Hedgehog signalling 17

1.1.1 Discovery of Hedgehog proteins 17

1.1.2 Role of the Hedgehog pathway in development 18

1.2 Molecular Mechanism of the Hh pathway in vertebrates 20

1.2.1 Processing of Hh ligands 21

1.2.2 HhN secretion and trafficking 23

1.2.3 The role of lipid modifications in Hh signalling 25

1.2.4 Hh receptors and signal transduction in vertebrates 26

1.2.5 Intracellular Hh signalling 30

1.3 Structural overview of the Hedgehog pathway 31

1.3.1 Structure of Hh ligands 31

1.3.2 Interaction of Hh ligands with receptors 34

1.3.2.1 Cdo 35

1.3.2.2 Hhip 37

1.3.2.3 Patched and E51 inhibitory antibody 40

1.3.2.4 Glycosaminoglycans 43
1.4 The Hedgehog pathway in disease 46
1.4.1 Developmental defects 46
1.4.2 Tissue homeostasis and regeneration 48
1.4.3 Cancer 50
1.4.3.1 Ligand-independent mechanism 50
1.4.3.2 Ligand-dependent mechanism 52
1.5 Thesis overview and aims 55

2 Structural characterisation of the Shh-Robotnikinin complex 57

2.1 Abstract 58
2.2 Introduction 59
2.3 Results 65
2.3.1 Molecular dynamics studies and docking of Shh 65
2.3.1.1 Global docking 65
2.3.1.2 Molecular dynamics simulations of ShhN 70
2.3.1.3 Docking to ShhN structures from molecular dynamics simulations 73
2.3.2 Crystallisation of HhN-Robotnikinin complex 76
2.3.2.1 Expression and purification of ShhN and DhhN . 76
2.3.2.2 Crystallisation trials 78
2.3.2.3 ShhN-Robotnikinin structure 80
2.4 Discussion 85

3 Targeting Sonic Hedgehog: fragment-based hit identification 88

3.1 Abstract 89

3.2 Introduction 90

3.3 Results 96

3.3.1 ShhN-SOS complex structure 96

3.3.2 Crystallographic fragment screening of commercial libraries against ShhN 100

3.3.3 Binding site 1 105

3.3.4 Binding site 2 110

3.3.5 Binding site 3 113

3.3.6 Binding sites 4-8 118

3.3.7 Crystallographic tables 121

3.4 Discussion 125

4 Targeting Sonic Hedgehog: Hit Validation and Lead Generation 130

4.1 Abstract 131

4.2 Introduction 132

4.3 Results 138

4.3.1 Hit validation 138

4.3.2 Further hit validation of fragment FRG 155 142

4.3.3 FRG 155: structure-activity relationship 143

4.3.4 *In silico* lead generation 151

4.4 Discussion 156

5 Structural and functional characterisation of Skinny Hedgehog 160

5.1 Abstract 161

5.2 Introduction 162

5.2.1 N-palmitoylation reaction 164

5.2.2 Skinny Hh topology 167

5.2.3 Targeting Skinny Hh 168

5.3 Results 171

5.3.1 Construct expression and detergent screen 171

5.3.2 Thermostability and additive screen 174

5.3.3 Large-scale purification and initial crystallisation trials . . 179

5.3.4 Purification detergent optimisation and activity 185

5.3.5 Construct engineering 189

5.3.6 Construct activity 192

5.3.7 Drug stability screen 194

5.4 Discussion 196

6 Conclusions 199

6.1 Molecular basis for Hh signalling inhibition by Robotnikinin . . . 201

6.2 Fragment-based screening for novel ShhN inhibitors 203

6.3 Skinny Hh structure and function 204

7 Materials and Methods 208

7.1 Methods for Chapters 2-4 209

7.1.1 Molecular Biology 209

7.1.1.1 Construct design and expression vectors 209
7.1.1.2 Transformation 209
7.1.2 Protein Production 209
7.1.2.1 Protein expression 209
7.1.2.2 Protein purification 210
7.1.2.3 Protein crystallisation 210
7.1.3 Crystallographic fragment-screen 211
7.1.3.1 Fragment Libraries 211
7.1.3.2 Crystal soaking 212
7.1.3.3 Crystallographic data collection, processing and refinement 213
7.1.3.4 PanDDAs hit identification 213
7.1.4 Hit validation assays 215
7.1.4.1 Commercially available compounds 215
7.1.4.2 Surface Plasmon Resonance (SPR) 215
7.1.4.3 Isothermal titration calorimetry (ITC) 218
7.1.4.4 Gli-reporter assay 219
7.1.5 *In silico* docking and molecular dynamics 220
7.1.5.1 Molecular structures and modelling of missing residues 220
7.1.5.2 Molecular dynamics simulations 220
7.1.5.3 Molecular docking 221
7.1.5.4 Lead generation and consensus docking 221
7.1.5.5 Molecular analysis 223

7.2 Methods for Chapter 5 223
7.2.1 Molecular Biology 223
7.2.1.1 Construct design 223
7.2.1.2 Restriction digests, ligations and transformations 223
7.2.2 Protein production and analysis 224
7.2.2.1 Small-scale expression 224
7.2.2.2 Small-scale purification 225
7.2.2.3 DNA purification 225
7.2.2.4 Large-scale expression 226
7.2.2.5 Large-scale purification 226
7.2.2.6 Western Blot 228
7.2.2.7 Thermostability and additive screens 228
7.2.3 Crystallisation 229
7.2.3.1 Lipidic cubic phase 229
7.2.3.2 Vapour diffusion 230

Appendices **231**

1.1 Skinny Hh sequence alignment 232
1.2 List of fragments and compounds 233
1.3 Membrane protein additive screen 234

Bibliography **235**

Chapter 1

Introduction

1.1 Hedgehog signalling

The Hedgehog (Hh) pathway is a fundamental morphogen pathway active during embryonic development. Originally discovered in flies, the Hh pathway is conserved across the metazoa and is essential for tissue patterning, organogenesis and proper symmetry formation of all bilaterians. Hh morphogens are secreted proteins that signal alongside other morphogens like Wnts and bone morphogenetic proteins (BMPs) in a gradient-dependent manner. The Hh pathway remains active in adults where it contributes to cell regeneration and tissues homeostasis. Malfunction of the pathway leads to developmental defects, most commonly holoprosencephaly and brachydactyly, and cancers, such as basal cell carcinoma and medulloblastoma. The Hh pathway is an important drug target in cancer research resulting in extensive efforts towards novel inhibitor discovery.

1.1.1 Discovery of Hedgehog proteins

The *hedgehog* (*hh*) gene was first identified in a mutation screen of *Drosophila* genes that affected the patterning and segment polarity of the larvae [1, 2]. The *hh* knockout phenotype manifested in duplication of denticles and deletion of naked cuticle in each segment [1]. As a result, *hh* mutant larvae were shorter than the wild-type (WT) and narrowly spaced denticles formed a dense lawn giving them a stubby 'hedgehog-like' appearance, hence the name hedgehog. The *hh* knockout phenotype is only apparent in homozygous embryos, which are zygotic lethal [1]. The role of *hh* in patterning and its dependence on other segment polarity genes,

wingless and *engrailed*, was further confirmed after the *hh* gene was cloned and its expression pattern examined in greater detail [3, 4, 5, 6, 7].

Hh genes were subsequently identified in vertebrates including mouse, zebra, fish and chick [8, 9, 10]. Unlike invertebrates that only had one gene, vertebrates possess multiple *hh* genes. In mice, three genes encoding homologous proteins, Sonic (Shh), Desert (Dhh) and Indian (Ihh) hedgehog, were discovered [9, 8, 11, 12].

1.1.2 Role of the Hedgehog pathway in development

Morphogens provide positional information to the cells within the embryo, specifying their place and role within the tissue and the body [13]. Embryonic cells require this positional information to undergo a particular fate, begin the differentiation into specific cell lines and develop into different tissues and organs. Morphogens are proteins that are secreted at a concentration gradient from the source and exhibit different signalling effects depending on the concentration threshold along the gradient [13] resulting in organ and tissue patterning.

In *Drosophila*, Hh was initially found to regulate the patterning of the posterior compartment of the wings and imaginal discs of legs and eyes [4, 14, 15]. Multiple roles of Hh were subsequently identified, including the stimulation of the development of gonads [16, 17], abdomen [18], gut [19], optic lamina [20, 21], tracheal system [22] and the migration of germ cells [23]. Hh is an essential protein for proper embryonic development of *Drosophila* larvae.

In vertebrates, Shh is the most widely expressed of the three Hh ligands, while Dhh and Ihh have more specialised roles. Dhh is expressed in gonads and peripheral nerves where it stimulates the development of the male germ line [24] and peripheral nerve sheath [25], respectively. Ihh primarily directs the development of bones and cartilage [26, 27] and plays a role in haematopoiesis and vasculogenesis [28].

Shh was first discovered in the notochord and the floor plate of the neural tube where it affected ventral-dorsal polarity, and in the zone of polarising activity (ZPA, limb bud) where it affected anterior-posterior polarity of the limb (Figure 1.1) [9, 10]. Ventral expression of Shh in the notochord resulted in the expression of floor plate genes on the ventral side of the neural tube (Figure 1.1, a) [9, 8, 12]. This indicated that Shh has a major role in the patterning of the

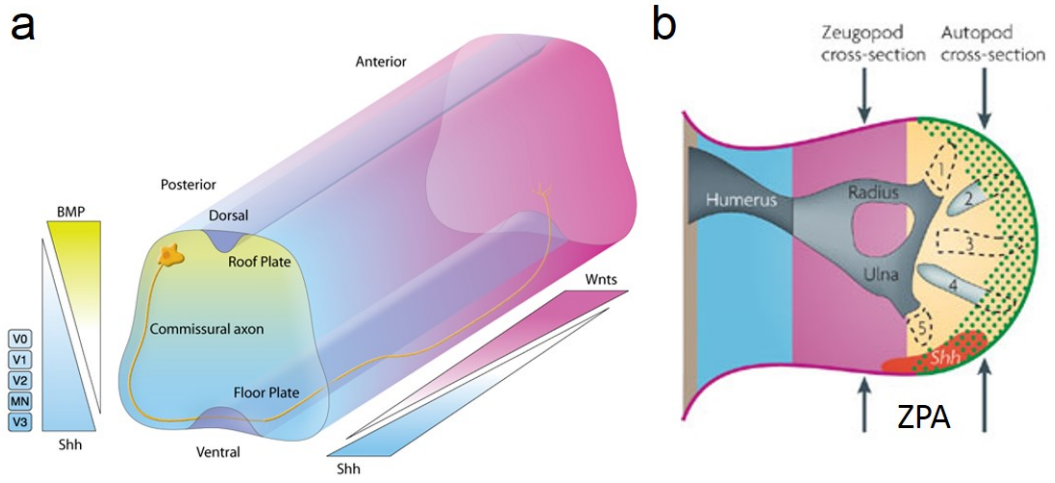


Figure 1.1: **The role of the Hh pathway in development.** a) Shh is important for determining ventral-dorsal polarity of the neural tube. It is secreted from the floor plate and signals together with other morphogens (Wnt and BMPs). Adapted from [29]. b) Shh is also involved in the development of limbs where it is secreted from the ZPA to determine anterior-posterior polarity and contributes to patterning. Adapted from [30].

central nervous system (CNS). The ZPA consists of a group of mesenchymal cells on the posterior part of the budding limb. When ZPA cells are grafted on the anterior side of the limb bud, the limb pattern duplicates resulting in the formation of double digits [10]. A similar phenotype is observed when Shh is expressed on the anterior side of the limb bud indicating that on the molecular level patterning of the limb is controlled by Shh in the ZPA (Figure 1.1, b) [10, 11]. This finding was the first evidence of a common molecular mechanism underlying patterning of the two distant regions in the body. Since then, Shh has been found in most tissues in the body where it affects cell fate, survival and proliferation, depending on the concentration gradient, and acts in both short- or long-range signalling [31]. Shh directly influences the lateral asymmetry of the body [32, 33] and is expressed in most tissues including the cerebellum [34], muscle [35], lungs [36], gut [37], heart [33], blood [38], bones [39] and many others [31].

1.2 Molecular Mechanism of the Hh pathway in vertebrates

The Hh signalling pathway is conserved from flies to humans. However, some aspects of the molecular mechanism have diverged and differ significantly. For the purpose of clarity, only vertebrate Hh signalling is described in this section. The three vertebrate ligands Shh, Dhh and Ihh are referred to as Hh since what is described applies to all ligands. The Hh signalling pathway can be divided into

3 phases: i) Hh processing and secretion, ii) interaction with receptors and signal transduction at the plasma membrane and iii) intracellular signalling.

1.2.1 Processing of Hh ligands

Hh is synthesised as a $\sim$ 45 kDa precursor protein and is processed into a mature and fully functional N-terminal signalling domain (HhN) in the Endoplasmic reticulum (ER). The precursor protein consists of a signal sequence, the signalling HhN domain and the autosplicing C-terminal domain (HhC). The Hh precursor protein undergoes two major reactions: i) autocleavage and attachment of cholesterol to the C-terminus of HhN, and ii) palmitoylation of the N-terminus of HhN (Figure 1.2).

Once transported into the ER, the signal sequence is cleaved off (residues 1-23 in human, 1-24 in murine Shh) and the precursor protein is autocleaved to liberate the $\sim$ 20 kDa signalling domain HhN. The autocleavage is accompanied by the addition of cholesterol to the C-terminus of HhN. Both reactions, autocleavage and the addition of cholesterol, are catalysed by intein homologue HhC, which is subsequently degraded via the ERAD degradation pathway [40, 41, 42]. The reaction proceeds in two steps: i) the thiol side chain of the conserved Cys198 residue located at the cleavage site acts as a nucleophile to attack the carbonyl group of the preceding amino acid, forming a thioester bond [41]; ii) the 3β-hydroxyl group of cholesterol acts as a nucleophile to attack the thioester bond forming a new thioester bond between HhN and the cholesterol, releasing HhC.

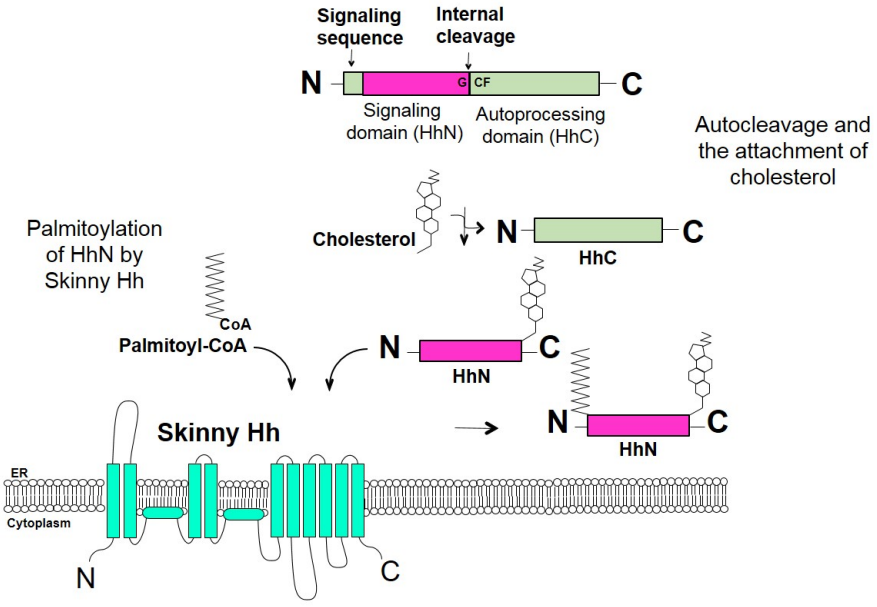


Figure 1.2: **Processing of Hh precursor protein.** Hh precursor protein is targeted to the ER where the signalling sequence is cleaved off. The protein is then autocleaved to liberate the signalling N-terminal Hh domain (HhN). In the same reaction catalysed by the C-terminal Hh domain (HhC), a cholesterol is attached to the C-terminus of HhN. Skinny Hh palmitoylates HhN making it fully functional and ready for secretion.

Similarly to other inteins, HhC contains conserved active site residues Cys198, His270, Thr267 and Asp243 [42]. Cys198, His270 and Thr267 are directly involved in initiating the cleavage, while Asp243 acts to deprotonate and to activate the 3β-hydroxyl group of cholesterol [43]. In addition to the intein homologue domain, HhC contains a hydrophobic C-terminal domain that most likely acts as the cholesterol binding domain [43].

The first amino acid of the signalling domain HhN is a conserved Cys residue (Cys24 in human, Cys25 in murine Shh), which is essential for palmitoylation by Skinny Hh acyltransferase (Skinny Hh or Hhat) [44]. This unique irreversible modification involves the attachment of palmitate to the N-terminus of HhN via an amide bond [45]. Autosplicing and palmitoylation are two independent

processes and the Hh precursor protein can undergo palmitoylation without being processed into HhN [46]. The palmitoylation mechanism and nature of Skinny Hh is discussed in greater detail in Chapter 5.

1.2.2 HhN secretion and trafficking

The dual-lipidated HhN is targeted to the plasma membrane where it associates with sterol-rich lipid rafts before it is secreted [47, 48, 49]. HhN was found in the extracellular environment in a monomeric state and as part of a larger multimeric complex (Figure 1.3) [50].

Dispatched (Disp), a member of the resistance-nodulation-division (RND) transporter family of lipid transporters [51, 52, 53], and Scube2, a secreted glycoprotein, are membrane associated proteins that aid secretion of soluble HhN monomers [54, 55]. Disp and Scube2 bind separate regions of the HhN cholesterol moiety via their sterol sensing domains (SSD) [51, 54]. While it is not clear how Disp and Scube2 recruit and release HhN, Scube2 remains associated with HhN after it is secreted in order to maintain HhN solubility (Figure 1.3, 1) [56, 55, 57]. An alternative mechanism whereby lipid modifications are shed off to release HhN has been proposed [58]. However, this is unlikely to happen as lipid modifications are essential for trafficking and receptor interactions [59, 49, 60]. HhN can also self-associate into multimers via hydrophobic interactions of cholesterol and palmitate (Figure 1.3, 2) [49, 61]. HhN lacking lipid modifications was found unable to form multimers and signal long-range, indicating that the lipid

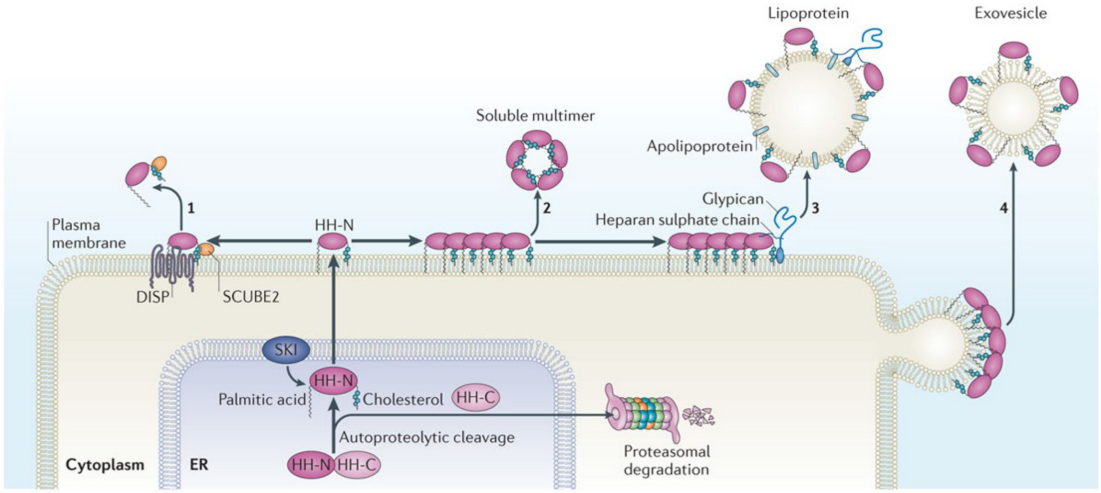


Figure 1.3: **Secretion of HhN in monomeric and multimeric forms.** The Hh protein is processed in the ER and the signalling domain HhN is secreted. 1) HhN associates with membrane proteins Disp and Scube2, which bind to the cholesterol moiety and aid ShhN secretion. Scube2 remains associated with HhN monomer and contributes to solubility. 2) HhN can assemble into soluble multimers via hydrophobic assembly of lipid modifications. However, the precise mechanism of how this occurs is not known. 3) Glypican, a GPI-anchored HSPG, promotes association between HhN and lipoprotein thereby facilitating the assembly of lipoprotein particles. 4) HhN can also be secreted as part of exovesicles and might be transported alongside other proteins like intergrins. Adapted from [64].

modifications are important for multimer formation and long-range signalling. [49, 62, 63].

HhN can also be secreted as part of lipoprotein particles and extracellular vesicles. Heparan sulphate proteoglycans (HSPGs) are proteins decorated with linear glycosaminoglycan (GAG) chains like heparan sulphate or chondroitin sulphate (CSPG). Lipoprotein particle formation is driven by GPI-anchored HSPGs, the Glypicans (GPCs) [65, 66, 67, 68]. GPCs associate with a scaffolding lipoprotein (apolipoprotein) via their GAG chains and promote the interactions between the lipoprotein and HhN (Figure 1.3, 3) [69]. It is still not understood how lipoprotein particles are released. Finally, there is some evidence that suggests that

HhN is secreted as part of the extracellular vesicles alongside other proteins like integrins (Figure 1.3, 4) [70, 71].

1.2.3 The role of lipid modifications in Hh signalling

Both palmitate and cholesterol have major roles in secretion and are essential for proper Hh signalling [72, 73, 68, 61]. Cholesterol has a dual role in promoting both short- and long-range Hh signalling. On one hand, cholesterol promotes HhN tethering to membranes after secretion thereby restricting diffusion and enabling short-range signalling [74, 48, 45]. However, cholesterol also plays a crucial role in the formation of soluble multimers or lipoprotein particles that are essential for long-range signalling [59, 75, 76]. Unprocessed Hh retains some activity *in vitro* and *in vivo*, but is mostly restricted to short distance cell-to-cell signalling [77]. Furthermore, mutations that prevent autoprocessing and cholesterol attachment cause Hh specific developmental defects like holoprosencephaly [78, 79], indicating that both processes are essential for proper signalling.

Palmitoylation was also found to be essential for HhN secretion and proper Hh signalling. The absence of palmitoylated HhN in flies abolished Hh signalling [80, 81]. In mice, the ShhN C25S mutant defective in palmitoylation was found to significantly reduce Hh signalling and to restrict it to short-range signalling and Skinny Hh knockout mice were found to have significantly reduced Hh signalling in the limb and neural tube [80, 72]. Palmitoylation was not necessary for membrane targeting of ShhN. However, it was essential for the formation of

soluble multimeric complexes, long-range signalling and gradient formation. Furthermore, *in vitro* studies showed that palmitoylated Shh is 40× more active than non-palmitoylated Shh [45], indicating the importance of lipid modifications on Hh signalling. Mutations in Skinny Hh or HhN mutations inhibiting palmitoylation result in developmental disease, supporting the importance of palmitoylation for proper Hh signalling [82, 79].

1.2.4 Hh receptors and signal transduction in vertebrates

In vertebrates, Hh signalling is tightly linked to the primary cilium, a microtubule-based projection on the surface of most vertebrate cells [83, 84]. Primary cilia are non-motile structures comprised of 9 outer microtubule doublets contained within the basal body.

The first HhN receptor discovered was Patched (Ptch) [85, 86, 87, 88]. In the absence of HhN, Ptch localises at the base of the primary cilium and constitutively represses the downstream signalling ability of the Hh signal transducer and G protein-coupled receptor (GPCR) Smoothened (Smo). Smo located in the intracellular vesicles in its deactivated state, thereby keeping the pathway off (Figure 1.4, a) [83, 89]. When HhN binds to Ptch, this inhibition is released resulting in endocytosis and degradation of HhN-Ptch complex and accumulation of Smo at the tip of the primary cilium (Figure 1.4, b) [90, 91]. The endocytosis of the HhN-Ptch complex also functions as a negative-feedback loop whereby Ptch sequesters HhN to stop it from further signalling [92]. All Hh signalling

is initiated via Ptch and Smo and both are essential for Hh signalling to occur. However, how Ptch and Smo communicate remains a mystery.

Studies have shown that Ptch inhibits Smo at 1:50 ratio indicating that the inhibition of Smo is indirect and might be controlled by a small-molecule [89, 94].

Smo belongs to the Frizzled-class GPCRs that consist of a cysteine-rich extracellular domain (CRD) and a 7-transmembrane (TM) domain, both of which contain distinct small-molecule binding sites [95]. Smo binds a number of small molecules that have both agonistic and antagonistic effect on Hh signalling. Inhibitors like SANT-1, Vismodegib and naturally occurring cyclopamine bind to the 7-TM domain [95, 96, 97], while oxysterol modulators, such as cholesterol and 20(S)-

hydroxycholesterol (20(S)-OHC) bind to the CRD [95, 98, 99, 100]. The small molecule-mediated allosteric interaction between the CRD and TM is important for Hh signalling regulation [101]. In particular, cholesterol was found to activate and stabilise Smo in the extended conformation when bound at the interface between the CRD and TM, and its displacement or binding of an inhibitor result in a conformational between the two domains [95, 102]. Furthermore, Smo was found to be inhibited by pro-vitamin D3 [103], leading to the hypothesis that pro-vitamin D3 and cholesterol are regulated by Ptch thereby modulating Hh pathway activation [104, 105].

Ptch is a 12-transmembrane integral membrane protein structurally similar to the RND family of bacterial permeases and also contains a sterol sensing domain (SSD) that was shown to bind cholesterol [106, 86]. Mutations in the RND or SSD

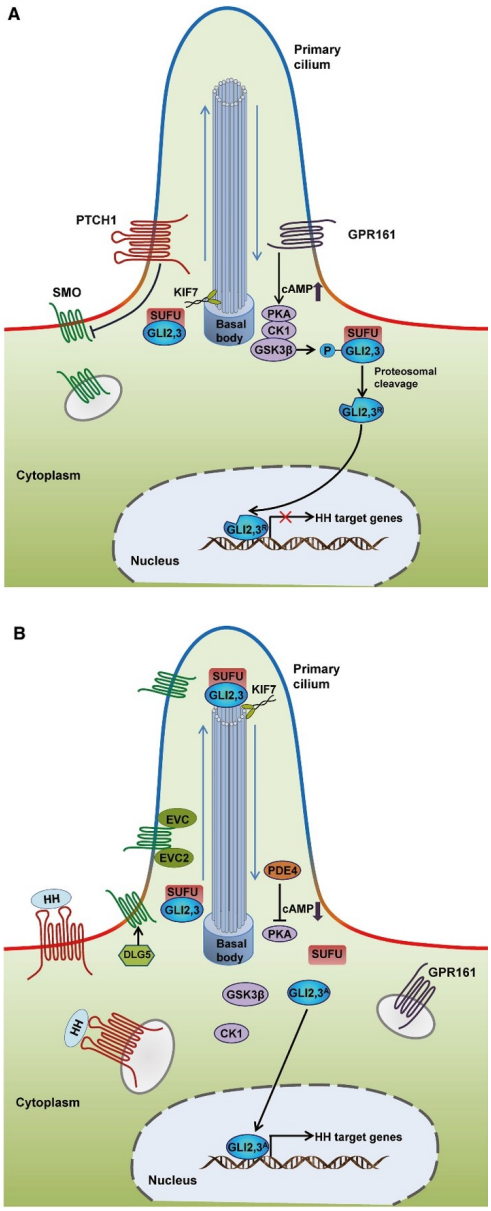


Figure 1.4: **Hh signalling mechanism and signal transduction.** a) In the absence of HhN, receptor Ptch localises at the base of the cilium, while Smo is contained within intracellular vesicles. Sufu forms a complex with Gli2/3-FL thereby preventing their translocation into the nucleus. High cAMP levels stimulate Gli2/3 phosphorylation by PKA, CK1 and GSK3β, which are then processed into the GLI-R repressor and translocated into the nucleus. b) When HhN is present, it binds to Ptch which in turn releases the inhibition over Smo. Ptch is internalised and degraded while Smo is transported to the membrane where it associates with EVC and accumulates at the tip of the cilium. Sufu-Gli2/3 complex is transported to the tip of the cilium by Kif7, where it dissociates. GLI2/3 are thought to undergo postranslational modifications to become GLI-A, and then travel to the nucleus to promote the transcription of Hh dependent genes. Adapted from [93].

domains result in continuous activation of Smo, indicating that both domains are important for proper function of Ptch [89]. One hypothesis for Ptch function is that it transports vitamin D3 in and cholesterol out of the cell to inhibit Smo in the absence of Hh [107]. Upon HhN binding, transport is reversed resulting in Hh pathway activation. However, extensive studies are still required to elucidate the true mechanism of Smo inhibition and Hh pathway activation.

In addition to Ptch, three agonistic co-receptors Cell adhesion molecule-related, Downregulated by Oncogenes (Cdo), Brother of Cdo (Boc) and Growth Arrest Specific 1 (Gas1) were found to directly bind to HhN and activate signalling [108, 109, 110]. Initial reports suggested that receptors compete with Ptch [111]. However, more recent studies indicate that the three co-receptors independently form multimeric complexes with Ptch to bind HhN [112, 113]. The absence of individual co-receptors results in a decrease in Hh signalling. However, the absence of all three co-receptors completely abolishes signalling [112, 113].

Hedgehog interacting protein (Hhip) is a negative regulator receptor that represses Hh signalling by sequestering HhN and targeting it for degradation [114]. Hhip can be membrane-associated or secreted, and binds HhN with nanomolar affinity [115]. Its expression is upregulated in response to HhN signalling creating a negative-feedback loop similar to Ptch [116]. Hhip represses Hh signalling when overexpressed in mice, while Hhip knockout results in increase in Hh signalling [116, 114]. In addition to Hhip, the glypican GPC3 also acts as a negative receptor for Hh that binds HhN ligands [117]. GPC3 forms a complex with HhN proteins and mediates their internalisation via endocytosis to promote reduction

of Hh signalling. This is supported by the observation that GPC3 knockout mice have higher Hh signalling activity despite maintaining similar amounts of HhN transcripts as the wild-type [117].

1.2.5 Intracellular Hh signalling

How Smo transduces the signal intracellularly is still not completely understood. Smo is thought to associate with Disc large homologue 5 (DLG5) before being recruited to the primary cilium [118], and with Ellis-van Creveld syndrome (EVC) protein once in the cilium, to promote signal transduction (Figure 1.4, b) [119]. Upon pathway activation, Smo accumulates at the tip of the cilium. The activation of Hh-dependent genes is ultimately controlled by Gli transcription factors, Gli 1, 2 and 3. Full-length Gli 2 and 3 precursor proteins (Gli-FL) can be processed into activators (Gli-A) or repressors (Gli-R), while Gli1 only acts as an activator [120]. In the absence of the Hh ligand, suppressor of fused (Sufu) associates with Gli-FL preventing it from translocating into the nucleus [121, 122]. While in the cytosol, Gli-FL is primed for phosphorylation by protein kinase A (PKA), glycogen synthase kinase 3 (GSK3) and casein kinase 1 (CK1). In addition, GPCR protein GPR161 stimulates an increase of cAMP, which in turn stimulates PKA activity. Phosphorylated Gli-FL is directed for ubiquitination and degradation to generate Gli-R that acts as a repressor of Hh target genes in the nucleus [123, 124]. When HhN is present, the Sufu-Gli-FL complex is thought to accumulate at the tip of the cilia with the help of Kif7, a

kinesin protein that promotes transport of proteins from the base to the tip of primary cilia [125, 122, 126, 127]. cAMP levels decrease, Gli-FL dissociates and is suspected to undergo post-translational modifications that prime it to become activated (Gli-A). Gli-A is then translocated into the nucleus where it initiates the transcription of Hh-dependent genes that regulate patterning, tissue growth and tissue homeostasis (Figure 1.4, b).

1.3 Structural overview of the Hedgehog pathway

1.3.1 Structure of Hh ligands

The atomic structure of the murine N-terminal signalling domain of Shh (ShhN) was first solved in 1995 by X-ray crystallography at 1.7 Å resolution by a muti-wavelength anomalous dispersion experiment [128]. Subsequently, crystal structures of all vertebrate HhN paralogues (ShhN, DhhN and IhhN) have been determined. HhN family members have almost identical structures, and have since been solved multiple times alone and in complex with various receptors. ShhN consists of an α/β sandwich comprised of two α-helices and a 6-stranded antiparallel β-sheet separated by extensive loop regions (Figure 1.5, a).

The hallmark of the structure is a conserved zinc-binding site reminiscent of the zinc hydrolases' catalytic site, such as that of thermolysin and carboxypeptidase A [129]. In the crystal structure the Zn^{2+} ion is coordinated by three

residues His141, His183, Asp148 and a water molecule (Figure 1.5, a). Catalytic residues characteristic of hydrolyases, His181 and Glu177, and other conserved residues His135 and Glu127 do not coordinate zinc but are found within the ShhN zinc-binding cleft.

The initial hypothesis suggested that ShhN hydrolyses the C-termini of other ShhN proteins based on the crystal packing where residues Ala194 and Lys195 stack against Zn^{2+} of a neighbouring molecule (Figure 1.5, b) [128]. However, subsequent crystal structures of ShhN where the N-terminal heparin-binding Cardin-Weintraub (CW) motif (residues 33-39) stacked against the hypothetical catalytic site, led to the hypothesis that the N-terminus of ShhN could be a substrate for hydrolysis (Figure 1.5, b) [130]. Both of these models are considered crystallographic artefacts as no other evidence has ever been obtained to indicate catalytic activity for any HhN ligand [90]. The zinc-binding site is hence often referred to as a pseudo-active site. The presence of a zinc ion was found to be crucial for stability and signalling activity of all HhN proteins [131, 90], and interaction with receptors [132, 115]. Zinc is also conserved across all HhN proteins with the exception of fly HhN, suggesting a possible divergence of the signalling mechanism.

The ShhN sequence and structure are highly conserved across species, human and mouse ShhN differ in only one amino acid, while human and fly ShhN share 66% sequence identity and 80% sequence similarity. Sequence identity between human ShhN and other metazoan ShhN is >95% (Figure 1.6, a). Furthermore, the sequences of vertebrate HhN proteins ShhN, DhhN and IhhN are highly

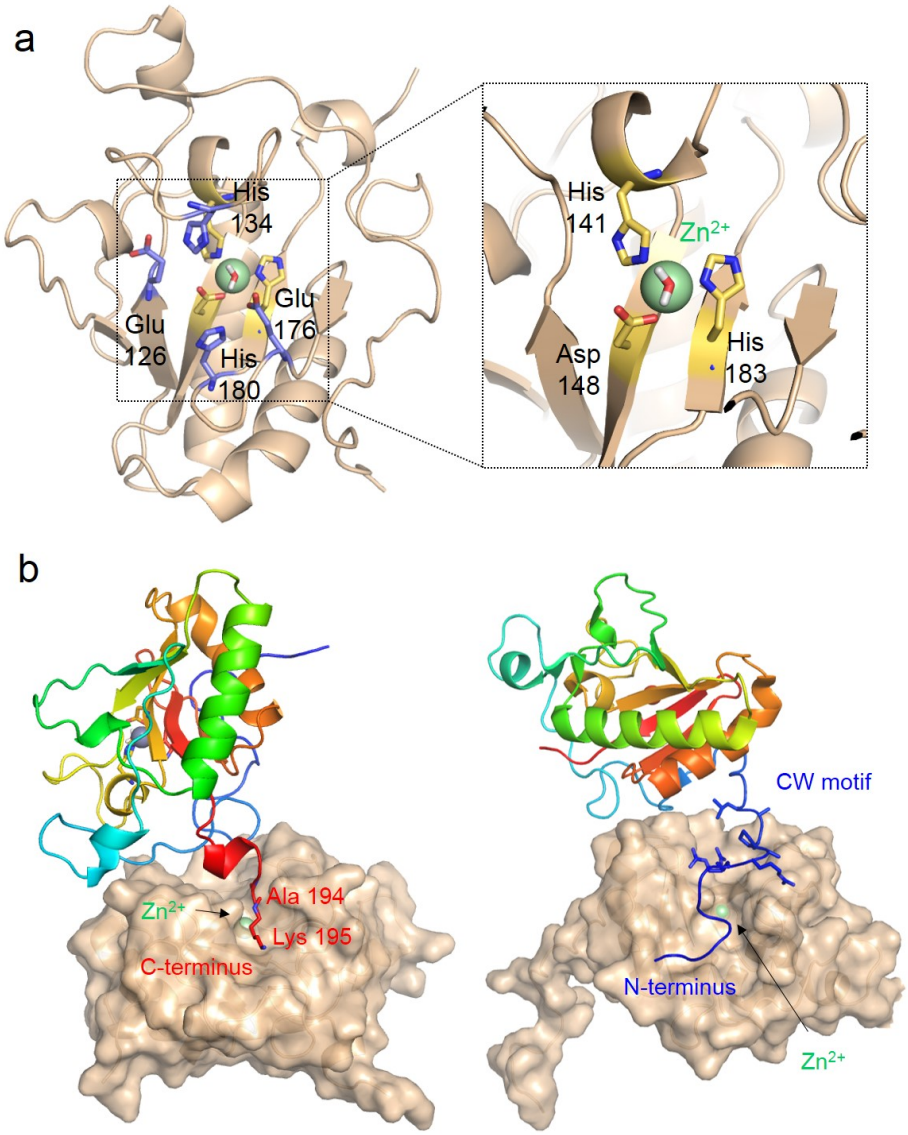


Figure 1.5: **Structure of murine ShhN.** a) ShhN structure (pdb: 1VHH) contains a conserved pseudo-active zinc-binding site similar to zinc hydrolases. The zinc ion is coordinated by His 141, Asp 148, His 183 (yellow) and a water molecule, equivalent to His 140, Asp 147 and His 182 in human ShhN. In addition, conserved hydrolase residues His 134, Glu 126, Glu 176 and His 180 (light blue) are found in close proximity of the zinc binding site. b) ShhN was thought to act as a hydrolase based on crystal structures where C- (red) or N-termini (blue) stack against the active site (pdb: 1VHH and 3M1N). However, no catalytic activity was ever detected for any HhN ligand.

conserved with 74% sequence identity between ShhN and DhhN, and 88% between ShhN and IhhN, while sequence similarity between all is >90% (Figure 1.6, b). Human and mouse ShhN, DhhN and IhhN protein structures solved to date have a similar overall fold (RMSD < 0.3-0.5 Å) and a conserved zinc-binding site.

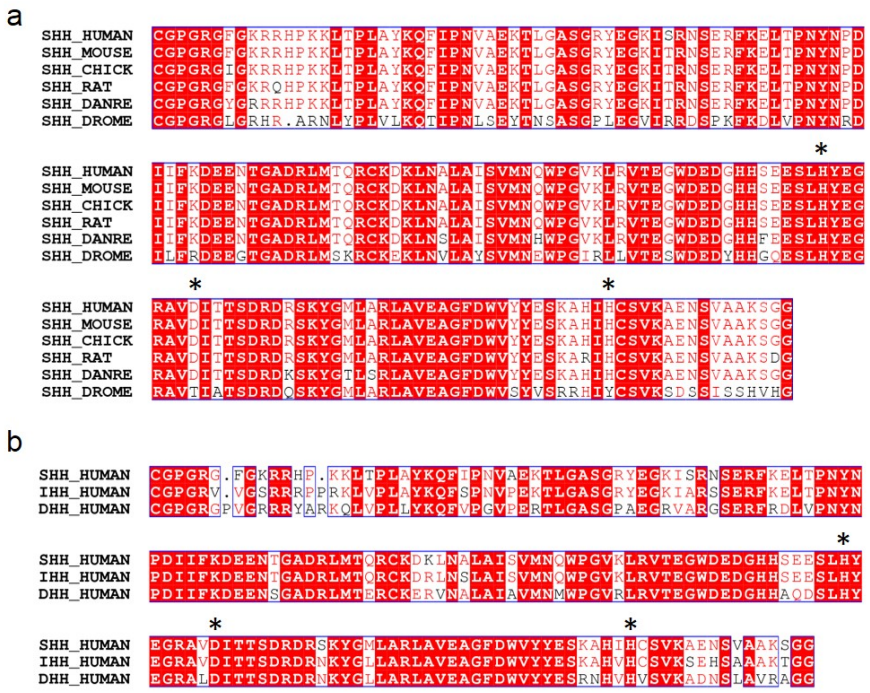


Figure 1.6: **HhN sequence alignment.** a) The ShhN sequences from human, mouse, chick, rat, zebrafish (DANRE) and fly (DROME) were aligned. The sequences are highly conserved with >80% sequence similarity. All sequences share >95% sequence identity when compared to human, except fly ShhN which shares 66%. b) Human ShhN, DhhN and IhhN are also highly conserved, with ShhN and IhhN sharing 88%, and ShhN and Dhh sharing 74% sequence identity. Conserved residues are highlighted in red and the zinc coordinating residues are labelled with *. The sequences were aligned using Multalin [133].

1.3.2 Interaction of Hh ligands with receptors

Structural studies of HhN family members with their receptors revealed another low affinity ion-binding site, namely a double Ca^{2+} site located adjacent to the

zinc site. This revealed the importance of the ShhN ion-binding cleft for receptor and GAG interaction and signalling potency.

1.3.2.1 Cdo

The vertebrate paralogues Cdo and Boc, homologues of the *Drosophila* receptors Ihog and Boi, are single-pass transmembrane receptors that bind directly to ShhN protein and act as positive regulators of Hh signalling [109]. The extracellular domain of Cdo and Boc consists of five immunoglobulin (Ig) domains and three fibronectin (Fn) type III domains, followed by a transmembrane region and a cytoplasmic tail.

Surprisingly, the crystal structure of the Cdo-ShhN complex revealed a different binding mechanism to that of *Drosophila* Ihog-HhN complex (Figure 1.7, a) [111]. *Drosophila* Ihog Fn1, which binds to heparin and mediates interaction with HhN, is homologous to Cdo Fn2. However, purified ShhN did not interact with either Fn2, Fn3 or Fn23 tandem of Cdo, even in the presence of heparin. Further studies indicated that ShhN binds to Cdo Fn3 at a 1:1 ratio with a dissociation constant of $\sim$ 1 µM only in the presence of 1 mM calcium. Calcium was found to bind to ShhN with very weak affinity (100 µM) hence its absence from previous structures. However, the Ca^{2+} concentration in the extracellular space is $\sim$ 1 mM indicating that Ca^{2+}-bound Shh is the physiological form.

In the crystal structure of the ShhN-CdoFn2,3 complex, two Ca^{2+} ions were found within the same binding cleft as Zn^{2+} (Figure 1.7, a) [111]. Overall, the receptor binding interface on the ShhN side encompasses the Ca^{2+} binding site,

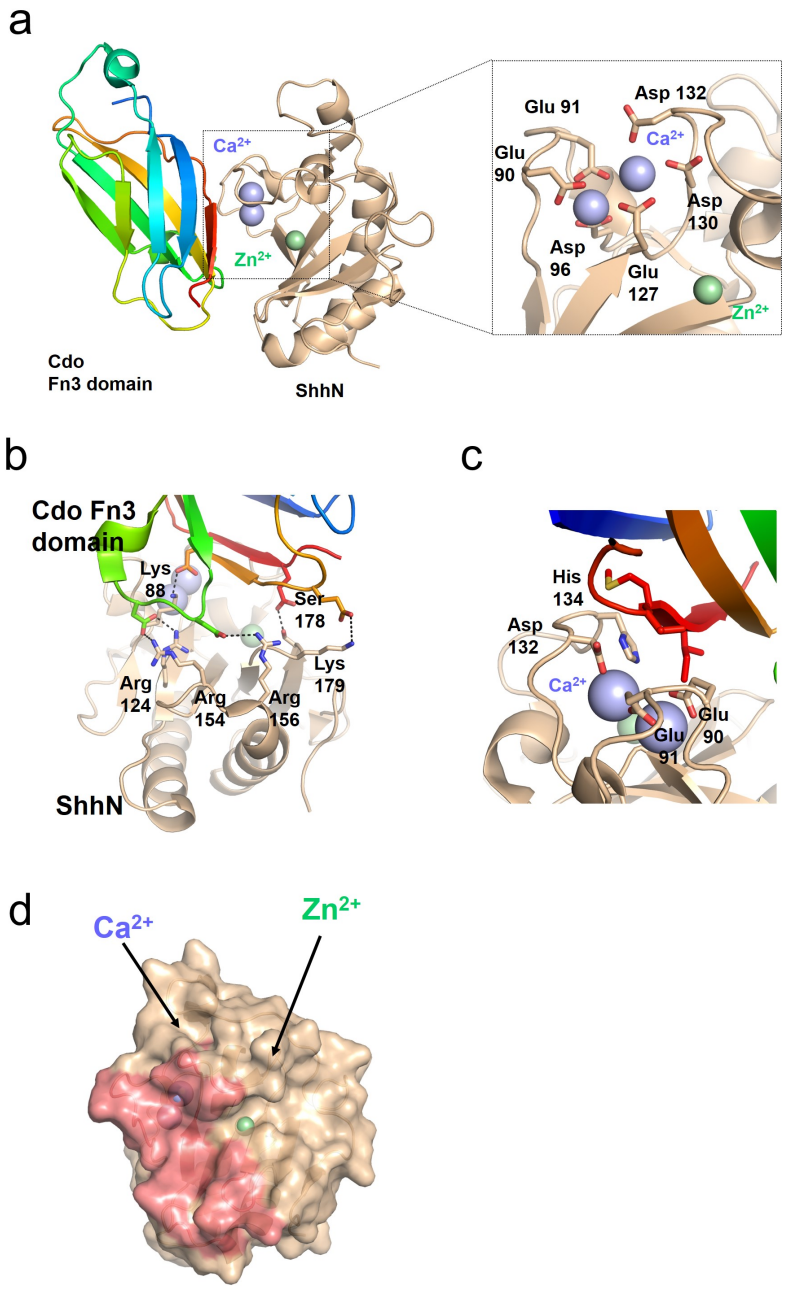


Figure 1.7: **Structure of Cdo-ShhN complex.** a) ShhN binds to Cdo Fn3 domain only in the presence of calcium ions that are coordinated by three Glu residues 90, 91 and 127, and three Asp residues 96, 130, 132 (pdb: 3D1M). b) ShhN-Cdo interactions are largely hydrophillic and involve the interaction with positively charged ShhN residues Lys88, Arg124, Arg154, Arg156 and Lys179, and Ser178. c) Although CdoFn3 does not make direct contact with calcium ions, the ions stabilise the loop of negatively charged residues increasing complementary of the interface. d) The interaction surface was mapped to ShhN and highlighted in deep salmon colour. The interface maps onto the ion binding cleft and is shifted towards calcium ions.

a positively charged area adjacent to Ca^{2+} and partially the Zn^{2+} binding site (Figure 1.7, d). The ShhN-CdoFn2,3 interface comprises mainly of hydrophilic interactions involving ShhN residues Lys88, Arg124, Arg154, Arg156, Ser178 and Lys179 and hydrophobic interactions involving His134 and Glu90 that are found around the calcium site (Figure 1.7, b).

The Ca^{2+} binding site on ShhN is novel and comprises three aspartate (Asp 96, Asp 130, Asp 132) and three glutamate residues (Glu 90, Glu 91 and Glu 127) that coordinate Ca^{2+} and that are conserved across all HhN proteins including flies (Figure 1.7, a). The oxygen atoms of coordinating residues were all found within 2.5 Å from ions, characteristic of Ca^{2+} coordination.

Even though Cdo does not interact with Ca^{2+} directly, Ca^{2+} ions stabilise the loop that allows Glu90, Asp132 and His134 to form an interaction with Cdo residues (Figure 1.7, c). Boc interacts with HhN in a similar manner [132]. In addition to binding to Cdo, calcium is essential for Gas1 binding and it increased binding to Hhip and to a lesser extent to Ptch, indicating the importance of the calcium-binding site in interactions with receptors [111].

1.3.2.2 Hhip

Hhip is a vertebrate specific negative regulator of the Hh pathway, which is upregulated in response to Hh pathway activation [114, 134]. It is composed of an N-terminal Cysteine Rich Domain (CRD) and a six-bladed β-propeller domain followed by two Epidermal Growth Factor (EGF) domains (Figure 1.8, a). Hhip can exist as both membrane-associated or in a secreted form depending on the

local GAG concentration, and can bind all HhN proteins with high affinity [114].

The complexes between Hhip and ShhN, and Hhip and DhhN were characterised structurally in the presence and absence of calcium to reveal a binding interface that also utilises the Zn^{2+} binding cleft (Figure 1.8, b) [52, 135]. Binding studies indicated that Hhip binds to HhN ligands with nanomolar affinity only in the presence of Zn^{2+}, and that binding is abolished in the presence of metal chelator EDTA. In contrast to Cdo, the presence of Ca^{2+} is not essential but it increases binding affinity of ShhN-Hhip more than 10-fold.

The different roles of ions might be explained by the molecular interactions between HhN ligands and Hhip. ShhN binds to Hhip loops BL1 and BL2 that protrude from blade 3 of the β-propeller (Figure 1.8, a). The Hhip-Asp383 side chain occupies the free coordination sphere on the zinc replacing the water molecule thus coordinating the Zn^{2+} ion directly (Figure 1.8, a). The role of Ca^{2+} ions in the ShhN-Hhip interaction is to reduce the repulsion between the negatively charged Hhip loop BL1 and the negatively charged ShhN loop (residues Lys88-Gly94) by stabilising it, thereby allowing for additional interactions to be formed (Figure 1.8, a).

Overall, the ShhN interaction surface is small (759 $Å^2$) and encompasses similar interface as Cdo. However, the higher, nanomolar affinity can be explained by the coordination of Hhip loop BL2 to directly interact with ShhN zinc-binding site. The Hhip-ShhN interface is more central with respect to the ion binding groove compared to the Cdo-ShhN, involving both the Zn^{2+} and Ca^{2+} binding sites (Figure 1.8, b).

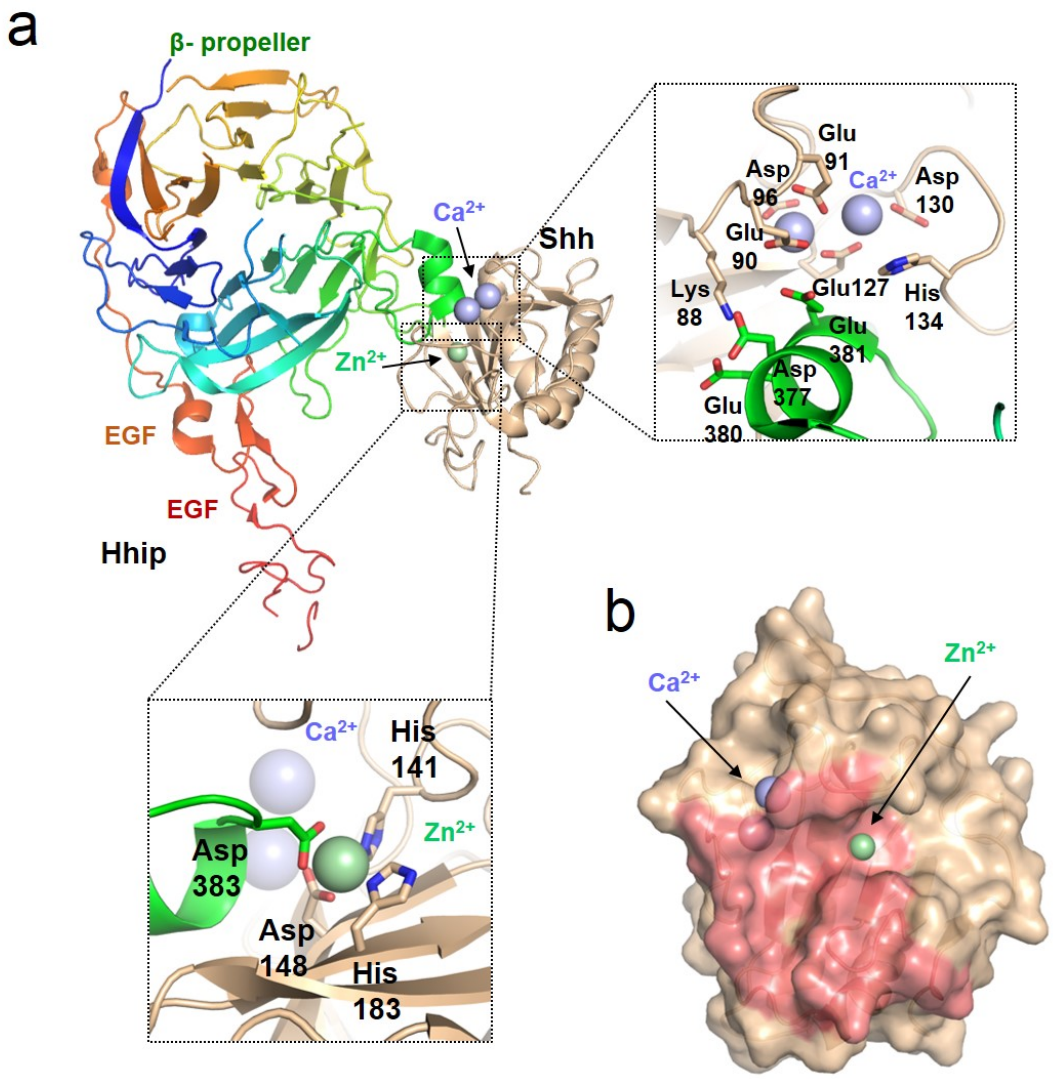


Figure 1.8: **Structure of the ShhN-Hhip complex.** a) Hhip consists of a β-propeller and two EGF domains and it interacts with ShhN via the third blade of the β-propeller (pdb: 2WFX). Hhip residues Asp 383 replaces the water molecule and directly coordinates the zinc of ShhN, while the calcium ions form hydrophilic interactions with negatively charged loop of ShhN thereby shielding it from the negatively charged patch of Hhip loop BL1. b) The Hhip interface was mapped onto ShhN and highlighted in deep salmon colour. The interface overlaps with Cdo, however, it is relatively more central with respect to the ion binding groove encompassing zinc as well as the calcium binding sites.

1.3.2.3 Patched and E51 inhibitory antibody

Despite Ptch being the principle HhN receptor, there has been no structural information on the HhN-Ptch interaction to date. Our understanding of how HhN might interact with Ptch comes primarily from mutation studies and interactions with other competing receptors like Hhip and antibody 5E1.

Binding to and inhibiting Ptch were found to be two separate events involving different areas of ShhN protein [136]. Palmitoylated Δ22Shh peptide (residues 23-45) was found to be enough to at least partially inhibit Ptch and activate Hh signalling. ShhN lacking the N-terminal peptide could still bind Ptch with nanomolar affinity but not have an inhibitory effect. This indicated that the globular domain of ShhN mediates high affinity interaction with Ptch while the palmitoylated N-terminus has a role in inhibiting Ptch [136]. Furthermore, a ShhN mutation that contains two N-terminal isoleucine residues (Cys24II) replacing the palmitoyl moiety was shown to be 8× more active than the non-palmitoylated Shh [137]. This suggested that it is the hydrophobicity rather than the specific hydrophobic moiety at the N-terminus of ShhN that induces Ptch inhibition [137]. It was further shown that although Ptch-mediated endocytosis of the HhN-Ptch complex is independent of palmitoylated N-terminal peptide, the globular protein alone is not sufficient to recruit Smo to the primary cilia and trigger Hh signalling [136]. Together, these findings suggest a two-step interaction between ShhN and Ptch where the globular domain of ShhN is responsible for mediating high-affinity binding, while the palmitoylated N-terminus

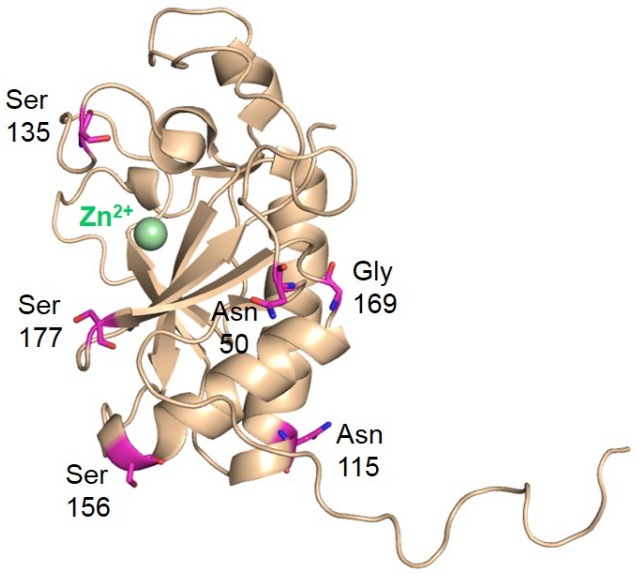


Figure 1.9: **ShhN residues involved in Ptch binding.** Mutation studies indicated that residues Ser156, Asn50, Ser177, Asn115, Gly169, Ser135 are important for binding to Ptch. The residues are highlighted in magenta and shown as sticks.

of ShhN inhibits Ptch.

The exact interface between globular ShhN and Ptch is not known. However, mutation and steric interference studies indicated that ShhN residues Ser156, Asn50, Ser177, Asn115, Gly169, Ser135 were in proximity of the Ptch binding site (Figure 1.9) [138]. This indicated that the ShhN-Ptch binding interface is likely to overlap with those of Hhip but includes additional residues.

Our understanding of the ShhN-Ptch interaction also comes from experiments using the Hh-specific antibody 5E1 that competes with Ptch for HhN binding, binds all three vertebrate HhN proteins with nanomolar affinity and inhibits Hh signalling [135]. This antibody is now used for studying the role of the Hh pathway in embryonic development and cancer. 5E1 competes with Ptch and Hhip for ShhN and was shown to have higher affinity for ShhN in the presence of zinc and calcium ions [135], indicating that the interaction surfaces might overlap.

The crystal structure of the 5E1 Fab fragment in complex with human ShhN,

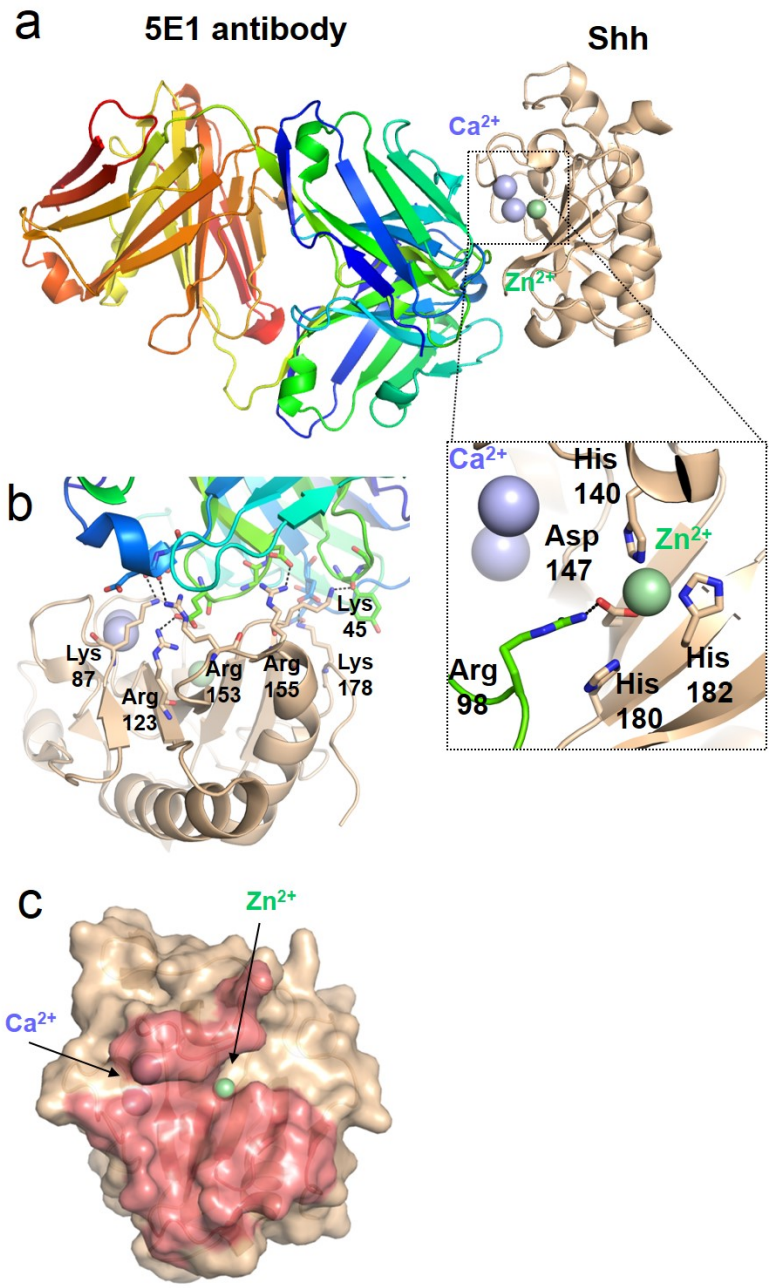


Figure 1.10: **Structure of the 5E1-ShhN complex.** a) 5E1 light and heavy chains interact with ShhN at the pseudo-active site (pdb: 3MXW). Although 5E1 does not interact directly with zinc, 5E1 residue Arg98 interacts with Asp147 that coordinates zinc. b) The binding interface is hydrophilic and comprises positively charged residues Lys87, Arg123, Arg153, Arg155, Lys178 and Lys45. c) The 5E1-ShhN interface was mapped onto ShhN and highlighted in deep salmon. The interface resembles Hhip and is located centrally relative to the ion binding groove.

confirmed that the interface encompasses the ion binding cleft and the pseudo-active site, and overlaps with Hhip and to an extent with Cdo (Figure 1.10, c) [139]. All six complementary determining regions (CDRs) from 5E1 contribute to the interaction with ShhN (Figure 1.10, b), resulting in a surface area similar to the interfaces observed in other Fab-antigen complexes. The majority of interactions are formed between a positively charged 'hotspot' of ShhN (Lys45, Lys87, Arg 123, Arg 153, Arg 155 and Lys 178), and negatively charged residues of 5E1 CDRs (Figure 1.10, b). Even though 5E1 residues do not interact with Zn^{2+} directly, there is an interaction with Zn^{2+}-coordinating residue Asp147 (Figure 1.10, a). In addition, van der Waals interactions are formed with Zn^{2+}-neighbouring residues Ser177, Ala179 and His180. As with Hhip, addition of Ca^{2+} was shown to increase the binding affinity [139]. Although no direct interactions between 5E1 and Ca^{2+} were observed, similarly to Hhip, Ca^{2+} stabilise the negatively-charged patch of residues that coordinate ions thereby increasing charge compatibility between Shh and 5E1 (Figure 1.10).

1.3.2.4 Glycosaminoglycans

The Cardin-Weintraub (CW) motif was identified as a conserved sequence found in heparin-binding proteins [140]. The sequence consists of a combination of 6-7 basic (B) and aliphatic (X) residues (-X-B-B-X-B-X- or -X-B-B-B-X-X-B-X-). HhN ligands contain a CW motif at the N-terminus (RRHPKKL, residues 33-39), which was shown to mediate interactions with GAGs, and to regulate Hh signalling [140, 141]. The importance of GAGs in Hh signalling has been

described in HhN secretion and gradient formation aided by the glypican, GPC3 [69]. Furthermore, various glypicans have been shown to act as both positive and negative Hh regulators [142, 143]. The interactions with GAGs were thought to be mediated by the CW motif. However, mutations in the CW motif did not completely abolish the interaction with GAGs [141, 144, 145].

In addition to CW motif, crystal structures of ShhN in complex with heparin and Chondroitin 6 sulphate (C6S) revealed another GAG binding site (Figure 1.11) [99]. Both GAGs bind to the receptor interface 'hotspot' consisting of a patch of positively charged residues (Arg124, 154 and 156, Lys88 and 179) that interacts with the negatively charged sulphate groups of heparin and C6S (Figure 1.11, b and c). The zinc and calcium ions were not directly involved in binding. However, they do shield negatively charged residues thereby stabilising the GAG-ShhN structure. Although the binding mechanism is similar, more sulphated heparin promoted dimerisation of ShhN molecules while less sulphated C6S only bound to monomeric ShhN (Figure 1.11, a and c). This structure not only revealed an alternative GAG binding site on ShhN, but also presented a possible mechanism by which multimersation of ShhN in the extracellular matrix is regulated by sulphation patterns of HSPGs thereby promoting gradient formation. Furthermore, this GAG binding site would allow for HhN proteins to be incorporated in lipoprotein particles to interact with CSPGs and HSPGs, and to form a concentration gradient.

Overall, structural information contributed to our understanding of HhN and receptor interactions indicating the importance of the ion binding cleft and the

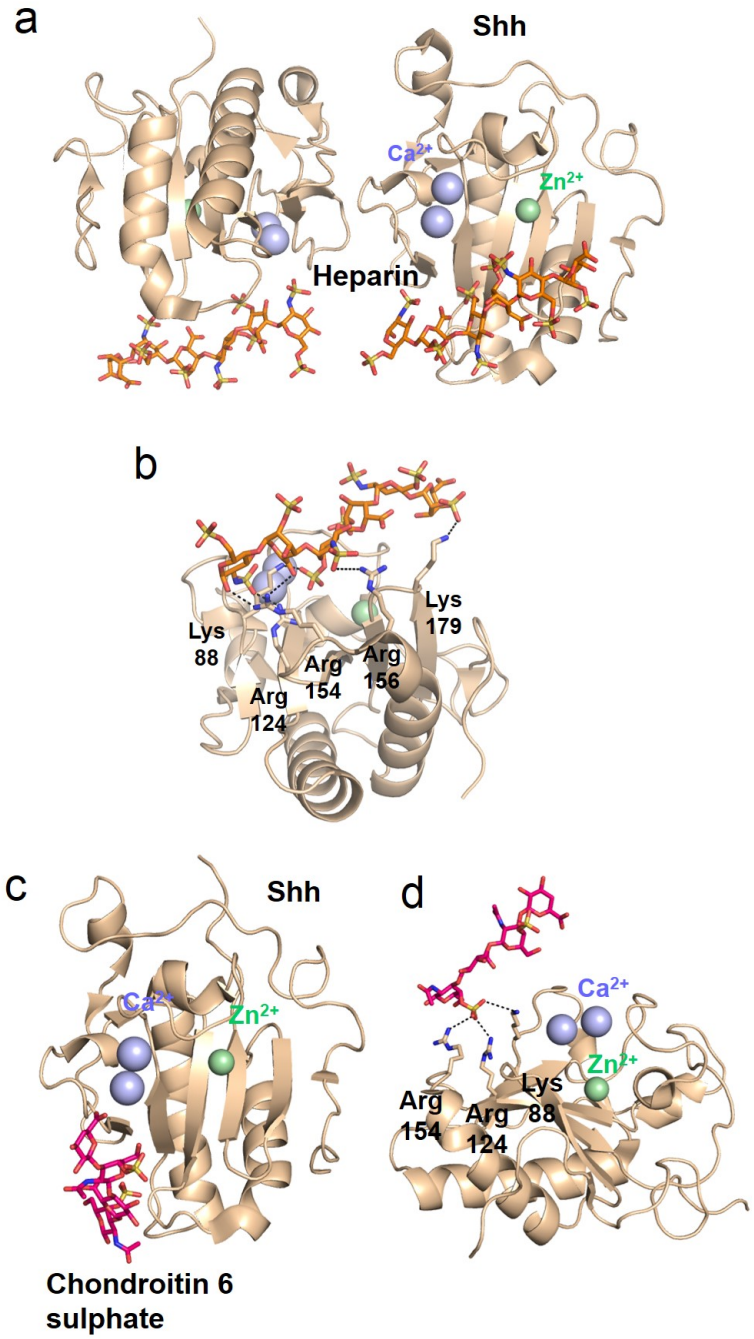


Figure 1.11: **ShhN interaction with GAGs.** a) ShhN interacts with highly sulphated heparin which promotes dimerisation (pdb: 4C4N) and binds to the receptor interface 'hotspot'. b) Positively charged residues Lys 179, Arg 156, Arg 154, Arg124 and Lys 88 interact with the negatively charged sulphate groups of heparin. c-d) ShhN interacts with less sulphated C6S as a monomer (pdb: 4C4M) forming on interactions with C6S via residues Lys 88, Arg 154 and Arg 124.

positively charged 'hotspot'. However, more molecular information is needed to understand the interactions between HhN and Ptch, and how this correlates to the interaction with co-receptors Cdo, Boc and Gas1.

1.4 The Hedgehog pathway in disease

In addition to embryonic development, the Hh pathway remains active in adults where it contributes to tissue homeostasis and regulates stem cell division, cell growth and proliferation. Dysfunction of the Hh pathway has been correlated with developmental defects, neurological disorders and cancer. From the therapeutic point of view, the Hh pathway is intriguing because it can be modulated both positively and negatively in different conditions. Hh-dependent diseases can either be caused by mutation-independent dysregulation of the pathway, or by mutations in HhN ligands, receptors or intracellular downstream signalling.

1.4.1 Developmental defects

Malfunction of the Hh pathway has been linked to developmental defects such as holoprosencenphaly (HPE) [146], ciliopathies [147], Pallister-Hall syndrome [148], Smith-Lemli-Opitz syndrome [149] and brachydactyly type a1 (BDA1) [150].

HPE is the most common forebrain development disorder, which is caused by incomplete separation of the forebrain into two hemispheres. The most extreme form of HPE is cyclopia, which manifests in incomplete symmetry formation and separation of the forebrain and eyes, and a lack of a nose [151]. Depending on

the severity, other physical manifestations can include midface hypoplesia with cleft lip, ethmocephaly (closeness off eyes) and single-nostril nose [152].

Hh pathway downregulation caused by mutations in Shh has been a major cause of HPE (Figure 1.12) [153]. In addition to frame-shift mutations leading to protein truncation, mutations of ShhN that contribute to protein stability and interactions with its receptors were correlated to the disease [146, 154].

Mutations of ShhN residues Pro26 and Gly31, essential for processing and palmitoylation of the N-terminus, and residues His140, Cys183 and loop Glu176-Lys178, located in close proximity of the pseudo-active site, have been discovered in patients with HPE [79, 155, 153, 78]. Furthermore, residue Asp88, which is important in stabilising the conformation of calcium binding loop thereby aiding binding to receptors Cdo and Gas1, was mutated to Val in patients with HPE [154]. In addition, more than 200 mutations in HPE have been associated with ShhN, which affect signal sequence processing, lipid modifications, secretion and receptor interaction [79]. These findings highlighted the need for the intact ion binding cleft of ShhN for proper Hh signalling [79].

BDA1 is a hereditary disorder that manifests in short or absent phalanges. Mutation in Ihh equivalent to Glu90, Asp95, Arg123, Thr125, Glu126 and Thr149 residues in ShhN [150, 156], located around the calcium coordination site, affect the function of IhhN in proper bone development leading to BDA1 (Figure 1.12). In addition to ShhN and IhhN, mutations in various other Hh pathway components including Disp, Ptch, Smo, Cdo, Skinny Hh and Gli proteins were associated with abnormal embryonic development [157].

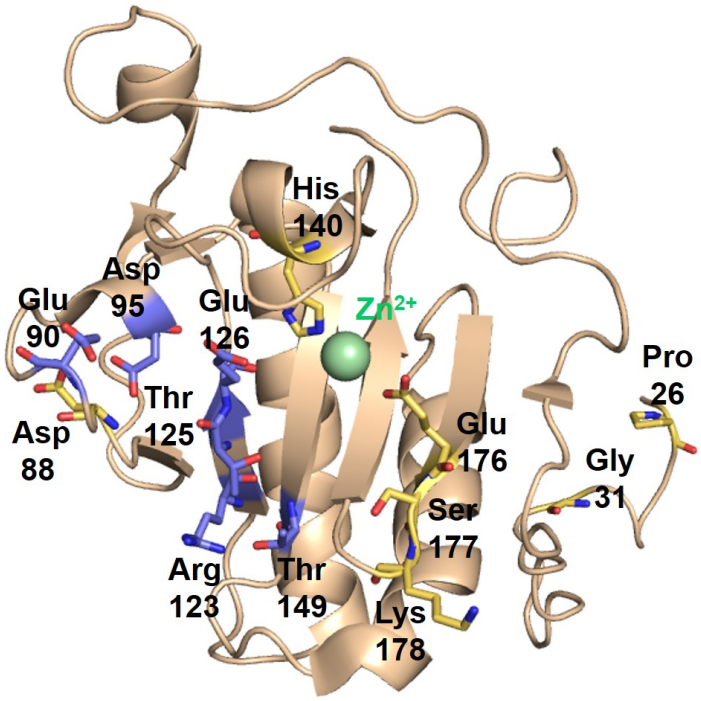


Figure 1.12: **Mutations of ShhN associated with developmental diseases.** Mutations of residues Asp88, His140, Glu176-Lys178, Gly31 and Pro26 (yellow) result in HPE, while mutations of Arg123, Thr125, Glu126, Asp95 and Glu90 (blue) cause BDA1.

1.4.2 Tissue homeostasis and regeneration

The Hh pathway remains active in adult organs and plays an important role in tissue self-renewal and healing [158].

As in embryonic development, the Hh pathway is particularly significant in neurogenesis of adult mammalian brains. Self-renewing neural stem cells (NSC) located in lateral ventricles and the dentate gyrus are converted into transient-amplifying cells, which act as committed progenitor cells that are further expanded into cell populations. This process occurs in all tissues with quiescent stem cells in response to injury, and is continuous in self-renewing tissues like brain and skin. During embryonic development, lack of Hh signalling results in

apoptosis of stem cell regions indicating that Hh is required for the formation of NCS population [159, 160]. In the adult brain, the Hh pathway has a role in the maintenance and self-renewal of NSC [161]. Activation of ShhN, the only Hh ligand in the brain, was correlated with neurogenesis in both lateral ventricles and dentate gyrus [161], while loss of ShhN was correlated with inhibition of NCS proliferation [162]. Although the exact mechanism of how Hh modulates stem cell renewal is not completely understood, ShhN was shown to modulate production of glia cells, oligodendrocytes and astrocytes [163, 164]. This highlighted the importance of ShhN in neuronal maintenance as well as regeneration in response to injury. In addition to neuronal tissue, Hh signalling also contributes to the maintenance of quiescent stem cells and contributes to regeneration of skin, teeth, hair, bone, muscle and to haematopoiesis [158].

Research into the application of Hh agonists has been explored for stimulating tissue regeneration and treatment of neurodegenerative diseases [165]. Smo agonist SAG was shown to improve proliferation of neural and glia cells, and improve synaptic plasticity in the hippocampus thereby influencing memory and learning [166, 167]. Purmorphamine, another Smo agonist, was shown to induce neuronal differentiation, to restore neurological damage after a stroke and to protect neurons from oxidative damage [168, 169]. Although this is a fairly new area of research, the Hh pathway is a potential target in tissue and organ regeneration.

1.4.3 Cancer

Although the Hh pathway is important in tissue regeneration, constitutive hyperactivation has been correlated with a number of cancers such as skin, brain, pancreatic, breast, lung, prostate and others [170]. Tumour growth and metastasis require activation of Hh, Wnt, TGF-β and Notch pathways that stimulate cell division, growth, survival, proliferation and vascularisation. The Hh pathway is exploited by tumour cells either through excess secretion of Hh ligands, mutation in Hh receptors or malfunction of the transcription regulation. Hh signalling initiates cellular processed that contribute to growth and metastasis of tumours.

The Hh pathway has an important role in shaping the tumour microenvironment and instigates proliferation, inflammation, metabolism deregulation, angiogenesis, genomic instability, invasion and metastasis, all processes necessary for cancer survival [171]. In addition, Hh was found to play a role in evasion of apoptosis, immune response and growth suppression, thereby enabling cancer growth [171].

Hh pathway overactivation in various cancers was found to be due to ligand-independent or ligand-dependent mechanisms, depending on the mode of action.

1.4.3.1 Ligand-independent mechanism

The Hh pathway can be continuously activated in cancer as a result of mutations in proteins downstream of Hh ligands like Ptch, Smo and Sufu. This mechanism is ligand-independent as Hh pathway activation does not depend on the presence

of Hh ligands. Ligand-independent mechanisms have been identified in basal cell carcinoma (BCC, skin cancer) [172], and the pediatric cancers medulloblastoma (MB, brain cancer) [173] and rhabdomyosarcoma (RMS, muscle cancer) [174]. The Hh pathway was first related to Gorlin syndrome (nevoid basal-cell carcinoma) after a discovery of a loss-of-function mutation in Ptch [175]. Gorlin syndrome manifests in defects in skin, eyes, bones and nervous system, and carries a predisposition to BCC, Keratocystic odontogenic tumour (KCOT, epidermal cysts) and MB [176]. Mutations in Ptch were found in over 90% of BCCs [177], 20% of MB [178] and 30% KCOT [179]. Furthermore, loss-of-function mutations in Sufu have been reported in 10% of MBs [180], and gain-of-function mutations of Smo were identified in BCC and MB [172, 181].

The best established target of the Hh pathway is Smo with dozens of modulators developed over the years resulting in 2 FDA approved compounds, Vismodegib and Sonidegib, for the treatment of advanced BCC based on 30-40% tumour response rate [182, 183]. Despite the initial excitement, Vismodegib performed poorly in the clinic with only 20% of patients achieving complete response [184]. Side effects such as hair loss, muscle spasms, fatigue and anorexia resulted in >50% drop-out rate [185]. Cancer recurrence was also common as a result of acquired resistance in Smo and Gli [185]. Structural studies revealed that the common Smo mutations map to the TM drug-binding site thereby preventing drug binding [96]. Over 25 Smo inhibitors have been developed to date and are being tested in clinical trials for the treatment of BCC, MB, leukaemia and pancreatic cancers [93]. Hh pathway inhibition is now a validated strategy for

inhibiting tumour growth and additional drugs targeting Gli transcription factors are also being researched as an alternative treatment.

1.4.3.2 Ligand-dependent mechanism

Even though Hh ligand-independent mechanisms were discovered first and are better understood, Hh was found to be active in a much broader number of cancers such as pancreatic, breast, lung, prostate and other cancers. However, no mutations were associated with Hh overactivation [170, 93]. There are three ligand-dependent mechanisms by which Hh can induce cancer: autocrine, paracrine and inverse paracrine.

In autocrine signalling, tumour cells express and secrete an excess of HhN which binds to the tumour cell creating a positive feedback loop. This mechanism helps sustain the activity of Hh target genes that control proliferation and cell survival, and was found in breast [186], lung [187], prostate [188], stomach [189] and colorectal cancers [190]. In paracrine signalling, tumour cells express and secret Hh ligands that bind to the surrounding stromal cells, which in turn stimulate tumour growth by secreting insulin-like growth factor (IGF), Wnt, interleukin-6 (IL-6) and vascular endothelial growth factor (VEGF) [191, 192]. This paracrine signalling mechanism was identified in pancreatic and colorectal cancers [193, 192]. In an inverse paracrine scenario, stromal cells secrete Hh ligands that bind to surrounding tumour cells and induce growth. Inverse paracrine signalling was found in myeloma and lymphoma [93].

Furthermore, cancer stem cells (CSC) which make up a minor subpopulation

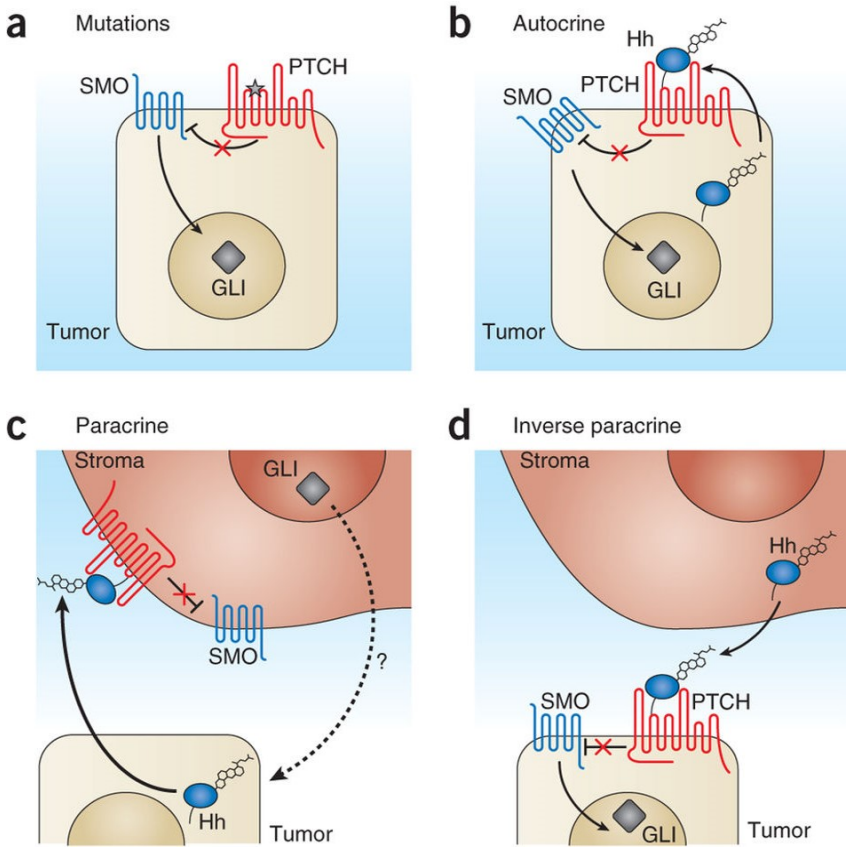


Figure 1.13: **Ligand-independent and ligand-dependent Hh mechanisms in cancer.** Hh pathway was found to be constitutively active in multiple cancers where it contributes to cell growth, survival and proliferation. a) A ligand-independent mechanisms is most commonly caused by loss-of-function mutations in Ptch and is found in BCC. Three ligand-dependent mechanisms (b-d) were also identified where the Hh pathway is overactivated in the absence of any mutation. b) In autocrine ligand-dependent mechanism, tumour cell secrets HhN ligand that binds to the same cell creating a positive feed-back loop. c) In paracrine mechanism, the tumour cell secrets Hh ligand that binds to neighbouring stroma cells which in turn secret IGF, Wnt, IL-6 and VEGF that stimulate growth. d) In inverse paracrine, stroma cells secrete Hh ligand that binds to the tumour cells. Adapted from [170].

of tumour cells were shown to be maintained by ligand-independent activation of the Hh pathway [194]. CSC are correlated with drug resistance and metastasis, therefore inhibiting their proliferation would have an impact on disease prognosis. Hh dependency was found in multiple cancers including breast, lung, glioblastoma and others. However, further studies are needed to establish the role of Hh

pathway and the viability of targeting.

Although the major Hh drug target is Smo, the discovery of ligand-independent mechanisms coupled with the poor performance of Smo inhibitors resulted in the possibility of targeting upstream Hh proteins, primarily HhN ligands and Skinny Hh. Inhibiting HhN interaction with Ptch and other receptors was probed with 5E1 antibody. Treatment of tumour cells with 5E1 was shown to inhibit the growth of pancreatic, liver and colon cancers cells [192, 195, 196]. In addition, a small-molecule inhibitor Robotnikinin that directly binds to ShhN with micromolar affinity was discovered and was shown to inhibit Hh signalling in human skin grafts [197]. The two ShhN inhibitors have been used as a proof of concept, but have not reached the clinic. Skinny Hh is essential for the production of palmitoylated HhN ligands required for proper Hh signalling. Small-molecule inhibitors of Skinny Hh called 'RUSKI' have been generated and were shown to inhibit palmitoylation of HhN and Hh pathway signalling in cells [198, 199]. Furthermore, inhibition of Skinny Hh was shown to reduce growth of pancreatic and non-small cell lung carcinoma in a cellular model [200, 201]. Research into targeting upstream components of Smo is still at a beginning stage. However, with ligand-dependent mechanisms being identified in far more cancers than ligand-independent, the true potential of alternative Hh targets and their scope in treating tumours is yet to be discovered.

1.5 Thesis overview and aims

To better understand the potential of targeting HhN and Skinny Hh in ligand-dependent cancers, more and better inhibitors are required. The aim of this thesis was to explore ShhN and Skinny Hh using structural biology techniques to understand their function, with the ultimate goal of developing novel more potent inhibitors.

ShhN has been extensively studied structurally and functionally. Despite the wealth of knowledge on receptor interactions, only one small-molecule and one antibody inhibitor have been identified to date. The absence of more inhibitors could be due to the challenging nature of inhibiting protein-protein interactions. However, this presents an opportunity for more potent and drug-like inhibitors to be developed using structural biology techniques.

Skinny Hh represents a challenging target due to being an integral membrane protein. Although significant progress has been made to understand topology and biochemistry of Skinny Hh, its molecular structure and precise mechanistic function remain unknown. Further advances have been made in developing palmitoylation assays, which led to the discovery of potent inhibitors. However, the limiting factor remains the inability to produce and isolate high-quantities of stable, monodisperse Skinny Hh to be used in crystallisation and drug screening trials. Understanding the mechanism of Skinny Hh will help develop novel more potent inhibitors.

Aims:

1. The first aim was to establish how Robotnikinin, small-molecule inhibitor interacts with ShhN, using X-ray crystallography and molecular docking. These results can help us to better understand the molecular mechanism of ShhN inhibition and to develop novel inhibitors, and are discussed in Chapter 2.

2. The second aim was to develop novel inhibitors for Shh using crystallographic fragment screening. Combining crystallographic, biophysical and cellular data, the aim was to develop better and more potent Shh inhibitors. These results are discussed in Chapters 3 and 4.

3. The third aim was to develop expression and purification protocols for the production of high quantities of thermostable, monodisperse Skinny Hh, with the ultimate goal of determining the crystal structure. These results can lead to a better understanding of the molecular mechanism of Skinny Hh and the development of novel molecules modulating Hh signalling, and are discussed in Chapter 5.

Chapter 2

Structural characterisation of the Shh-Robotnikinin complex

2.1 Abstract

The Hh pathway is essential for embryonic development and is an important drug target for a wide range of cancers. In addition to targeting GPCR Hh signal transducer Smo, other Hh proteins including the ShhN morphogen are potential targets for ligand-dependent Hh-related cancers. Robotnikinin is the only small-molecule inhibitor known to bind directly to ShhN and inhibit the Hh pathway in cell assays.

In this chapter, the interaction between HhN-Robotnikinin were researched using structural biology approaches: molecular docking and X-ray crystallography. In *in silico* studies, Robotnikinin was docked to all available crystal structures of HhN ligands indicating that it likely binds to the pseudo-active site of ShhN, interacting with residues Asp148, His183, His134, Ser135, Glu177 and His181. Using molecular dynamics, we identified that the ShhN loop containing residues 131-141 is important for Robotnikinin interactions. Furthermore, our preliminary crystallographic data further confirmed that a fragment of Robotnikinin, most likely corresponding to the macrocyclic portion, binds to the zinc-binding site. Together, this data supports the finding that Robotnikinin binds at the ion binding cleft of ShhN and not the adjacent receptor-binding 'hotspot'.

2.2 Introduction

The Hedgehog (Hh) pathway is an essential morphogen pathways that affects organogenesis and symmetry formation during embryonic development. In adults, the Hh pathway remains active contributing to tissue regeneration and homeostasis. Malfunction of the Hh pathway causes developmental defects while overactivation has been correlated with cancers, making it an important therapeutic target [202]. Significant effort has been put into developing novel Hh pathway inhibitors that would halt the growth and proliferation of tumours.

The first inhibitors developed for the Hh pathway targeted the Hh signal transducer Smoothened (Smo), which acts downstream of the Hh receptor Patched (Ptch) to activate Hh signalling [203]. In the majority of basal cell carcinomas (BCC) and medulloblastomas (MB), the cancer is caused by a loss-of-function mutation in the receptor Ptch or an activating mutation in Smo. Thus targeting Smo, which is downstream of Ptch was an effective strategy [177, 204, 205]. However, clinical use of the two FDA-approved Smo inhibitors, Vismodegib and Sonidegib, revealed adverse effects that resulted in high dropout rates and cancer recurrence due to acquired drug resistance. More recent insight into ligand-dependent requirement of certain Hh-related cancers [191], like pancreatic and breast cancer, led to revisiting other targets of the Hh pathway such as the Hh morphogen ShhN itself and its palmitoyltransferase Skinny Hh, both acting upstream of Smo [170].

Two inhibitors of ShhN were developed to date: 5E1 antibody and a small-

molecule Robotnikinin. 5E1 antibody was shown to inhibit Hh pathway in cancer cell lines [206] and was structurally characterised in complex with ShhN [139]. Robotnikinin was shown to bind to ShhN and inhibit the Hh pathway [197], but the precise mechanism of how it interacts with ShhN is not known. Neither of the two inhibitors have reached clinical trials.

Robotnikinin is the only small-molecule inhibitor developed that binds to ShhN directly (Figure 2.1, a) [197]. It was discovered in a small-molecule microarray (SMM) screen of approximately 10,000 diversity-oriented synthesis compounds and natural products. Initial hits comprised of a number of structurally related macrocycles that bound to ShhN with low micromolar affinities, subsequently established in SPR assays [197, 207]. Robotnikinin was developed through structure-activity relationship (SAR) optimisation that increased the affinity to 3.1 µM as determined by SPR kinetics analysis [197] (Figure 2.1, a and b). Robotnikinin also showed moderate inhibition of the Hh pathway in luciferase assays using Shh-LIGHT2 cells in the presence of palmitoylated Shh (Figure 2.1, c). Robotnikinin was shown to act upstream of Smo in $Ptch^{-/-}$ cell lines where both Ptch alleles were replaced by β-galactosidase (β-gal) reporter [197]. While small molecules targeting Smo could ablate β-gal activity, addition of Robotnikinin in combination with Shh had no effect on β-gal activity, indicating that it binds upstream of Smo. Robotnikinin was also shown to inhibit differentiation of mouse mesenchymal stem cells, normally accomplished by the addition of Shh, and to inhibit the Hh pathway in a concentration-dependent manner in primary human keratinocytes (Figure 2.1, d).

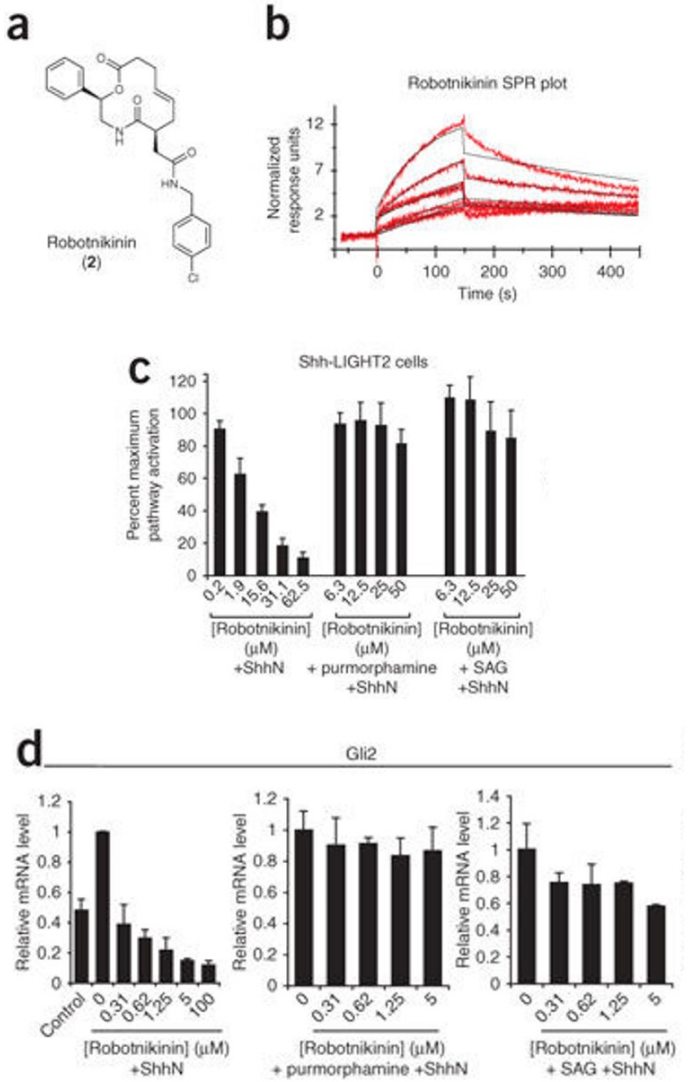


Figure 2.1: **Robotnikinin binds ShhN with micromolar affinity and inhibits Hh pathway in a concentration dependent manner.** a) Structure of Robotnikinin. b) SPR sensorgrams and kinetic analysis of ShhN and Robotnikinin binding. Robotnikinin was tested in a concentration range of 25 to 1.5 μM resulting in a K_d of 3.1 μM. c) Robotnikinin inhibits Hh signalling in a concentration dependent manner in a luciferase reporter assay. The inhibitory effect of Robotnikinin is overridden by Smo agonists purmorphamine and SAG. d) Treatment of primary human keratinocytes with Robotnikinin resulted in inhibition of the Hh pathway and a decrease of endogenous Gli2 in a QPCR assay. The inhibitory effect was ablated with Smo agonists purmorphamine and SAG. Adapted from reference [197].

Following the discovery of initial hits alongside Robotnikinin, attempts have been made to optimise the hits for better potency [208]. When the positions of the macrocyclic oxygen and nitrogen were reversed, an increase in potency

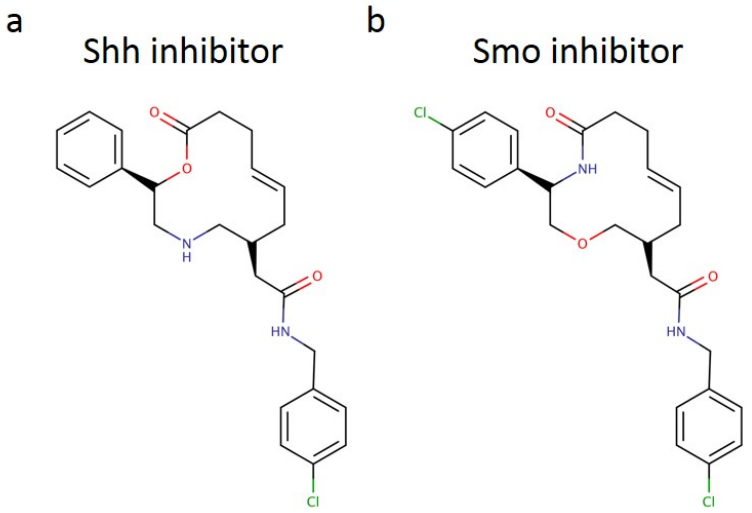


Figure 2.2: **Structure of Robotnikinin and SAR optimised macrocycle** a) Structure of ShhN inhibitory macrocycle Robotnikinin. b) Structure of SAR optimised compound with reverse macrocyclic oxygen and nitrogen, and a 4-chloroarene substituent. Although the optimised compound has higher potency in inhibiting the Hh pathway it changed specificity from ShhN to Smo.

was observed, as demonstrated in stem cell proliferation and Gli assays [208]. Furthermore, the potency was improved by changing arene at 2' position of the macrocycle for a 4-chloroarene substituent. Interestingly, the novel compound, although resembling Robotnikinin and inhibiting the Hh pathway, changed specificity from ShhN to Smo (Figure 2.2).

In silico methods have also been used in an attempt to screen for the next generation of inhibitors using pharmacophore modelling [209]. In this study, the binding site of Robotnikinin was assumed to be the ShhN pseudo-active site. This was based on the ShhN-Hhip complex structure and the sequence similarities between Hhip L2 and Ptch L2 loops, both of which interact with ShhN (Figure 2.3) [115, 135]. Ptch L2 and Robotnikinin were docked to the Hhip L2 binding site on ShhN to generate Shh-Ptch L2 and Shh-Robotnikinin

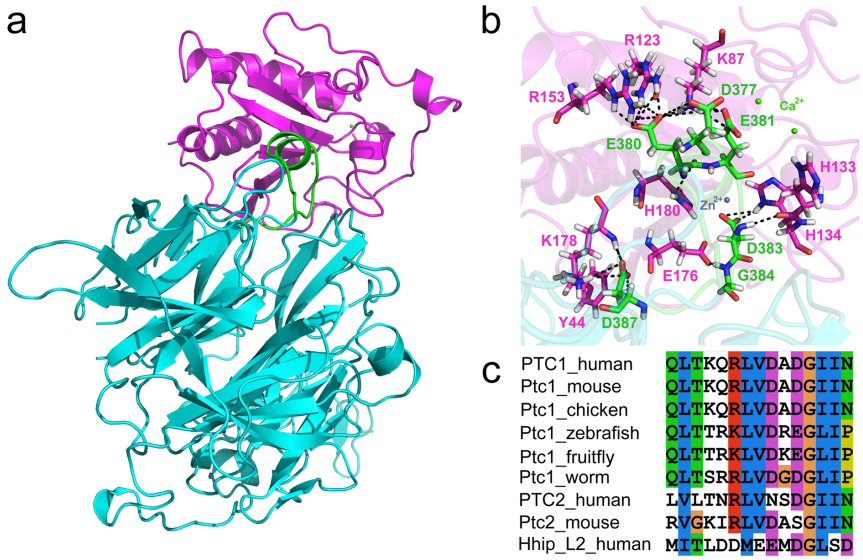


Figure 2.3: **Structure of ShhN-Hhip complex and Hhip L2: Ptch L2 sequence analysis.** a) Structure of ShhN-Hhip complex in cartoon presentation. ShhN is coloured in magenta, Hhip is cyan and Hhip L2 is green (pdb: 3HO5). b) Interactions between Hhip L2 and ShhN. Side chains are shown as sticks and hydrogen bonds are coloured black. c) Alignment of Ptch L2 sequence showing conservation between species and similarity to Hhip L2. Adapted from [209].

complexes. After molecular dynamics (MD) simulations was performed on each complex, two binding sites corresponding to Ptch L2 and Robotnikinin were defined. These binding sites were used to construct two similar pharmacophore models. The ASINEX library ($>$ 200,000 compounds) was filtered using these two models and favourable drug-like properties (Lipinski rules and ADMET). The remaining $\sim$ 500 compounds were docked to Ptch and Robotnikinin binding sites and assessed by consensus scoring. The highest ranking poses were examined for high affinity towards both the Ptch L2 and Robotnikinin binding sites and the two most potent hits were selected (Figure 2.4). When docked to Ptch L2 binding site, both hits bound to the ion binding cleft where they stacked against zinc and calcium ions. When docked to the Robotnikinin binding site, the compounds

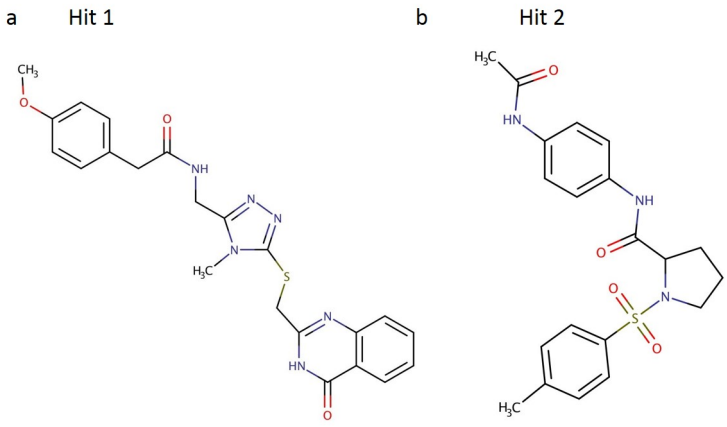


Figure 2.4: **Structure of ShhN hits from *in silico* screen.** Hits 1 (a) and 2 (b) were discovered in a docking screen of ASINEX library. The compounds were screened based on ShhN-Robotnikinin and ShhN-Ptch L2 pharmacophore models [209].

bound closer to the calcium site and no longer interacted with the Hh-bound zinc. However, neither compound, nor their binding modes were ever tested in binding or cellular Hh signalling assays.

The molecular mechanism of ShhN-Robotnikinin interaction remains unknown. Based on structures of ShhN in complex with receptors Cdo and Hhip, and 5E1 antibody and Heparin/C6S, Robotnikinin is expected to bind in one of the two areas: the ion binding cleft or the adjacent receptor-binding 'hotspot'. The first aim of this thesis was to structurally characterise HhN-Robotnikinin complex using molecular docking and X-ray crystallography. Crystallographic analysis of the HhN-Robotnikinin complex will unravel the binding site of the small-molecule inhibitor and the molecular interactions between the two. This will form the basis for structure-guided improvements of next generation inhibitors, as well as contribute to development of novel compounds with greater potency.

2.3 Results

2.3.1 Molecular dynamics studies and docking of Shh

2.3.1.1 Global docking

Robotnikinin was first docked to the 10 available crystal structures of ShhN or DhhN (human or mouse) retrieved from the PDB [210] (Figure 2.5, a) using Gold docking software [211]. All structures contained a zinc ion while calcium was only present in six. Human and mouse ShhN differed in only one amino acid (a conserved exchange from serine to threonine), while human DhhN has 74% sequence identity to human ShhN. All structures were also very similar with the root mean square deviation (RMSD) values of approximately 0.3-0.5 Å for over 150 equivalent C_α atoms when aligned to the first ShhN structure (pdb: 1VHH) (Figure 2.5, b).

Based on ShhN interactions with receptors and other binding partners, Robotnikinin is expected to bind within the ion binding cleft or the receptor-binding 'hotspot' (Figure 2.6, a). However, to avoid introducing bias, we first performed 'blind' docking by sampling the entire protein surface. Docking was conducted in three different conditions: Zn^{2+} and Ca^{2+}, only Zn^{2+}, and no ions. Each run was repeated three times on all 10 structures. The docking results were analysed in Maestro and were categorised per ion condition for every structure separately. In total, 654 docking hits were obtained, of which 474 were in the presence of ions. Overall, more than 90% of hits were docked into the ion-binding groove

a

PDB	Name	Organism	Resolution (Å)	Ions	Complex
1VHH	ShhN	*Mus musculus*	1.70	Zn^{2+}	-
2WFQ	DhhN	*Homo sapiens*	1.85	Zn^{2+}	-
2WFR	DhhN	*Homo sapiens*	1.95	Zn^{2+}, $2Ca^{2+}$	-
2WFX	ShhN	*Mus musculus*	3.20	Zn^{2+}, $2Ca^{2+}$	Hhip
2WG3	DhhN	*Homo sapiens*	2.60	Zn^{2+}	Hhip
2WG4	ShhN	*Mus musculus*	3.15	Zn^{2+}	Hhip
3D1M	ShhN	*Mus musculus*	1.70	Zn^{2+}, $2Ca^{2+}$	Cdo
3HO5	ShhN	*Homo sapiens*	3.01	Zn^{2+}, $2Ca^{2+}$	Hhip
4C4M	ShhN	*Mus musculus*	1.74	Zn^{2+}, $2Ca^{2+}$	chondroitin-4-sulphate
4C4N	ShhN	*Mus musculus*	2.36	Zn^{2+}, $2Ca^{2+}$	heparin

b

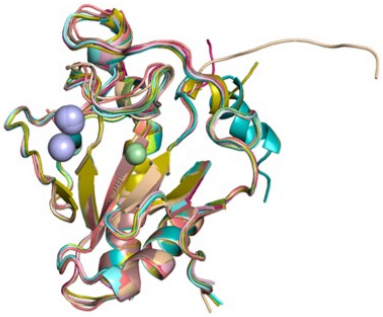


Figure 2.5: **Structures used for docking and their structural superposition.** a) Table containing all ShhN and DhhN structures used in the initial docking experiments, indicating origin, ions, structure resolution and other molecules in complex structures. b) Superpositions of all available HhN structures were carried out in program PyMOL against the first determined ShhN structure (pdb:1VHH); dark yellow: pdb 1VHH, dark green: pdb 2WFQ, dark blue: pdb 2WFR, dark brown: pdb 2WFX (dark brown), light brown: pdb 2WG3, brick red: pdb 2WG4, light pink: pdb 3D1M, cyan: pdb 3HO5, pink: pdb 4C4M and light green: pdb 4C4N. Calcium ions are shown as light blue spheres, while zinc is shown in pale green.

(Figure 2.6, a and b) indicating that this is the most likely binding site. Within the ion binding cleft, two discrete binding sites could be distinguished: the zinc, and the calcium-binding site (Figure 2.6, c and d).

In the absence of ions, ∼85% of the hits still bound to the ion binding groove, implying that the ions did not directly contribute to interactions with Robotnikinin, but maintained structural features of the protein, essential for the interaction. To examine the effect of ions in more detail, we would need to repeat the docking following molecular dynamic simulations, where structural changes would reflect the absence of ions.

In this docking experiment, there was not a single consensus pose that could determine the correct binding mode. Instead, certain poses that were obtained multiple times throughout docking trials were considered as the most reliable

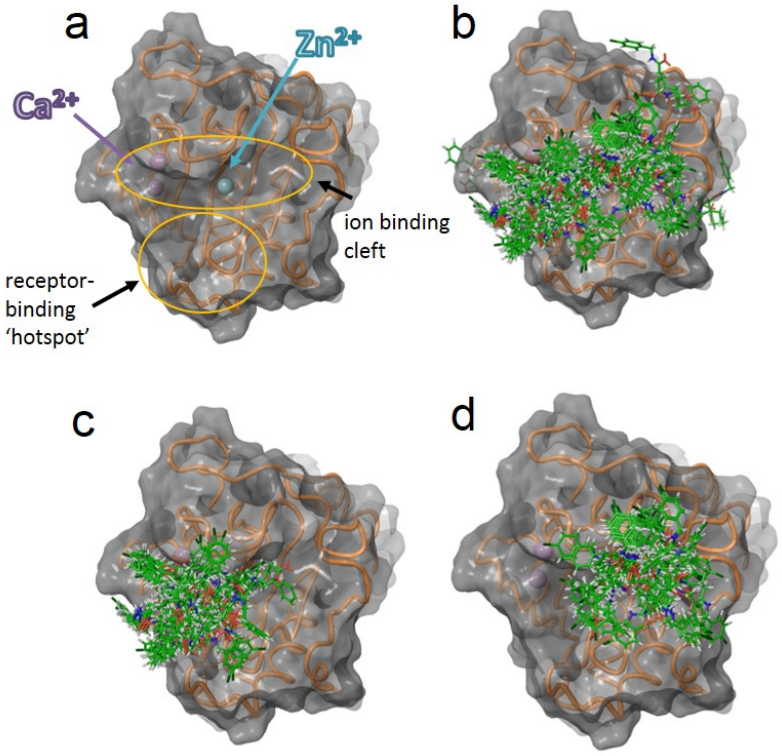


Figure 2.6: **Blind docking results for DhhN crystal structure (pdb: 2WFR).** a) Surface presentation of DhhN indicating the ion binding cleft and the receptor-binding 'hotspot'. The zinc ion is shown in green and calcium ions in purple. b) All the docking results to human Dhh (pdb: 2WFR) showing the majority of hits within the ion binding cleft. Within the ion binding cleft, two discrete sites: zinc-binding site (c) and calcium-binding site (d) could be distinguished. The best 10 and most diverse hits from each group were chosen for a more detailed analysis. Robotnikinin poses are shown as green sticks.

docking results. To examine these in more detail, similar poses for each structure were grouped and the twenty best ones were chosen, based on the frequency, docking score and number of interactions with the protein. Maestros Ligand Interaction Diagram tool was used to identify all the residues that made van der Waals, salt bridge or hydrogen-bonds with Robotnikinin. The best hits were selected separately for zinc and calcium binding sites.

Top poses were examined for the HhN residues they form interactions with. Ten HhN residues were identified as being involved in interaction around the zinc

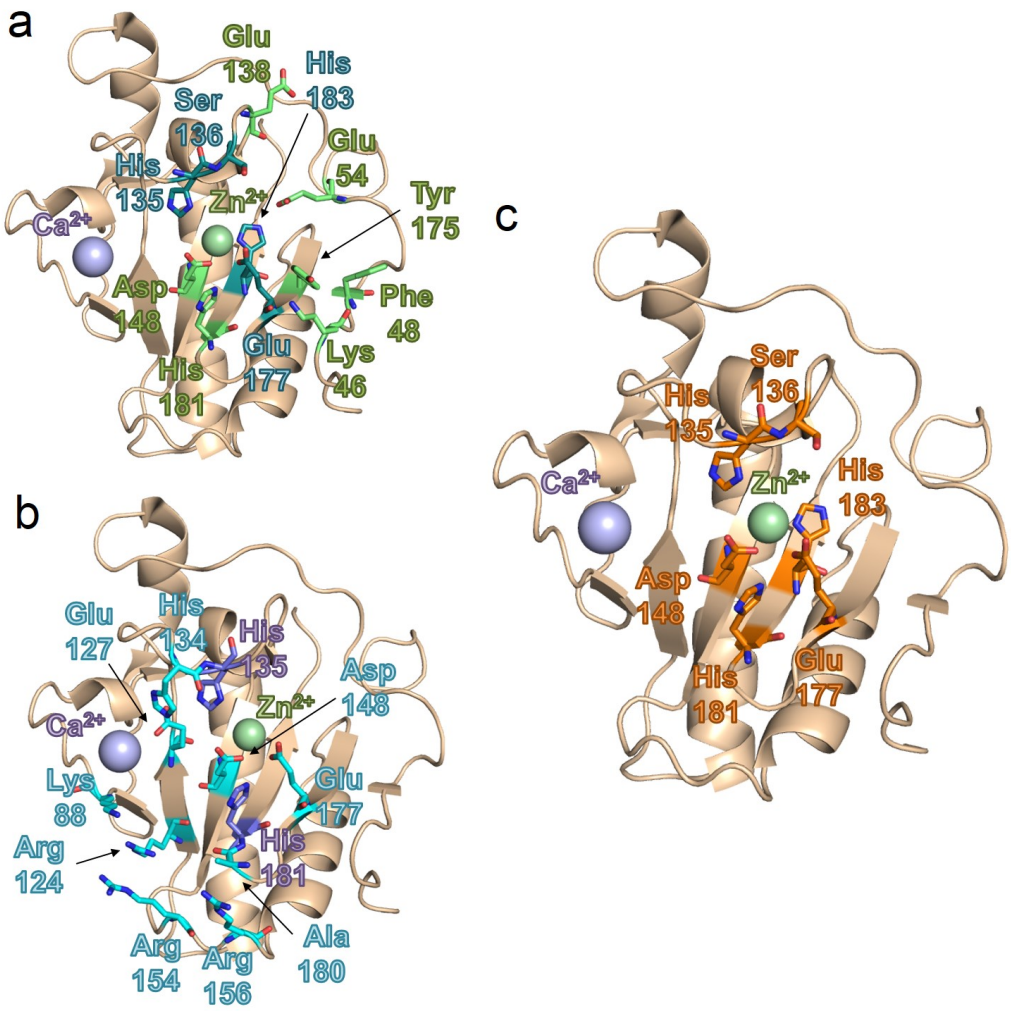


Figure 2.7: **Residue analysis of Robotnikinin docking.** a) The ten best hits from the zinc binding area were chosen based on frequency, docking score and interactions with the protein. All the interacting residues are shown as light green sticks while the most common ones are coloured dark green. b) The same analysis was performed for the calcium binding area. Residues found to interact with Robotnikinin are shown in cyan while the most common residues are shown in purple. c) Overall, residues Asp148, Glu177, His135 and His181 were common for both areas and are highlighted in orange.

binding site: Lys46, Phe48, Glu54, His135, Ser136, Glu138, Asp148, Tyr175, Glu177 and His183, out of which the most frequent residues were His135, Ser136, Glu177 and His183 (Figure 2.7, a). Another 11 residues were identified in the calcium binding site: Lys88, Arg124, Glu127, His134, His135, Asp148, Arg154,

Arg156, Glu177, Asn180 and His181, with His135 and His181 being the most frequent (Figure 2.7, b). An interesting observation is that Asp148, Glu177, His135 and His181 residues are common for both sites (Figure 2.7, c). These residues map to the mid-region of the ion binding cleft and constitute the conserved pseudo-active zinc binding site, homologous to zinc hydrolases. Additional frequently interacting residues Ser136 and His183 also map to the mid region of the groove (Figure 2.7, c).

The blind docking results indicated that Robotnikinin most likely binds within the ion binding cleft area that forms part of the ShhN-receptor binding interface. This is in accordance with the importance of ions for ShhN function and interaction with receptors, GAGs and 5E1 antbody.

The limitations of docking were that a definitive binding pose could not be determined without experimental constraints. To further explore the role of individual residues in binding to Robotnikinin, we ran molecular dynamics simulations of ShhN and repeated the docking to three different areas of the ion-binding groove.

2.3.1.2 Molecular dynamics simulations of ShhN

Molecular dynamics simulations were ran using OPLS and AMBER fields for 100 ns using ShhN crystal structure (pdb: 3N1R) in the presence of Zn^{2+} and Ca^{2+} ions.

I first wanted to examine the behaviour of ions. In the ShhN crystal structure, the zinc ion is coordinated by His141, His183, Asp148 and a water molecule. This remains the case in the energy minimised structure and at the initial stages of the simulation. However, as the simulation progresses, His141 and His183 move $\sim$ 4 Å away from Zn^{2+}, which then becomes coordinated by Asp148, Glu177 and two water molecules instead (Figure 2.8, a). Calcium ions remain coordinated by the same residues as in the crystal structure Glu90, Glu91, Asp96, Thr126, Glu127 throughout the whole simulation (Figure 2.8, a).

The overall dynamics were analysed by looking at the RMSD value of the ShhN backbone in reference to the initial energy-minimised structure. Within the first 30 ns, there was an increase in RMSD from $\sim$1 Å to $\sim$ 4.5 Å indicating some dynamics. However, after 50 ns the RMSD value settled at $\sim$ 3.5 Å (Figure 2.8, a). Although this describes the overall structure, we wanted to obtain a better understanding of which parts of the structure were dynamic. The fluctuation of RMSD per residue was calculated and converted to B-factor values that were mapped onto the ShhN structure (Figure 2.8, b).

This indicated that N- and C-termini contributed to the dynamics of the structure. Furthermore, 131-141 residue loop was found to be dynamic, which

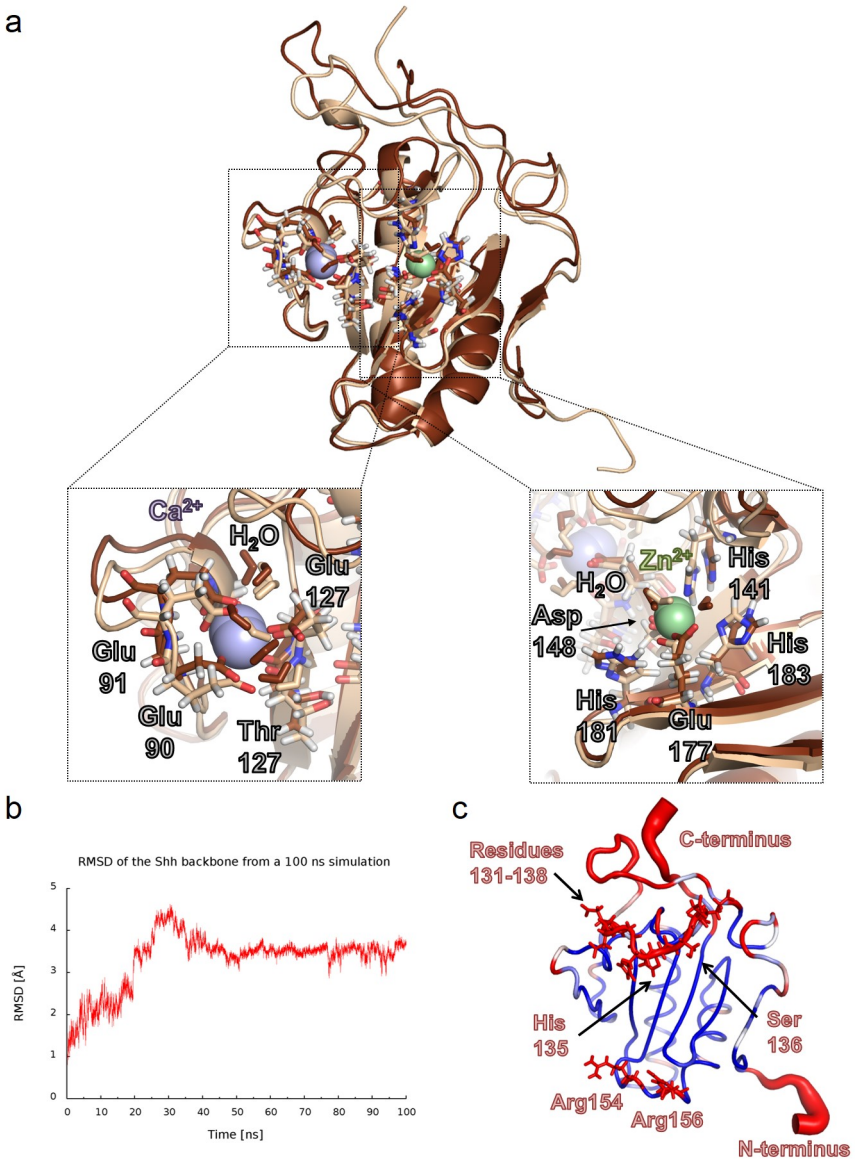


Figure 2.8: **Molecular dynamics simulation of ShhN (pdb: 3N1R).** Molecular dynamic simulation was run for 100 ns in OPLS-AA and AMBER force fields. a) ShhN structures before (wheat) and after (brown) the simulation were superposed. While calcium remains coordinated by the same residues as in the crystal structure, those that coordinate zinc, His141 and His183 move away from the ion. By the end of the simulation zinc is coordinated by Asp148, Glu177 and two water molecules. b) The RMSD of the backbone was plotted against time. Initially, the RMSD increases to $\sim$ 4.5 Å over 30 ns, after which it settles at $\sim$ 3.5 Å. c) The RMSD fluctuation was calculated per residue and converted to B-factors using Gromacs g_rmsf tool. The ShhN structure is shown as a putty on a blue to red spectrum, representing low and high B-factors, respectively. This indicated that the increase in RMSD is largely due dynamic N- and C-termini and the loop consisting of residues 131-141 including His135 and Ser136.

intriguingly included His135 and Ser136 residues identified in docking as important in Robotnikinin binding (Figure 2.8, b). To examine the importance of these residues, three conformations of His135 and Ser136 were isolated and used for docking (Figure 2.9). These include His135 and Ser136, pointing into the ion-binding groove and gradually moving away from it.

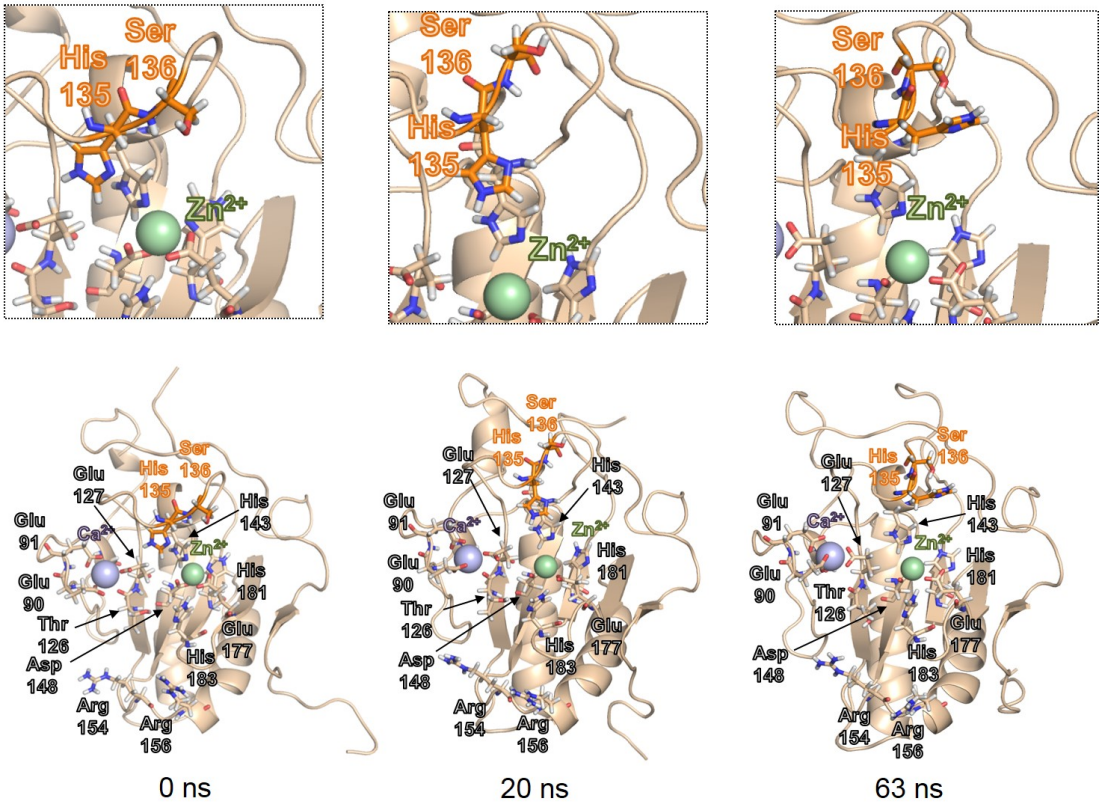


Figure 2.9: **Conformations of His135 and Ser136 residue throughout MD simulation.** The loop containing residues 131-141 was found to be dynamic due to its high RMSF value. Closer inspection of the individual residues from the loop indicated different conformations of His135 and Ser136 with respect to the ion binding cleft. Since His135 and Ser136 were identified as potentially likely to interact with Robotnikinin, three different Shh conformations were chosen to be used for docking.

2.3.1.3 Docking to ShhN structures from molecular dynamics simulations

Robotnikinin was docked to the following areas of the ShhN: zinc binding area (residues 41-55, 132-150, 173-187), calcium binding area (residues 85-96, 123-137, 146-158, 180-182) and mid-region (residues 133-137, 177-183). Each run was repeated 3 times and the results were analysed as described before. The interactions between the top hits and ShhN were analysed using Maestros Ligand Interaction Diagram tool to identify the most frequently interacting residues.

The first ShhN conformation at 0 ns had His135 and Ser136 pointing towards the ion binding cleft, similar to the crystal structure. The two most frequent binding modes spanned the middle of the groove and included the two best hits with most interactions (Figure 2.10). Overall, the most frequent binding residues were Glu90, His134, His135 and Ser136.

The second Shh conformation at 20 ns had His135 and Ser136 pointing slightly out of the groove. Docking resulted in two frequent binding modes at the mid-region of the ion binding groove (3 or more contacts with the protein) and at the receptor-binding 'hotspot' (one contact with the protein) (Figure 2.10). Hits with the most interactions, however, did not belong to either group and formed between 6 to 8 contacts at the mid-region of the ion bidning cleft. The most commonly interacting residues overall included Glu54, Lys88, His134, His135, His141, Trp173 and Tyr175.

The third Shh conformation at 63 ns had His135 and Ser136 pointing com-

pletely away from the ion binding groove. The most frequent binding mode remained the mid-region of the ion binding cleft. Three poses with the most interactions were identified. However, only one also belonged to the most frequent poses (Figure 2.10). Most frequent binding residues for the third ShhN conformation were residues Glu90, His134, His141 and Glu177.

The docking results for the first and third conformation appear to be more consistent in terms of the frequency and interactions of the hits, while the second conformation appears to have some discrepancies. This might indicate that certain conformations of His135 and Ser136 are preferred and that these residues are important for binding of Robotnikinin to ShhN. While this might be true, it also appears that when His135 and Ser136 point away from the ion binding cleft and are not able to interact with Robotnikinin, His134 and His141 are able to substitute these interactions. This is apparent in docking to the third conformation. Overall, the docking results indicate the importance of loop 131-141 rather than individual residues, since the interactions with the loop are common regardless of the ShhN conformation. Taken together, the most common residues remain within the same region as in the initial docking trial, and include Glu90, His134, His135 and His141.

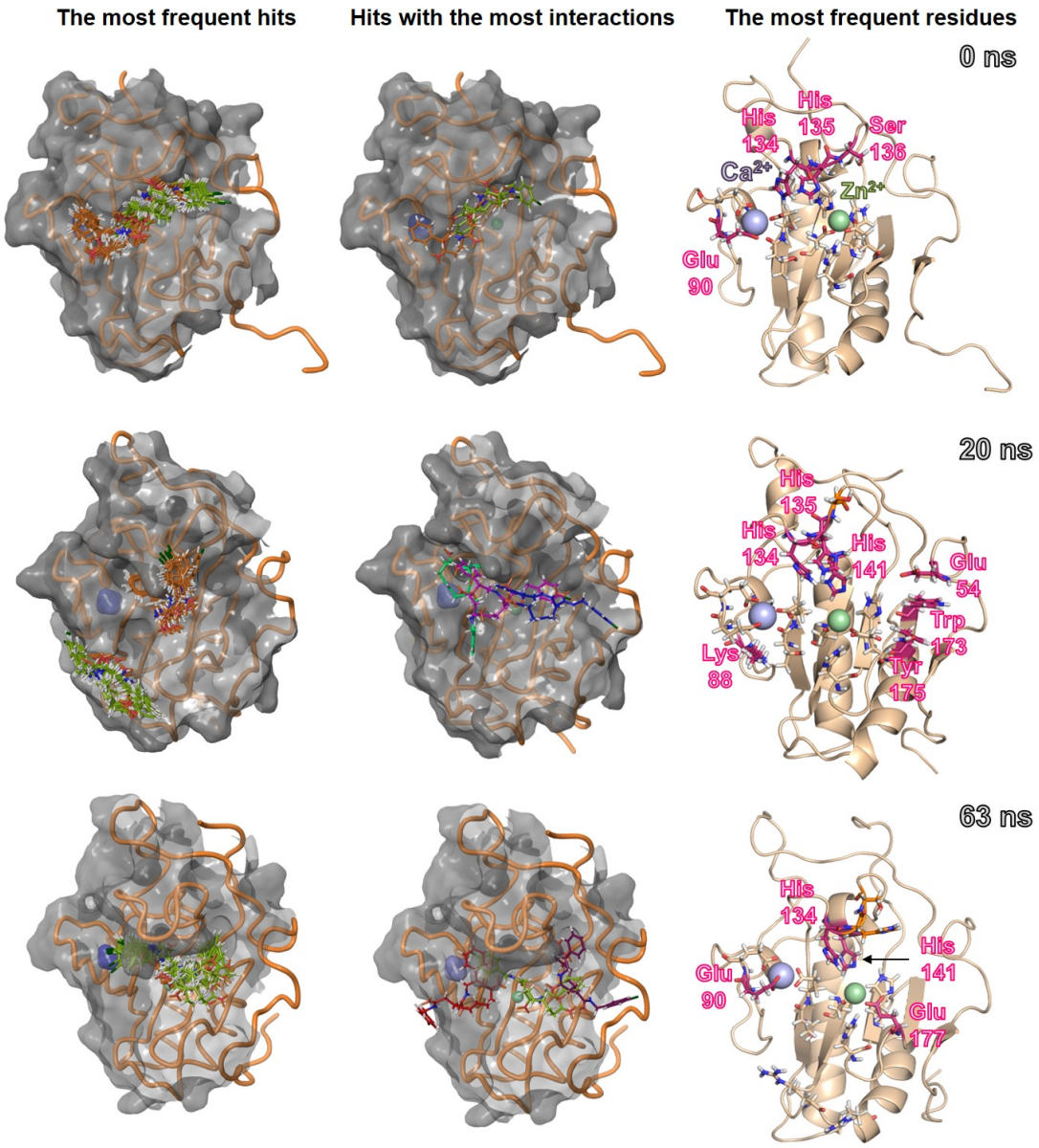


Figure 2.10: **Docking of Robotnikinin to three different ShhN conformations.** Different conformations of ShhN were chosen from MD simulations, depending on the orientation of His 135 and Ser 136 (shown as sticks in orange). The hits with the highest frequency and the highest number of interactions are shown for each ShhN conformation. The most common residues involved in interactions are labelled in pink with corresponding residues shown as sticks. Zinc and calcium are shown as spheres coloured in green and blue, respectively.

2.3.2 Crystallisation of HhN-Robotnikinin complex

2.3.2.1 Expression and purification of ShhN and DhhN

ShhN and DhhN proteins were first purified to be used in crystallisation trials. ShhN and DhhN constructs were previously designed and cloned into the pET22b expression vectors containing a 6xHis-tag at the C-terminus. We expressed two human ShhN constructs, the inactive, N-terminally deleted ShhNΔ39 (residues 39-192), and ShhN24II (residues 24-192) that contained two isoleucine mutations at the N-terminus replacing Cys24 that is palmitoylated *in vivo*. The Shh24II construct was shown to be 8× more active than the non-palmitoylated ShhN in cell assays [137]. We also expressed and purified one human DhhN construct DhhNΔ39 (residues 39-192). These constructs were considered the most suitable for HhN-Robotnikinin complex crystallisation because they readily crystallised and their respective structures have previously been determined in the Siebold group by X-ray crystallography. All constructs were expressed *E. coli* Rosetta (DE3) cultures and were purified using Talon affinity chromatography. The purification steps were analysed on SDS-PAGE gels and all constructs ran at their approximate predicted molecular weight: ShhN24II at $\sim$ 20.5 kDa, ShhNΔ39 at $\sim$ 18.7 kDa and DhhNΔ39 at $\sim$ 18.3 kDa (Figure 2.11, a-c). The eluate of each protein was run on size exclusion chromatography (SEC) and was exchanged into 10 mM Hepes 7.5, 150 mM NaCl buffer (Figure 2.11, a-c). All three proteins appeared as monodisperse peaks and were monomeric, as expected. We obtained high yields for both ShhN constructs, approximately 16 mg and 37

mg for ShhN24II and ShhNΔ39, respectively, and a lower yield of 1.1 mg for DhhNΔ39.

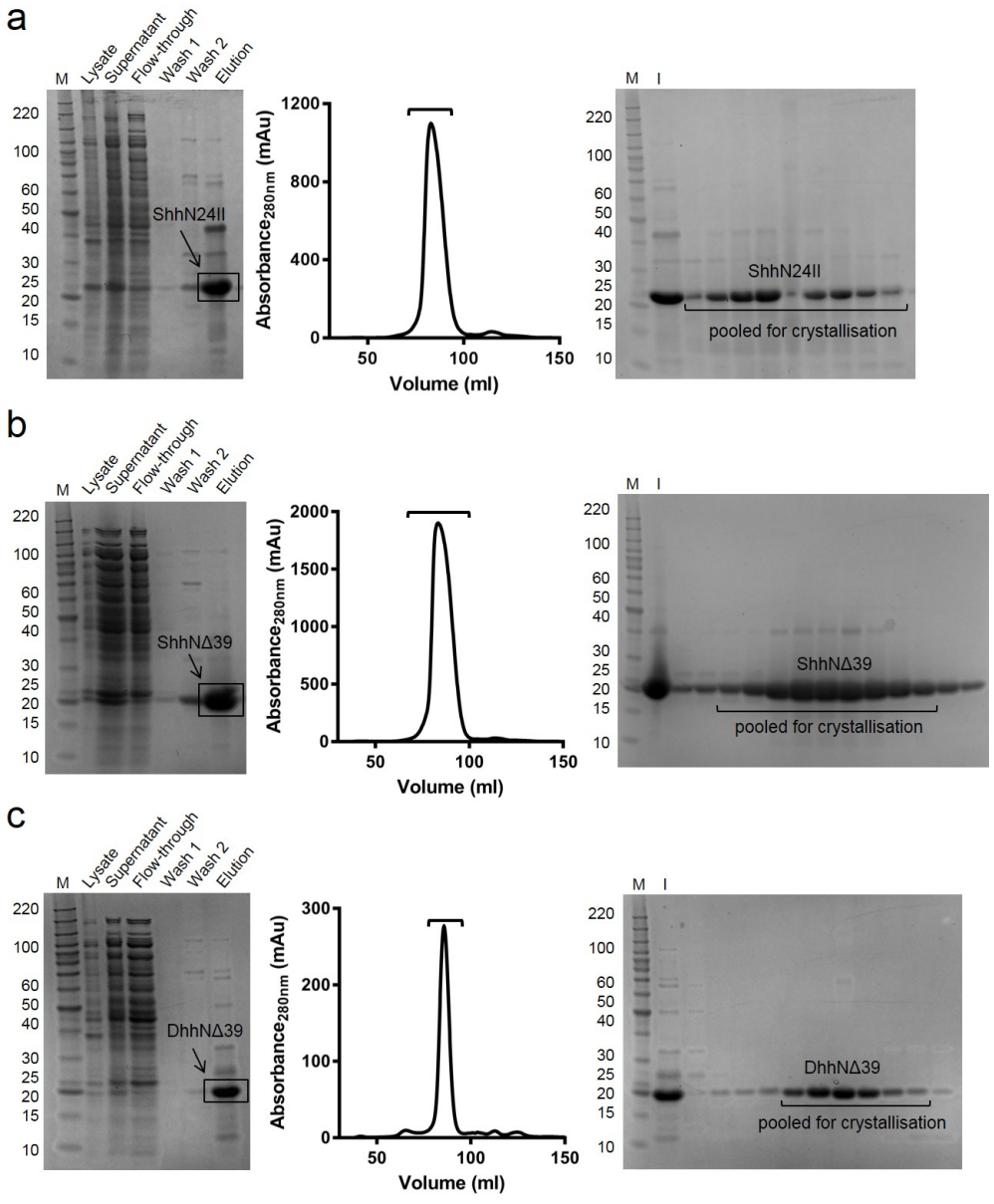


Figure 2.11: **Purification of ShhN and DhhN constructs for complex crystallisation.** Two ShhN and one DhhN constructs were purified by affinity chromatography using TALON beads and were run on SEC in 10 mM Hepes, 150 mM NaCl buffer. We obtained 16 mg and 37 mg of protein for ShhN24II (a) and ShhNΔ39 (b), respectively, while DhhNΔ39 (c) produced a lower yield of 1.1 mg.

2.3.2.2 Crystallisation trials

We attempted to crystallise ShhN and DhhN proteins in complex with Robotnikinin with the aim of identifying the binding site and determining molecular mechanism of their interaction.

Robotnikinin is a poorly soluble compound that was dissolved in DMSO at 10 mM. From previous SPR studies, we knew that it bound ShhN with a low micromolar affinity of $\sim$ 3 µM [197]. Crystallising a protein in complex with an insoluble ligand requires consideration of ligand solubility, protein stability in the presence of DMSO and affinity of a protein-ligand complex.

There is a trade-off between protein stability and its ability to crystallise, and ligand solubility. Often, proteins do not tolerate or crystallise in DMSO concentrations higher than 5% while ligands often precipitate at equivalent solvent concentrations, in particular when dissolved at 100 µM or higher. The affinity of protein-ligand complexes can influence the molar ratio at which the complex is set up for crystallisation. In cases where the K_d of the complex is lower than the concentration of protein, having a stoichiometric molar ratio of the two would suffice. However, in instances where the binding is weaker and the K_d is closer to, or higher than the concentration of the protein, up to a 10-fold excess of ligand is used.

Our approach was to first identify the protein concentration to be used based on previous experience of ShhN and DhhN crystallisation in the Siebold group. We used high protein concentrations between 400-500 µM. Considering that the

affinity of the complex is only 3 µM, we wanted to achieve a protein:ligand ratio where the ligand would be at its highest possible excess. However, this would be done at the expense of high DMSO concentrations, which were tolerated by ShhN crystals in the past but might interfere with the crystallisation process. We also used $CaCl_2$ to introduce Ca^{2+} as we know from previous studies that Ca^{2+} has a role in receptor binding and the same might be the case with small-molecule binding. We attempted three different approaches to co-crystallise the ShhN/DhhN-Robotnikinin complex.

First, I mixed ShhNΔ39 with Robotnikinin at a 1:5 protein:ligand molar ratio for a final ratio of 400uM to 2mM ShhNΔ39:Robotnikinin. Even though this resulted in moderately high DMSO concentration of 20%, Robotnikinin nevertheless precipitated. Despite that, I set up two screens in MRC Swiss-sci plates at three different drop ratios (2:1, 1:2, 1:1). However, none yielded crystals. We decided to change the approach to avoid Robotnikinin precipitation.

Second, I set up crystallisation plates by dispensing protein:reservoir-liquor: ligand at 1:1:1 drop ratios. ShhNΔ39 and ShhN24II were concentrated at $\sim$ 500 µM while Robotnikinin was dispensed from a 5 mM stock. In total 20 crystallisation screens were set up for each construct both in the presence and absence of 3 mM $CaCl_2$. However, no crystals were obtained, most likely due to the high DMSO concentration which interfered with crystallisation.

My third approach tried to overcome high DMSO concentration by dissolving Robotnikinin directly into the protein solution. I expected that, despite low solubility, enough ligand would solubilise to bind protein and achieve high enough

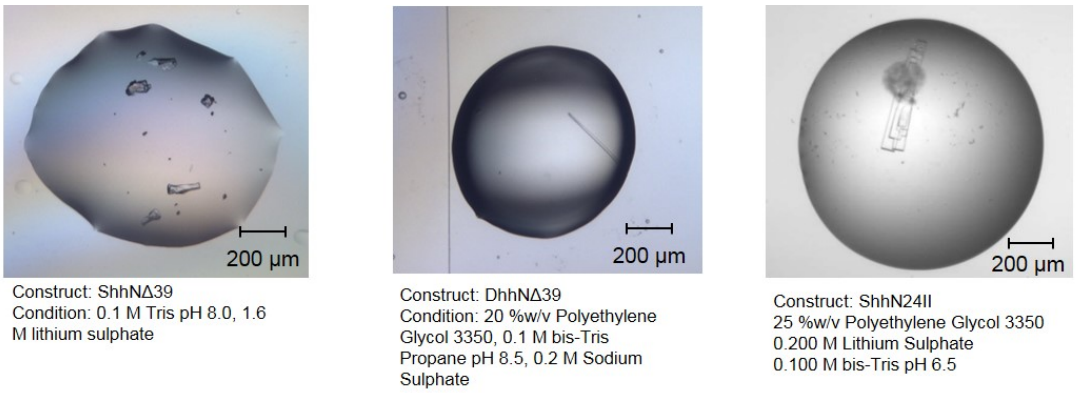


Figure 2.12: **ShhN and DhhN crystals.** ShhNΔ39, ShhN24II and DhhNΔ39 constructs were mixed with 1 mg of Robotnikinin at 1:10, 1:20, and 1:15 molar ratios, respectively, overnight at 4°C. Crystals were obtained for all three constructs and three conditions for which structures were solved are shown.

ligand occupancy. I mixed 400 µM ShhNΔ39 with 1 mg Robotnikinin at 1:10 and 1:20 ratio both in the presence, and absence of $CaCl_2$ allowing us to set up 10 and 11 crystallisation screens, respectively. I repeated the same approach with, ShhN24II, and with DhhNΔ39 at 200 µM at 1:15 molar ratio for 15 screen. Finally, I obtained crystals in multiple conditions and collected complete data sets on three crystals (Figure 2.12). All structures were solved by molecular replacement. However, density of a ligand that could correspond to Robotnikinin was observed in only one structure.

2.3.2.3 ShhN-Robotnikinin structure

We observed a large unaccounted electron density blob, which we could not attribute to terminal peptides or small-molecules from crystallisation condition (Figure 2.13, a). The potential Robotnikinin density blob appeared at the pseudo-active site in close proximity of the zinc-coordinating residues His140, Asp147 and

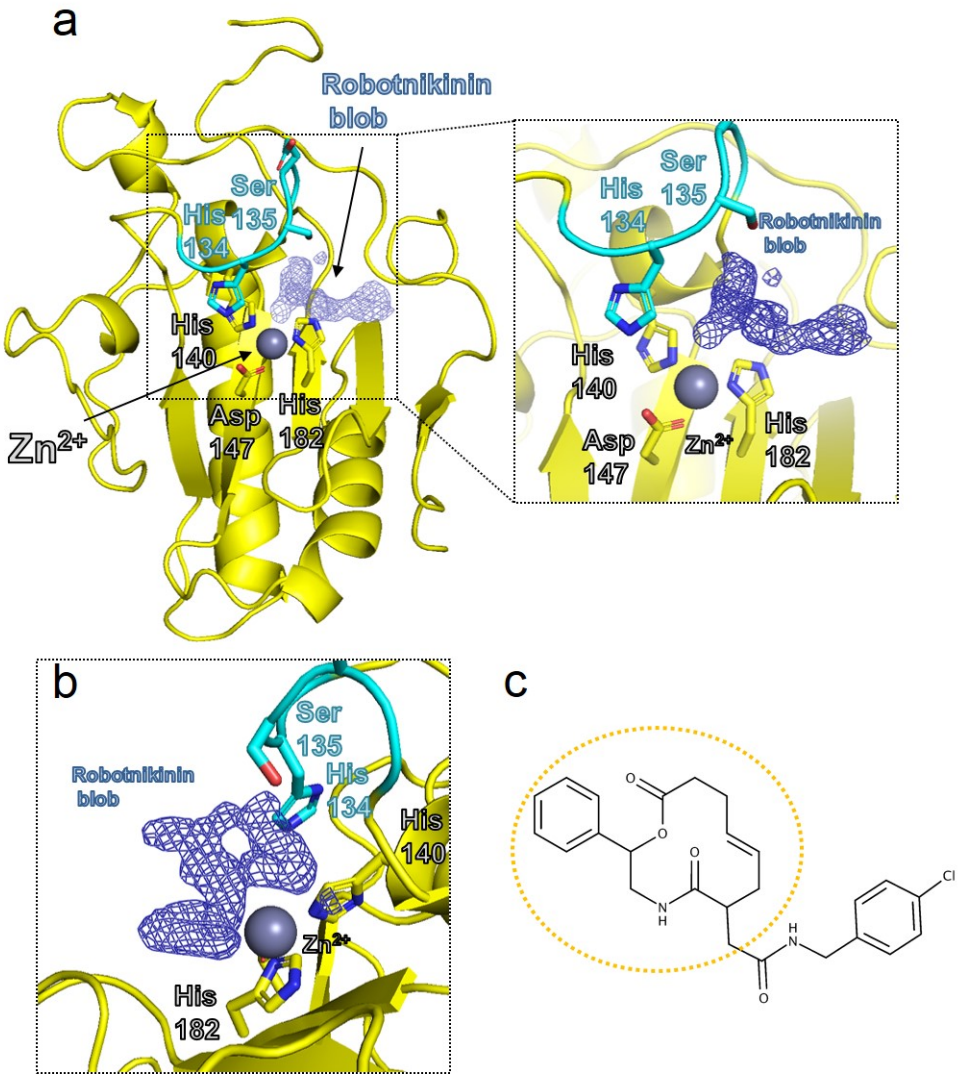


Figure 2.13: **ShhNΔ39 structure with unaccounted electron density blob.** a) ShhNΔ39 structure revealed an unaccounted density that could not be attributed to terminal peptides or small-molecules from the crystallisation condition. The SigmaA-weighted Fo-Fc electron density blob was located at the pseudo-active site in close proximity of zinc coordinating resides His140, Asp147 and His182 (yellow). In close proximity were also residues His134 and Ser135, identified as important for Robotnikinin binding in docking studies. Based on the shape of the density (b) we speculate that a cleavage product containing the macrocyclic portion of Robotnikinin (c, marked in yellow) corresponds to the bound fragment.

His183, and loop resides His134 and Ser135, which were identified in docking as important for Robotnikinin binding. The density did not correspond to the whole of Robotnikinin. However, based on the density shape we suspect that the it cor-

responds to the macrocyclic cleavage product of Robotnikinin (Figure 2.13, b and c). We further suspect that the overnight incubation of Robotnikinin with the protein solution resulted in ligand hydrolysis. Although the structure did not yield a complete HhN-Robotnikinin complex structure, it was in agreement with *in silico* docking results.

Furthermore, on examining the remaining two datasets collected, the difficulty of HhN-Robotnikinin complex crystallisation became apparent (Figure 2.14). The structures of ShhN24II and DhhNΔ39 contained the C-terminus from a neighbouring symmetry molecule inserted into the ion binding cleft (Figure 2.14). Although the crystal packing was different than previously published ShhN structures [128, 130], it appears to be a common occurrence with HhN crystallisation. Not surprisingly, these interactions would prevent a small-molecule from binding within the binding cleft resulting in HhN only crystals. Conversely, binding of a small-molecule to the pseudo-active site might prevent crystal contact formation and, therefore, crystallisation all together. The crystallographic data for three datasets collected is shown in Table 2.1.

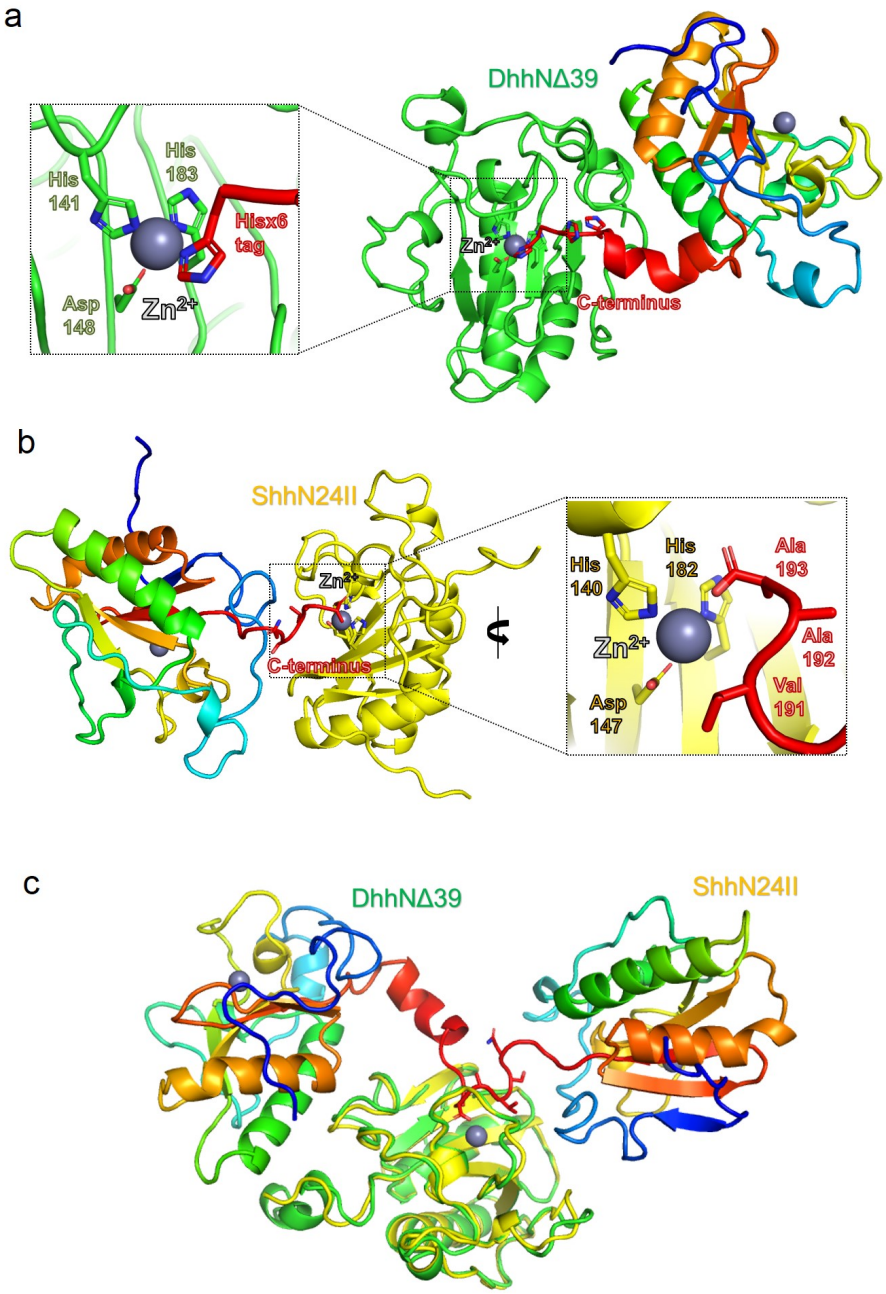


Figure 2.14: **Structures of ShhN24II and DhhNΔ39.** a) Structure of DhhNΔ39 revealed the stacking of the C-terminal His-tag of the neighbouring molecule within the pseudo-active site. The terminal His residue directly coordinates zinc ion in this structure. b) Structure of Shh24II also revealed stacking of the C-terminus from the neighbouring molecule against the pseudo-active site. c) The two structures were superimposed using program PyMOL indicating that despite different crystal packing, the C-termini of neighbouring proteins stack within the ion binding cleft thereby generating crystal contact.

	DhhNΔ39	ShhN24II	ShhNΔ39-Robotnikinin
DATA COLLECTION			
Beamline	DLS-I04-1	DLS-I02	DLS-I04-1
Wavelength [Å]	0.91741	0.97949	0.91741
Resolution [Å]	42.01-1.61 (1.65-1.61)	29.73-1.10 (1.13-1.10)	44.82-1.56 (1.60-1.56)
Space group	P21	C2	P212121
Cell dimensions	a = 42.1 Å b = 40.4 Å c = 83.0 Å α = 90° β = 91.3° γ = 90°	a = 106.6 Å b = 53.7 Å c = 35.5 Å α = 90° β = 90.7° γ = 90°	a = 44.8 Å b = 51.3 Å c = 58.6 Å α = β = γ = 90°
Unique reflections	34741 (2439)	75052 (3399)	19832 (1447)
Completeness [%]	95.7 (94.1)	92.5 (57.1)	99.8 (99.7)
R_{merge} [%]	7.1 (65.4)	5.9 (59.3)	5.7 (61.4)
R_{pim} [%]	4.7 (44.1)	2.9 (41.7	2.7 (28.4)
$CC_{1/2}$	0.997 (0.638)	- -	- -
I/σI	12.1 (1.9)	14.7 (2.3)	16.7 (2.8)
Multiplicity	3.3 (3.1)	6.2 (4.1)	6.4 (6.5)
REFINEMENT			
Resolution range [Å]	42.01-1.61	29.13-1.10	38.63-1.56
Number of reflections (test set)	33021 (1708)	71273 (3779)	19790 (1029)
R_{work}/R_{free} [%]	22.4/26.9	15.8/17.1	16.4/18/6
Number of atoms			
Protein	2638	1262	1289
Ligand	2	26	22
Water	135	233	94
B factors [Å^2]			
Protein	22.5	12.9	24.3
Ligand	16.0	30.9	37.6
Water	23.5	29.0	30.3
r.m.s.d. bonds [Å]	0.019	0.019	0.007
r.m.s.d. angles [deg]	1.928	1.963	0.970
Ramachandran statistics [%]			
Favoured	93.4	95.5	96.2
Disallowed	0.6	0.6	0

Table 2.1: Crystallographic data collection and refinement statistics of DhhNΔ39, ShhN24II and ShhΔ39-Robotnikinin structures. Rfree equals to the R-factor against 5% of the data (test set). Ramachandran statistics were calculated using MolProbity.

2.4 Discussion

The Hh pathway is an important drug target that was correlated with various cancers by ligand-dependent and ligand-independent mechanisms. Two inhibitors targeting Hh GPCR signal transducer Smo are being used as a treatment of BCC, which is caused by ligand-independent activation of the Hh pathway. The discovery of ligand-dependent mechanism of Hh related cancers resulted in alternative Hh targets being explored, including HhN morphogens. Robotnikinin has been identified as the only small molecule that binds directly to ShhN and inhibits the Hh pathway in cell assays. However, the molecular mechanism of ShhN-Robotnikinin interaction is not known.

The first objective of this thesis was to structurally characterise HhN-Robotnikinin interactions and understand the underlying molecular mechanism of small molecule inhibition of HhN. Based on previous structural information on HhN interaction with receptors and the 5E1 antibody, two parts of the HhN receptor interface could be identified: the ion binding cleft and the adjacent receptor-binding 'hotspot' comprised of positively charged resides (Figure 2.15). The Robotnikinin binding site is likely to overlap with one of these two areas thereby occluding the ShhN receptors interface and preventing the interaction between ShhN and the receptors.

To address the question of how HhN and Robotnikinin interact I conducted blind docking studies that indicated that the most probable binding site for Robotnikinin is the ion binding cleft. Furthermore, the residues around the

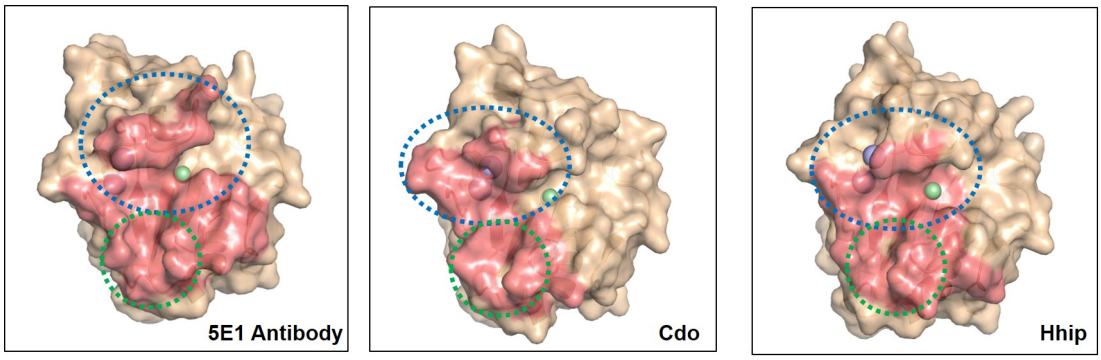


Figure 2.15: **ShhN receptor interfaces.** The interfaces between 5E1-ShhN (pdb: 3MXW), Cdo-ShhN (pdb:3D1M) and Hhip-ShhN (pdb: 3HO5) were mapped onto the crystal structure of ShhN by selecting residues within 4 Å of the receptor in PyMOL [212]. The residues involved in interaction are shown in deep salmon, the ion-binding cleft is circled in blue and the positively charged 'hotspot' in green.

zinc binding site Asp148 and His183, as well as His134, Ser135, Glu177 and His181 were identified as the most frequently interacting residues with docked Robotnikinin. MD simulations indicated that loop containing residues 131-141 is dynamic and docking Robotnikinin to ShhN structures that contained different conformations of His134 and Ser135 residues indicated that residues 131-141 are important for binding.

Next, I attempted to crystallise HhN-Robotnikinin through multiple attempts and various strategies for crystallisation. We finally obtained one dataset that contained an unaccounted electron density that could not be attributed to the protein termini or small-molecules from the crystallisation condition. We speculate that Robotnikinin undergoes cleavage as a result of overnight incubation with ShhN. Based on the shape of the density, we speculate that the macrocyclic portion of Robotnikinin is bound. Whether this binding mode is related to interactions that occur under physiological conditions is not known. However, it

is plausible to assume that Robotnikinin undergoes hydrolysis during a 24 h cell culture treatment at 37°C.

To determine the precise mechanism of interactions between Robotnikinin and HhN further crystallisation trials need to be conducted. For example, lowering the protein concentration for crystallisation would allow the Robotnikinin solution to be mixed at a much lower final concentration and reduce the chance of precipitation. This could also be attempted with solvents other than DMSO such as ethyleneglycol. Another approach would be to dispense Robotnikinin onto crystallisation plates and allow it to dry before proceeding to set up crystallisation screens as usual. Ultimately, construct optimisation that would shorten the termini and include protease cleavage site to remove the His-tag could improve the chances of crystallisation. In the absence of ShhN-structure, mutating residues in the ShhN loop 131-141 to perform binding studies could contribute to our understanding of ShhN-Robotnikinin interactions.

In conclusion, my preliminary data indicates that Robotnikinin binds to the pseudo-active site and not the positively charged receptors-binding 'hotspot'. To determine the precise molecular interactions, further studies, including crystallisation and *in vitro* binding studies, need to be done.

Chapter 3

Targeting Sonic Hedgehog:

fragment-based hit identification

3.1 Abstract

The Hh pathway is an important drug target in BCC and MB, currently targeted by Smo inhibitors. ShhN is a potential alternative drug target for ligand-dependent Hh-related cancers such as pancreatic, lung, liver, breast and many others. To date only one small molecule inhibitor targeting ShhN exists.

The aim of this chapter was to identify fragment hits with which to target ShhN. I conducted a crystallographic fragment screen using a novel ShhN-SOS complex structure where SOS bound to the receptor 'hotspot' and left the ion binding cleft exposed. The screening of over 1,200 unique fragments resulted in 13 novel hits and their respective structures in complex with ShhN. Five of the identified hits were found withing the ion binding cleft: three were found at the zinc-binding site (binding site 1), while two occupied the calcium binding site (binding site 2). Hits within binding site 1 were considered of the highest priority due to the importance of the pseudo-active zinc site in receptor binding. The three hits formed a similar network of electrostatic interactions involving residues Glu176, and His134-Ser136. The results reported here present a basis for the development of novel hits targeting the ShhN ion binding cleft and can potentially be developed into novel inhibitors.

3.2 Introduction

Traditional drug discovery is a multistep process that consists of four phases: i) target identification and validation, ii) lead generation and optimisation, iii) pre-clinical studies, and iv) clinical trials. The process is often lengthy, taking over 13 years from hit generation to market launch, and expensive, adding up to 1.8 billion dollars in costs for the launch of a single drug candidate [213]. The attrition rates are high and novel drug discovery methods are constantly being innovated to generate better drug candidates more efficiently, and in a cost-effective manner. Interestingly, only 8% of candidates that enter clinical trials will successfully reach the launch meaning that for every successful candidate 12.4 candidates must enter clinical trials [213]. The main causes of attrition appear to be low numbers of candidates, as well as reduced efficacy and safety of compounds [213].

The high cost and attrition rate in clinical trials could, in part, be reduced by generating a higher number of efficient leads through a more effective lead-discovery process. Lead molecules exhibit good potency in biological assays that can validate the targeted inhibition mechanism. Leads further need to adhere to a range of chemical, physical and biological properties that make them good drug candidates. For example, they should have moderate lipophilicity, low toxicity and high stability in plasma.

There are a number of different approaches to generating leads:

i) Optimising an existing lead to produce a so called 'best-in-class' compound

rather than novel molecular entity. This occurs by incremental optimisation of existing synthetic or natural compounds for better potency.

ii) High-throughput screening (HTS) is the dominant method for lead discovery today. In this approach hundreds of thousands of compounds are screened at a time in *in vitro* bioassay screens with the aim of selecting the most potent ones. The screened compounds are normally 300-500 Da in size and have affinities in µM-nM range. Despite high affinities, large compounds include functional groups that negatively affect binding. This approach is resource intensive, and high attrition rates in hit-to-lead generation require multiple iterations of chemical synthesis.

iii) Fragment-based lead discovery (FBLD) is a more recent method that has emerged over the past 20 years. Fragments are smaller building blocks of compounds, which are 150-300 Da in size with fewer than 20 heavy (non-hydrogen) atoms. Fragment libraries contain tens of thousands of fragments and in practice, only 500-5,000 fragments are screened in biophysical assays. Due to their small size and fewer interactions, the binding affinities are lower than HTS (mM-µM range) and, therefore, specialist techniques like NMR, SPR, X-ray crystallography, ITC or mass-spectrometry are needed to detect binding.

Fragment screening has been a particularly efficient method for discovering hits for more difficult targets such as protein-protein interactions. FBLD results in higher hit rate with less false positives than HTS [214]. Due to their small size, positive hits tend to be more efficient binders with more interactions per atom. In contrast to HTS, a much smaller number of fragments is required to

HTS	FBLD
MW range 300-500	MW range 150-300
Affinity μM-nM, focus on potency	Affinity mM-μM, focus on efficiency
Hits identified in bioassay screens	Hits identified in biophysical assays
Lead generation based on SAR	Lead generation based on structural data and SAR
Resource intensive and requires high-throughput chemistry	Up to 100 follow-up compounds required for lead generation
Compounds: 'rule of five'	**Fragments: 'rule of three'**
MW < 500 Da	MW < 300 Da
cLogP = 5	cLogP = 3
HBD ≤ 5, HBA ≤ 10	HBD, HBA ≤ 3
	Polar surface area = 60 Å^2
	Number of rotable bonds ≤ 3

Table 3.1: Summary of High-throughput screening (HTS) and Fragment-based lead discovery (FBLD).

sample the chemical space in comparison to larger compounds. This is because higher molecular weight compounds will inherently have more chances of forming mismatches with the target protein. Even though it is difficult to accurately quantify, it has been estimated that approximately 1,000 fragments can sample a chemical space more effectively that 1 million compounds [215].

The 'goodness' of a fragment is measured differently than that of compounds. In a landmark paper, Lipinski described physicochemical properties of the majority of successful drugs [216, 217]. These properties are also known as the rule of five and signify that the drug should: i) be less than 500 Da in size, ii) have lipophilicity or clogP less than 5, iii) have fewer than 5 hydrogen bond donors (HBD), and iv) have fewer than 10 hydrogen bond acceptors (HBA). In FBLD, these rules still apply to the final drug candidate. However, due to the different nature of fragments an adapted 'rule of three' has been proposed as a guidance

for lead generation. The rules stipulate that the fragments should: i) be less than 300 Da in size, ii) have clogP less than 3, iii) have fewer than 3 HBA and HBD, iv) have a polar surface of 60 $Å^2$ and v) have fewer than 3 rotable bonds [218].

High lipophilicity and molecular weight have an adverse effect on absorption, distribution, metabolism and secretion (ADMET) [219]. Therefore, the advantage of fragment screening is the ability to start from fragments that obey the 'rule of three' and have the control during lead generation to develop drug candidates that obey Lipinski's rules.

There are three approaches to generating leads following a fragment screen: i) fragment evolution: where fragment development is guided by protein structure to increase the number of interactions between the ligand and the protein, ii) fragment linking: where fragments that bind to separate binding sites but are in close enough proximity can be linked and iii) fragment self-assembly: where reactive fragments that are capable of linking in the presence of an additional template compound [214].

In contrast to HTS where the primary metric is potency, fragment leads are guided by efficiency in addition to affinity. Ligand efficiency is defined as:

$$LE = pIC_{50}/N \tag{3.1}$$

where pIC_{50} is log concentration of an inhibitor required for 50% inhibition of an enzyme, while N is the number of non-hydrogen atoms (heavy atoms). Once affinity and efficiency are achieved, leads are further optimised for favourable

physio-chemical and biological properties that make them more drug-like.

Crystallographic fragment-based screening is one of the most powerful methods for hit identification as it supports a range of affinities (nM-mM) and it provides an unrivalled amount of information with respect to ligand binding. The initial screen is done by medium- or high-throughput crystal soaking, followed by hit validation using SPR, ITC, NMR or bioactivity assay.

Hits are further validated in structure-activity relationship (SAR) studies, which are designed to identify chemical groups that contribute positively or negative to binding, and inform the chemists' strategy for lead generation. Throughout this process, individual aspects of hits are modified and the new compounds are assessed in biophysical or bioactivity assays of choice. The binding mode of the most promising ones are determined by X-ray crystallography. This is an iterative process that requires 20-100 follow-ups to be screened to obtain quality leads, which is significantly less than HTS [215].

The challenges of this approach include the need for advanced facilities that allow for medium- or high-throughput crystal soaking, harvesting and data collection. The production of large quantities of protein, reproducible crystallisation and high-resolution crystal diffraction are essential but can prove to be difficult for some targets. Crystal robustness is also needed as fragment libraries contain DMSO, which in some cases can affect crystal quality or even bind the target binding site [220]. However, other solvents like ethylene glycol or water can be used. In addition, data analysis can present a challenge not only due to sheer volume of data but also the difficulty of detecting low occupancy ligands, which

is a common issue with the procedure.

In this chapter, the focus is on targeting ShhN using hit generation by crystallographic fragment-based screening. High-quality crystals of ShhN that diffracted to 1.2 Å and had the ion binding groove exposed were obtained and optimised. Over 1,200 unique fragments were screened by soaking resulting in 13 hits over eight binding sites, five of which were localised within the ion binding groove. Here, ShhN-FRG complex structures are described and priorities for follow-up steps are proposed.

3.3 Results

3.3.1 ShhN-SOS complex structure

ShhN24II construct was purified as described in Chapter 2 and crystallised in the presence of 10 mM of the glycosaminoglycan mimic sucrose octasulphate (SOS) and 3 mM $CaCl_2$. Crystals appeared in multiple conditions after 2-3 days and crystallographic dataset was collected at Diamond Light Source (DLS) synchrotron, beamline I04. A novel high-resolution structure of ShhN-SOS was solved at 1.2 Å from crystals in 0.2 M ammonium sulphate, 0.1 M sodium cacodylate, pH 6.5, 30% w/v polyethylene glycol 8000 condition (Figure 3.1).

The overall fold and structure of ShhN is similar to previously published ShhN structures. When aligned to the first ShhN structure solved by X-ray crystallography (pdb: 1VHH) over 151 C_α atoms, the complex structure resulted in a RMSD of only 0.27 Å (Figure 3.1, a). High resolution allowed us to identify zinc, calcium and SOS binding sites (Figure 3.1, b-d). Zinc is coordinated by residues His140, His182 and Asp147, and a water molecule, while two calcium ions are coordinated by Glu89, 90, 126 and Asp95, 129, 131 as described before.

The SOS molecule bound beneath the ion binding cleft within the area that comprises ShhN-receptor interface (Figure 3.1, a). The binding site contained a number of positively charged side chains that formed the basis of the interaction with the negatively charged sulphate groups of SOS (Figure 3.2, a). The main points of contact are electrostatic interactions between sulphate groups and side chains of Lys87, Arg123, Arg153, Arg155 and His180 residues (Figure 3.2, a).

These interactions are similar to those described in previously published ShhN-heparin and ShhN-C6 complexes (Figure 3.2, d) [99]. In contrast to ShhN-Heparin and similarly to ShhN-C6, ShhN-SOS is monomeric and SOS does not mediate multimerisation. Even though the overall structure is similar to the apo ShhN structure, residues Lys87, Arg153 and Arg155 undergo conformational changes to be able to accommodate and interact with SOS in the complex structure

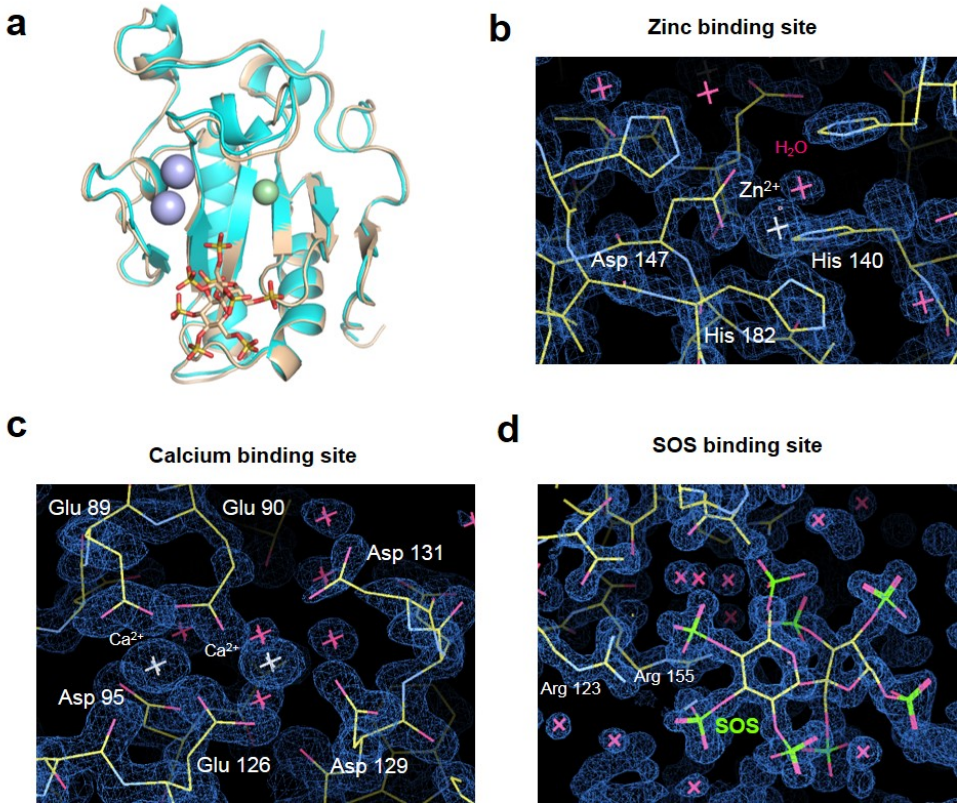


Figure 3.1: **ShhN-SOS complex structure.** a) Cartoon representation of the ShhN-SOS complex shown in wheat and previously published ShhN apo structure (pdb: 1VHH [128]) in cyan. Zinc is depicted as a green sphere and calcium ions as purple spheres, while SOS in stick representation. The two structures are similar and have a RMSD of 0.27 Å when aligned over 151 C_α atoms. In the complex structure, SOS binds below the ion binding cleft and occupies part of the positively charged receptor 'hotspot'. b-d) SigmaA-weighted 2Fo-Fc electron density map after the final round of refinement of the ShhN-SOS complex illustrating well resolved zinc, calcium and SOS binding sites. The map is contoured at 1.0σ.

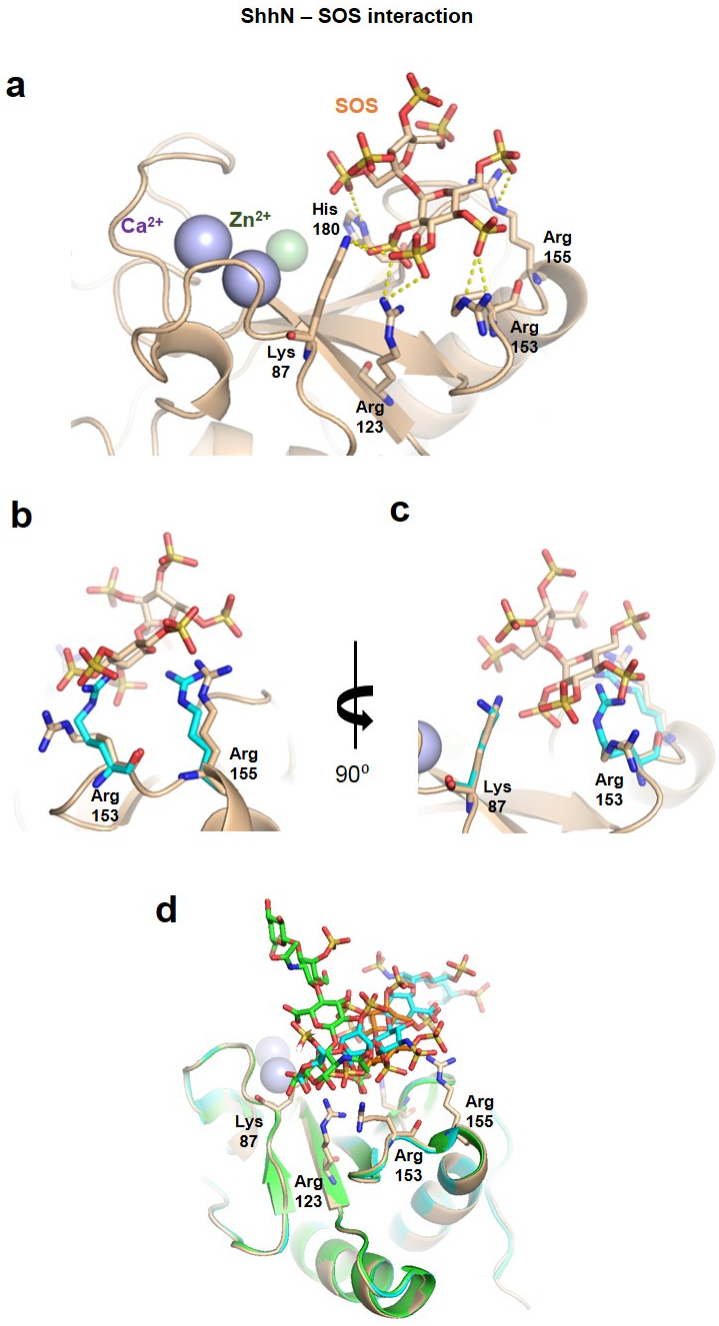


Figure 3.2: **SOS binds a positively charged receptor 'hotspot'.** a) SOS interacts with residues Lys87, Arg123, Arg153, Arg155 and His180 via electrostatic interactions. Residues are shown as sticks and interactions as yellow dashed lines. b) and c) Residues Arg153 and 155, and Lys87 undergo a conformational change upon the ShhN-SOS complex formation. ShhN residues and SOS are shown as sticks, the ShhN-SOS structure is coloured in wheat and apo structure (pd: 1VHH) in cyan. d) Superposition of ShhN-SOS complex (wheat), ShhN-heparin (cyan, pdb: 4C4N), and ShhN-C6S (green, pdb: 4C4M).

(Figure 3.2, b-c).

The unique aspect of this structure is the crystal form where the ion cleft is exposed to the solvent (Figure 3.3). In every previous structure, the ion cleft was occupied with the His-tagged C-terminal peptide of the neighbouring molecule making it unsuitable for soaking with small-molecule ligands. This structure enabled us to perform crystallographic fragment screen against ShhN by soaking fragment libraries.

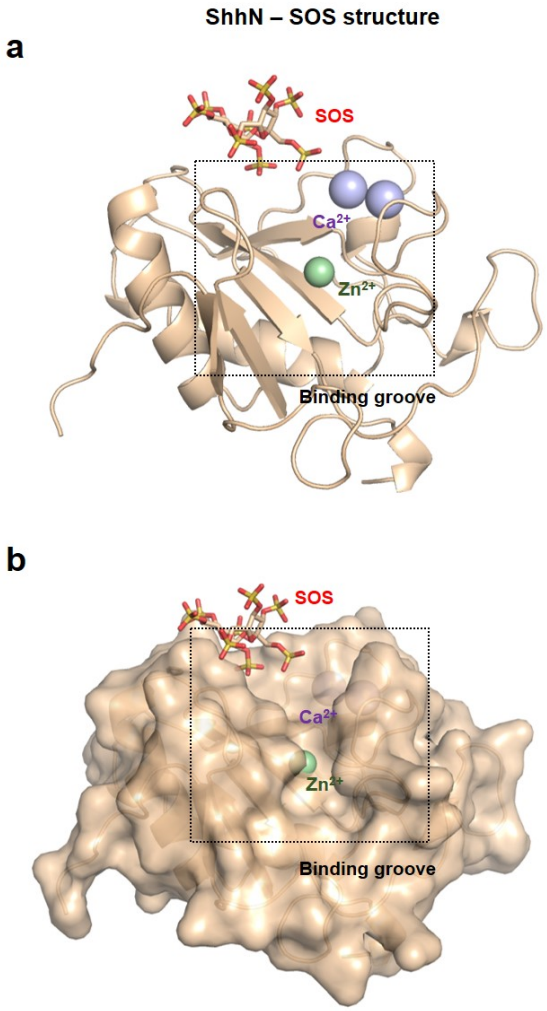


Figure 3.3: **The ShhN-SOS binding groove.** Previously crystallised ShhN structures had the binding groove occupied with the N-terminal peptide from neighbouring molecule in the crystal lattice. a) In the ShhN-SOS structure, the ion binding cleft is exposed and suitable for soaking experiments. b) ShhN is shown in surface representation emphasising the ion binding groove.

3.3.2 Crystallographic fragment screening of commercial libraries against ShhN

Crystallographic fragment screening was performed at the DLS XChem laboratory and data were collected at DLS beamline I04-1. Crystallisation was optimised in MRC Swiss-Sci plates to produce more than 100 crystals per plate. We optimised crystals in 2 conditions: 1) 48% 0.09 M NPS* , 0.1 M imidazole and MES pH 6.5, 30% v/v PEG 500* MME; 10% w/v PEG 20000 (NPS: 0.3 M sodium nitrate, 0.3 M sodium phosphate dibasic, 0.3 M ammonium sulphate), and 2) PEG 8000 30% w/v, 0.2 M sodium acetate trihydrate , 0.1 M sodium cacodylate pH 6.5. The fragment screening process is described in Figure 3.4.

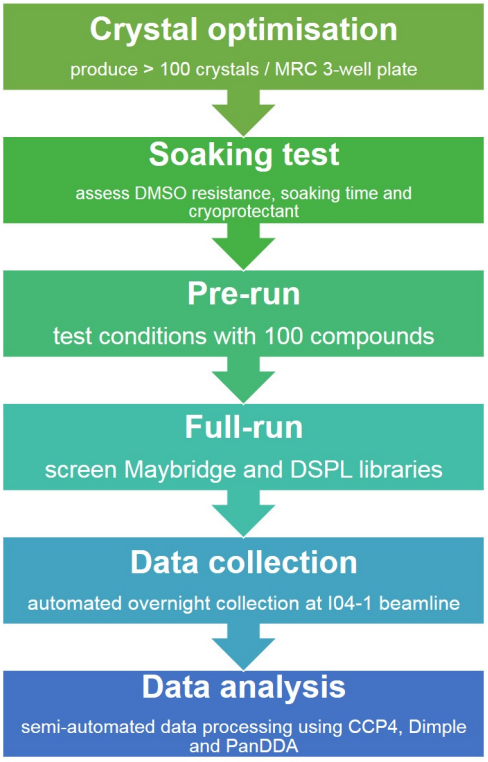


Figure 3.4: **Crystallographic fragment screen experimental overview.** The schematic illustrates steps taken from crystal optimisation to full library screening.

For fragment screening, ShhN crystals were first tested for DMSO resistance and suitable cryoprotectant. DMSO was tested in a concentration range of 0-40% for 1h, 3h and overnight with and without 20% ethylene glycol (EG) as cryoprotectant. The most suitable conditions for soaking were found to be 30%

Run	1	2	3	4	5	6	Total
Condition	Sulphate	Sulphate	Cacodylate	Cacodylate	Cacodylate	Sulphate	
Library	Maybridge	DSPL, Maybridge, AR	Maybridge	DSPL	AR	OxXChem	
Solvent	DMSO	DMSO	DMSO	Ethylene glycol	Ethylene glycol	DMSO	
Number of fragments screened	100	480	540	255	2x51	50	~ 1,200 (unique)
Datasets collected	78	440	490	208	100	47	1,363
Optimisation	-	Crystallisation condition	Solvent	-	Soaking time	-	
Hits	2	7	2	-	-	2	13

Table 3.2: Summary of all fragment screen runs. The table summarises 6 runs during which we screened over 1,200 unique fragments and have collected 1,363 datasets. Four different libraries were screened and various factors were optimised as indicated. This resulted in 13 identified hits.

DMSO, 1-2 h soaking time and 20 % EG as cryoprotectant. A pre-run of 100 test fragments was conducted to test the resistance of crystals. The test indicated that crystals diffracted highly and more than 1,200 unique fragments from 4 different libraries were subsequently screened as part of the main screen.

For the full-run two large libraries, DSPL (Diamond-SGC poised library) [221] and Maybridge 1000, were used and accounted for $\sim$ 1,100 fragments screened. Additionally, two libraries of 50 compounds each were provided by collaborators: AR, compounds with zinc-interacting groups, and OxXChem, 3D fragments sourced from ChemBioHub Oxford. All fragments were supplied in DMSO or ethylene glycol at concentrations 200-500 mM, apart from AR, which was at 10 mM.

In total, over 1,350 datasets were collected over 6 runs. Following each at-

tempt the data were analysed by examining the electron density maps and looking for additional density in Fo-Fc difference maps that could correspond to bound fragments. Initially, after examining data collected in runs 1 and 2, no fragments were identified.

Possible reasons for this could have included interference of molecules from crystallisation condition or DMSO with fragment binding. In particular, a sulphate/phosphate group originating from the crystallisation buffer could be observed coordinating zinc ion in addition to residues His140, Asp 147 and His182. Although we did not observe DMSO molecules close to the ion binding groove, it is common for DMSO to interfere with protein chemistry and affect protein-compound complex formation [222, 223]. Over the course of the next five runs the second crystallisation condition was used and the fragment solvent was exchanged for ethylene glycol. The soaking time was also prolonged to 3h and overnight, as soaking for longer periods can increase the likelihood of complex formation provided it does not affect crystal quality [223]. After observing Fo-Fc difference maps, we could not identify bound fragments.

All datasets were re-examined using Pan-dataset density analysis (PanDDA) software resulting in identification of 40 potential ligands [224]. PanDDA is software developed at the DLS XChem group to address the issue of low occupancy fragments, a common problem in crystallographic fragment screens. This method uses 40-200 electron density maps to calculate an average map. The map is then subtracted from each individual dataset (modified by background density correction factor) to produce an event map. This enhanced signal to noise ratio

event map highlights low-occupancy ligands that were previously unobservable in traditional 2Fo-Fc electron density maps.

Out of 1,363 datasets, PanDDa identified 202 events as significant. The majority, however, were the result of flexibility around SOS, the zinc region and the N-terminus. About 40 events were densities of potential ligands and this number was reduced to 13 after further examination. The overall hit rate was approximately 1%, significantly lower than the average of 10% for screens done at the XChem facility. Hits were identified based on having acceptable PanDDA scores, PanDDa density map and low crystal contact contribution (Figure 3.5).

In total, 13 hits were found across 8 binding sites, and 2 binding sites were within the ion binding cleft of ShhN, the previously described region for ShhN-receptor interactions [225]. Three hits were found adjacent to zinc and two hits adjacent to calcium ions. Another binding site with 3 hits was found behind the calcium binding site. We prioritised sites 1 and 2 due to the importance of the ion binding cleft in receptor binding, and site 3 due to higher hit rate and site novelty. All structures were very similar to the original ShhN-SOS structure (RMSD= 0.175 Å). They all contain a zinc ion coordinated by His140, Asp147, His182, and a sulphate/phosphate group, while the two calcium ions were coordinated as described before. All structures are presented along with PanDDa maps and relevant scores. The scoring methods are described in more detail in the Methods section. However, as a general guide real-space correlation coefficient (RSCC) should be > 0.7, real-space Z-difference score (RSZD) < 3, RMSD < 1 Å and B-factor ratio $\sim$ 1 (acceptable up to 1.5). Crystallographic tables (Table 3.3,

Table 3.4, Table 3.5 and Table 3.6) are provided in the final subsection 3.3.7 of the results chapter.

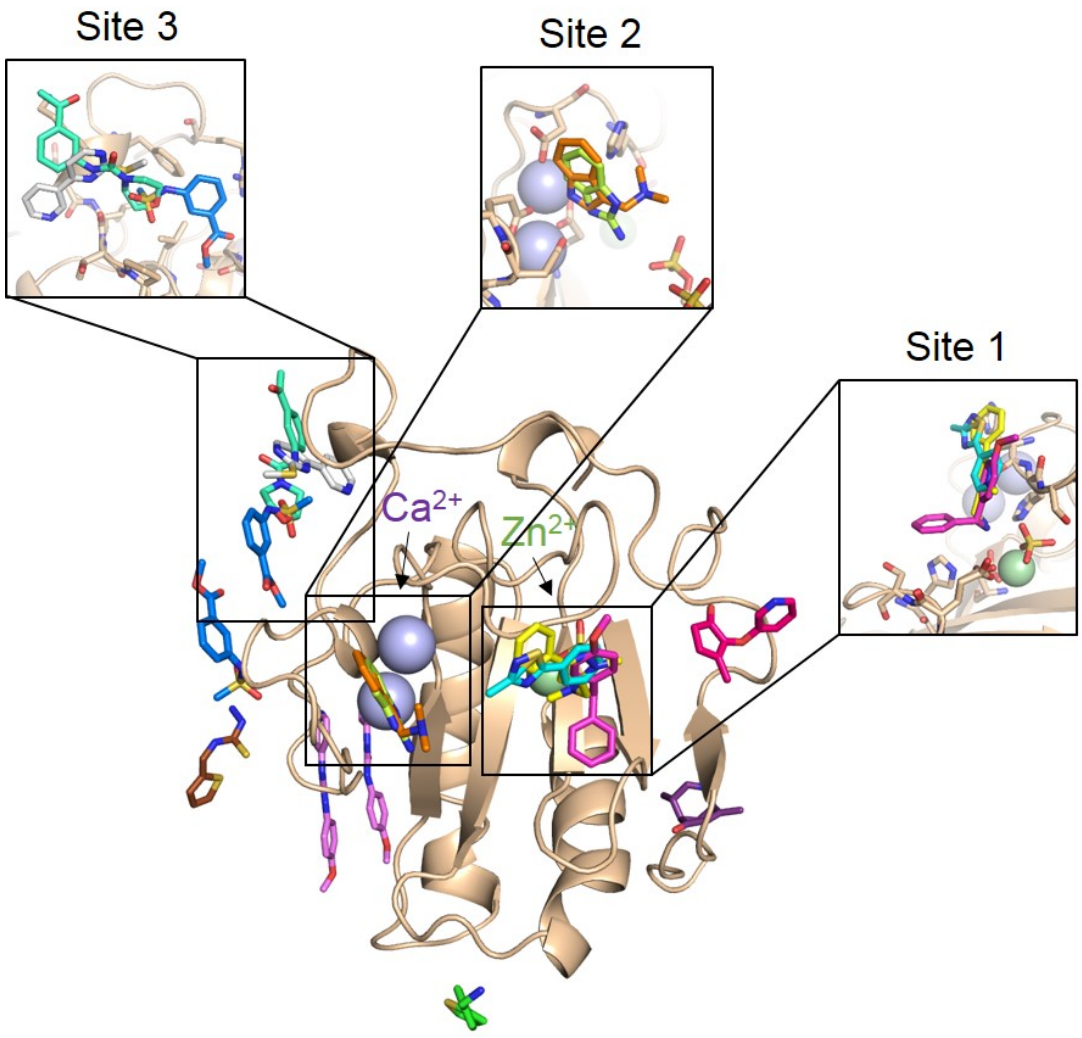


Figure 3.5: **Overview of hits identified from the fragment screen mapped onto the the ShhN structure.** In total, 13 hits were identified across 8 binding sites. Binding sites 1 and 2 are within the ion binding cleft and comprise 5 hits. The two sites together with binding site 3 were priorities for further follow-up studies. The remaining hits bind at various sites on the surface of Shh. SOS is present in all structures but was omitted for clarity.

3.3.3 Binding site 1

Three hits were identified within the zinc binding site: FRG 155, FRG 17 and FRG 458.

The ShhN-FRG 155 structure was solved at 1.1 Å resolution with a ligand occupancy of 0.35 (Table 3.3 and Figure 3.6, a). Fragment 155 is a benzenemethanamine derivative and is approximately 213 Da in size. The amine group of FRG 155 forms electrostatic interactions with residue Glu176 and sulphate/phosphate group (Figure 3.6, a). The amine group is located $\sim$ 4.6 Å from the zinc ion resulting in no direct interaction. FRG 155 stacks against the peptide bond between His134 and Ser135 (distance $\sim$ 3.8 Å) (Figure 3.6, a). Interestingly, these two residues appear to be dynamic and potentially involved in the interaction with Robotnikinin as shown in the MD and docking studies, previously discussed in Chapter 2.

Even though the ligand density was only partially visible in 2mFo-DFc map, the fragment density was clear in PanDDa maps (Figure 3.6, b-e). Following refinement, the fragment remained in place with a RMSD of 0.03 Å between the initial and refined models. All PanDDa scores were within range (Figure 3.6, e).

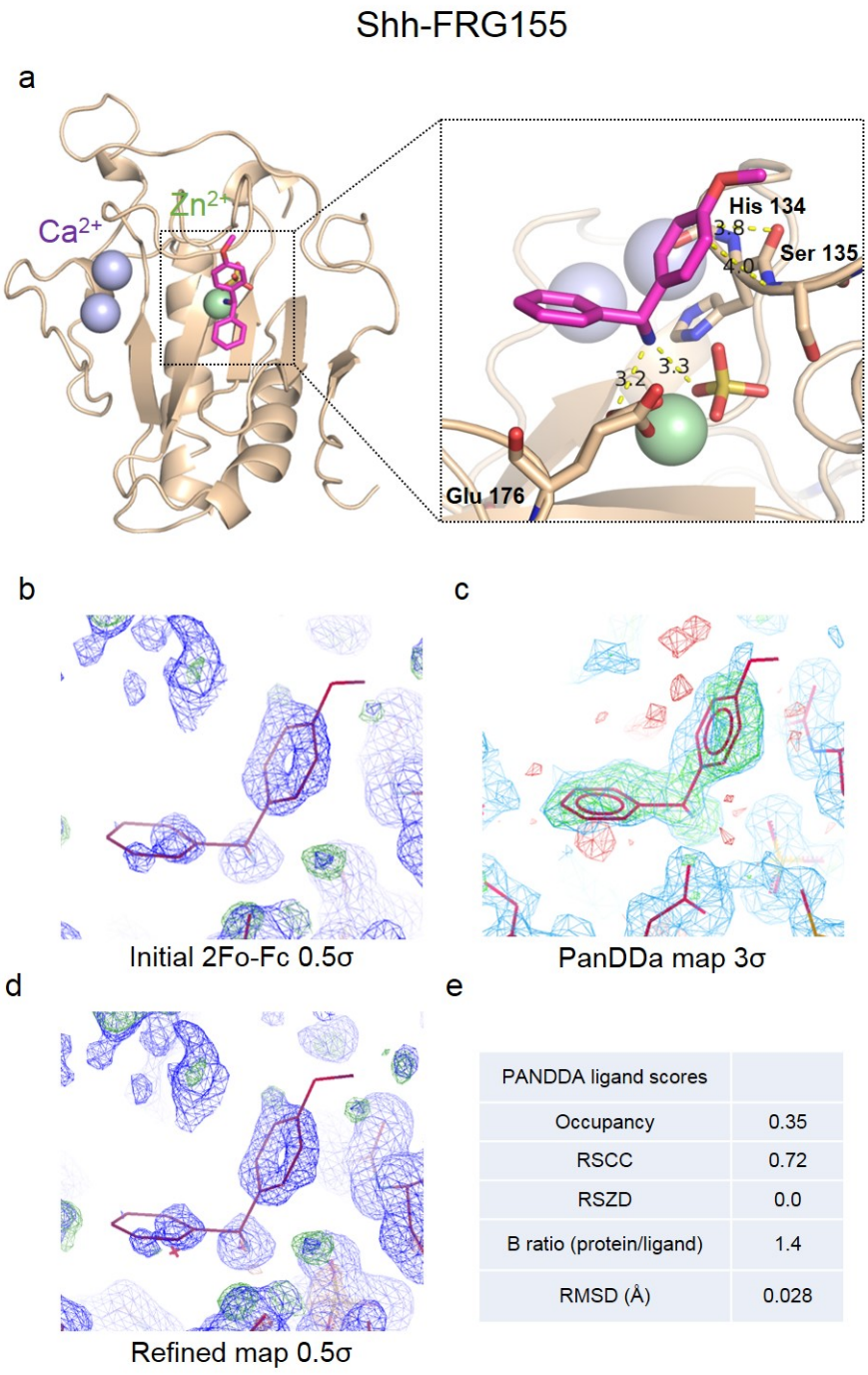


Figure 3.6: **ShhN-FRG 155 structure.** a) FRG 155 (shown as magenta sticks) binds across the zinc binding site forming stacking interactions with the peptide bond between residues His134 and Ser135. The amine group forms electrostatic interactions with Glu176 and sulphate/phosphate group. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.

The ShhN-FRG 17 structure was solved 1.2 Å with an occupancy of 0.30 (Table 3.3). FRG 17 is a methylamino phenyl-pyrazole which binds across the zinc site (Figure 3.7, a). The interactions between FRG 17 and ShhN are similar to ShNh-FRG 155 interactions (Figure 3.7, a). The methylamine group overlaps with amine from FRG 155, and forms electrostatic interactions with Glu176 and sulphate/phosphate group. In addition, the pyrazole moiety forms electrostatic interactions with Ser135, while the phenyl ring stacks against the peptide bond between His133 and His134 (distance $\sim$ 3.6 Å). The electron density for FRG 17 can be seen in SigmaA-weighted 2Fo-Fc maps with contour 0.5σ with minimal density noise around the fragment (Figure 3.7, b, d-e). FRG 17 is also well defined in PanDDa maps (Figure 3.7, c), with all PanDDa scores being acceptable.

The ShhN-FRG 458 structure was solved to a resolution of 1.1 Å with an occupancy of 0.34 (Table 3.5 and Figure 3.8, a). The fragment consists of a phenyl thiadiazol with a further methyl group modification. FRG458 binds to ShhN in a similar fashion to FRG 17, with the respective phenyl groups overlaying (Figure 2.10, a). The fragment could be observed in standard electron density maps when contoured at lower sigma and was well defined in the PanDDa maps (Figure 2.10, b-e). Initially, it appeared as if the fragment bound in two different orientations and the model was built and refined with two fragments, accordingly. However, after refinement one of the orientations moved out significantly. This indicated that the correct orientation is likely the one where the thiadiazol is closer to zinc site and the final model was refined accordingly, with only one ligand. In this orientation, the fragment stacks against the loop formed by residues

133-135, while the thiadiazol moiety forms electrostatic interactions with Ser135 ($\sim$ 3.8 Å).

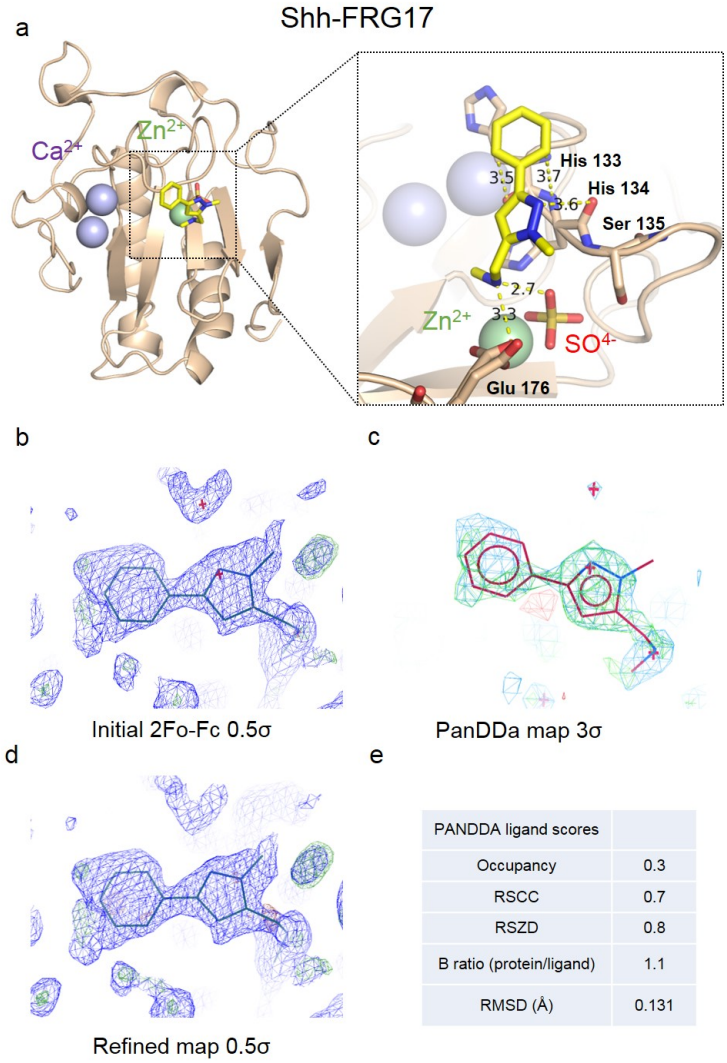
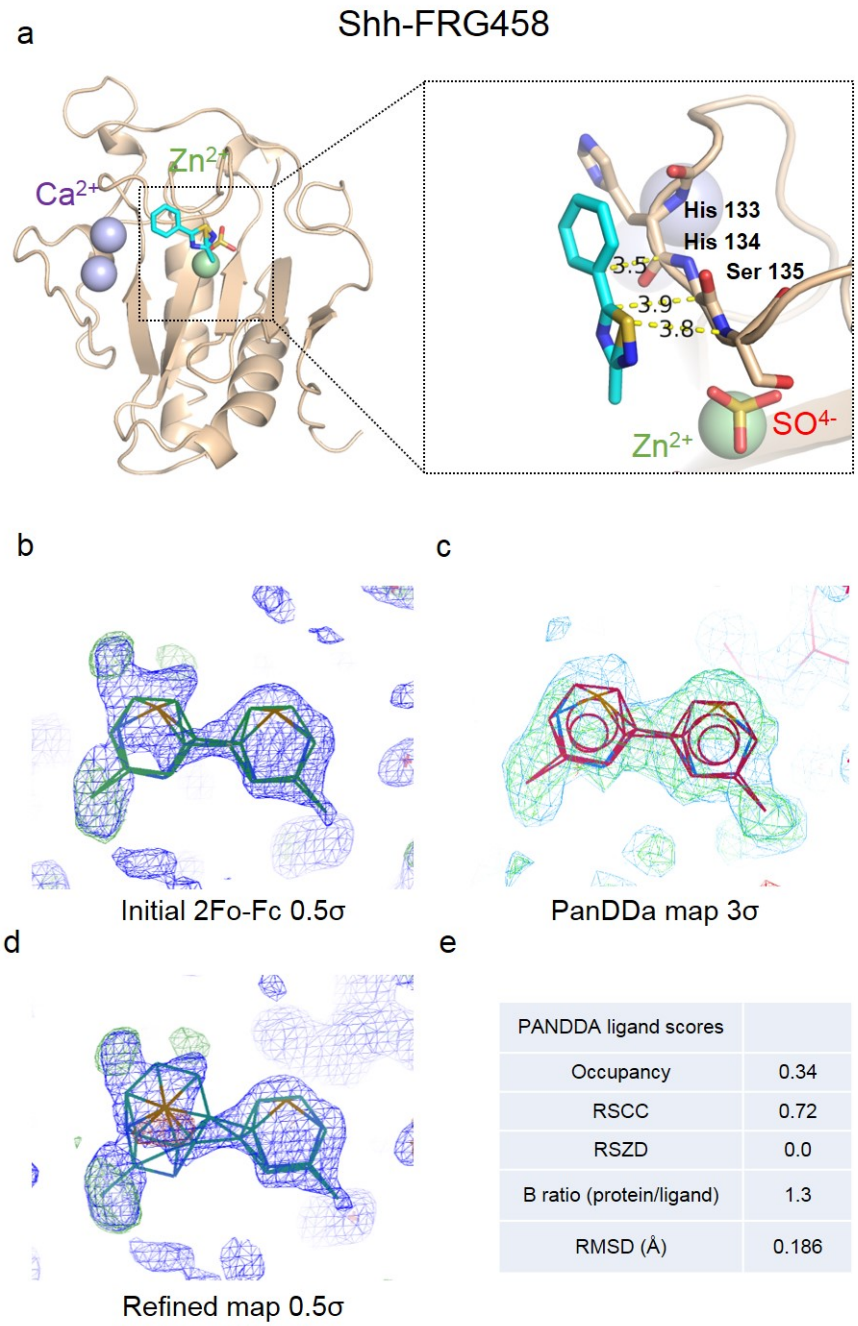


Figure 3.7: **The ShhN-FRG 17 structure.** a) FRG 17 (shown as yellow sticks) binds close to the zinc binding site stacking against the peptide bond between residues His133 and His134, while the pyrazole forms electrostatic interactions with the Ser135 backbone. The methylamine group forms electrostatic interactions with Glu176 and sulphate/phosphate group, similarly to FRG 155. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at ±3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.



PANDDA ligand scores	
Occupancy	0.34
RSCC	0.72
RSZD	0.0
B ratio (protein/ligand)	1.3
RMSD (Å)	0.186

Figure 3.8: **The ShhN-FRG 458 structure** a) FRG 458 (shown as cyan sticks) binds close to the zinc binding site stacking against peptide bond between residues His133 and His134, while thiadiazol forms electrostatic interactions with Ser135 backbone. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.

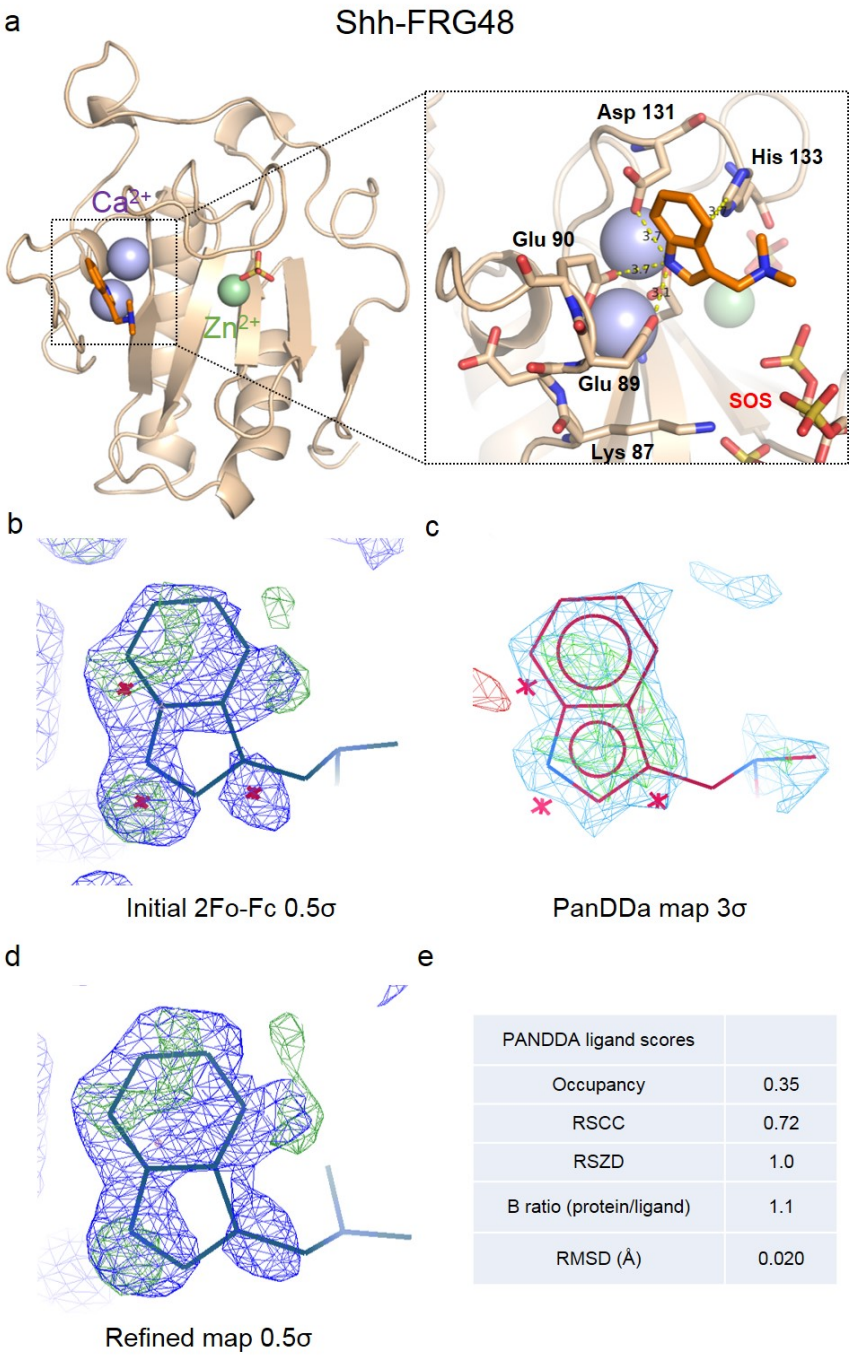
3.3.4 Binding site 2

Binding site 2 is adjacent to the calcium coordination site where two hits (48 and 505), both indole derivatives, bound.

The ShhN-FRG 48 structure was solved at 1.3 Å resolution with an occupancy of 0.35 (Table 3.3 and Figure 3.9, a). FRG 48 is a naturally occurring compound called Gramine, which is an indole methanamine derivative. Gramine is found in plants where it functions as a defence mechanism against herbivorous insects [226]. It was also found to be an agonist of adiponectin receptors and adrenergic re-uptake inhibitor at synanapses [227, 228].

Gramine binds to ShhN across the calcium binding site (Figure 3.9, a) where it forms electrostatic interactions with residues Glu89, Glu90, Asp131 and T-shaped pi-pi stacking interactions with His133. It also interacts with Ile111 via crystal symmetry. The fragment is less apparent in standard electron density maps. However, PanDDa maps present it clearly (Figure 3.9, b-e).

The ShhN-FRG 505 structure was solved at 1.2 Å resolution with 0.44 occupancy (Table 3.5 and Figure 3.10, a). FRG 505 is an aminobenzimidazole, similar in structure and binding mode to FRG 48. It forms electrostatic interactions with Glu89 and T-shaped pi-pi stacking to His133 (Figure 3.10, a). In addition, FRG 505 also interacts with SOS and Ile111 via crystallographic symmetry. FRG 505 can be observed in lower contoured electron density maps with minimal noise around, and is well defined in PanDDa maps (Figure 3.10, b-e).



PANDDA ligand scores	
Occupancy	0.35
RSCC	0.72
RSZD	1.0
B ratio (protein/ligand)	1.1
RMSD (Å)	0.020

Figure 3.9: **The ShhN-FRG 48 structure** a) FRG 48 (shown as orange sticks) binds close to the calcium binding site forming electrostatic interactions with residues Glu89, Glu90, and Asp131, and stacking interactions with His133. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.

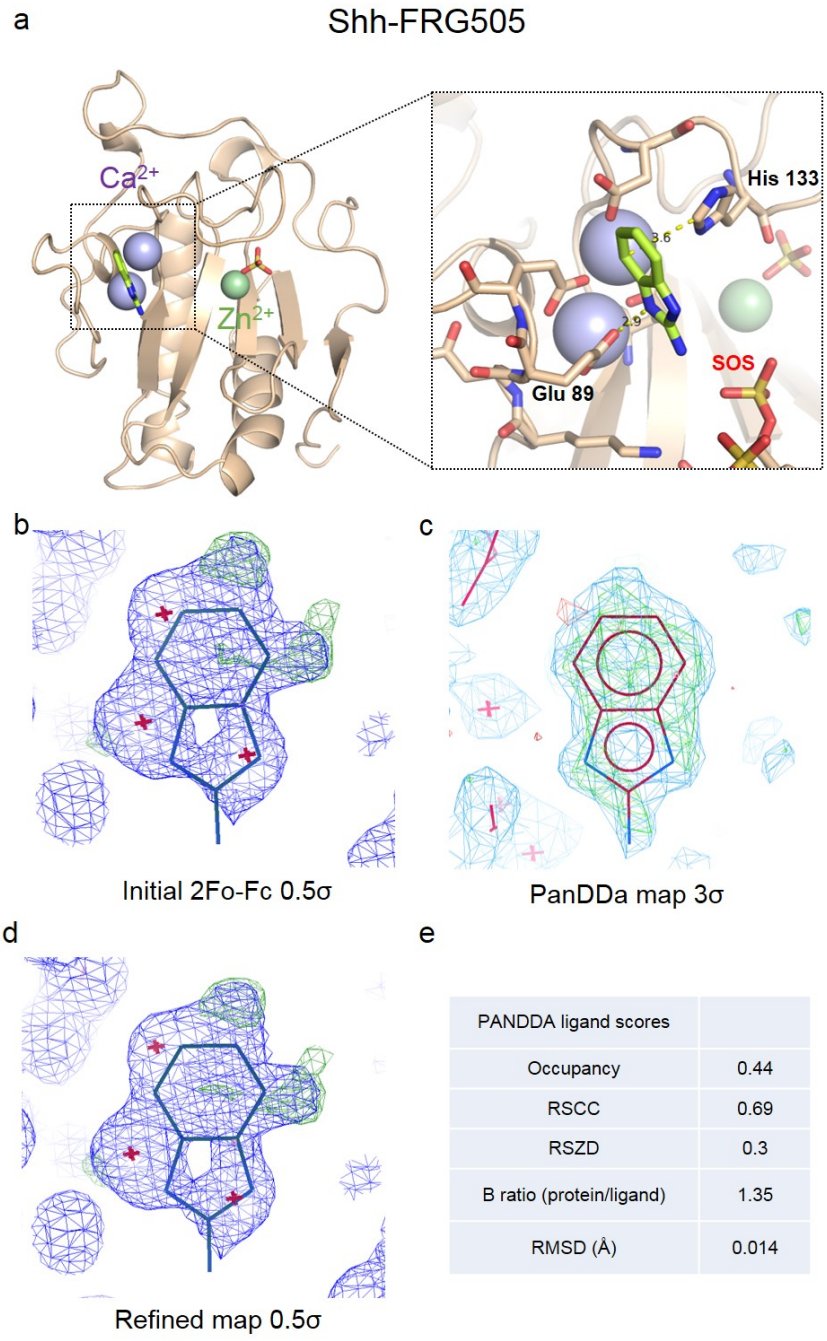
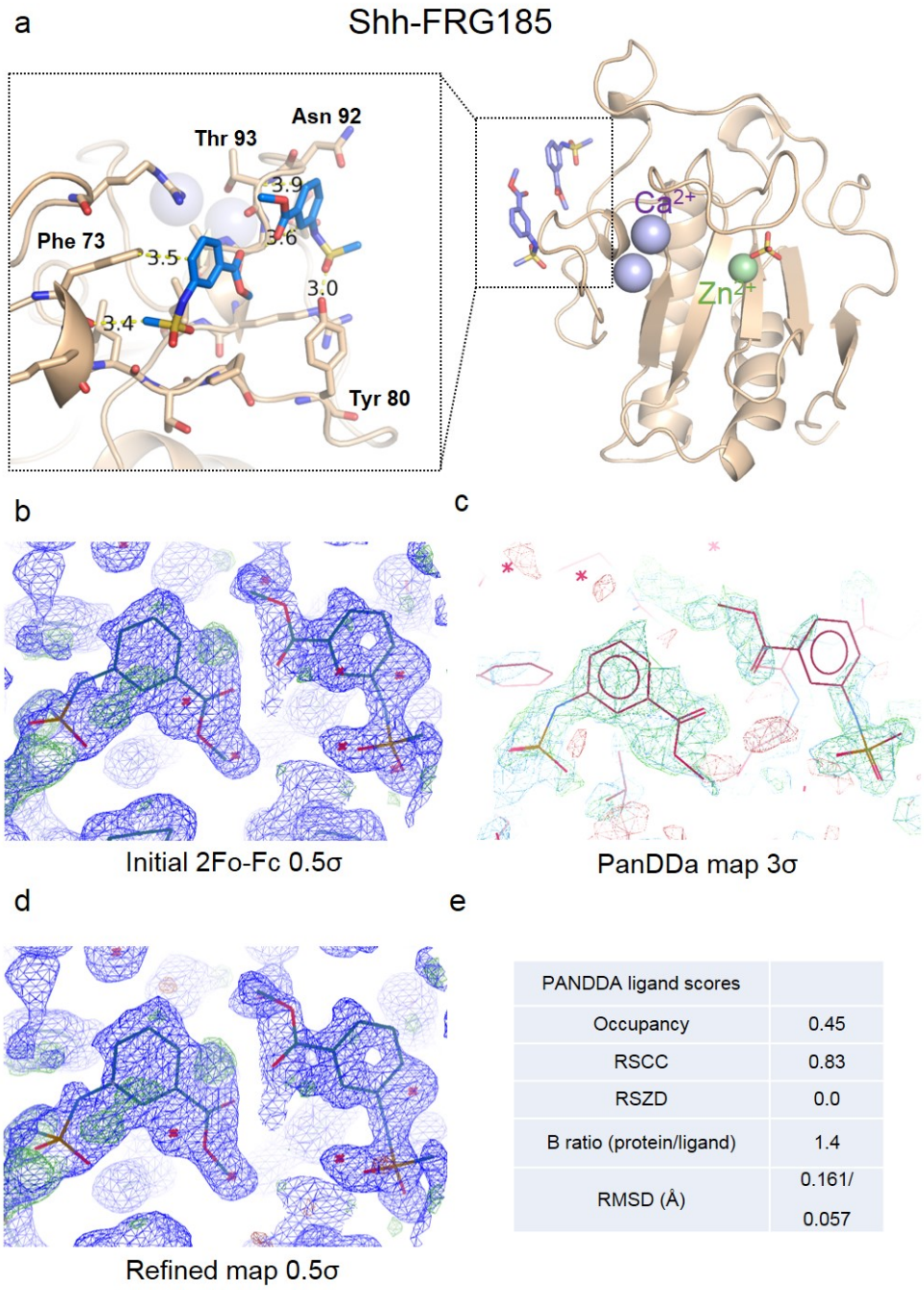


Figure 3.10: **The ShhN-FRG 505 structure** a) FRG 505 (shown as green sticks) binds close to the calcium binding site forming electrostatic interactions with residues Glu89, and stacking interactions with His133. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.

3.3.5 Binding site 3

Binding site 3 is located on the opposite side to the ion binding cleft of ShhN and contains three hits FRG 185, FRG 303 and FRG 484.

The ShhN-FRG 185 complex was solved to a resolution of 1.1 Å with a ligand occupancy of 0.45 (Table 3.4 and Figure 3.11, a). FRG 185 is a benzoic acid derivative with amino methylsulfonyl modification. In the structure, two FRG 185 fragments are binding in opposite orientations and interact with Phe73, Asn92 and Thr93 (Figure 3.11, a). There are also interactions with Phe47 via crystallographic symmetry. The electron density for both ligands is apparent in low contoured maps (Figure 3.11, b, d-e). However, PanDDa map appears to be better defined for the fragment located further away from the ion binding cleft (Figure 3.11, c).



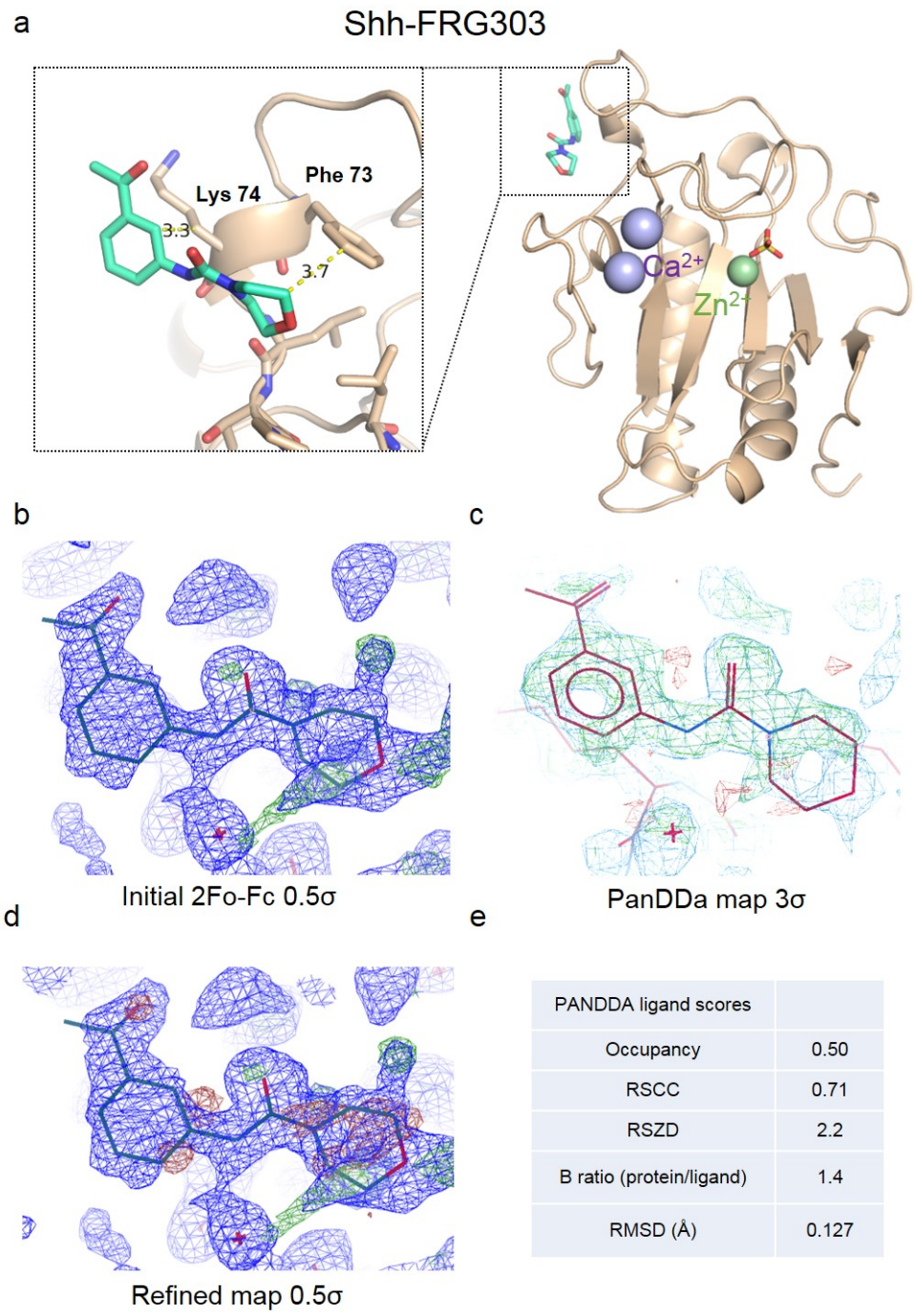
PANDDA ligand scores	
Occupancy	0.45
RSCC	0.83
RSZD	0.0
B ratio (protein/ligand)	1.4
RMSD (Å)	0.161/ 0.057

Figure 3.11: **The ShhN-FRG 185 structure** a) Two FRG 185 (shown as bluen sticks) fragments bind in tandem in opposite orientations at binding site 3. The fragments interact with residues Phe73, Asn92 and Thr93. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.

The ShhN-FRG 303 structure was solved at 1.1 Å resolution with occupancy of 0.50. The fragment is a two ring structure morpholinecarboxamide derivative. FRG 303 stacks against Phe73 and Lys74 side chains in close proximity to FRG 185 (Figure 3.12, a). Although less defined in standard electron density maps, the fragment is well defined in PanDDa maps (Figure 3.12, b-e).

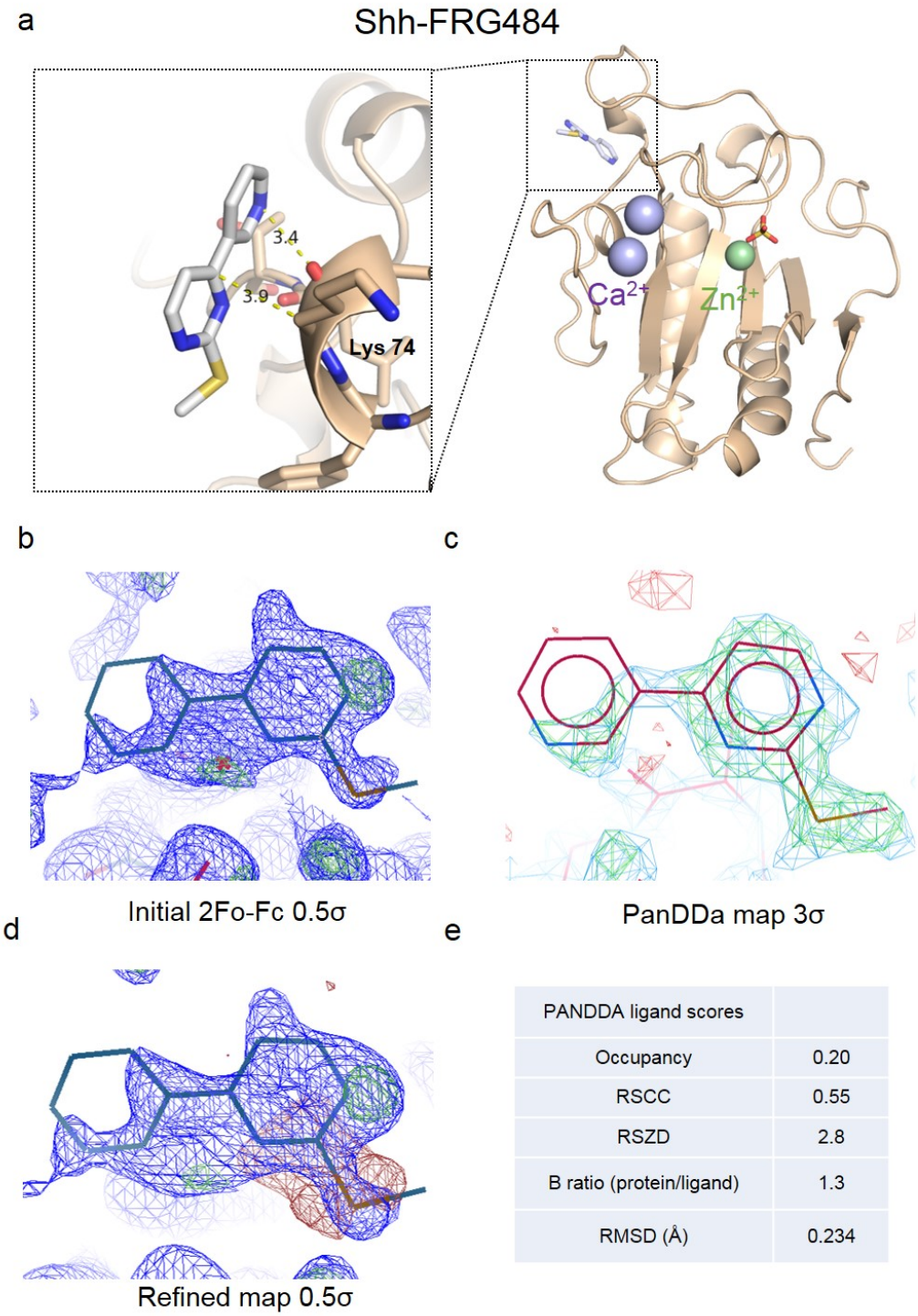
The ShhN-FRG 484 structure was solved at 1.1 Å resolution with occupancy of 0.20. The fragment consists of pyridine and pyrimidine rings with methylthio modifications. FRG 484 stacks against Lys74 in a similar fashion to FRG 303 (Figure 3.13, a). The fragment is less well defined in the standard electron density map but the density for most of the molecule is visible in the PanDDa map (Figure 3.13, b-e). Lack of complete density for pyridine ring might indicate compound hydrolysis.

All three fragments at this site also interacts with residues Val51 and Ala52 via crystal symmetry.



PANDDA ligand scores	
Occupancy	0.50
RSCC	0.71
RSZD	2.2
B ratio (protein/ligand)	1.4
RMSD (Å)	0.127

Figure 3.12: **The ShhN-FRG 303 structure** a) FRG 303 (shown as greencyan sticks) fragment interact with residues Phe73 and Lys74. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.



PANDDA ligand scores	
Occupancy	0.20
RSCC	0.55
RSZD	2.8
B ratio (protein/ligand)	1.3
RMSD (Å)	0.234

Figure 3.13: **The ShhN-FRG 484 structure** a) FRG 484 (shown as white sticks) fragment interact with Lys74. b) Initial SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. c) PanDDA map contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map. d) Post-refinement SigmaA-weighted 2Fo-Fc map contoured at 0.5σ is shown in blue, and the difference map Fo-Fc contoured at 3σ is shown in green/red. e) Summary of PanDDa scores.

3.3.6 Binding sites 4-8

The remaining fragments were bound to various sites across the ShhN surface (Figure 3.14).

The ShhN-FRG 159 structure was solved to 1.1 Å resolution with fragment occupancy of 0.20 (Table 3.4). FRG 159 is a benzenemethanamine derivative with a thiophene ring and is very clearly represented in PanDDa maps. It forms electrostatic interactions with SOS and the thiophene ring stacks against Asp154, with no additional interactions via crystal symmetry (Figure 3.15, a-c).

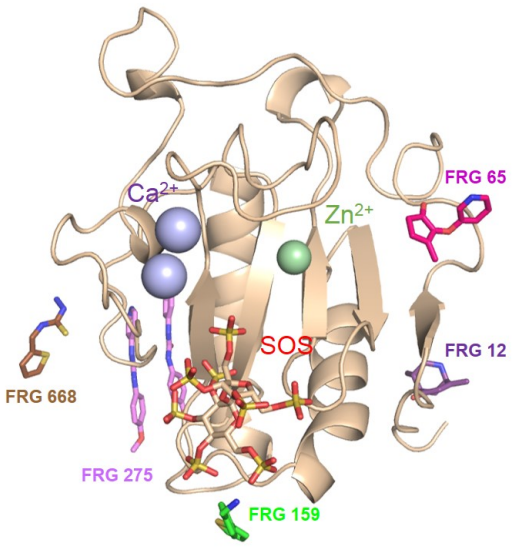


Figure 3.14: **Binding sites 4-8** Remaining fragments bind across the surface of ShhN in previously unidentified binding sites. Fragments are presented as sticks: FRG 159 (green), FRG 275 (violet), FRG 668 (brown), FRG 12 (purple), FRG 65 (hotpink).

The ShhN-FRG 275 structure was solved to 1.0 Å resolution with fragment occupancy of 0.44 (Table 3.4). FRG 275 is a methoxyphenyl pyridinylurea and it binds to ShhN in tandem (Figure 3.15, d-f). Both compounds are clearly visible in PanDDa maps and they form electrostatic interactions with residues Ile111 and Asp83. They form additional interactions with SOS and residue Glu89 via crystal symmetry.

The ShhN-FRG 668 structure was solved to 1.3 Å resolution with fragment

occupancy of 0.47 (Table 3.5). FRG 668 is a thiphene with azothiol extension. FRG 668 interacts with Tyr80 and SOS via crystal symmetry (Figure 3.15, g-j).

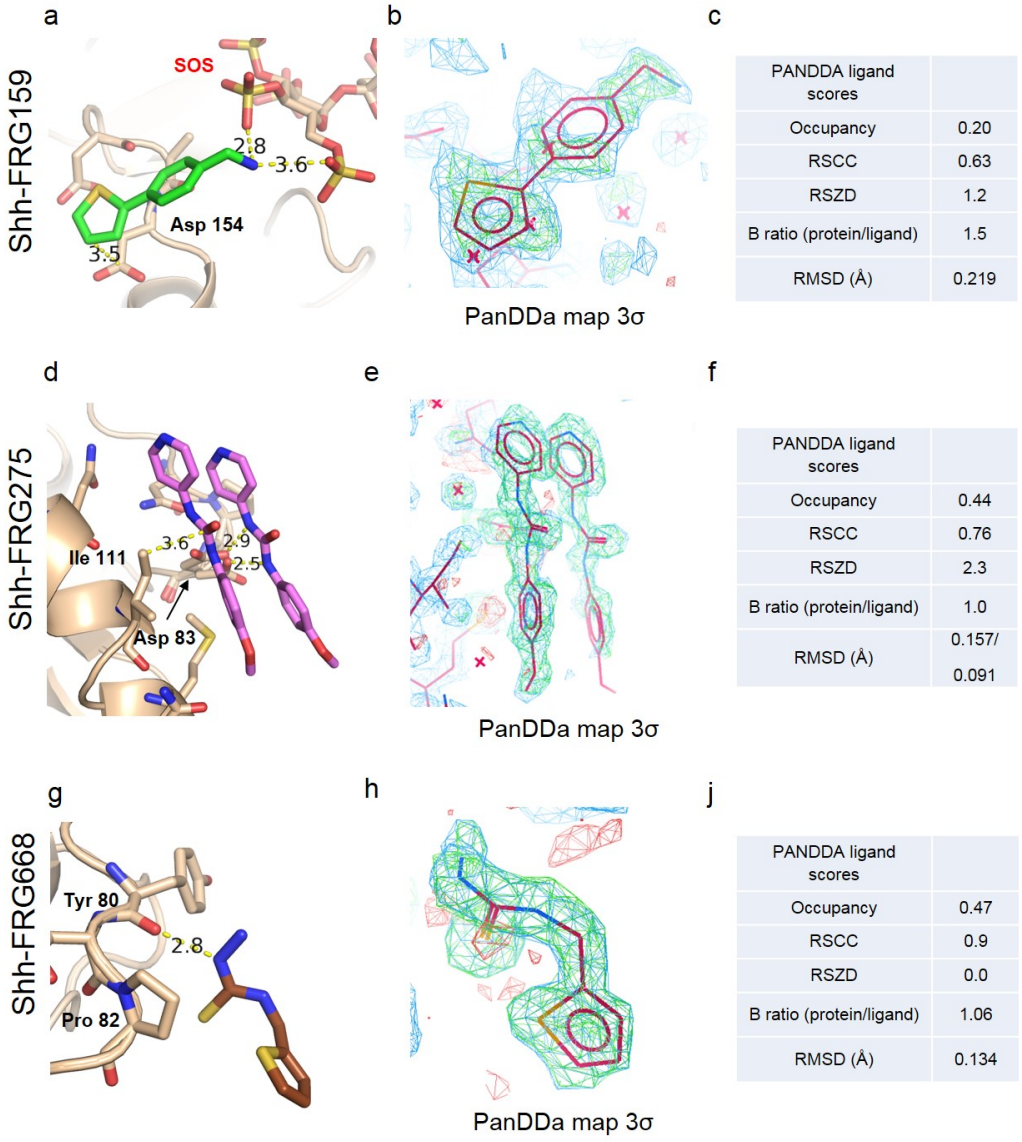


Figure 3.15: **Structures of the ShhN-FRG 159, ShhN-FRG 275 and ShhN-FRG 668 complexes.** a, b) FRG 159, presented as green sticks, interacts with SOS and stacks against Asp154. d, e) FRG275, presented as violet sticks, binds in a tandem of two fragments that interact with Ile111 and Asp83. g, h) FRG668, presented as brown sticks, makes interaction with Tyr80. c, f, j) PanDDA map for relevant fragments contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map.

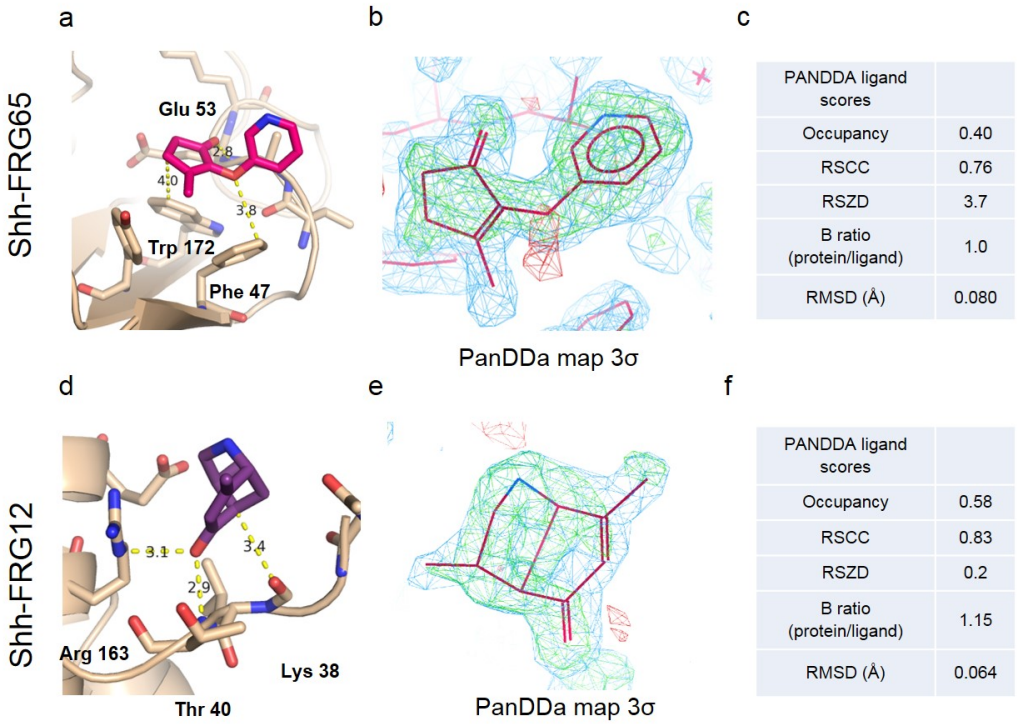


Figure 3.16: **Structures of Shh-FRG 65 and Shh-FRG 12.** a, b) FRG 65 binds close to the zinc site and interacts with residues Trp172, Phe47 and Glu53. d, e) FRG 12 bind close to the N-terminus and induces residues before Lys38 to be structured. In interacts with residues Arg163, Thr40 and Lys38. c, f) PanDDA map for relevant fragments contoured at 3σ shown in green, while the Z-map contoured at 6σ is shown in light blue. Z-map represents how significantly the data set deviates from the average map.

The ShhN-FRG 65 structure was solved at 1.1 Å resolution with an occupancy of 0.40 (Table 3.6). The fragment consist of pyridine and cyclopentane rings and it binds nearby the zinc binding site but not within the ion binding groove. FRG 65 interacts with residues Trp172, Phe47 and Glu53 (Figure 3.16, a-c).

The Shh-FRG 12 structure was solved to a resolution of 1.2 Å with an occupancy of 0.58 (Table 3.6). The fragment binds at the N-terminus, making it partially structured. FRG 12 forms electrostatic interaction with residues Arg163, Thr40 and Lys38 (Figure 3.16, d-f).

3.3.7 Crystallographic tables

	Shh-SOS	Shh-FRG17	Shh-FRG48	Shh-FRG155
CRYSTALLISATION				
Crystallisation condition	0.2 M ammonium sulphate, 0.1 M sodium cacodylate, pH 6.5, 30% w/v PEG 8000	Condition I	Condition I	Condition I
Library, compound	-	Maybridge 250 mM	Maybridge 250 mM	DSPL 500 mM (DMSO
Soaking conditions	-	30 % compound, 1h	30 % compound, 1h	30 % compound, 1h
Cryoprotectant	20 % ethylene glycol	20 % ethylene glycol	20 % ethylene glycol	20 % ethylene glycol
DATA COLLECTION				
X-ray source	DIAMOND-I04	DIAMOND-I041	DIAMOND-I041	DIAMOND-I041
Resolution	43.8-1.2 (1.2-1.2)	46.2-1.2 (1.2-1.1)	46.7-1.3 (1.4-1.3)	27.0-1.1 (1.1-1.1)
Space group	C 1 2 1	C 1 2 1	C 1 2 1	C 1 2 1
Cell dimensions				
a, b, c [Å]	93.5, 53.2, 35.4	98.9, 54.0, 35.9	98.7, 54.1, 35.9	98.8, 54.0, 36.0
α, β, γ [°]	90.0, 111.0, 90.0	90.0, 110.8, 90.0	90.0, 111.0, 90.0	90.0, 111.0, 90.0
Unique reflections	50054 (3476)	58652 (4243)	41308 (3042)	69011 (4609)
Completeness [%]	95.9 (68.0)	98.5 (95.9)	99.5 (99.5)	96.4 (87.3)
R merge [%]	6.0 (56.0)	5.5 (71.1)	6.7 (97.2)	3.1 (48.3)
Rpim		5.4 (69.8)	6.3 (89.7)	2.9 (44.6)
I/σI	6.3 (1.1)	9.0 (1.3)	9.1 (1.1)	13.8 (1.3)
CC (1/2)		1.00 (0.5)	1.00 (0.5)	1.00 (0.6)
Multiplicity	3.1 (2.2)	3.1 (2.3)	3.3 (3.2)	3.1 (2.0)
REFINEMENT				
Resolution range [Å]	43.8-1.2	46.6-1.2	46.7-1.3	46.6 -1.1
Number of reflections	47213	55671	39271	65513
No. of atoms (protein/Zn/Ca/ SOS/ H_2O/ LIG)	1202/1/2/55/ 103/-	1276/1/2/55/ 160/15	1294/1/2/55/ 171/13	1291/1/2/55/ 138/16
R work [(%)]	18.5	18.7	17.3	17.6
R free [%]	22.4	20.3	19.1	18.4
RMSD bonds [Å]	0.02	0.01	0.01	0.01
RMSD angles [deg]	2.8	1.4	1.4	1.3
Mean B factor (protein/ligand)	-	17.1/20.0	21.1/22.1	16.22 / 23.2
Ramachandran statistics				
Favoured [%]	95.8	98.0	98.0	98.6
Disallowed [%]	0	0	0	0

Table 3.3: Crystallographic tables for ShhN-SOS, ShhN-FRG17, ShhN-FRG48 and ShhN-FRG155.

	Shh-FRG159	Shh-FRG185	Shh-FRG275	Shh-FRG303
CRYSTALLISATION				
Crystallisation condition	Condition I	Condition I	Condition I	Condition I
Library, compound	DSPL/Maybridge	DSPL/Maybridge	DSPL	DSPL
Soaking conditions	30 % compound, 1h	30 % compound, 1h	30 % compound, 1h	30 % compound, 1h
Cryoprotectant	20 % ethylene glycol	20 % ethylene glycol	20 % ethylene glycol	20 % ethylene glycol
DATA COLLECTION				
X-ray source	DIAMOND-I041	DIAMOND-I041	DIAMOND-I041	DIAMOND-I041
Resolution	46.4-1.1 (1.1-1.1)	33.3-1.1 (1.1-1.1)	46.7-1.0 (1.1-1.0)	26.7-1.1 (1.2-1.1)
Space group	C 1 2 1	C 1 2 1	C 1 2 1	C 1 2 1
Cell dimensions				
a, b, c [Å]	99.2, 54.1, 35.9	98.0, 54.2, 35.9	99.2, 54.1, 35.9	98.8, 54.0, 35.9
α, β, γ [°]	90.0, 110.0, 90.0	90.0, 110.8, 90.0	90.0, 110.6, 90.0	90.0, 110.9, 90.0
Unique reflections	76982 (4924)	70960 (4605)	82250 (4045)	61878 (3951)
Completeness [%]	95.9 (82.9)	93.7 (82.3)	91.3 (52.0)	93.4 (81.1)
R merge [%]	4.3 (51.9)	3.2 (60.9)	4.1 (40.9)	4.0 (48.8)
Rpim	4 .0 (47.5)	3.1 (59.1)	3.9 (40.9)	4.0 (48.3)
I/σI	10.8 (1.2)	12.6 (1.2)	10.9 (1.2)	11.7 (1.5)
CC (1/2)	1 (0.6)	1 (0.58)	1 (0.7)	1 (0.7)
Multiplicity	2.9 (1.6)	3.0 (2.0)	2.8 (1.2)	3.2 (2.3)
REFINEMENT				
Resolution range [Å]	46.74 -1.0	46.73-1.0	46.74-1.0	46.64-1.1
Number of reflections	73054	67436	78108	58753
No. of atoms (protein/Zn/Ca/ SOS/ H_2O/ LIG)	12771/2/55/ 148/13	1266/1/2/55/ 199/30	1281/1/2/55/ 173/36	1262/1/2/55/ 138/18
R work [(%)]	18.5	18.3	17.5	17.77
R free [%]	19.6	20.2	18.6	18.9
RMSD bonds [Å]	0.01	0.01	0.01	0.01
RMSD angles [deg]	1.3	1.4	1.3	1.3
Mean B factor (protein/ligand)	15.5/23.0	16.8/22.2	15.3/16.7	15.5/22.0
Ramachandran statistics				
Favoured [%]	97.4	98.0	98.6	98.0
Disallowed [%]	0	0	0	0

Table 3.4: Crystallographic tables for ShhN-FRG159, ShhN-FRG185, ShhN-FRG275 and ShhN-FRG303.

	Shh-FRG458	Shh-FRG484	Shh-FRG505	Shh-FRG668
CRYSTALLISATION				
Crystallisation condition	Condition I	Condition I	Condition I	Condition II
Library, compound	Maybridge	Maybridge	Maybridge	Maybridge
Soaking conditions	30 % compound, 1h	30 % compound, 1h	30 % compound, 1h	30 % compound, 1h
Cryoprotectant	20 % ethylene glycol	20 % ethylene glycol	20 % ethylene glycol	20 % ethylene glycol
DATA COLLECTION				
X-ray source	DIAMOND-I041	DIAMOND-I041	DIAMOND-I041	DIAMOND-I041
Resolution	46.6-1.1 (1.1-1.1)	46.3-1.1 (1.1-1.1)	46.7-1.2 (1.3-1.2)	46.6-1.3 (1.3-1.3)
Space group	C 1 2 1	C 1 2 1	C 1 2 1	C 1 2 1
Cell dimensions				
a, b, c [Å]	99.0, 54.0, 35.9	99.0, 54.0, 35.8	98.8, 54.1, 35.8	98.9, 54.0, 36.0
α, β, γ [°]	90.0, 110.8, 90.0	90.0, 110.7, 90.0	90.0, 110.8, 90.0	90.0, 110.8, 90.0
Unique reflections	10501 (4607)	74202 (4721)	49043 (3572)	140055 (7090)
Completeness [%]	95.8 (84.8)	95.6 (82.3)	98.2 (97.0)	98.0 (95.2)
R merge [%]	3.4 (71.4)	2.7 (64.8)	7.1 (92.6)	3.8 (60.2)
Rpim	3.3 (70.4)	2.7 (62.6)	6.9 (87)	3.6 (55.4)
I/σI	11.9 (1.2)	14.2 (1.2)	8.5 (1.0)	12.9 (1.2)
CC (1/2)	1 (0.6)	1 (0.6)	1 (0.5)	1 (0.5)
Multiplicity	3.0 (2.1)	3.0 (1.9)	3.3 (2.8)	3.1
REFINEMENT				
Resolution range [Å]	46.6-1.1	46.6-1.1	46.7-1.2	46.6-1.3
Number of reflections	66927	70390	46610	43463
No. of atoms (protein/Zn/Ca/ SOS/ H_2O/ LIG)	1284/1/2/55/ 120/12	1294/1/2/55/ 201/14	1302/1/2/55/ 174/10	1276/1/2/55/ 150/11
R work [(%)]	18.6	17.9	18.3	18.6
R free [%]	19.3	18.9	20.0	20.3
RMSD bonds [Å]	0.01	0.01	0.01	0.01
RMSD angles [deg]	1.3	1.3	1.4	1.1
Mean B factor (protein/ligand)	16.4/23.3	16.6/21.8	19.4/24.2	18.5/19.6
Ramachandran statistics				
Favoured [%]	98.0	98.0	98.0	98.6
Disallowed [%]	0	0	0	0

Table 3.5: Crystallographic tables for ShhN-FRG458, ShhN-FRG484, ShhN-FRG505 and ShhN-FRG668.

	Shh-FRG65	Shh-FRG12
CRYSTALLISATION		
Crystallisation condition	Condition I	Condition I
Library, compound	OxChem (500 mM)	OxChem (500 mM)
Soaking conditions	30 % compound, 1h	30 % compound, 1h
Cryoprotectant	20 % ethylene glycol	20 % ethylene glycol
DATA COLLECTION		
X-ray source	DIAMOND-I041	DIAMOND-I041
Resolution	33.4-1.1 (1.2-1.1)	25.0-1.2 (1.2-1.2)
Space group	C 1 2 1	C 1 2 1
Cell dimensions		
a, b, c [Å]	98.7, 53.9, 35.9	99.1, 54.2, 35.7
α, β, γ [°]	90.0, 111.1, 90.0	90.0, 110.5, 90.0
Unique reflections	59014 (3491)	57815 (3950)
Completeness [%]	89.6 (72.3)	96.8 (88.5)
R merge [%]	5.7 (89.3)	4.1 (69.7)
Rpim	5.6 (88.3)	4.0 (69.1)
I/σI	9.9 (1.1)	10.8 (1.1)
CC (1/2)	1 (0.7)	1 (0.6)
Multiplicity	2.9 (2.0)	3.0 (2.1)
REFINEMENT		
Resolution range [Å]	46.6-1.1	46.8-1.2
Number of reflections	56026	54871
No. of atoms (protein/Zn/Ca/ SOS/ H_2O/ LIG)	1269/1/2/55/ 202/14	1304/1/2/55/ 121/12
R work [(%)]	20.3	18.6
R free [%]	21.9	10.5
RMSD bonds [Å]	0.01	0.01
RMSD angles [deg]	1.5	1.4
Mean B factor (protein/ligand)	19.3/14.4	16.5/20.0
Ramachandran statistics		
Favoured [%]	98.0	98.0
Disallowed [%]	0	0

Table 3.6: Crystallographic tables for ShhN-FRG65 and ShhN-FRG12.

3.4 Discussion

The aim of this chapter was to generate fragment hits that bind ShhN in a crystallographic fragment screen.

Firstly, we solved a high-resolution ShhN-SOS complex structure that allowed us to perform medium-throughput crystallographic screening. SOS is a sugar that contains six sulphate groups and binds to the positively charged region of ShhN, which is the binding site of other glucosaminoglycans (GAGs) such as Heparin and C-6. Like C6, SOS does not promote dimerisation [99]. The hallmark of this structure is the crystal form in which the ion binding groove of ShhN is exposed. ShhN is a pseudo-zinc hydrolase with a conserved zinc site, which is essential in maintaining the structural stability of ShhN. Zinc and calcium ions both play a crucial role in receptor binding and are, therefore, primary targets for inhibitor development. Zinc coordination site plays a key role in binding E51 antibody [139] and the receptor Hhip [115, 135], while calcium ions are essential for ShhN binding to positive regulator Cdo [111] and increase the affinity towards negative regulator Hhip [115].

Secondly, in a medium-throughput crystallographic fragment screen of 1,200 fragments from DSPL, Maybridge, AR and OxXChem libraries, we collected 1,363 datasets and identified 13 hits that bind ShhN, which were detected by the PanDDa software. Because of our initial low hit rate extensive optimisation of the crystallisation condition, compound solvent and soaking time were carried out. Further optimisation should include construct engineering, with the main focus on

the N-terminus. This domain appeared to be semi-ordered in some structures and could be observed close to the zinc site of a neighbouring molecule. Constructs that have various lengths of the N-terminus should be expressed and crystallised. In addition, optimising a new crystal form without SOS should be attempted. SOS occluded a part of the binding site and introduced flexibility that could have had an impact on the low hit rate.

In total, 13 hits that bound to 8 different sites were identified. Five hits were found within the ion binding groove, three in a novel binding site behind the calcium coordination site, and the remaining 5 were found across the Shh surface.

Binding site 1 encompassed the zinc coordination site and contained three hits: FRG 155, 17, and 458. FRG 155 is a benzenemethanamine derivative with tetraheadral geometry and two benzene rings. FRG 17 is a methylamino phenyl-pyrazole and FRG 458 is a phenyl thiadiazol with a further methyl group modification. The three fragments share similar features and interactions, in particular an amine group in FRG 17 and FRG 155, and a phenyl in FRG 17 and FRG 458, suggesting importance of these features (Figure 3.17). The crucial interactions common to fragments appear to be the network of electrostatic interactions between amine and Glu176 and sulphate/phosphate groups that coordinate zinc. The fragments also stack against residues His133-Ser135, which were previously predicted to be significant for Robotnikinin binding in docking studies. None of the fragments interact directly with the zinc ion but rather via a sulphate/phosphate group that coordinates zinc. This feature could be used

in lead development by incorporating a sulphate or phosphate group into leads. Binding site 1 hits are the highest priority sites for performing further structure-activity relationship (SAR) studies. This site is crucial for receptor binding and was predicted as the binding site for Robotnikinin in MD and docking studies.

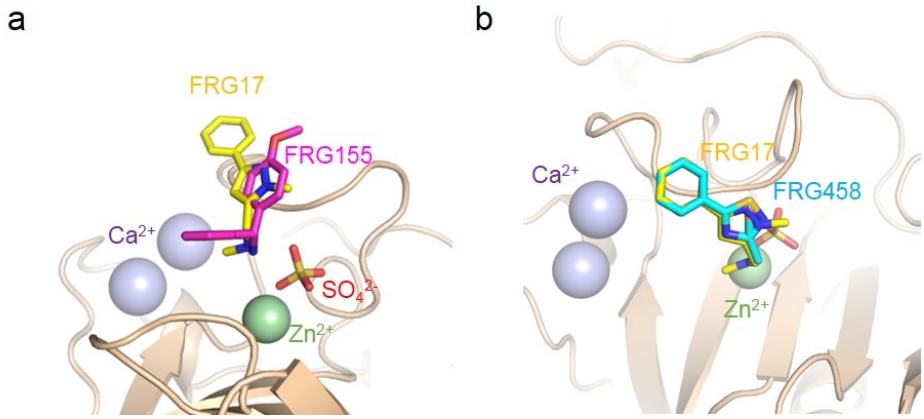


Figure 3.17: **Binding site 1 fragments** a) FRG 17 and FRG 155 contain an overlapping amine moiety that forms electrostatic interactions with Glu176 and sulphate/phosphate group. b) FRG 17 and FRG 458 both stack against residues His133-Ser135 and have overlapping phenyl and pyrazoole/thiadiazol groups.

Binding site 2 encompasses the calcium coordination site and is of high importance for developing follow-up leads. Calcium is an important feature in ShhN-receptor binding [111, 115]. Binding site 2 included two indole derivatives FRG 48 and FRG 505 that interacted in a similar manner with ShhN. (Figure 3.18). The two fragments interacted with Glu89, 90 and Asp131, residues that coordinate calcium, and with His133. Exploring whether fragment binding is calcium dependent would be part of hit validation. Furthermore, exploiting interactions

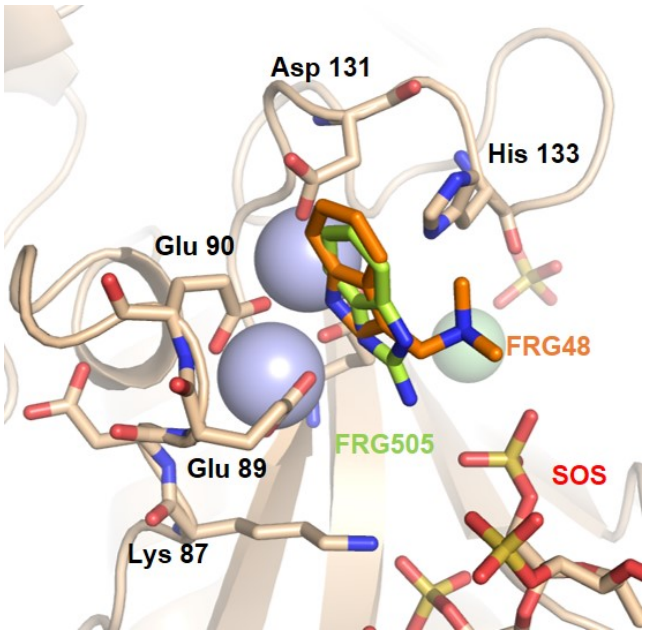


Figure 3.18: **Binding site 2 fragments** FRG 48 and FRG 505 are similar in structure and bind in similar ways across the calcium binding site. The follow-up compounds could include exploring different indole modifications to encourage interactions with neighbouring residues.

with nearby residue Lys87, which has been shown to be essential for receptor binding [111, 135, 115, 139], should be explored by fragment modifications as part of the follow-up strategy.

Binding site 3 is a novel site and residues in this area have not been previously reported as potential receptor binders. However, considering the identification of 3 hits within this area as well as a close proximity of sulfone groups of FRG 484 and FRG 185, follow-up compounds could be explored to test this areas as a possible binding site.

The remaining hits are of lower priority but should be further validated together with high priority hits. In particular, FRG 159 which binds close to

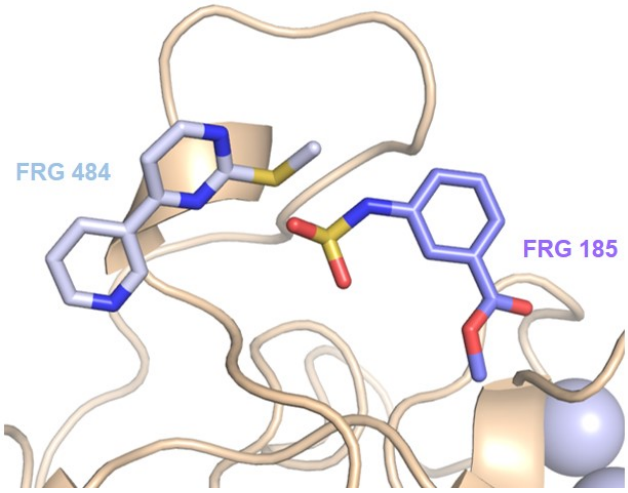


Figure 3.19: **Binding site 3 fragments.** Binding of FRG 185 and FRG 484 with their sulfone groups in close proximity could allow for fragment linking as a potential follow-up strategy.

residues 153-155, which are important for binding to Cdo and GAG molecules [111, 99], and FRG 12, which binds to the N-terminus just after the Heparin binding Cardin-Weintraub motif (residues 33-39) [140].

In conclusion, we successfully identified 13 fragment hits that bind to ShhN through crystallographic fragment screening and generated 13 high-resolution ShhN-FRG structures. Further *in vitro* binding experiments need to be done in order to validate the binding of each hit, and inform the SAR studies and lead generation strategy.

Chapter 4

Targeting Sonic Hedgehog: Hit Validation and Lead Generation

4.1 Abstract

Targeting protein-protein interfaces is challenging due to the flat nature of the protein surfaces, however, fragment screening has been a successful tool in addressing this. As discussed in the previous chapter, crystallographic fragment screening yielded 13 novel fragment hits that bound to ShhN, five of which localised to the ion binding cleft.

In this chapter, ten fragments were purchased and validated using a biophysical assay. In total, six hits were found to be true binders with two low micromolar binders, fragments FRG 155 and FRG 17. The tightest binder, FRG 155, which localised to the pseudo-active binding site of ShhN and formed electrostatic interactions with residues Glu176 and His134-His135, was further characterised. The binding affinity was further confirmed by ITC, and the effect on the Hh pathway was tested in cell assays, where FRG 155 was shown to inhibit Hh signalling in a concentration-dependent manner. Furthermore, preliminary structure-activity relationship data suggested possible modifications that could contribute to higher affinity compounds. These results revealed the true hits identified in the fragment screening and provided a basis for lead generation.

4.2 Introduction

In drug discovery, lead generation is a phase that comes after hit identification and it includes iterative cycles of designing and synthesising hit analogues, evaluating them in *in vitro* or *in silico* assays and optimising them further (Figure 4.1). This chapter is a follow-up to hit identification described previously in Chapter 3, and focuses on hit validation and lead generation (Figure 4.1).

In Chapter 3, we identified 13 hits that bind to ShhN in a crystallographic fragment screen of $\sim$ 1,200 fragments. The aim of this Chapter is to validate these hits in *in vitro* biophysical and cellular assays, and determine which hit is the most promising for follow-up studies. Furthermore, the aim is to conduct an initial structure activity relationship (SAR) assay with the goal of proposing further SAR and lead generation strategy.

SAR is a relationship between the structure of a compound and its effect on the target. It is an essential part of the hit-to-lead phase of drug development, as it allows chemists to identify the specific groups responsible for both desirable and undesirable effects. Compounds used for SAR can be 'off the shelf' (commercially available) or synthesised, and the assays used to test these would depend on the specific target. However, they are often a combination of biophysical binding assays and cellular assays. The most potent compounds are selected and further characterised structurally using NMR or X-ray crystallography, if the target structure is available. Follow-up compounds can also be tested in *in silico* screens as purchasing or synthesising compounds can be expensive.

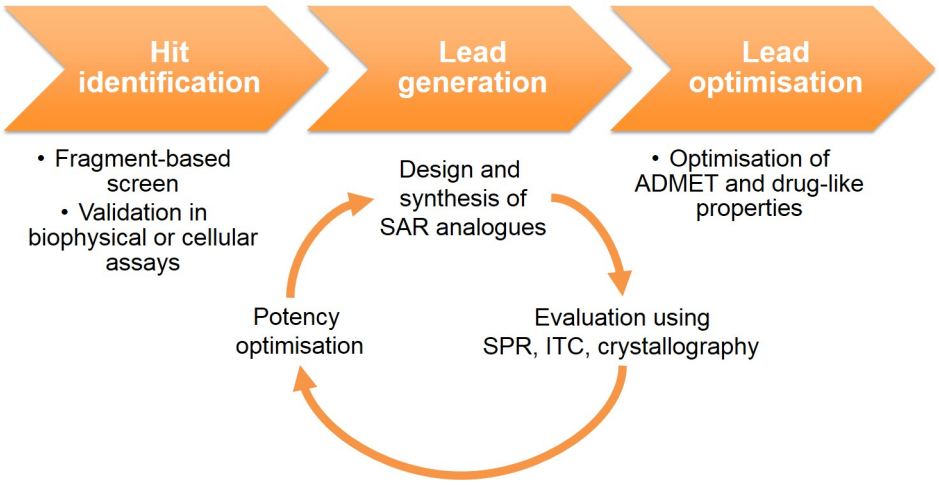


Figure 4.1: **Schematic of fragment-based drug discovery.** Hit identification is done via a medium-throughput screen of fragments, followed by hit validation in biophysical or cellular assay. In the next phase, the most promising hit is further chemically elaborated in iterative cycles of structure-activity relationship assays. Once promising leads are generated, they are further optimised for ADMET and other drug-like properties.

These screens normally include virtual docking of a hit analogue library obtained via a substructure search of online databases like ZINC [229], SciFinder [230] or PubChem [231]. Compounds can also be modified via *in silico* synthesis using methods like core or scaffold hopping in Schrödinger [232] or DOGS (design of genuine structures) [233].

The challenging aspect of SAR in fragment screening is the small size and low affinity of hits identified, and the availability of assays that can be used to detect them. To validate ShhN hits and perform SAR we predominantly used surface plasmon resonance (SPR) to measure binding affinities. We also used a novel *in silico* lead generation software Xpoise, to generate and rank potential leads.

SPR was traditionally developed to measure binding between molecules larger

than 500 Da. However, recent developments in sensor chip design and Biacore machines has made it possible for fragments (100-500 Da) to be screened in a high-throughput fashion. The response units recorded in SPR depend on the mass of the ligand and analyte, and in the case of fragment screening, very low responses (1-10 RU) are detectable. The advantage of using SPR in addition to crystallography is that it enables binding affinities (pM-mM) and kinetics of interaction to be measured using very little protein ($<$ 100 µg). The challenges of working with fragments include the immobilisation of enough ligand ($>$ 10,000 RU), non-specific interactions at high compound concentrations and false-positives due to interference of solvents like DMSO. Problems with ligand immobilisation can be overcome by using various sensor chips including a high capacity CM7 chip specially designed for fragment-screening. To avoid non-specific binding, protocols for affinity or kinetic analysis have been developed to screen a range of concentrations as opposed to a single concentration for each compound and, if a positive control is available, to include controlled Rmax curve fitting that would exclude any non-specific response. For small molecule screening, solvent correction is a necessary step to account for the high refractive index of solvents such as DMSO, and any discrepancies between the buffer and sample concentration. A difference of 1% DMSO between the sample and running buffer can result in an error of 1000 RU. Solvent correction curves from a series of DMSO concentrations is plotted and used to correct the data, in addition to subtracting controls the blank sample and the reference flow-cell.

Although SPR is a valuable technique for SAR, purchasing and synthesis-

ing compounds can be costly and complex. We used *in silico* lead generation software Xpoise, developed by the XChem lab (DLS) as a follow-up to fragment screening. The lead generation builds on the idea of poised fragment libraries (Maybridge and DSPL), which are being used in crystallographic screening at the XChem lab (DLS). These libraries are assembled by selecting fragments that can be broken down into synthons and can undergo common chemical reactions such as amide coupling, N-arylation, ester synthesis, ether synthesis etc. [221]. These compounds are selected as poised and, if identified as hits, are used as input for Xpoise software. Xpoise virtually breaks each fragment into synthons, it applies a filter for a particular reaction that the synthon can undergo, and it screens a library of commercially available fragments against it to 'synthesise' novel compounds. Any compounds that yield two or more products are removed. Unlike other software that mainly relies on core hopping, Xpoise is customised for poised libraries and is built with the purpose of making the follow-up chemistry simple.

Once a library of follow-up compounds has been generated, virtual docking was used to select the leads that are most likely to successfully bind to ShhN. Docking happens in two steps: first, the compounds are docked into the binding site to generate a set of binding poses, and second, the poses are scored by having their respective affinities calculated in the form of binding energy. Although docking can be a quick and inexpensive way to screen compounds, the pose prediction and scoring can often be inaccurate, with the former influencing the latter. In fact, the accuracy of common docking programs like Autodock and Vina

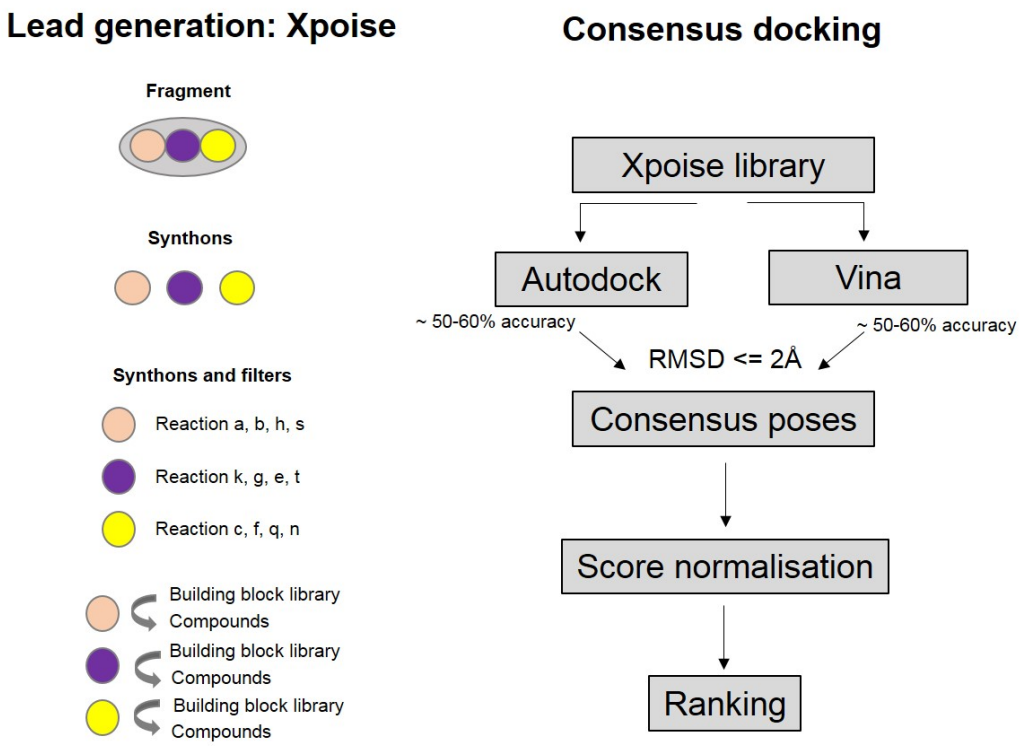


Figure 4.2: **Schematic of Xpoise lead generation and consensus docking.** a) Hits identified in fragment screening are broken down into individual synthons, which are then matched with reaction filters. A commercially available fragment library is then screened against synthons to produce novel compounds. b) In consensus docking, compound libraries are docked using two programs such as Autodock and Vina in parallel. Only the poses that are similar between the two (RMSD $<$ 2 Å) are selected as correctly docked. The binding scores are then normalised by taking the cubic root of heavy-atoms of the particular hit, and ranked in order, from lowest to highest binding energies.

are 55 % and 64 %, respectively when used on their own. As a result, consensus docking was developed to improve docking accuracy [234, 235]. In consensus docking, compound libraries are docked to the protein binding site using two different programs (e.g. Autodock and Vina) in parallel. The poses that are docked similarly by the two programs (RMSD $<$ 2 Å) are selected. This approach improves the accuracy to 82 % when compared to crystal structures [234]. Despite accurate pose prediction with consensus docking, the scoring algorithms were

found to be biased by the physical properties of compounds (molecular weight, cLogP, HBA, HBD, rotable bonds) as opposed to chemical interactions with the target [236, 237]. For example, larger compounds were scored as better because of a higher probability to form hydrogen bonds, even though that was not the case based on protein interactions. To account for this, score normalisation was introduced, where the docking score is divided by the cubic root of non-hydrogen atoms (heavy atoms) of the individual compound [236].

In this chapter, the focus is on hit validation by SPR to identify true binders from fragment-based screening, followed by SAR results of the most promising hit. Lead generation using Xpoise is described further followed by consensus docking with score normalisation using Autodock and Vina, and the highest ranked hits for future synthesis.

4.3 Results

4.3.1 Hit validation

Ten out of thirteen fragment hits were purchased to be used in the validation screen. Fragments 458, 275 and 484 were not available and were, therefore, excluded from further tests.

We validated hits in SPR affinity binding screens and determined the corresponding binding constants. Purified Shh24II was covalently immobilised onto a CM5 Biacore chip and fragments were injected as analytes in a concentration series. CM5 chips were used in the screening stage due to their low cost and ability to capture up to 7,000 RU of ShhN24II. We obtained SPR binding curves for five fragments FRG 155, 17, 48, 505 and 159, while FRG 303, 668, 12 and 65 did not show specific binding to ShhN24II (Figure 4.3).

FRG 155 and 17 were localised to the zinc-binding site, and bound to ShhN24II with relatively high affinity binding constants of 5 µM and 30 µM, respectively (Figure 4.3, b and c). The low micromolar affinity is reflected in ShhN-FRG 155 and ShhN-FRG17 crystal structures, where both fragment stacked against loop His133-Ser135 and formed electrostatic interactions with Glu176 and sulphate/phosphate groups.

In contrast, FRG 48 and 505 bound with very different affinities of 12 µM and 200 µM, respectively, despite similarities in compound structures (Figure 4.3, d and e). This is consistent with the interactions observed in the crystal structures of respective complexes, where FRG 48 forms electrostatic interactions with

Glu89, Glu90, Asp131 and His133 ShhN residues, while FRG 505 only interacts with Glu89 and His133. Further experiments need to be done to examine the calcium dependence of FRG 48 and 505 binding.

FRG 185 binds to ShhN with an 80 µM binding affinity and FRG 159 with 180 µM affinity (Figure 4.3, f and i). However, the binding affinity of FRG 185 is more difficult to interpret since the fragment binds as a dimer in the ShhN complex structure, positioned between two symmetry related ShhN molecules. It is likely that only one of the binding sites is specific, and that the binding curve is a mixture of specific and non-specific binding.

Our results indicate that hits FRG 155, 17, 48, 505, 185 and 159 are true binders. The remaining fragments that were tested (FRG 303, 668, 12 and 65) did not bind or bound non-specifically to the sensor chip surface.

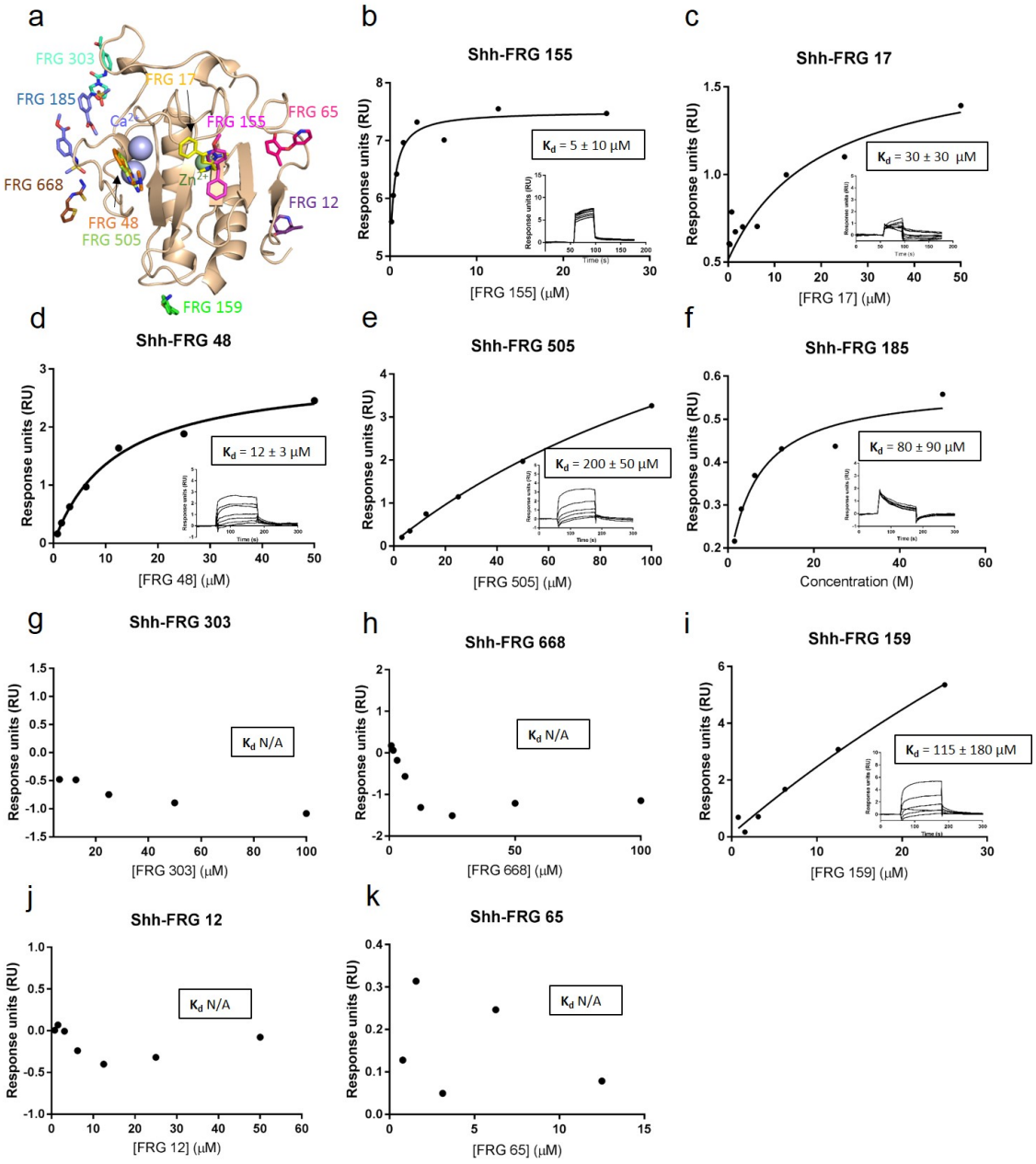


Figure 4.3: **SPR binding screen of ShhN hits.** a) Crystal structure of ShhN with mapped hits identified in the crystallographic fragment screen. ShhN is shown in cartoon representation while the hits are shown as sticks. b-k) SPR equilibrium binding curves and sensorgrams for ShhN intercations with FRG 155, 17, 48, 505, 185, 303, 668, 159, 12 and 65, with respective binding constants.

Fragment-protein interactions cannot be determined as accurately as protein-protein interaction due to their small size and weaker binding affinities of fragments resulting in quite large errors in K_d calculations. The values reported here are averages of multiple runs and should be an indication of a binding affinity range rather than a precise value.

Based on our crystal structures, binding data and costs of the follow-up chemistry, our primary focus was to explore structure-activity relationship of FRG 155 and suggest a possible strategy for lead development. In addition, the fragment was ideally positioned within the pseudo-active site of ShhN in the ShhN-FRG 155 complex structure, which is essential for ShhN interaction with receptors Cdo and Hhip, and antibody 5E1.

4.3.2 Further hit validation of fragment FRG 155

FRG 155 is the first compound that, to our knowledge, binds ShhN with low micromolar affinity, in addition to previously reported ShhN Robotnikinin [197]. Binding of FRG 155 to ShhN was further confirmed using isothermal titration calorimetry (ITC) and the effect on Hh signalling were assessed using a cellular assay (Gli-assay).

Binding affinity of FRG 155 to ShhN was measured in ITC using a concentration ratio of 200 µM of FRG 155 to 20 µM of ShhN24II over 26 injections in solution. The interaction was exothermic with 1.14 binding sites, indicating a 1:1 interaction. The binding affinity was 2 µM, consistent with SPR studies (Figure 4.4, a).

Hh pathway activity was also measured in a QPCR based Gli-assay performed on NIH 3T3 cells (Figure 4.4, b). In this assay, cells are plated onto 96-well plates and are treated with ShhN and proteins or compounds for 24h. After the cells are harvested and lysed, all RNA is reverse-transcribed in RT-PCR step. cDNA fragments corresponding to Gli1, Hh transcription activator, are then amplified and detected with SYBR green fluorescent dye in QPCR amplification step. The amount of Gli1 cDNA present corresponds to the activity of the Hh pathway in cells.

In this assay, the cells were treated with ShhN24II and FRG 155 at a final concentration of 0.5 % DMSO. In the presence of ShhN24II alone the Hh pathway activity increases, while in the presence of the positive control Sant-1, a

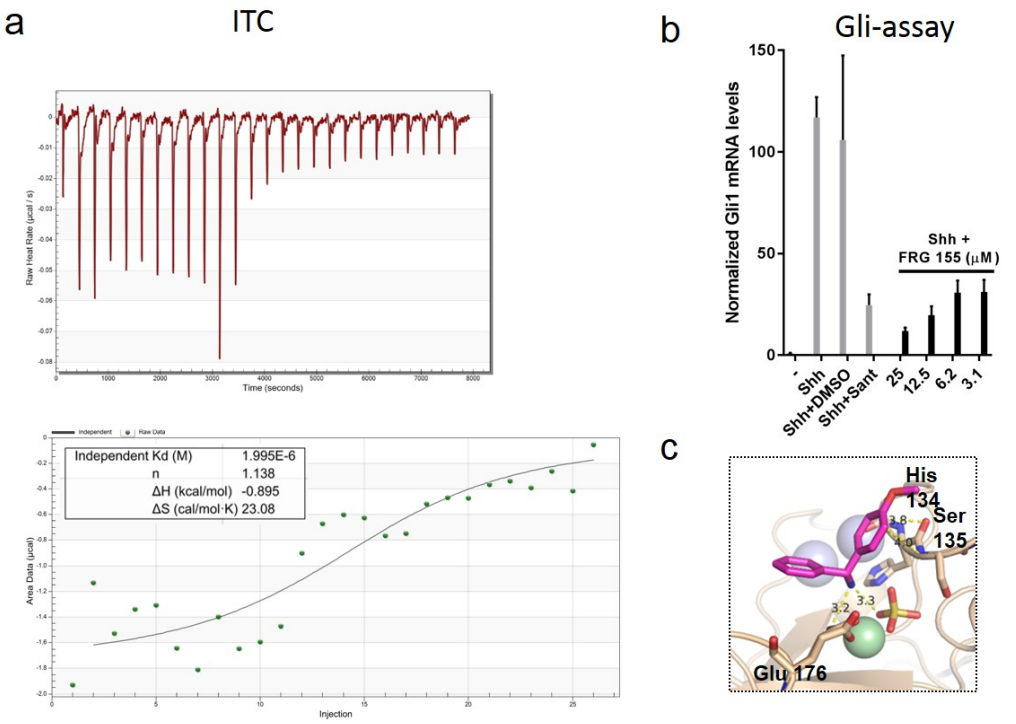


Figure 4.4: **FRG 155 hit validation.** a) The raw and processed ITC data for ShhN24II-FRG 155 binding. The stoichiometry of the interaction was determined as 1:1 with a K_d of 2 µM, consistent with previous SPR results. b) In a Hh-specific Gli-assay, FRG 155 showed concentration dependent inhibition. DMSO and Smo inhibitor Sant-1 were used as controls. c) Crystal structure of the ShhN-FRG 155 binding site is shown. Zinc is shown as a green sphere and FRG 155 as magenta sticks.

Smoothened inhibitor, activity decreases. When treated with FRG 155, the Hh pathway is inhibited in a concentration-dependent manner. These results further validate FRG 155 as a true hit, confirming the binding affinity and demonstrating the inhibitory effect on a cellular level.

4.3.3 FRG 155: structure-activity relationship

We wanted to examine the structure-activity relationship of FRG 155 and propose a strategy for further lead development. For the initial round, 17 commercially available FRG 155 analogues were purchased and tested for ShhN binding using

SPR. The compounds were selected to examine the effects of the amino group, the presence of methoxy groups and the tetrahedral shape of FRG 155. In addition, we wanted to test whether the addition of various functional groups to the amino group would alter affinity and whether we could mimic a merger between FRG 155 and 17 (Figure 4.5, Compound C7). The features of all compounds are summarised in Table 4.1, and their structures and modifications in reference to the original FRG 155 are highlighted in Figure 4.5. Most follow-up compounds fall within the fragment, rather than compound category based on their size (< 300 Da). The majority have favourable properties that obey the rule of three, although with generally lower polar surface area.

Compound	MW (Da)	cLogP	HBA	HBD	Rotable bonds	Polar surface (Å²)
FRG 155	213.3	3.0	2	2	4	35.3
21	183.25	2.9	1	2	3	26.0
20	243.3	3.2	3	2	5	44.5
3	227.3	3.5	2	2	5	35.3
7	215.3	0.8	4	2	4	64.9
27	197.3	2.5	1	2	4	26.0
28	227.3	2.6	2	2	5	35.3
25	241.3	4.1	3	2	5	49.3
30	198.3	2.8	1	1	4	20.2
31	214.3	2.6	2	1	4	29.5
33	244.3	2.8	3	1	5	38.7
2	242.3	2.8	3	1	4	46.5
1	227.3	2.3	3	2	4	49.3
22	197.2	3.2	2	1	3	32.6
29	227.3	2.8	2	2	5	35.3
23	181.2	3.1	1	2	1	26.0
24	195.22	3.3	2	1	1	32.6
26	227.3	2.7	3	3	4	63.3

Table 4.1: List of all FRG 155 analogues used in SAR compounds and their properties. The list includes molecular weight (MW), partition coefficient (cLogP), hydrogen bond acceptors (HBA) and donors (HBD), rotable bond and polar surface.

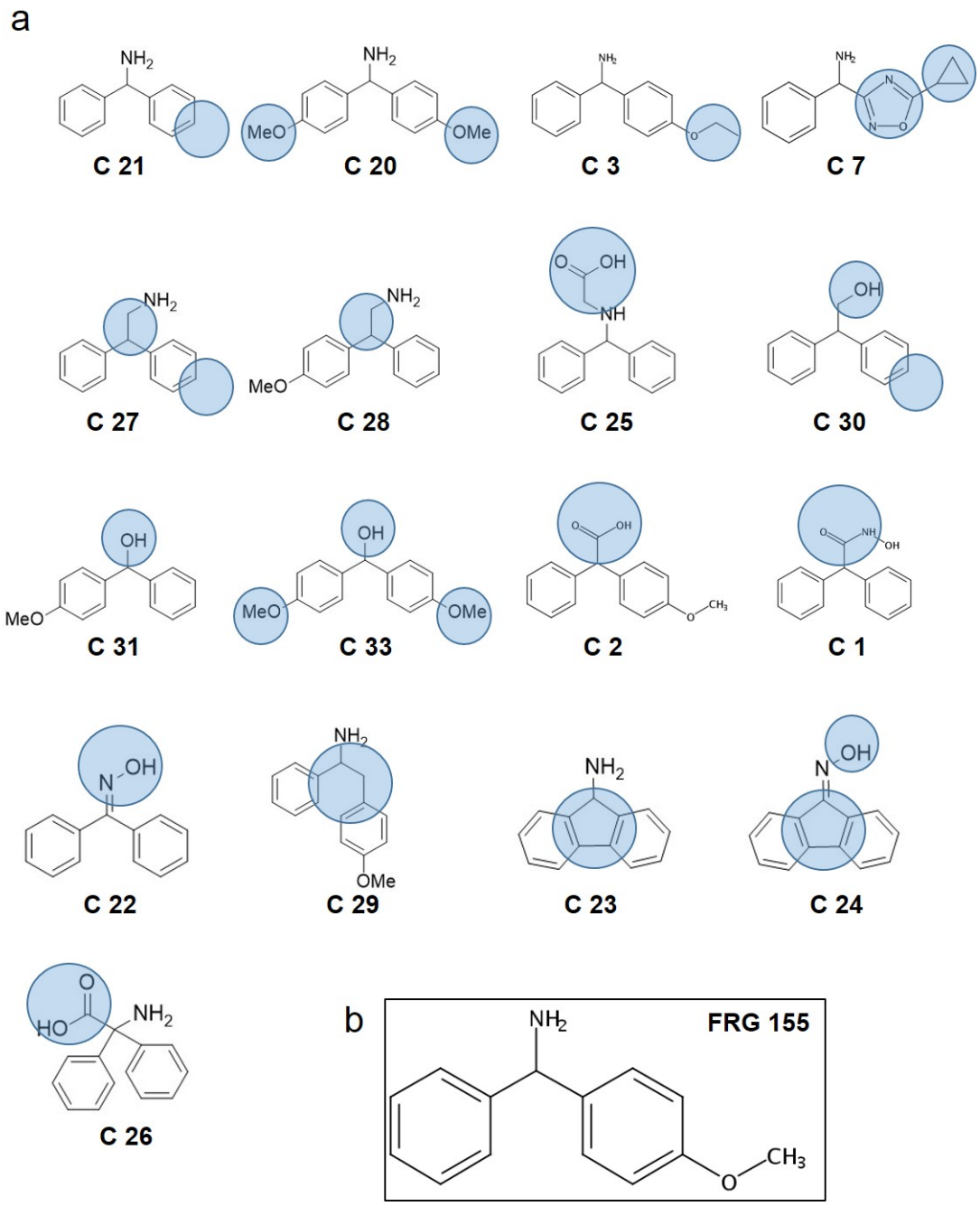


Figure 4.5: **Structures of all FRG 155 analogues used in SAR.** a) Structures of all FRG 155 analogues purchased to be used in SPR assays. The structural differences relative to FRG 155 are highlighted with a blue circle. b) Structure of FRG 155 shown for reference.

Based on the ShhN-FRG 155 crystal structure we would anticipate that the amino group as well as the tetraheadral shape of the compound are essential for binding to maintain all observed interactions. Compounds that do not have these features would be expected to not bind or bind with lower affinity. The SPR affinity screen was done as described before and the binding curves are shown in Figure 4.6 and Figure 4.7.

The most significant findings were the tight binding of compound C 20 (K_d= 5 µM), which had both phenyl groups modified with methoxy groups, and compound C 25 (K_d=4.5 µM), which had an additional carboxyl group as an extension to the amino group. From the crystal structure of the ShhN-FRG 155 complex, it was not clear whether binding is specific to only one of the enantiomers of FRG 155. However, compound C20 is symmetrical and still binds with high affinity indicating that either enantiomer of FRG 155 can bind to ShhN. In addition, it further indicates that both phenyl rings can be elaborated with a methoxy group. The tight binding of C 25 might imply that the fragment can now directly coordinate zinc with a negatively charged carboxyl group and, therefore, generate more interactions. The exact binding modes of each compounds should be confirmed by X-ray crystallography, and a compound that combines the features of the two should be synthesised.

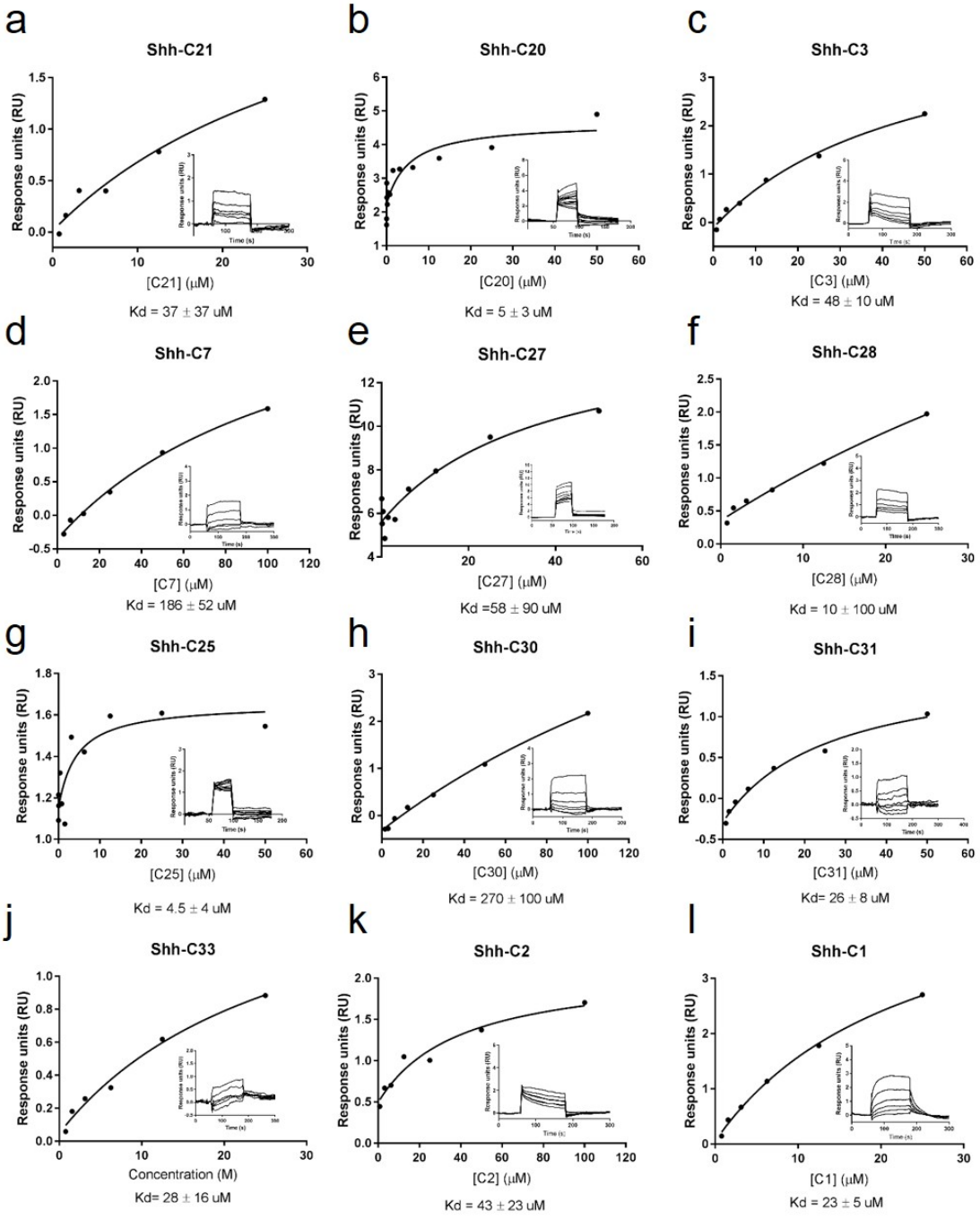


Figure 4.6: **SPR binding screen FRG 155 analogues.** a-l) SPR sensorgrams and binding curves of follow-up compounds with respective binding affinities.

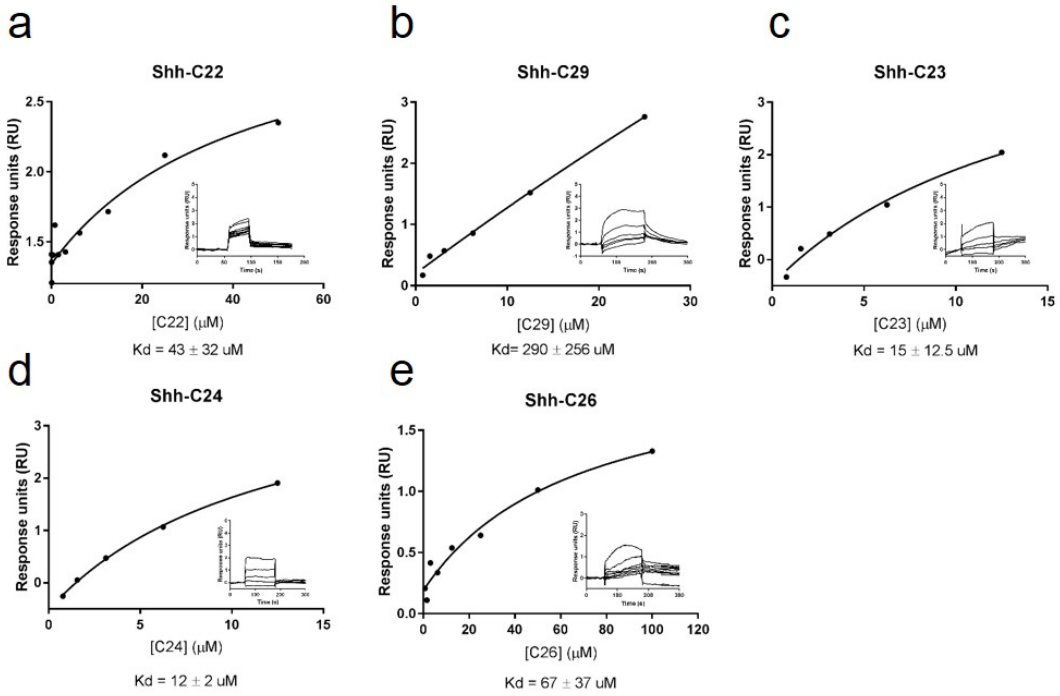


Figure 4.7: **SPR binding screen FRG 155 analogues.** a-e) SPR sensorgrams and binding curves of follow-up compounds with respective binding affinities.

The most significant negative interactions were of compounds: C 7 (K_d= 186 µM), C 30 (K_d= 270 µM), and C 29 (K_d= 290 µM). These results indicated that substituting one of the phenyl rings with oxadiazolidine (C 7), substituting the amino group with a hydroxyl group (C 30) or disturbing the tetrahedral shape of FRG 155 (C 29) all result in a significant decrease in affinity, as expected.

The binding affinities of the remaining compounds to ShhN are lower than that of FRG 155, but not as significantly as compounds C 7, C 30 and C 29. From the binding assay, it appears that any modification to methoxy groups, apart from symmetrical addition of two methoxy groups on each of the phenyl rings, decreases binding affinity to ShhN. Removing the methoxy group from FRG 155 (C21) results in 7x lower affinity (K_d= 37 µM), while methoxy group

extension with another methyl group (C 3) resulted in binding affinity of 48 µM. The presence of the amino group and its position was found to be critical for high-affinity binding. Any substitutions of the amino group (C 31, C33, C 2, C1 and C 22) results in decreased affinity.

Furthermore, if the position of the amino group was altered by the addition of a methylene bridge the affinity also decreases independently of methoxy groups (C 27 and C28). Surprisingly, the flat compounds that had the tetrahedral shape disturbed (C 22, C 23 and C 26) could still bind to ShhN but with decreased affinity. We would expect to observe no binding for these compounds, however, these result do not indicate that the compounds bind to the same binding site as FRG 155.

Further investigation of both positive and negative binders should be performed by ITC, to confirm binding constants, and by co-crystallising the ShhN-C complexes, to confirm the binding site and binding mode.

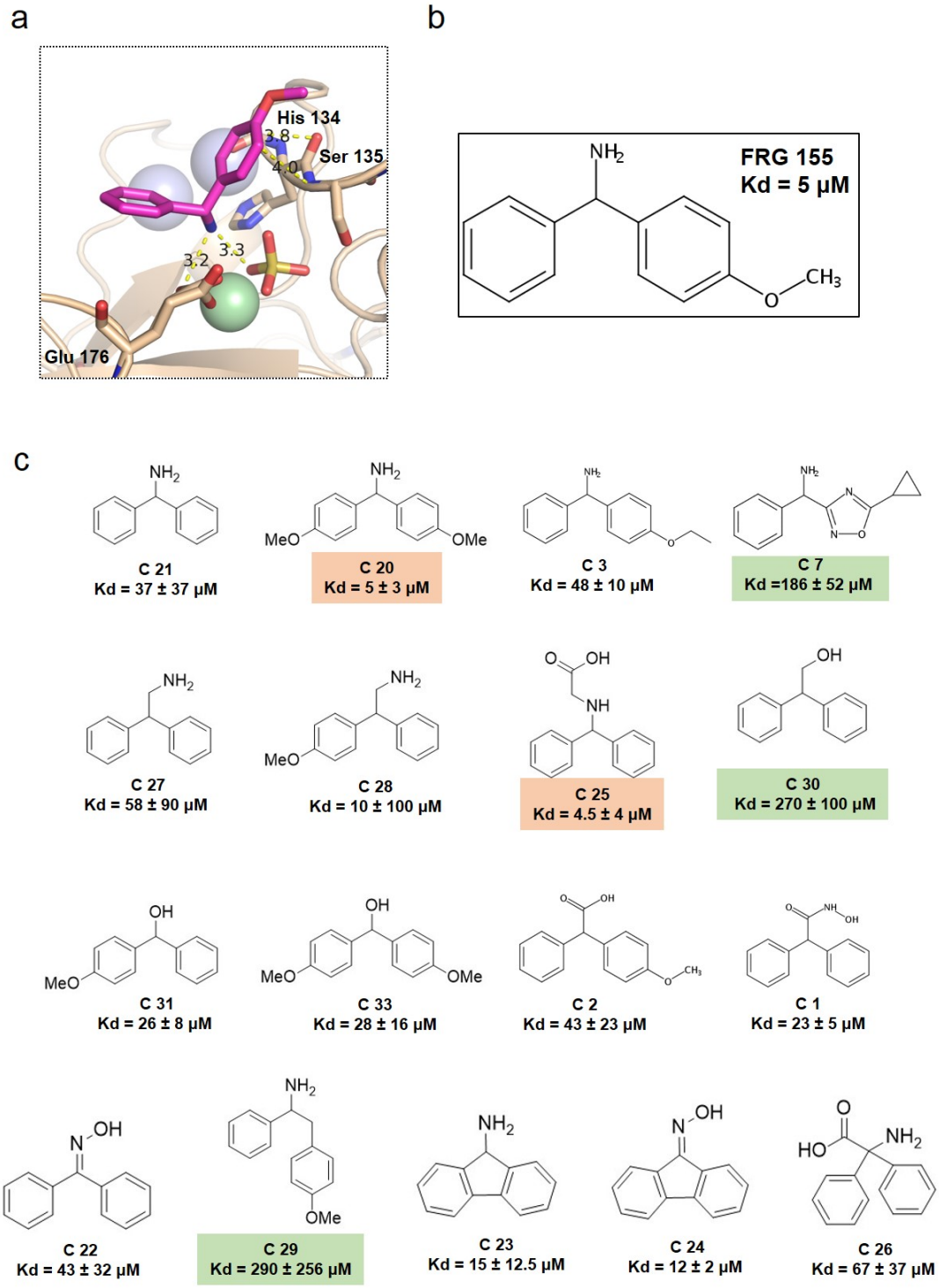


Figure 4.8: **Summary of SAR results.** a) The ShhN-FRG 155 binding site from the crystal structure. FRG 155 is shown as magenta sticks, zinc is shown as a green sphere and sulphate/phosphate is shown as yellow sticks. b) Chemical structure of FRG 155 and ShhN binding affinity. c) Structures of follow-up compounds with ShhN binding affinities obtained in SPR. The tightest binders are highlighted in red and lowest binders in green.

4.3.4 *In silico* lead generation

In addition to commercially available compounds, we wanted to explore synthesis of novel compounds using *in silico* lead generation.

We used Xpoise to generate 8,856 follow-up compounds using FRG 155 as the input. Compounds were generated from 3 reaction filters: reductive amination of the amino group (2,436 compounds), N-alkylation of the amino group (498 compounds) and ether coupling of the methoxy group (two synthons: 2,105 and 3,817 compounds each).

To select the best hits, we performed consensus docking using two different docking programs Autodock and Vina. We docked all the fragments to the ShhN-FRG 155 structure by defining the binding site box around FRG 155 ($15 \times 15 \times 15$ Å). We compared the poses and selected 86 compounds, which had the RMSD between Autodock and Vina < 2 Å (consensus). The binding energies were normalised by dividing each score with the cubic root of heavy atoms of that hit, and the compounds were ranked in order from the lowest binding energy to the highest.

We examined the top five hits by looking at the particular binding mode and compound features. Interestingly, the top three hits contained either a phosphate or a sulphate group, which were also present in the crystal structure of the original FRG 155 hit.

Hit 1 and Hit 3 contain a modified amino group with an additional phosphate and amino sulfone groups, respectively (Figure 4.9, a and c). In comparison to

the binding of the original FRG 155 hit, the amino group and phosphate/sulfone groups overlap with the amino group of the FRG 155 and the sulphate/phosphate group that co-ordinated the zinc in the crystal structure. The overall orientation of the compound changed so that only one phenyl ring aligns to the FRG 155 crystal structure. Both compounds directly coordinate zinc with the phosphate/ amino sulfone groups and form electrostatic interactions with Glu176 via the amino group. The phenyl ring, however, no longer stacks to loop His133-Ser135, instead pointing towards the calcium ions. Hit 5 also contains a modification to the amino group (Figure 4.10, b). The core of the compound is docked in a similar manner to hits 1 and 3, and the modification fits into the binding site. However, no extensive interactions were made by the additional modification.

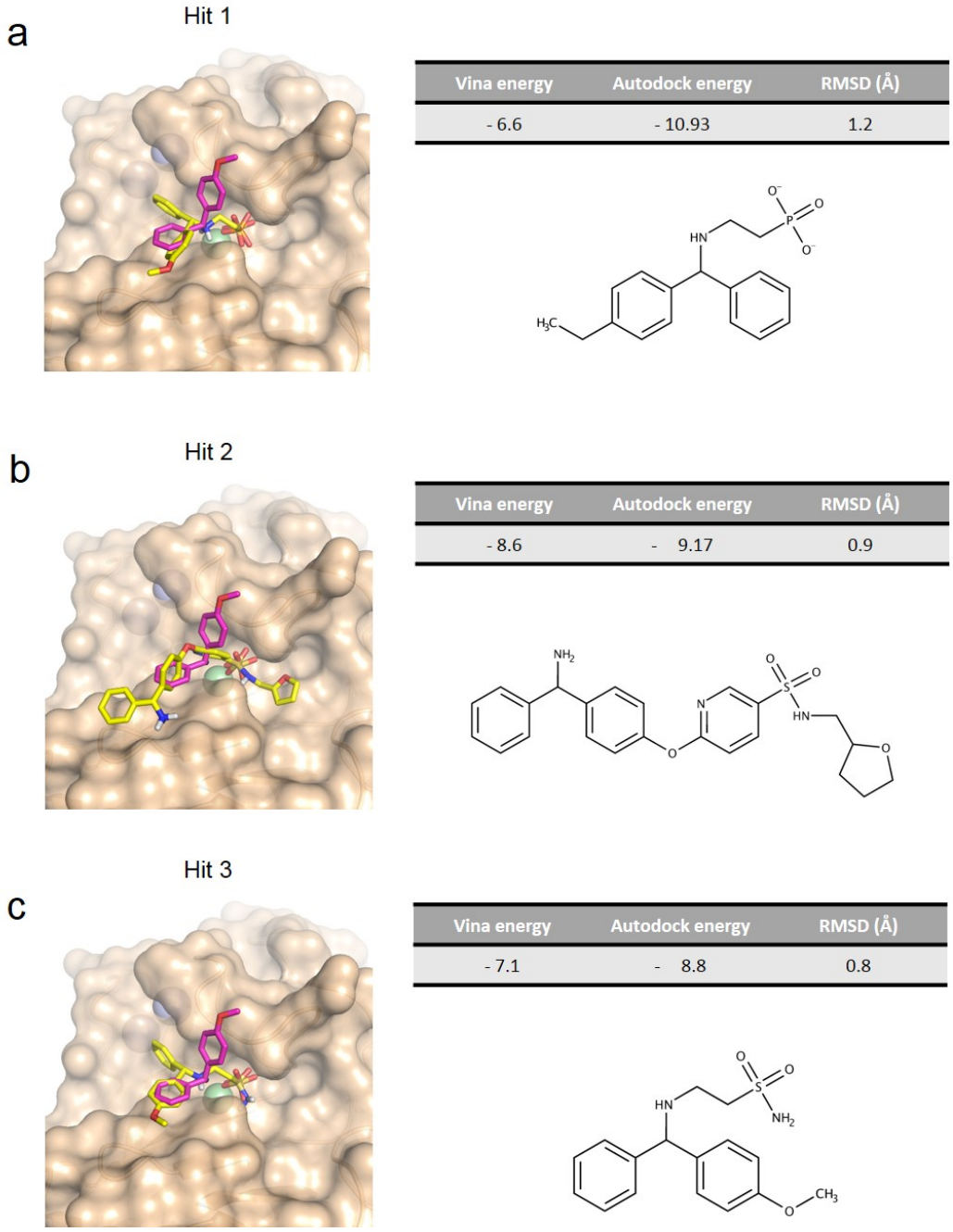


Figure 4.9: **Top 3 hits ranked after consensus docking of Xpoise generated hits.** Hits 1 (a) and 3 (c) contain an additional phosphate and amino sulfone groups, respectively, while Hit 3 (b) has a fragment added to methoxy group. In all cases, sulphate and phosphate moieties coordinate zinc and overlap with sulphate/phosphate group in the ShhN-FRG 155 crystal structure. ShhN is shown as wheat surface, FRG 155 us shown as magenta sticks and docked hits are shown as yellow sticks.

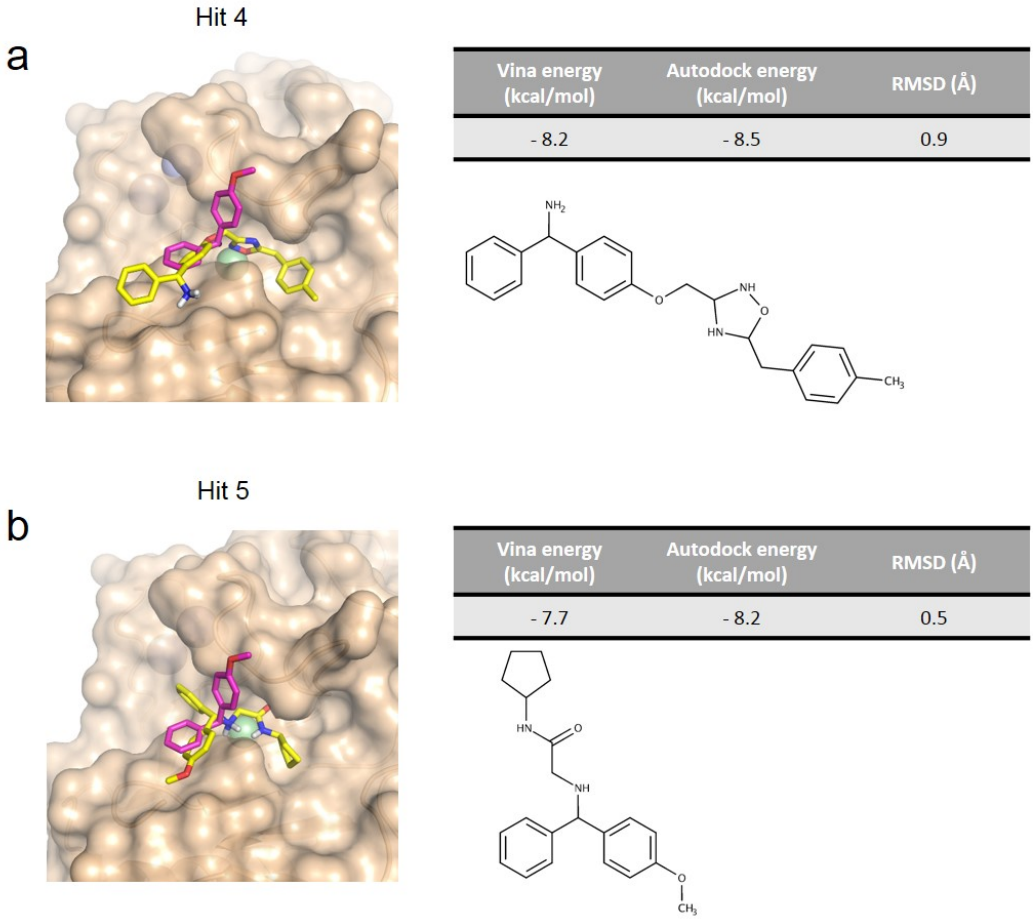


Figure 4.10: **Hits 4 and 5 ranked after consensus docking of Xpoise generated hits.** a) Hit 4 contains a modification to the methoxy group and is displaced from the binding site. b) Hit 5 contains a modification to the amino group that fits into the zinc binding pocket but does not coordinate zinc and does not make significant interactions with the protein. ShhN is shown as wheat surface, FRG 155 is shown as magenta sticks and docked hits are shown as yellow sticks.

Hits 2 and 4 contain more extensive modifications, where an additional fragment has been added to the methoxy group (Figure 4.9, b and Figure 4.10, a). The binding mode is completely different from FRG 155 in the crystal structure for both hits. It appears that the core of FRG 155 has been displaced from the zinc binding site to allow zinc coordination by sulfone group (hit 2) or oxadiazolidine (hit 4).

These results suggest that the most promising modification to the original hit would be the addition of a sulphate or a phosphate group that would directly coordinate zinc, as shown in the crystal structure. In case of hits 1, 3 and 5 one of the phenyl rings is buried in the binding cleft, pointing towards the calcium binding site instead of stacking against residues His133-Ser135, like FRG 155 in the ShhN-FRG 155 crystal structure. This might suggest that modified compounds have a different binding mode that could allow modifications to the phenyl ring to extend the compound towards the calcium binding site.

4.4 Discussion

In this chapter, we tested 10 out of 13 hits identified in the crystallographic fragment screen. We established that FRG 17, 155, 48, 505, 185 and 505 are all true ShhN binders, binding with a K_d of 30, 5, 12, 200, 80, 115 µM, respectively, while FRG 668, 12 and 65 did not bind ShhN.

The tightest identified binder, FRG 155, bound directly across the ShhN zinc binding site and had a K_d of 5 µM in a SPR equilibrium binding experiment. This is, to our knowledge, the only small molecule, with the exception of Robotnikinin, that binds to ShhN with low micromolar affinity. Furthermore, FRG 155 binds more efficiently than Robotnikinin resulting in a similar binding affinity with approximately half of the molecular weight. The binding of FRG 155 was further confirmed in solution using ITC resulting in a 1:1 stoichiometry and a K_d of 2 µM. The hit also had a dose-dependent inhibitory effect in a cellular Hh signalling assays, although further experiments need to be done with a broader concentration spectrum.

In the preliminary SAR study, 17 commercially available analogues of FRG 155 were screened in SPR binding assays (Figure 4.11). Findings indicated that the amino group in the same position as in FRG 155 is essential for binding and substitution with another group, such as a hydroxyl group (C 31, 33), results in decreased binding. The tetrahedral shape also proved to be of crucial importance as shown in lower binding affinities of C 29 and C 26 that impair the tetrahedral arrangement. Both findings are consistent with the ShhN-FRG 155 crystal

structure, where the amino group and a tetrahedral shape are important for binding. The tightest binders were C 20 and C 25, which contained methoxy groups on both phenyl rings and a carboxyl group after the amino group, respectively. The lowest affinities were obtained for C 7 (phenyl substitution), C 30 (amino substitution) and C 29 (flexible shape) compounds.

Our current and future work is focused on verifying the two most promising hits C 20 and C25 in ITC and in cell assays to confirm the binding constants and examine the effect on Hh signalling. Our objective is also to confirm the binding site for these compounds and determine the binding mode of each compound by co-crystallisation. We are also pursuing further investigation into the binding of two negative control compounds C 23 and C 33 in ITC, and are attempting to co-crystallise the complex to observe the effect of modification on binding mode. In particular, we would expect that C 23 compound no longer binds to the zinc

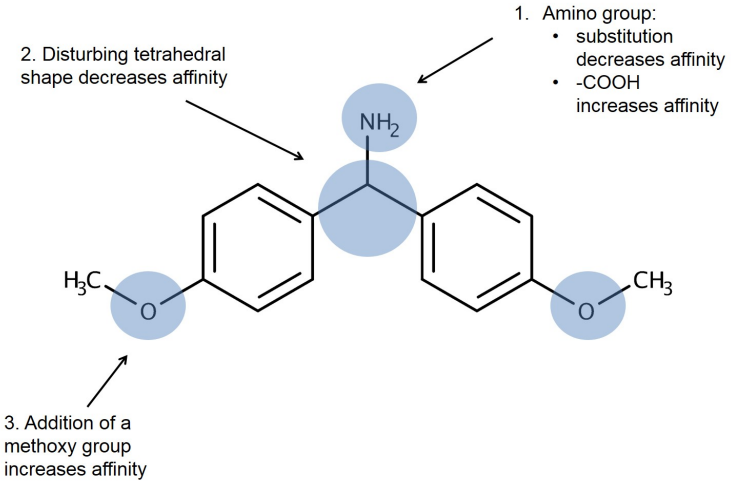


Figure 4.11: **Summary of SAR findings.** The substitution of the amino group as well as a lack of the tetrahedral shape of FRG 155 results in lower affinities. The addition of a carboxyl group to the amino group, and methoxy groups on both phenyl rings result in high affinity binding.

binding site. In the absence of structural information, we are exploring the use of NMR for small molecule-protein interactions via the Instruct network. We are also currently purchasing and synthesising the second round of SAR compounds including a cross-over compound (C 20/C 25), in collaboration with Prof. Angela Russell and OxXChem group. These compounds will be screened in SPR assays and further characterised structurally if needed.

Our work also included *in silico* leads generation, which we selected and ranked via consensus docking. The top hits contained modifications to the amino group or methoxy group. The hits that contained sulphate or phosphate groups scored highly and overlapped with the crystal structure. Further docking analysis with more sophisticated software like Schrödinger will allow us to identify the most efficient hits. We are also pursuing the synthesis and *in vitro* characterisation of these compounds, which would allow us to compare the sulphate and carboxyl group modification to the amino group.

Although the compounds being screened are still expected to have a degree of non-specific and off-target binding, future work will also include tests for specificity. In particular, we want to be able to measure whether the compounds that we are producing are specific to ShhN or whether these bind to any other zinc-containing protein. For this, we will test binding and the effects of our compounds on zinc-dependent histone deacetylases (HDACs) in collaboration with SGC Toronto.

In this chapter, we validated hits from the crystallographic fragment screen and determined that one of our priority hits FRG 155 binds with low micromolar

affinity. An initial round of SAR using commercially available compounds and *in silico* lead generation and docking indicated important features for ShhN-FRG 155 interaction and established a good starting point for future lead generation attempts.

Chapter 5

Structural and functional characterisation of Skinny Hedgehog

5.1 Abstract

Skinny Hedgehog (Skinny Hh) is a membrane-bound acyltransferase essential for the palmitoylation of HhN ligands and proper Hh signalling. It catalyses the formation of an amide bond between HhN Cys24 and a palmitoyl moiety via an unknown mechanism. Recent studies have revealed the topology of Skinny Hh, which consists of 10 transmembrane domains and two re-entrant loops, and have identified essential catalytic residues. Furthermore, the development of novel palmitoylation assays led to the discovery of more potent and specific inhibitors of Skinny Hh. However, the molecular function of Skinny Hh remains unknown due to difficulties of studying membrane proteins using structural biology approaches.

In this chapter, protocols for expression and purification of Skinny Hh have been optimised to produce milligram quantities of the pure and stable Skinny Hh. The full-length Skinny Hh construct was found to be exceptionally active in *in vitro* palmitoylation assays. Additive screens have been conducted to improve the thermostability of Skinny Hh for crystallisation trials. Construct engineering was used to identify regions that affect the stability and activity of Skinny Hh, most significant being the first luminal loop which is essential for activity. Furthermore, the effect of inhibitors on Skinny Hh thermostability was examined. The first identified RUSKI inhibitors were found to be poor stabilisers. However, a novel TLH 45 inhibitor improved stability significantly. The work described here provides a basis for structural characterisation and drug screening for novel inhibitors of Skinny Hh.

5.2 Introduction

The active Hh morphogen is synthesised as a precursor protein and is transported into the ER where the signal sequence (residues 1-23) is cleaved off. The precursor protein is then autocleaved into an N-terminal signalling domain (HhN), that has been shown to be sufficient to mediate all Hh signalling functions (Figure 5.1). The cleavage via an intein-mediated reaction is catalysed by the Hh C-terminal domain (HhC) (Figure 5.1). In the same reaction a cholesterol molecule is attached to the C-terminus of HhN, while the N-terminal Cys24 residue is palmitoylated. Skinny Hh is the membrane-bound acyltransferase enzyme responsible for the palmitoylation of HhN in the ER and Golgi [44, 46, 238].

Skinny Hh belongs to a subgroup of membrane-bound *O*-acyltransferase (MBOAT) group of enzymes [239]. MBOATs can acylate sterols, lipids or peptides and proteins, and are predicted to have 8-12 transmembrane domains. They have characteristically conserved catalytic residues consisting of an invariant histidine residue surrounded by hydrophobic amino acids, and an asparagine/aspartate residue situated in a partially hydrophobic environment [240]. Mutations of these active site residues result in reduced enzymatic activity across all MBOAT family members, implying that these residues are essential for catalysis [44, 241, 242, 243, 244, 245]. In addition to Skinny Hh, MBOAT proteins that can acylate secreted proteins include Porcupine [246], which palmitoyleoylates Wnt morphogen proteins, and Ghrelin O-acyltransferase (GOAT) [245], which catalyses octanoylation of ghrelin, an appetite regulating hormone. Even though

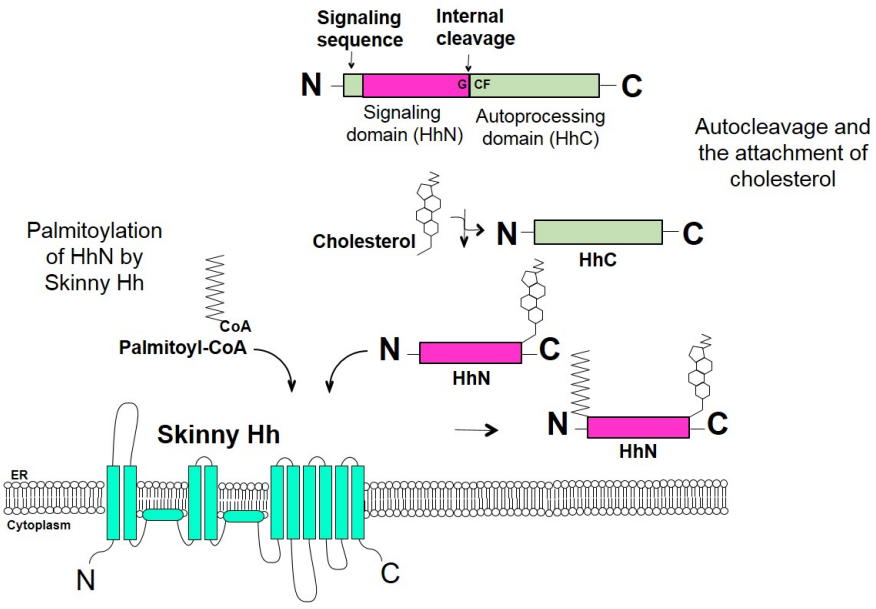


Figure 5.1: **Processing of the Hh precursor protein.** The Hh precursor is transported into the ER where the signal sequence is cleaved off to expose the N-terminal Cys. The protein is then autocleaved to produce the N-terminal signalling domain HhN. Cholesterol is attached to the C-terminus of HhN, while the N-terminal is palmitoylated by Skinny Hh before the protein is secreted.

their overall topology is thought to be similar, the structures and molecular mechanism of MBOATs remain unknown.

Skinny Hh has been discovered in different species and has been referred to using various names including Central missing [53], Rasp [247] and Sightless [248] in *Drosophila*, and Hhat [46] in vertebrates. In both flies and mammals, palmitoylation is essential for proper Hh signalling. Skinny Hh or Rasp knockout closely resemble the phenotype of Shh Cys24Ser mutant described in the Chapter 1, Section 1.2.3. [72]. The *Drosophila* homologue Rasp is essential for Hh maturation and proper Hh signalling in flies [44, 53, 248]. Murine embryonic fibroblasts (MEF) lacking Skinny Hh have significantly less palmitoylated Shh while Skinny

Hh knock-out mice have reduced Hh signalling in the limb and neural tube [72]. Loss-of-function mutations of Skinny Hh result in severe developmental defects including holoprosencephaly with acrania and agnathia [82], and disorder of sex development [249].

In addition to Hh, numerous other proteins are palmitoylated: intracellular proteins like Src tyrosine kinases [250, 251] and Ras GTPases [252], and secreted extracellular proteins such as Wnt morphogen ligand, which is S-palmitoyleoylated by Porcupine [246], and the *Drosophila* epidermal growth factor-like protein Spitz, which is N-palmitoylated by Rasp [253]. Although Skinny Hh appears to be specific to Hh ligands, another 40 proteins or peptides have been identified as potential substrates in a bioinformatics analysis [254]. None, however, have been verified experimentally.

5.2.1 N-palmitoylation reaction

Skinny Hh localises to the ER and Golgi membranes and catalyses the attachment of palmitic (C16:0) fatty acid from palmitoyl-CoenzymeA (P-CoA) to the first N-terminal residue of HhN Cys24, both *in vivo* and *in vitro* [46]. The unique aspect of this reaction is the formation of an amide bond between the palmitate moiety and the Cys24. Most palmitoylation reactions in cells are formed via S-palmitoylation whereby a palmitate is attached to the side chain of a Cys residue via a thioester bond. This is a reversible reaction that can lead to cycles of palmitoylation and depalmitoylation similar to phosphorylation [255].

In contrast, Skinny Hh attaches the palmitate to the N-terminus of Cys24 via an amide linkage. This reaction is irreversible and can withstand treatment with hydroxylamine. Although the exact mechanism of N-palmitoylation remains unknown, two hypotheses exist: an S-to-N shift, or direct N-palmitoylation (Figure 5.2). The S-to-N shift is a two-step reaction where palmitate is first linked to the side chain of Cys24 via a thioester linkage before being transferred to the

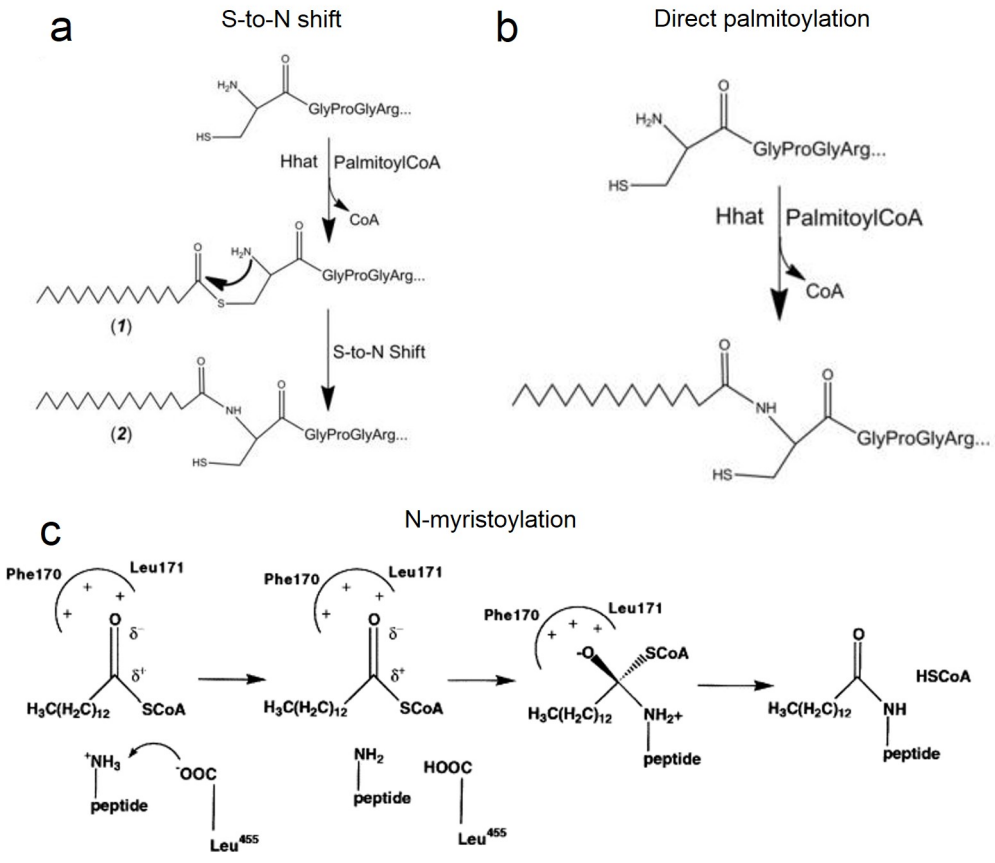


Figure 5.2: **Potential mechanisms of N-palmitoylation by Skinny Hh.** a) In the S-to-N shift reaction, palmitate is first attached to the side chain of Cys via a thioester bond, after which it is transferred to the N-terminus to form an amide bond. b) Direct palmitoylation would involve the attachment of palmitate to the N-terminus in a one step reaction. This would happen similarly to N-myristoylation (c) where the N-terminal residue and myristoyl-CoA are activated allowing for electrophilic attack by NH_2 and the formation of an amide bond. Adapted from [238] (a and b), and [256] (c).

amino terminal via an S-to-N shift to form an amide bond [257, 45] (Figure 5.2, a). Direct palmitoylation would likely proceed similarly to N-myristoylation, where the acyl-CoA moiety and the N-terminus are activated, which allows for electrophilic attack and direct formation of an amide bond [258, 259] (Figure 5.2, b and c). Even though the precise mechanism remains unknown, no thioester intermediates or double palmitoylated HhN have been identified to support the S-to-N shift [46]. Skinny Hh can palmitoylate both the Hh precursor or HhN, indicating that palmitoylation is independent of autoprocessing and attachment of cholesterol [46].

Palmitoylation of HhN can be conducted *in vitro* in a low salt condition at pH 6.5, in the presence of P-CoA [46, 260]. Skinny Hh does not bind free fatty acids. However, it can use a range of fatty acyl-CoAs as a substrate [46]. In fact, ShhN ligands modified with a number of different fatty acids have been identified in cells grown in different conditions [261]. The prevalence of palmitoylated HhN is thought to be the result of higher P-CoA levels in the ER relative to other acyl-CoA. Skinny Hh is specific for Hh proteins and the minimal sequence requirement includes the initial 6 amino acids of HhN [254]. Mutations of amino acids do not have an effect on palmitoylation. However, a positively charged residue is required within the first 7 amino acids but not adjacent to Cys24 [254]. Interestingly, while the first Cys residue is essential for palmitoylation in *Drosophila*, Skinny Hh can palmitoylate both Cys and Ser residues. However, only 25-50% of the Hh is palmitoylated in the Cys24Ser mutant [254, 136].

5.2.2 Skinny Hh topology

Skinny Hh was predicted to have 10-12 transmembrane loops. However, the exact topology could not be predicted using algorithms alone. The combination of TM prediction analysis and membrane permeabilisation studies concluded that Skinny Hh has 10 transmembrane domains and 2 re-entrant loops with cytoplasmic N- and C-termini (Figure 5.3) [262, 263, 264].

Residues in the loop between TM 9 and TM 10 were predicted to contain a nucleotide-binding Walker motif (G-xxxx-GK-[TS] consensus sequence) [265]. Although the role of the Walker motif for Skinny Hh is not known, it might be involved in binding the phosphoadenosine of its substrate palmitoyl-CoA [266]. The C-terminal domain was found to contain the conserved catalytic residues

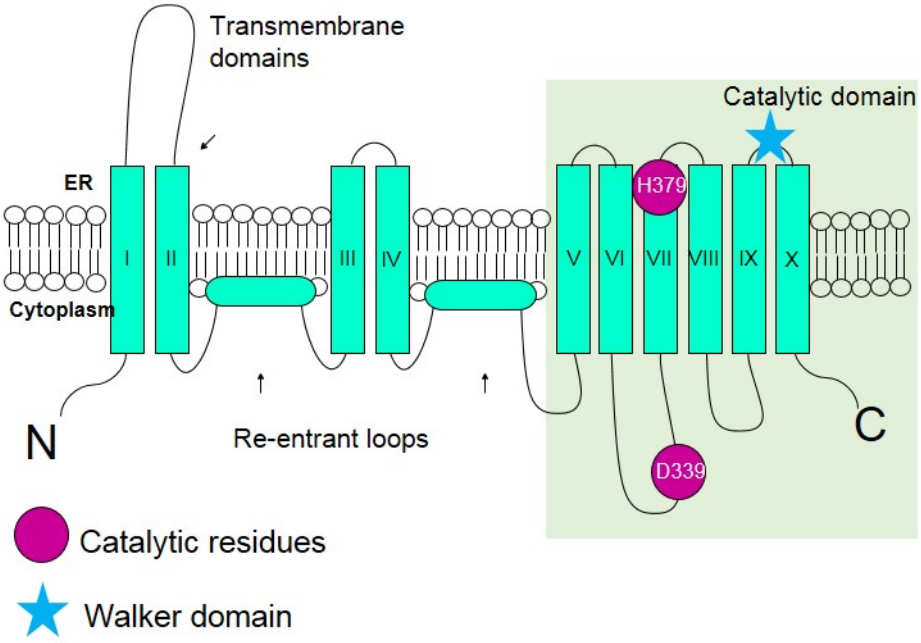


Figure 5.3: **Skinny Hh topology.** Skinny Hh is comprised of 10 transmembrane helices and 2 re-entrant loops. The C-terminal domain contains two key catalytic residues His379 and Asp339, and a predicted nucleotide-binding Walker domain. The exact role of the remaining portion of the protein is not known.

His379 located at the luminal verge of TM 7, and Asp339 in the cytoplasmic loop between TM 6 and TM 7 [263]. While mutation of His379 results in a 50% decrease in activity, mutations of Asp339 practically abolished activity [263]. Mutation of residues Gly217, Phe338 and Trp378 affect the activity, but not the stability of Skinny Hh [262]. Residues 194-233, which include the second re-entrant loop, are conserved within the MBOAT family. However, the role of this region is not known [262]. Skinny Hh is itself palmitoylated on 4 cytosolic Cys residues, which are thought to have an effect on the stability and conformation [263].

5.2.3 Targeting Skinny Hh

The Hh pathway has been implicated in a range of cancers including medulloblastoma, basal cell carcinoma, pancreatic, lung, breast and many other cancers [203]. Inhibition of the Hh pathway is now a validated approach to inhibiting cancer growth and metastasis. The discovery of ligand-dependent mechanism for Hh cancers led to the exploration of Hh targets other than Smo, such as Skinny Hh, to be targeted in cancer.

The first small molecule inhibitors of Skinny Hh were found in a peptide-based assay of 63,885 unique compounds, resulting in identification of 4 so-called 'RUSKI' inhibitors: 39, 41, 43, 50 (Figure 5.4) [198]. RUSKI-43 was shown to inhibit palmitoylation of ShhN *in vitro* and in cells, and was shown to reduce Hh signalling in fibroblasts and pancreatic cancer cell line [267]. Despite its apparent

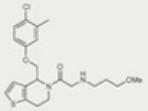
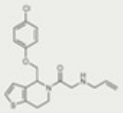
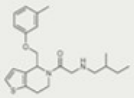
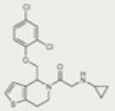
Number	Compound ID	Structure	IC_{50} (μM)
1	RU-SKI 39		3.07
2	RU-SKI 41		0.18
3	RU-SKI 43		0.85
4	RU-SKI 50		5.44

Figure 5.4: **First small-molecule inhibitors of Skinny Hh.** Inhibitors were identified in a peptide-based screen against Skinny Hh. The best hits included RUSKI 39, 41, 43 and 50. Adapted from [198]

specificity, RUSKI-43 was shown to have off-target cytotoxic effects [260, 199].

Novel inhibitors in the RUSKI series were developed including RUSKI 101 and 201 [268, 269]. In particular, RUSKI-201 was the most potent, exhibiting specificity and no cytotoxicity, and was shown to reduce Hh signalling in breast, lung and pancreatic tumour cell lines making it a more potent and less toxic alternative to RUSKI-43 (Figure 5.5) [199, 270]. In addition, Skinny Hh knock-down in pancreatic and non-small cell lung cancer cells resulted in decreased Hh paracrine signalling [201].

In this chapter, the focus is on Skinny Hedgehog, an enzyme that became a prominent Hh target due to its importance in HhN palmitoylation. Overall, little is still known about the structure and mechanism of Skinny Hh primarily due to common difficulties of working with membrane proteins. Although recent advances have made it possible to isolate small amounts of Skinny Hh [46],

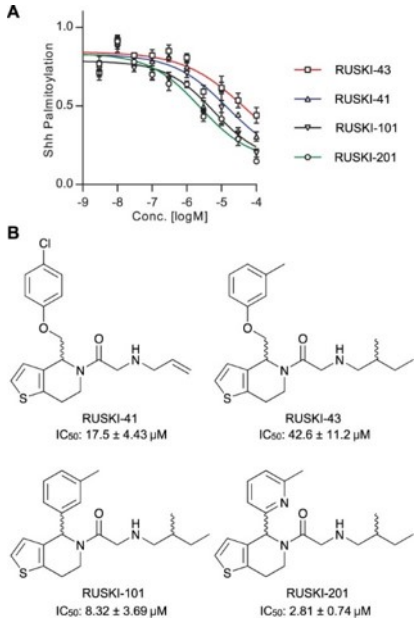


Figure 5.5: **Structure and potency of novel RUSKI inhibitors** a) ShhN palmitoylation assay in the presence of 4 inhibitors, including the novel RUSKI 101 and 201, which were more potent in inhibiting Skinny Hh. b) Structures and corresponding IC_{50} of RUSKI compounds. Adapted from[260] .

and more sophisticated microfluidic mobility shift (MSA) palmitoylation assays have been developed [260], protein yield remains an issue for structural and drug screening studies. The goal of this chapter was to address this issue and optimise the production and stability of Skinny Hh. Further aims include determining the structure of Skinny Hh using X-ray crystallography and performing a 100,000 compound drug screen. These findings would not only help us understand the mechanism of Skinny Hh and MBOATs, but would aid in the development of novel more potent inhibitors.

5.3 Results

5.3.1 Construct expression and detergent screen

Skinny Hh from fly and zebrafish share 20% and 48% amino acid sequence identity, respectively, with the human Skinny Hh (Appendix, Section 1.1). All three proteins were tested in expression trials because of their sequence diversity.

H. sapiens (human, h), *D. melanogaster* (fly, f) and *D. rerio* (zebrafish, z) full-length (FL) *ski* genes were cloned into pHL-sec vectors containing either a Rho1D4 (1D4) tag or a 1D4-mono Venus (mV) tag [271]. Proteins were expressed in a small-scale HEK293T cell culture, purified and analysed on a western blot (WB) and fluorescence size-exclusion chromatography (FSEC) (Figure 5.6). Both hSkinny Hh constructs (1D4 and mV tagged) could be expressed. However, the

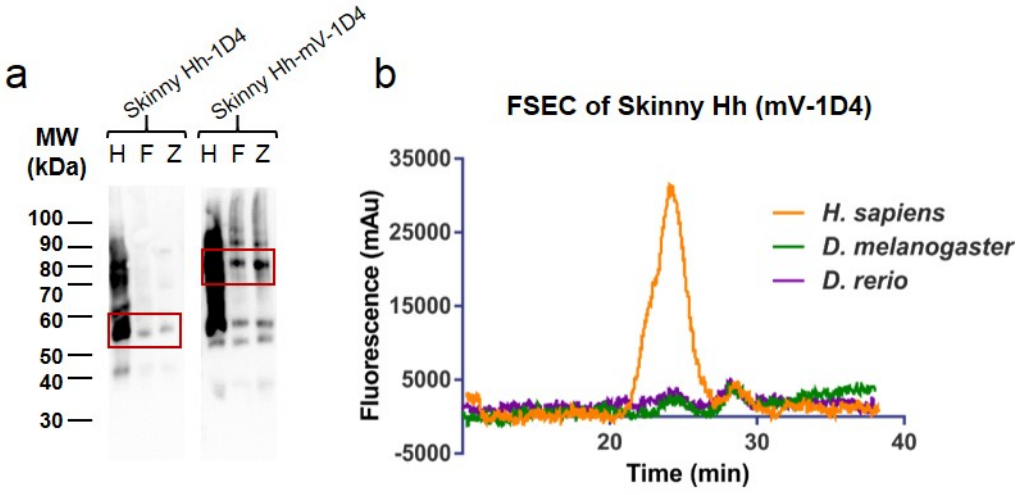


Figure 5.6: **Expression test of Skinny Hh proteins.** a) Western blot showing the expression of full-length *H. sapiens* (human, h), *D. melanogaster* (fly, f) and *D. rerio* (zebrafish, z) containing 1D4 and mV tags. Bands corresponding to FL-Skinny Hh are highlighted in a red box. b) FSEC trace of hSkinny Hh, fSkinny Hh and zSkinny Hh after purification. Only hSkinny Hh expressed well and ran as a monodispersed peak.

expression is higher in the presence of mV tag. hSkinny Hh (mV) protein ran as a single peak in FSEC. Fly and zebra fish constructs expressed poorly, therefore hSkinny Hh was used in all further experiments.

Next, FL-hSkinny Hh (mV) was expressed in HEK293T small-scale culture and purified using different detergents to test for suitable solubilisation conditions. The following detergents were tested: Octyl Glucose Neopentyl Glycol (OGNG), Decyl Maltose Neopentyl Glycol (DMNG), *n*-dodecyl-β-D-maltopyranoside (DDM), Neopentyl Glycol (NG) and Octyl β-D-glucopyranoside (β-OG), alone and in the presence of 20:1 and 40:1 (molar ratio) cholesterol hemisuccinate (CHS) in 20 mM Hepes pH 7.2, 300 mM NaCl buffer (MB buffer). The purified samples were analysed in FSEC (Figure 5.7). NG and β-OG failed to solubilise hSkinny Hh (Figure 5.7, d and e), while DDM, DMNG and OGNG were able to extract hSkinny Hh from the membrane and produce a monodispearse peak in gel filtration (Figure 5.7, a-c). In the initial screen, OGNG 40:1 CHS produced the highest yield and a monodispearse peak, and was used as a detergent of choice for purifying hSkinny Hh in subsequent experiments.

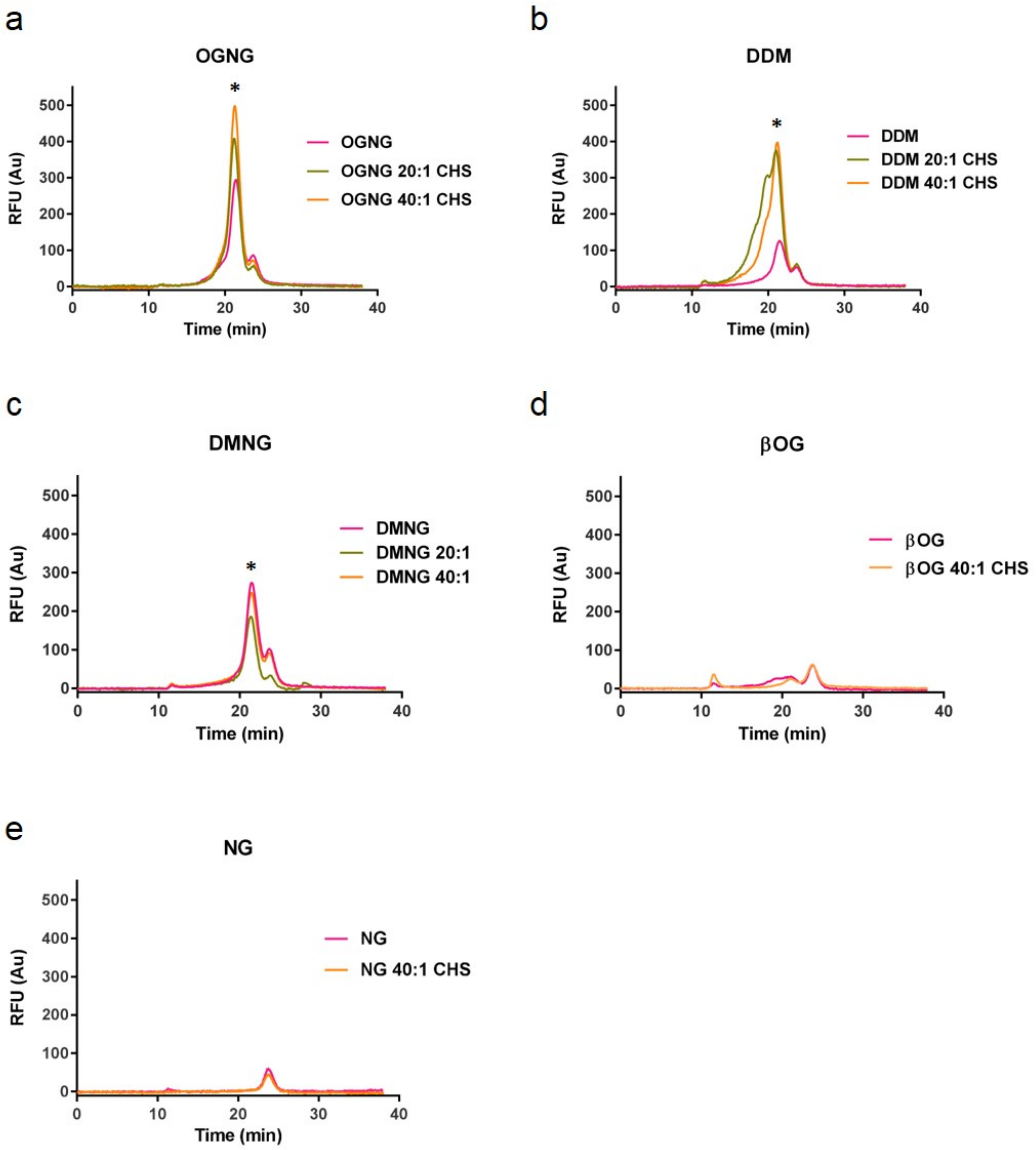


Figure 5.7: **FSEC traces from the detergent screen for FL-hSkinny Hh (mV).** Skinny Hh was purified in detergents OGNG (a), DDM (b), DMNG (c), β-OG (d) and NG (e), with and without 20:1 and 40:1 cholesterol hemisuccinate (CHS), and was analysed in FSEC. The initial screen indicated that a combination of OGNG and CHS at a ratio of 40:1 was the best condition for solubilising FL-hSkinny Hh. The peak corresponding to FL-hSkinny Hh is marked with an asterisk (*).

5.3.2 Thermostability and additive screen

Protein stability is one of the key factors that contribute to protein crystallisability [272]. The stability of FL-hSkinny Hh (1D4) was tested by incubating protein samples at different temperatures for 1 h and analysing them on SEC. FL-hSkinny Hh produces a single monodispearse peak in SEC at 4°C. Gradual denaturation of Skinny Hh at higher temperatures results in protein precipitation and a decrease in the peak in SEC. The absorbance of the peak height that corresponds to FL-hSkinny Hh was plotted against the temperature to produce a 'denaturation curve'. FL-hSkinny Hh was first incubated at 4-80°C resulting

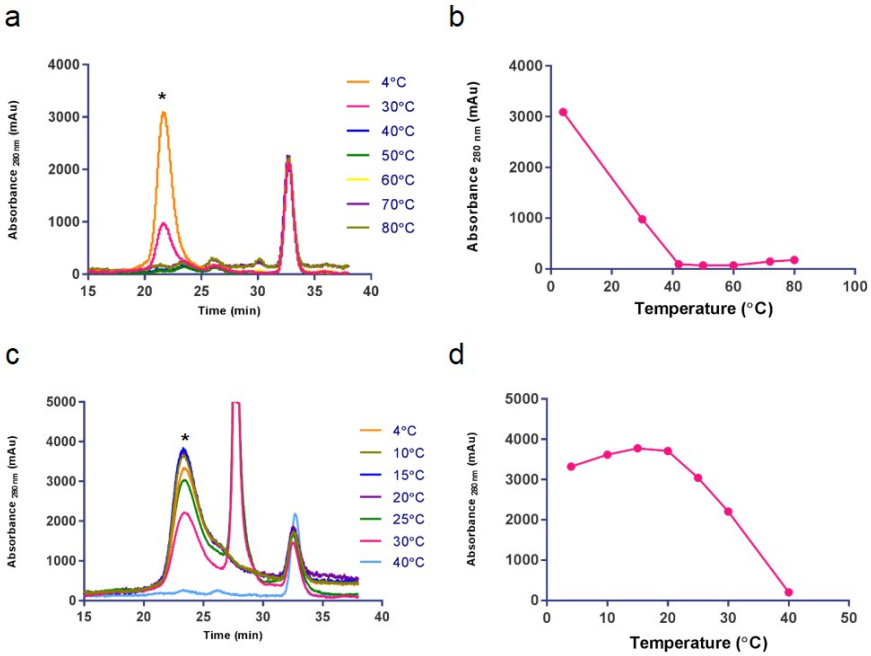


Figure 5.8: **SEC-based thermostability assay of FL-hSkinny Hh (1D4).** a) and b) Protein samples were incubated at different temperatures and analysed on FSEC, indicating that denaturation of FL-hSkinny begins to occur at 30°C. c) and d) Further analysis of hSkinny indicated that the protein is stable up to $\sim$ 20°C. Peak corresponding to FL-hSkinny is marked with an asterisk (*).

in partial and complete denaturation at 30°C and 40°, respectively (Figure 5.8, a and b). The protein stability was further tested in the range 4-30°C, indicating that FL-hSkinny Hh-1D4 (Skinny Hh-1D4) is stable at up to $\sim$ 20°C. The approximate melting temperature (Tm) corresponding to 50% absorbance signal was $\sim$ 30°C (Figure 5.8, c and d).

One approach to stabilising proteins is the addition of additives. The stability of Skinny Hh-1D4 was screened in the presence of 96 additives prepared for membrane proteins, which included detergents, lipids, pH conditions, salts, sugars, metals and other chemicals (in-house prepared conditions are shown in the Appendix, Section 1.3). The initial screen involved mixing Skinny Hh-1D4 with additives and incubating them at 30°C for 1h followed by SEC analysis. The following detergents improved the stability of Skinny Hh-1D4: LMNG, CYMAL-6-NG and CYMAL-7-NG, and lipid POPC in DMNG. These were further analysed in a full-temperature screen (Figure 5.9). This indicated that LMNG, CYMAL-7-NG and POPC were the most promising additives for Skinny Hh-1D4 stabilisation (Figure 5.9, a and c).

We further examined the effects of Skinny Hh substrates: P-CoA, non-hydrolysable P-CoA analogue (16:0 Ether Coenzyme A), Shh peptide (residues 24-38, CGPGRGFGKRRHPKK) and the product, palmitoylated Shh peptide (palmitoyl-CGPGRGFGKRRHPKK), and their ability to stabilise Skinny Hh at 30°C (Figure 5.10). Surprisingly, Shh peptide did not stabilise Skinny Hh, nor did the palmitoylated Shh peptide (Figure 5.10, a and b). Despite these findings, both peptides could still be binding to Skinny Hh but not stabilise its struc-

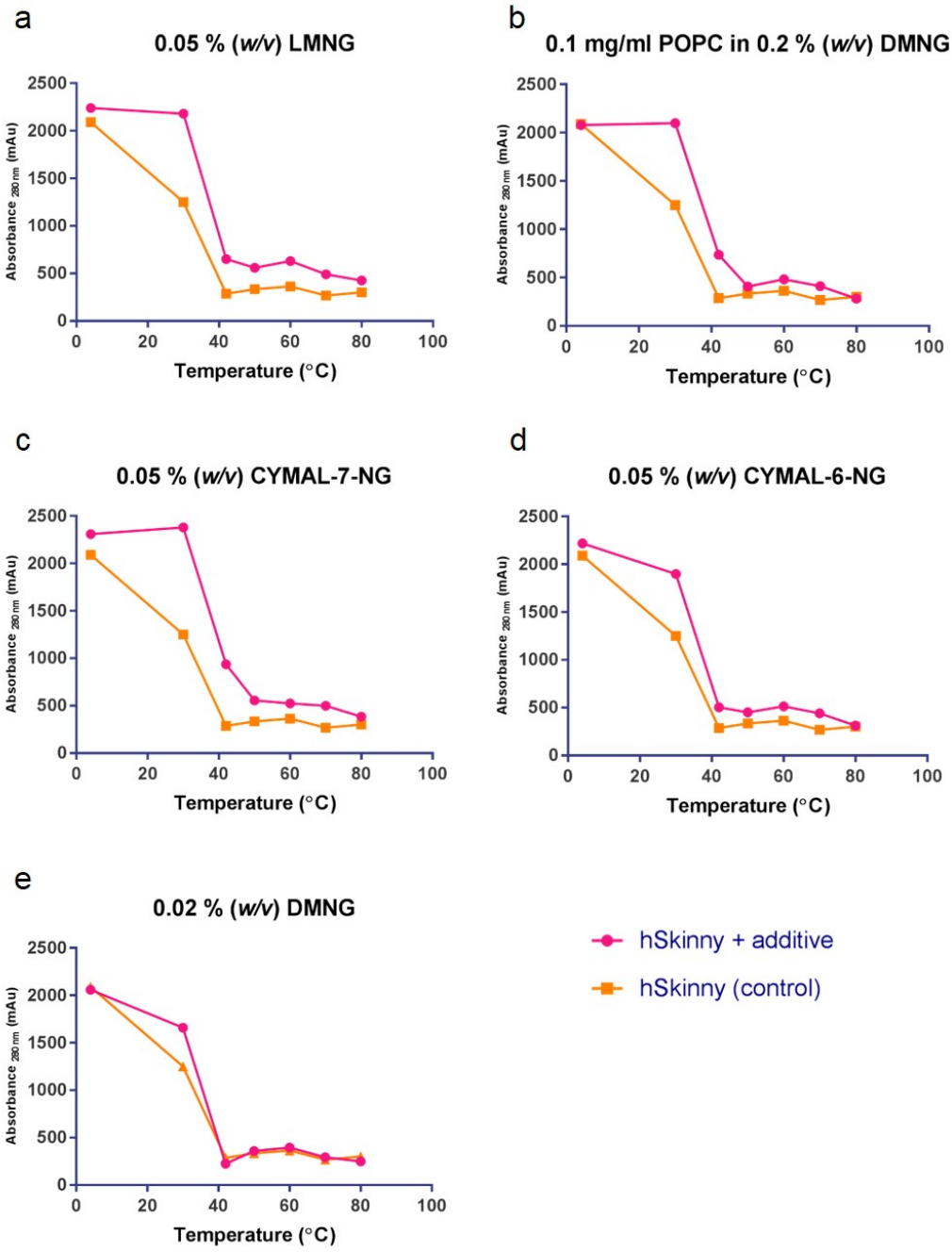


Figure 5.9: **Membrane protein additive screen (MPAS) for FL-hSkinny Hh.** MPAS comprised of 96 additives that included different pH conditions, salt, metals, detergents and lipids. The additives LMNG (a), POPC in DMNG (b), CYMAL-7-NG (c), CYMAL-6-NG (d) and DMNG (e) were further tested in a complete temperature screen. LMNG and CYMAL-7-NG were found to be the best additives for the stability of Skinny Hh.

ture. In contrast, P-CoA and the non-hydrolysable analogue did stabilise Skinny Hh, in accordance with previous findings that Skinny Hh uses P-CoA as a substrate [46] (Figure 5.10, c and d). The binding of non-hydrolysable P-CoA to Skinny Hh would allow for crystallisation of Skinny Hh in the presence of Shh peptide and P-CoA without initiating catalysis. This could potentially provide structural information relating to substrate binding and the reaction mechanism. From the sequence analysis of Skinny Hh, a nucleotide-binding Walker motif has been identified between TM helices 9 and 10, on the luminal loop. We wanted to examine whether nucleotides Guanosine 5'-[β,γ-imido]triphosphate (GTP-β,γ, non-hydrolysable analogue) and Adenosine 5'-[γ-thio]triphosphate (ATP-γ, non-hydrolysable analogue) improve the stability of Skinny Hh. However, neither had an effect (Figure 5.10, e and f). In addition, we wanted to examine whether the addition of the lipids phosphatidylcholine (PC) and phosphatidylethanolamine (PE), which were found to be abundant in the ER membrane [273], would have an effect on Skinny Hh, and indeed the addition of both lipids increased the stability of Skinny Hh at 30°C (Figure 5.10, g and h).

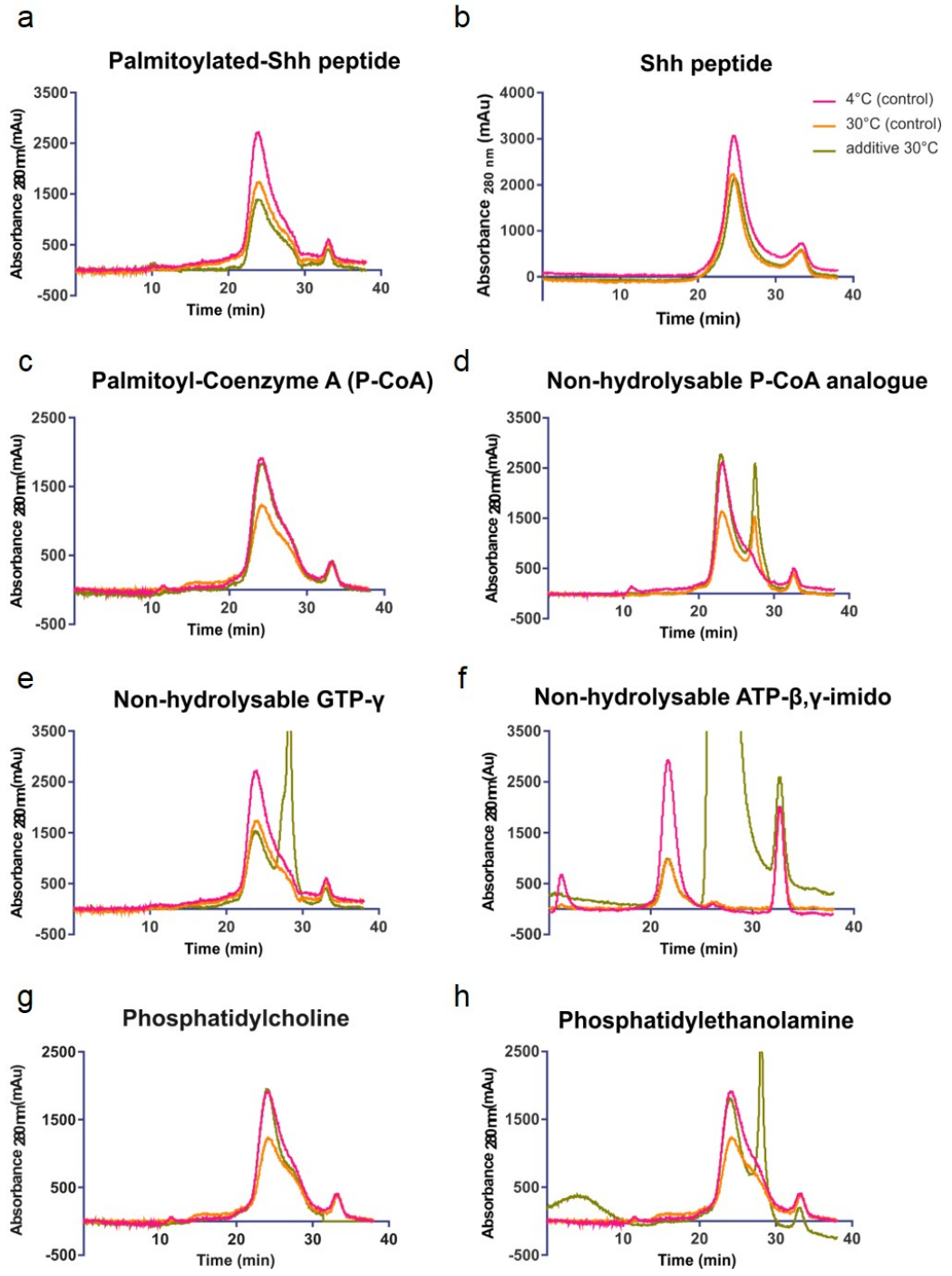


Figure 5.10: **Thermostability assay with Skinny Hh specific additives.** The addition of palmitoylated (a) and non-palmitoylated Shh (b) peptides, as well as non-hydrolisable GTP (e) and ATP (f) analogues did not stabilise Skinny Hh. However, the addition of substrate P-CoA (c) and non-hydrolisable P-CoA (d) and lipids PC (g) and PE (h) did contribute to the stability of Skinny Hh at 30°C.

5.3.3 Large-scale purification and initial crystallisation trials

FL-hSkinny was expressed on a large-scale from 6L of HEK293 GNTI$^-$ (HEK293S) suspension cell culture, and purified via α1D4-antibody affinity chromatography followed by SEC in OGNG 40:1 CHS detergent (Figure 5.11, a and b). As in small-scale expression trials, Skinny Hh ran as a single monodisperse peak with very little aggregate (Figure 5.11, b).

The predicted molecular weight of Skinny Hh-1D4 is 58 kDa. However, SDS-PAGE and Western blots indicate three bands of approximately 35 kDa, 60 kDa and 90 kDa both in reducing and non reducing conditions (Figure 5.11, c and d). These bands could represent different Skinny Hh aggregates formed upon addition of SDS, with the fully denatured protein running below its predicted molecular weight at 35 kDa [274].

The remaining 60 kDa and 90 kDa bands could also represent a dimer and a trimer. However, this is unlikely because different aggregation states would be apparent in a SEC trace. The bands could potentially represent a differentially palmitoylated Skinny Hh.

Alternatively, the 60 kDa band could correspond to the Skinny Hh-1D4, while the 35 kDa band may represent a cleaved fragment and the $\sim$ 90 kDa band could possibly be an association of the two. Since all three bands appear in Western blot when probed with α1D4-antibody which binds to the C-terminal 1D4 tag of Skinny-Hh constructs (Figure 5.11, c), the cleaved 35 kDa fragment

would correspond to the C-terminal domain. Furthermore, the three domains would probably remain associated in solution via hydrophobic interactions of the transmembrane regions as only one peak is detected in SEC.

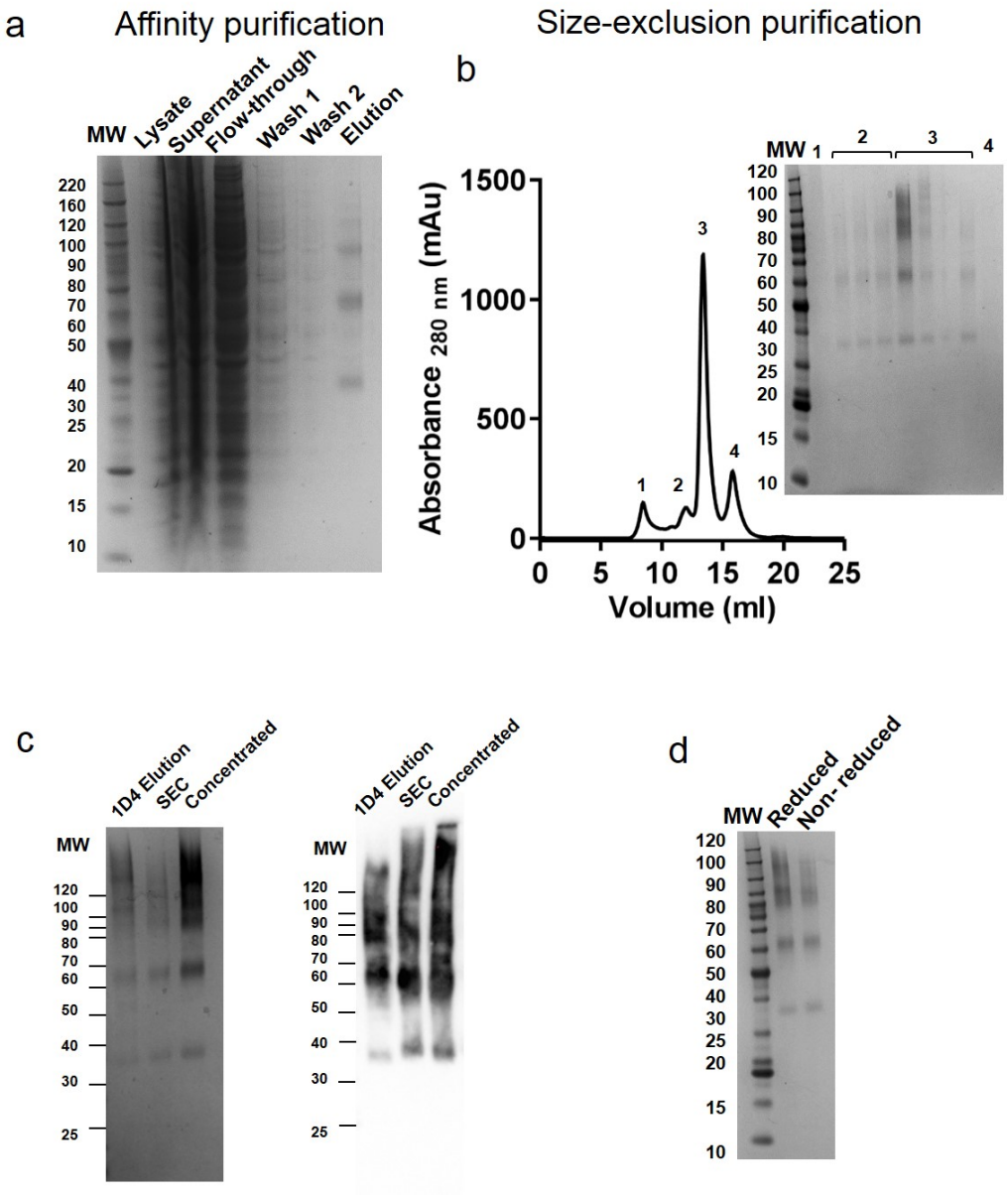


Figure 5.11: **Purification of hSkinny Hh.** a) The protein was first purified using α1D4 antibody affinity purification and the samples corresponding to each step were run on an SDS-PAGE gel. b) The protein was then concentrated and analysed by SEC resulting in one main monodisperse peak, and the samples were run on an SDS-PAGE gel. c) Skinny Hh appears as three bands in SDS-PAGE gels and WB, under both d) reducing and non-reducing conditions.

The initial stages of purification yielded an extremely low yield of $\sim$ 0.2 mg from 6 L of culture. The expression of Skinny Hh was optimised by testing valproic acid (VPA) and sodium butyrate (NaBu), which are added to mammalian cell cultures to increase protein expression [275]. In addition, they reduce the cell pellet produced thereby preserving the amount of detergent needed for protein solubilisation. Although the precise mechanism of how they enhance protein expression is unknown, they both act as histone deacetylases (HDAC) inhibitors [276, 277]. In addition, two different temperatures 30°C and 37°C were tested. HEK293S suspension cells were transiently transfected and treated with 5 mM VPA or NaBu, and incubated at 30°C or 37°C. Samples were collected every 24

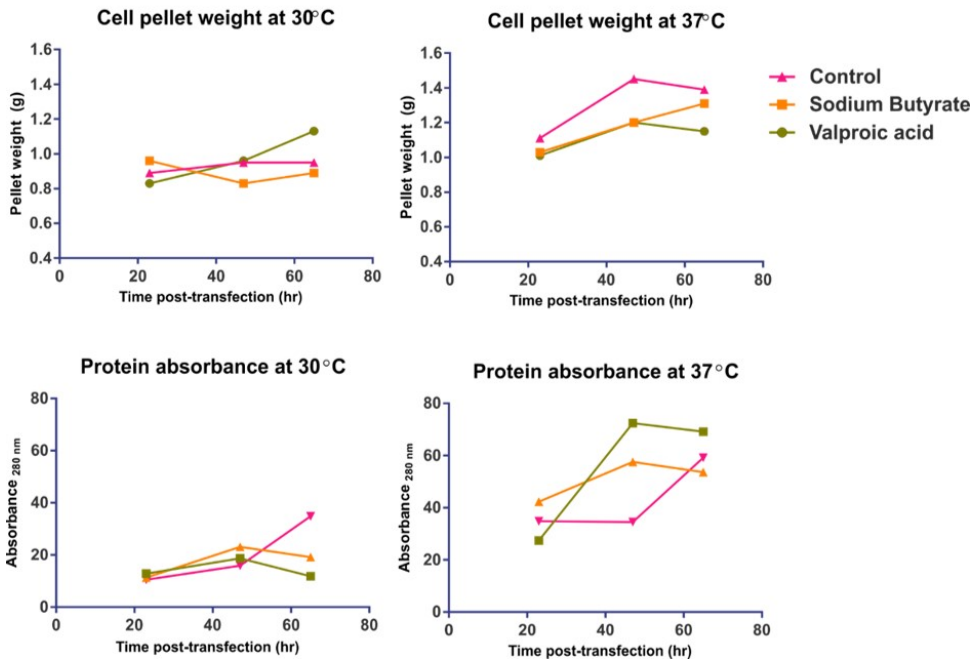


Figure 5.12: **Expression optimisation for hSkinny Hh.** The protein was tested at 30°C and 37°C in the presence of Valproic acid or Sodium Butyrate. The optimal condition was found to be expression at 37°C in the presence of Valproic acid for 48 h, resulting in the highest yield and lowest cell pellet mass.

h and the pellet weight was measured. The protein was then purified from each sample and was run on SEC. The pellet weight and the protein peak absorbance were plotted against time post-transfection (Figure 5.12).

Protein expression was significantly higher at 37°C than at 30°C. The effects of VPA and NaBu on expression were quite apparent, in particular at 48 h post-transfection. The addition of VPA resulted in $\sim$ 100% increase in protein expression at 37°C 48 h post-transfection when compared to the control, while NaBu resulted in $\sim$ 60% increase. Although the difference is minimal at 72 h, the use of VPA would allow for shorter expression time. VPA and NaBu also had a positive impact on the cell pellet resulting in $\sim$ 20% decrease after 48 h. Overall, the expression at 37°C in the presence of 5 mM VPA for 48 h was found to be optimal for expression of hSkinny Hh in HEK293S suspension cell culture.

We further optimised the use of α1D4-antibody beads in affinity purification. The optimal amount to be used was found to be 100 µl of solid beads per 0.7 mg of cell pellet, resulting in $\sim$ 50% increase in protein yield when compared to the use of 50 µl of beads (Figure 5.13).

After optimisation of expression and purification, and the use of additives (non-hydrolysable P-CoA and Shh peptide) the yield increased 5-fold from $\sim$ 0.2 mg to $\sim$ 1.0 mg per 6 L of culture. Despite a reasonable yield for a membrane protein, the main issue remained the inability to concentrate the protein due to precipitation. This resulted in a huge loss of protein during the two concentration steps prior to SEC and crystallisation plate set-up. In addition, a significant amount of protein could not be eluted and was retained on the α1D4-antibody

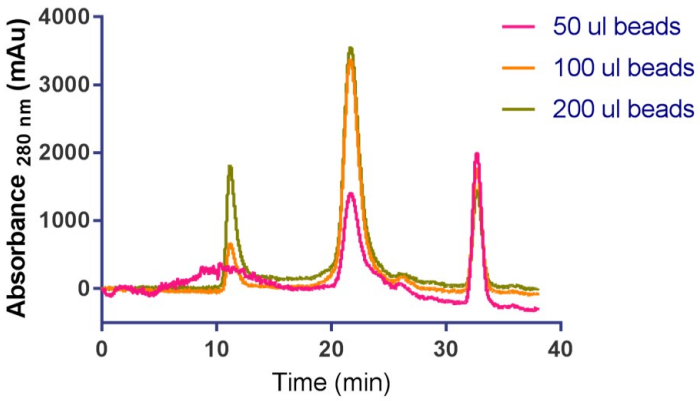


Figure 5.13: **Bead test for Skinny Hh purification.** The optimal amount of α1D4-antibody affinity beads used for extracting Skinny Hh is 100 ul per 0.7 mg of cell pellet.

affinity beads, which was apparent due to the dark yellow colour of the protein.

From two large-scale purifications, the protein could be concentrated in the presence of Shh peptide and non-hydrolysable P-CoA to 2.9 mg/ml. Despite the low concentration, this was enough to set-up 6 lipid cubic phase crystallisation (LCP) screens at 4°C, and room temperature (RT). The screens yielded crystals in two conditions. However, no protein diffraction was detected (Figure 5.14, a). In addition, protein was concentrated to 1.9 mg/ml and after performing a pre-crystallisation test (PCT), traditional vapour diffusion crystallisation screens were set-up at 4°C. In spite of the low concentration, there was significant precipitation across the majority of conditions. The screens did not yield protein crystals, however, small detergent crystals were obtained in multiple conditions (Figure 5.14, b).

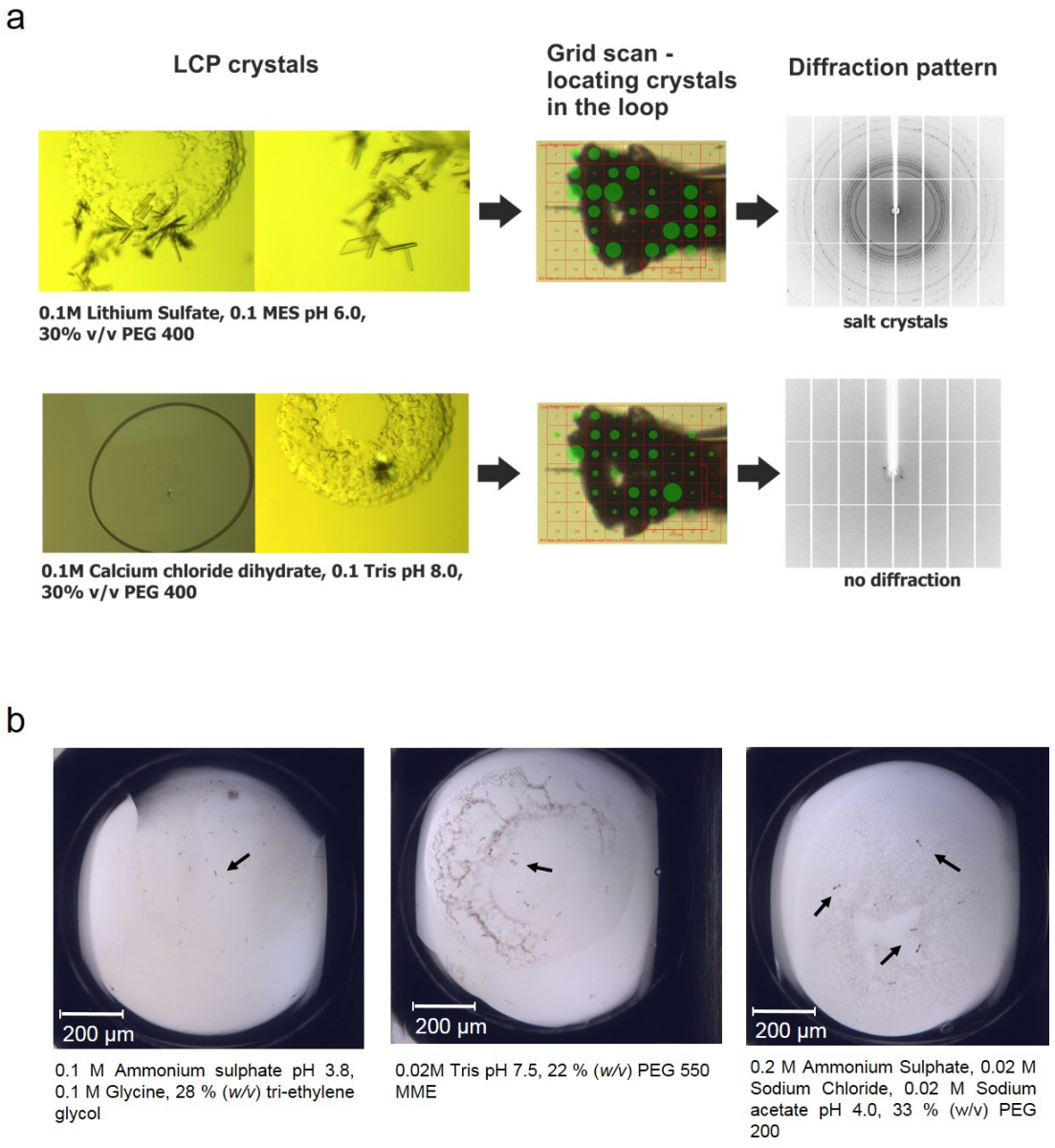


Figure 5.14: **Crystallisation of Skinny Hh-1D4 in the presence of non-hydrolysable P-CoA and Shh peptide.** a) Lipidic cubic phase (LCP) yielded crystals in two conditions. Crystals were tested for X-ray diffraction using a grid scan due to the inability to locate the crystal. Neither crystal yielded protein diffraction. b) Screens were set up in vapour diffusion format, however, only detergent crystals were obtained, indicated with an arrow.

5.3.4 Purification detergent optimisation and activity

The purification of hSkinny Hh was further optimised by screening other detergents that might be more suitable for concentration and crystallisation. LMNG was identified as a stabiliser in the additive screen. However, when used as a primary detergent to extract and solubilise Skinny Hh the protein precipitated during the elution step of the affinity purification. The protein could not be stabilised upon addition of other additives or glycerol. LMNG was found to stabilise Skinny Hh only at extremely high concentration (> 10 CMC) and was found unsuitable for purification.

Based on previous research, DDM and DMNG were identified as the most successfully used detergents used for membrane protein crystallisation [278, 279]. In addition, the majority of proteins that were crystallised in LCP were in DDM [280].

hSkinny Hh (1D4) was purified on a medium-scale by solubilisation using OGNG 40:1 CHS followed by detergent exchanged into DMNG 40:1 CHS and DDM 40:1 CHS (Figure 5.15). Glycerol and L-arginine were added to the buffer from the solubilisation step to increase the stability of the protein. Furthermore, Shh peptide, non-hydrolysable P-CoA, CYMAL-7-NG and POPC were added after elution from affinity purification and prior to concentration for SEC. The initial yield post-elution was comparable between different conditions: $\sim$ 0.33 mg for OGNG and DMNG, and $\sim$ 0.39 mg for DDM per 15 g of pellet. Upon concentration hSkinny Hh in DDM remained stable resulting in 0.30 mg of pro-

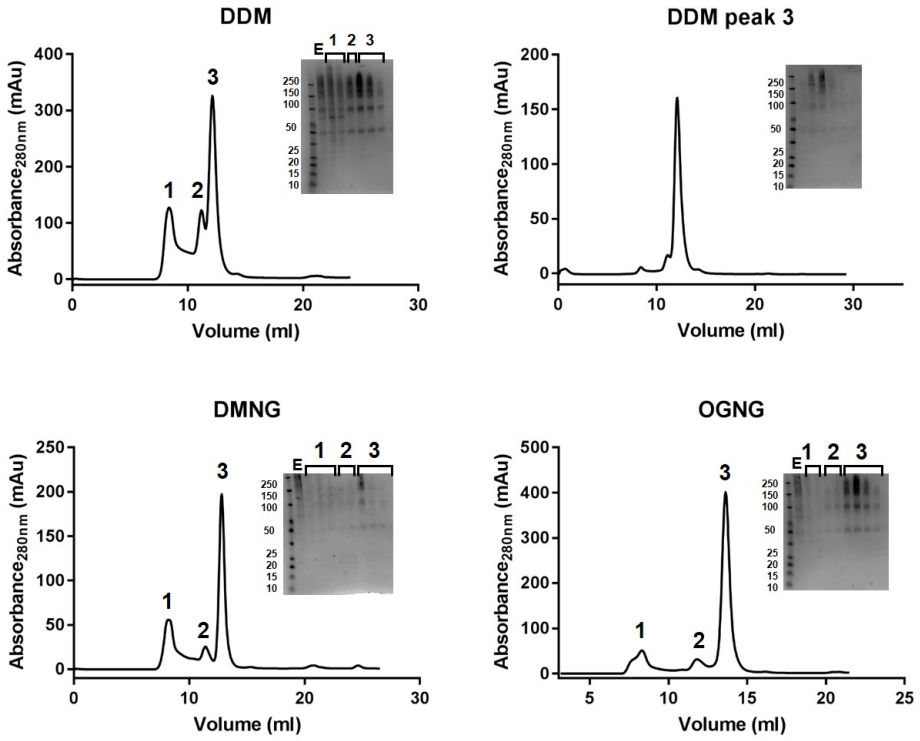


Figure 5.15: **Test purification of hSkinny Hh.** The protein was purified using DDM, DMNG and OGNG supplemented with 40:1 (molar ratio) CHS. The SEC traces resulted in three peaks for each condition. Peak 3 from DDM prep was concentrated and re-ran resulting in a single monodispersed peak. This test indicated that Skinny Hh purified in DDM and in the presence of additives results in high yield and stability throughout the concentration process.

tein, in contrast to 0.19 mg for OGNG and 0.12 mg for DMNG. All samples were examined on SEC and resulted in one main peak (peak 3) and two aggregate peaks (peak 1 and 2) (Figure 5.15). Peaks 2 and 3 from the DDM purification were not resolved and, therefore, peak 3 was concentrated and re-ran on SEC, resulting in a single monodisperse peak with almost no aggregation.

This purification indicated that we could successfully purify Skinny Hh by solubilising it in OGNG 40:1 CHS and exchanging the detergent into DDM 40:1 CHS, and by using stabilising additives during concentration steps to produce quantities suitable for crystallisation trials. The protein could also be concen-

trated without significant loss resulting in a yield of approximately 0.2 mg/L of cell culture. Furthermore, the use of DDM is better for crystallisation than OGNG.

We also wanted to test the activity of hSkinny Hh in microfluidic mobility shift assay (MSA). For this, the protein was purified in DDM 40:1 CHS with no additives other than glycerol. In this case, hSkinny Hh appeared to aggregate more resulting in one large aggregate peak and two protein peaks (Figure 5.16, b). Both protein peaks were pooled separately and tested for activity. Surprisingly, both hSkinny Hh peaks were highly active (Figure 5.16, c) and approximately 10× more active than what was previously reported by our collaborators (Figure 5.16, a) [260]. These samples were more active when diluted to 17µg/ml (10-fold dilution) (Figure 5.16, c) than the previously reported Skinny Hh activity assay performed at 45 µg/ml (Figure 5.16, a) [260]. Furthermore, the samples exhibited significant activity even when diluted 1000-fold to 170 ng/ml and remained active after 8h (Figure 5.16, d). Taken together, these findings provide a platform for high-throughput drug screening against Skinny Hh as a target. High protein purification yield and high activity allow for a small-molecule screen to be carried out.

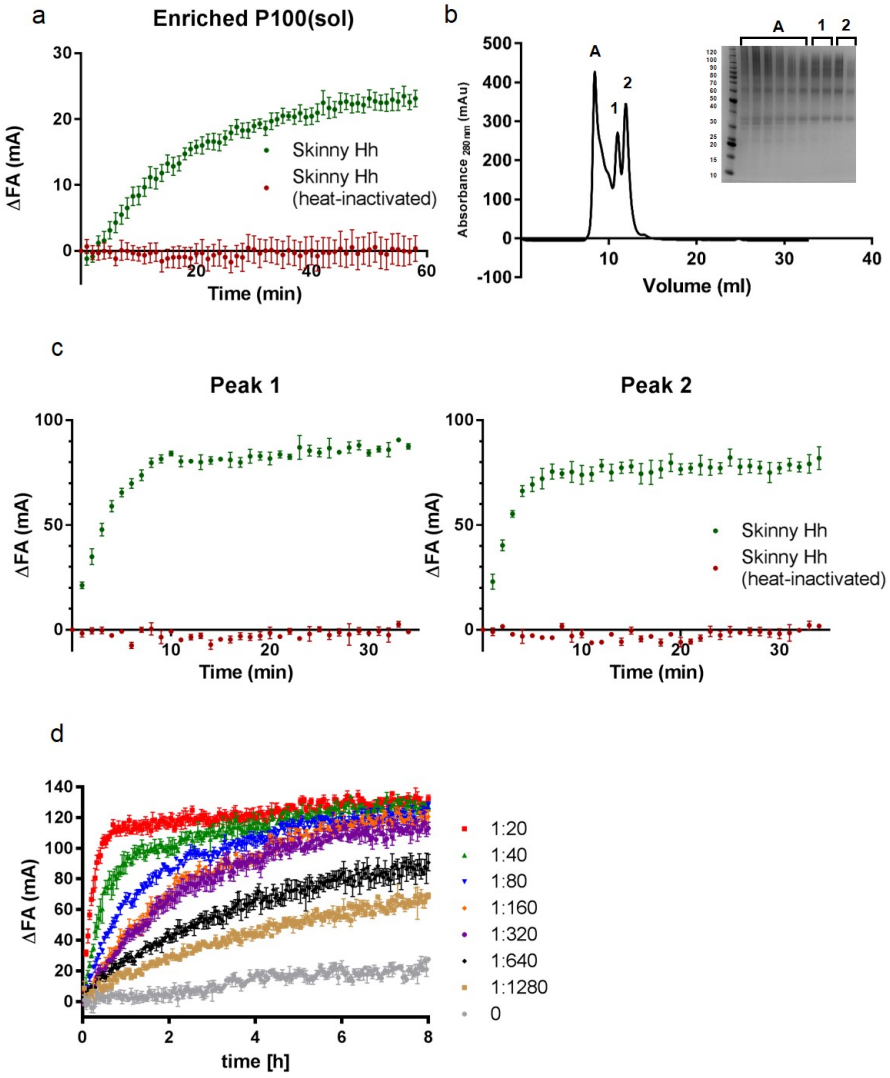


Figure 5.16: **Skinny Hh palmitoylation activity assay.** a) Skinny Hh sample from our collaborators is moderately active when measured at 45 µg/ml in MSA palmitoylation assay, reaching signal of 20 mA over 40 min, as reported before [260]. b) Skinny Hh was purified in DDM 40:1 CHS resulting in more aggregation. Two protein peaks 1 and 2 were separately pooled at tested for activity in MSA palmitoylation assay. c) Skinny Hh was diluted 10× to 17 µg/ml and tested, reaching peak signal of 80 mA after only 10 min. d) Skinny Hh sample remains active even when diluted > 1000× to 170 ng/ml for 8 h. The experiments were performed by Thomas Lanyon-Hogg, Imperial College London.

5.3.5 Construct engineering

Novel hSkinny Hh constructs were engineered based on previous topology and mutation studies with the aim of improving expression, stability and propensity to crystallise [262, 263]. Bril is a thermostabilised apocytochrome b562 comprised of a four-helix bundle often used to aid expression, stability and crystallisation of GPCR proteins, analogous to lysozyme [281]. Initially we cloned, expressed and tested for thermostability the following constructs: Δ30-54, Δ30-54-Bril, 44-Bril, 157-Bril, 194-Bril and Δ231-233 (Figure 5.17). It appeared that the modification to the first and second luminal loops of Skinny (N-terminal) increased the stability but decreased the expression significantly (Figure 5.17, b-e), while the addition of Bril at loops 4 and 5 resulted in high expression (Figure 5.17, f-g). The combination of these features, Δ30-54 Skinny 194-Bril and Δ30-54 Skinny Δ231-233-Bril, resulted in stable protein but low and medium expression, respectively (Figure 5.18, b and c). We further produced clones comprised of the central and C-terminal portion of the protein, important for activity of Skinny Hh. In addition, the 35 kDa band that is evident on SDS-PAGE gels could correspond to one of the C-terminal domains. We produced Δ194 and Δ127 constructs, in combination with Bril inserts at positions 194 and Δ231-233. All constructs with the exception of Δ127 Skinny 194-Bril yielded a stable construct (Figure 5.17). However, only Δ194 Skinny Hh resulted in medium expression (Figure 5.17, d). The most promising constructs for crystallisation included Δ30-54 Skinny Δ231-233-Bril and Δ194 Skinny Hh. However, further protein engineering and purification

optimisation needs to be done to increase the expression of more thermostable constructs.

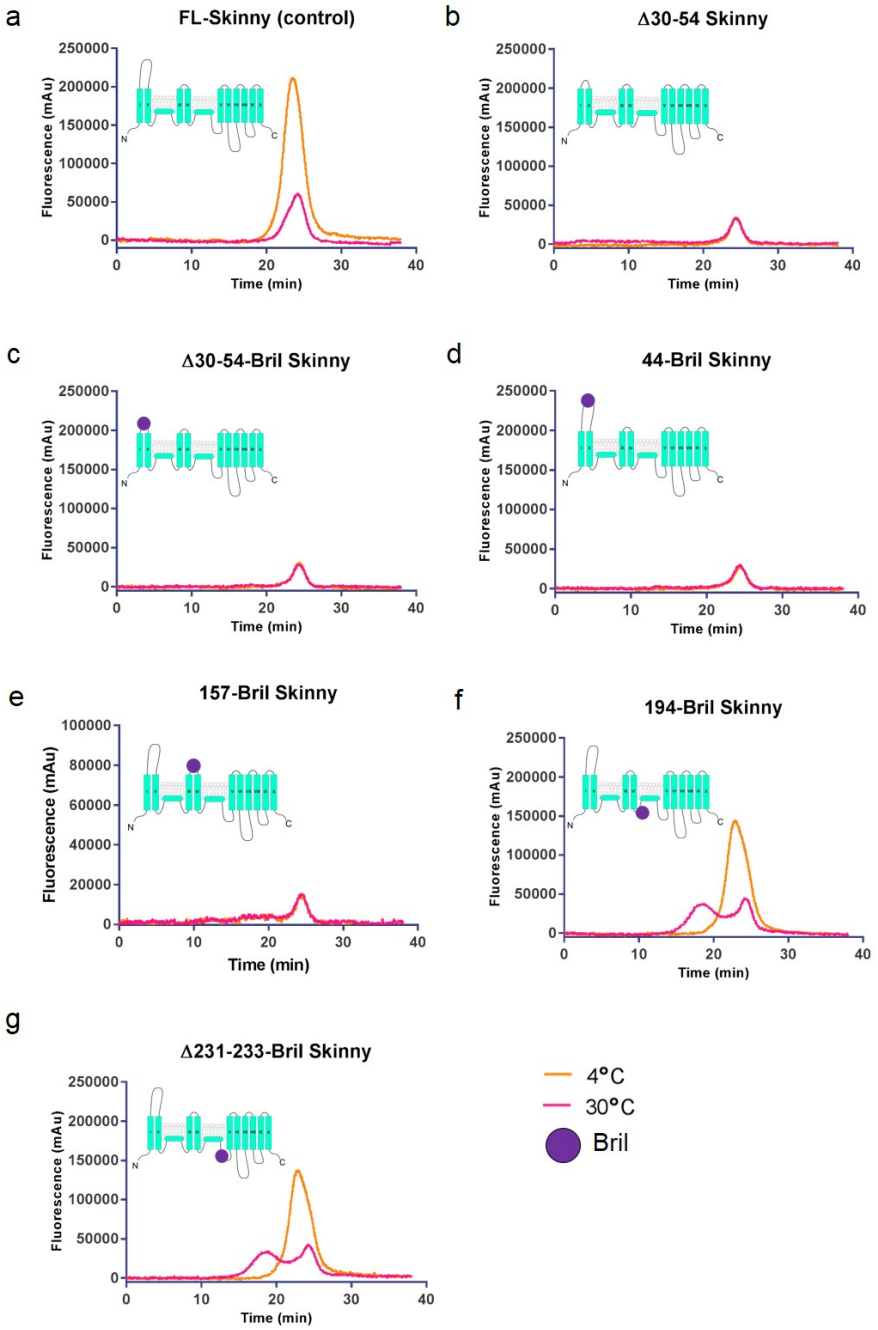


Figure 5.17: **hSkinny Hh construct engineering.** FSEC traces of thermostability test of Skinny Hh constructs at 30°C. The modification to the first and second loop (luminal side) of Skinny Hh (b-e) results in increased stability and decreased expression, while addition of Bril to 194 and 231-233 site increases expression when compared to the control FL-Skinny Hh (a).

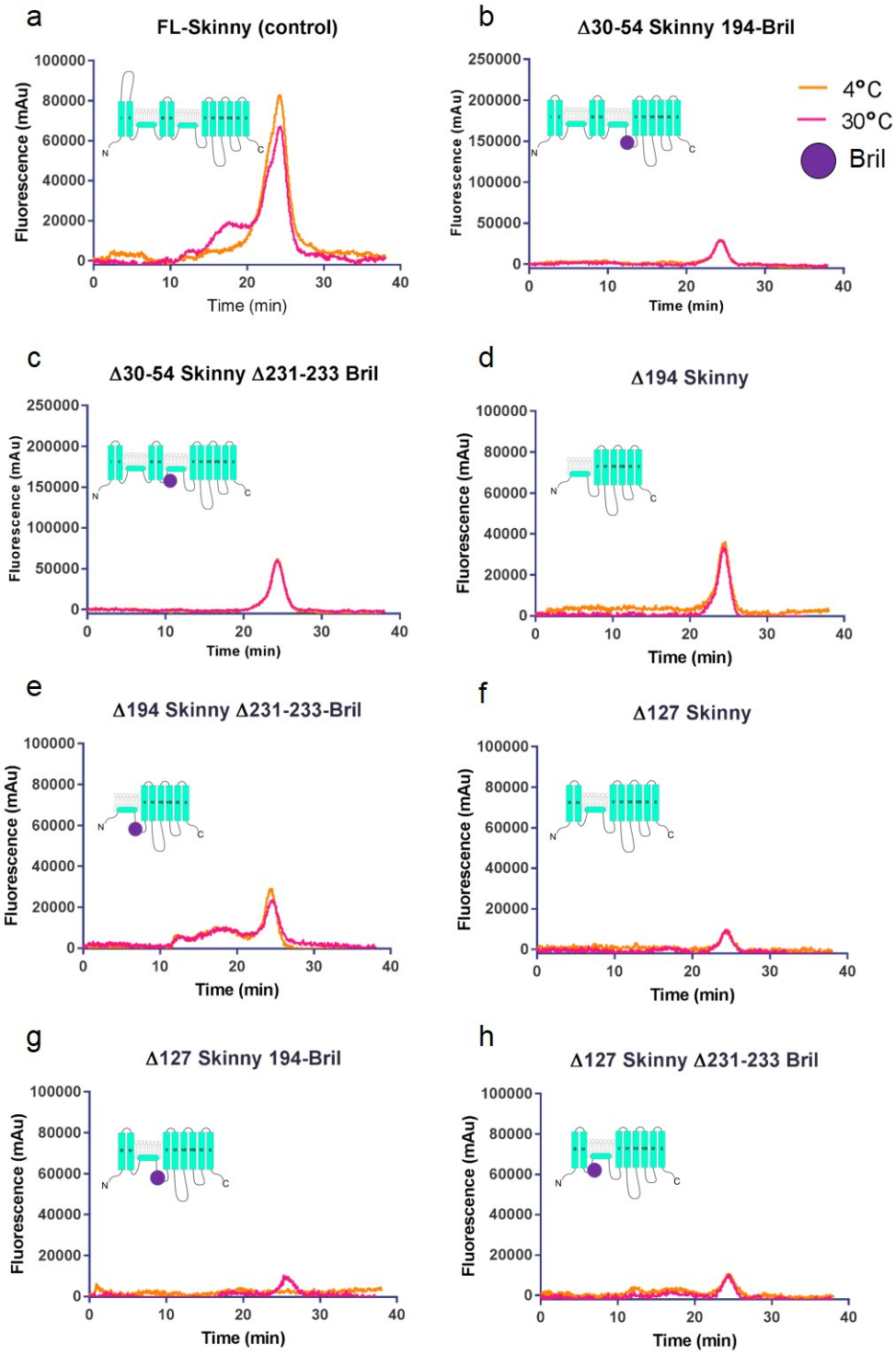


Figure 5.18: **hSkinny Hh construct engineering.** FSEC traces of thermostability test of Skinny Hh constructs at 30°C. The combination of first loop excision and addition of Bril (b-c) resulted in increased stability and medium expression when compared to the control (a). The C-terminal domain constructs (d-h) resulted in increased stability and low expression for all apart from Δ194 Skinny Hh.

5.3.6 Construct activity

Five Skinny Hh constructs: Δ30-54 Skinny Hh, 194-Bril Skinny Hh, Δ231-233 Bril Skinny Hh, Δ194 Skinny Hh and Δ30-54 Skinny Δ231-233-Bril were tested in MSA palmitoylation assay and their activity was compared to the FL-Skinny Hh (Figure 5.19, a). Interestingly, more thermostable constructs lacking the 30-54 loop: Δ30-54 Skinny Hh, Δ194 Skinny Hh and Δ30-54 Skinny Δ231-233-Bril showed no activity in the palmitoylation assay (Figure 5.19, b). The two constructs that only had a Bril insert, 194-Bril Skinny Hh and Δ231-233-Bril Skinny Hh, exhibited 68% and 88% relative activity, respectively, when compared to the WT FL-Skinny Hh (Figure 5.19, b). Although removing 30-54 stabilises Skinny Hh in thermostability assays, it completely abolishes its activity. These results indicated that residues in region 30-54 are essential for the activity of Skinny Hh. Furthermore, the two partially active constructs 194-Bril and Δ231-233 Skinny Hh are preferred for crystallisation trials due to their ability to interact with substrates and potentially generate crystal contact via Bril inserts.

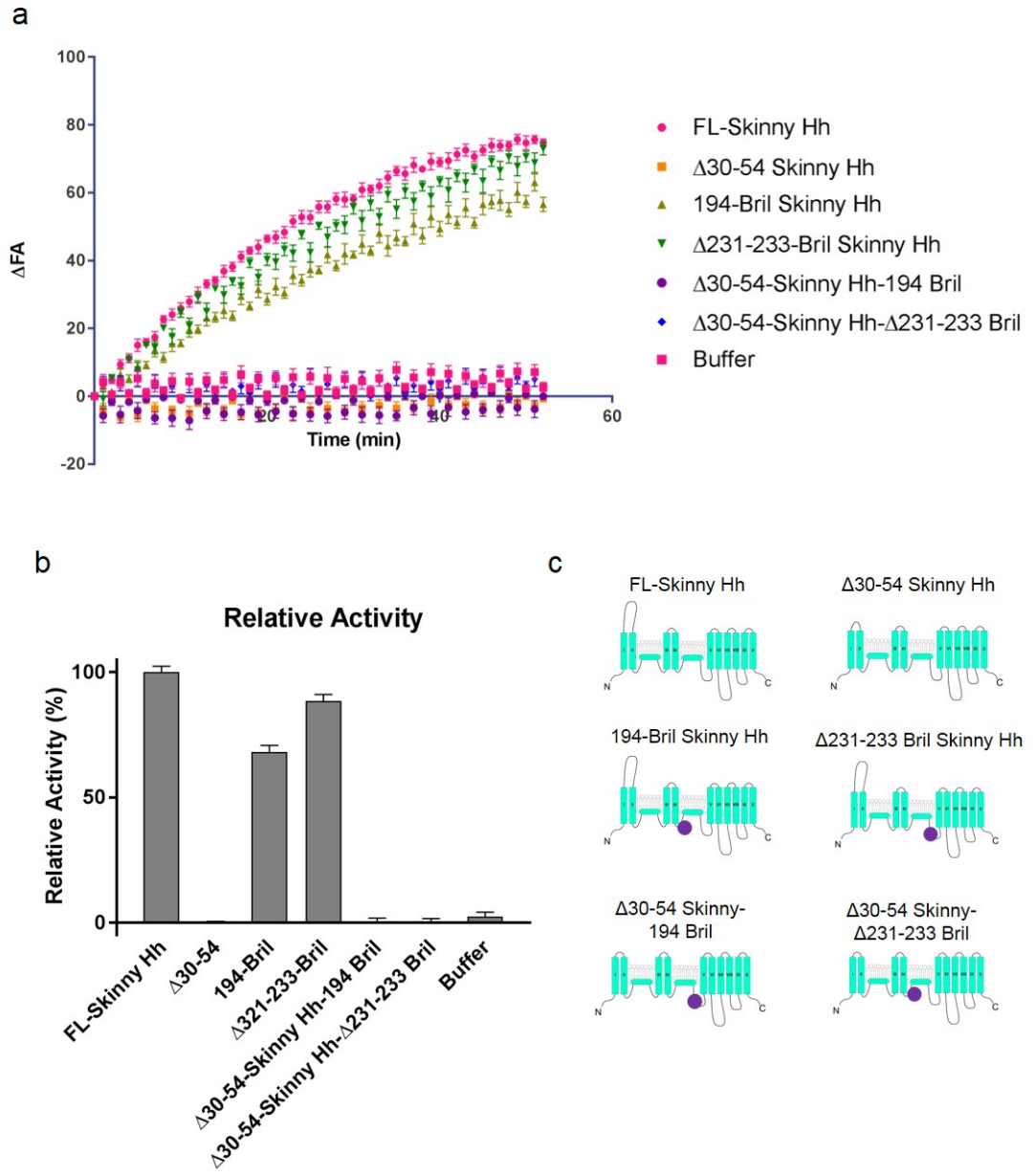


Figure 5.19: **Activity assay of novel Skinny Hh constructs.** a) Five constructs were tested in MSA palmitoylation assay, indicating that only 194-Bril Skinny Hh and Δ231-233-Bril Skinny Hh constructs were active. b) The relative activity of 194-Bril Skinny Hh and Δ231-233-Bril Skinny Hh constructs was 68% and 88%, respectively, when compared to the FL-Skinny Hh (100%). c) Visual schematic of tested constructs. The experiments were performed by Thomas Lanyon-Hogg, Imperial College London.

5.3.7 Drug stability screen

In addition to additives, the effect of Skinny Hh inhibitors were tested in a thermostability test. Surprisingly, RUSKI-43 [198] the first commercially available Skinny inhibitor and its analogue RUSKI-41 identified in the same study, had very little effect on the stability of Skinny Hh (Figure 5.20, a and b). However, the newly developed analogues RUSKI-101 and 201 increased the stability of Skinny Hh significantly (Figure 5.20, c and d). The most significant effect was observed with a novel inhibitor TLH-45 (Figure 5.20, e), which was developed and synthesised by our collaborator Tom Lanyon-Hogg (Prof Ed Tate group, Imperial College London, not published). These results indicate that TLH-45 is the best additive that can be used to stabilise Skinny Hh for crystallisation trials in addition to substrates. Crystallisation of the Skinny Hh-TLH45 complex would provide valuable information to unravel the mode of inhibition of Skinny Hh and provide guidance for rational drug design.

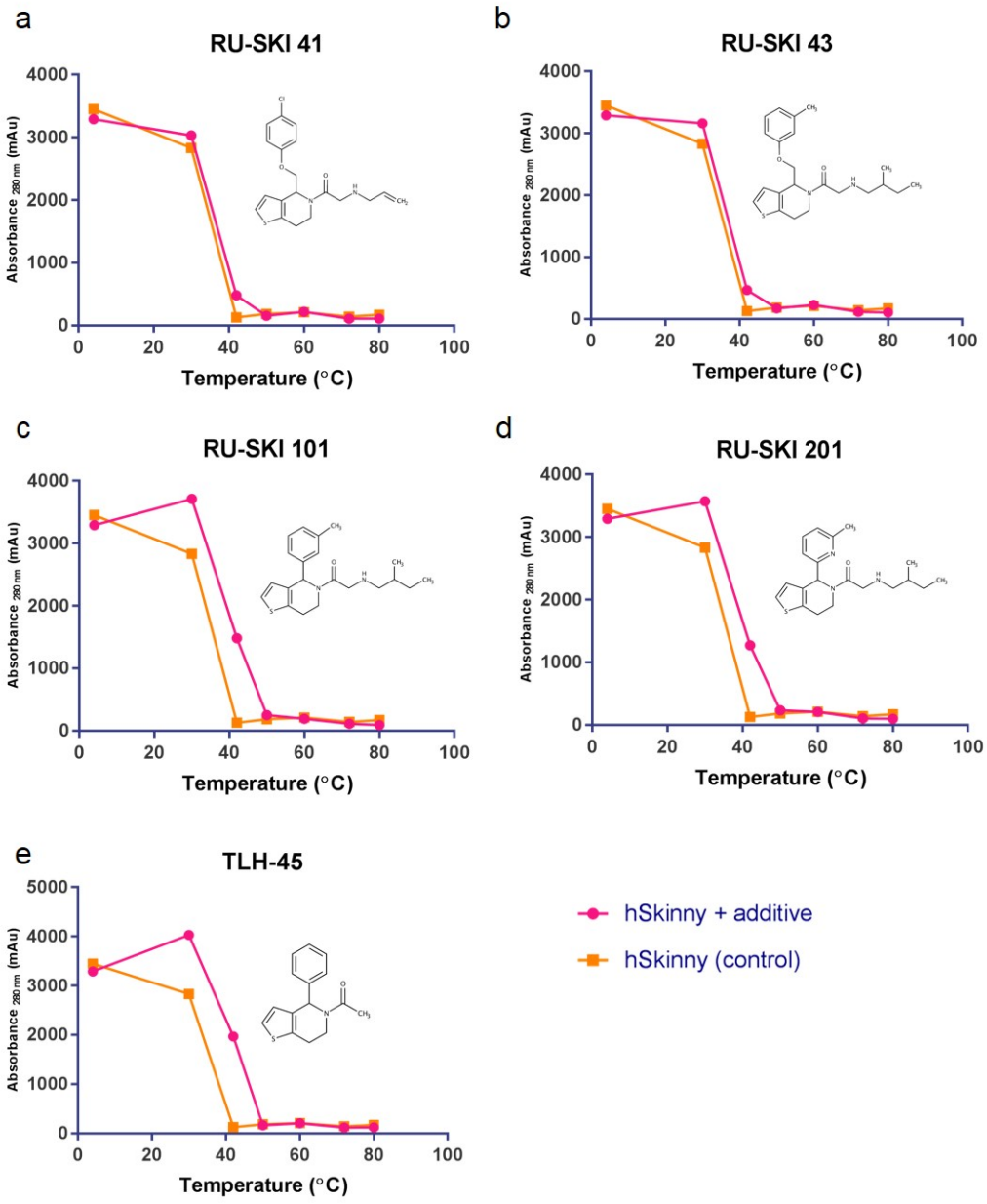


Figure 5.20: **Small-molecule thermostability screen.** RUSKI 41 (a) and 43 (b) inhibitors identified in the first small-molecule screen against Skinny Hh had very little impact on the stability of Skinny Hh. The improved RUSKI 101 (c) and 201 (d) had a significant effect on protein stability, comparable to substrate P-CoA. TLH-45 (e) had the most significant impact on the stability.

5.4 Discussion

Skinny Hh is an enzyme that palmitoylates all Hh ligands before they are secreted and is essential for proper Hh signalling in flies and vertebrates [44]. Skinny Hh knockouts and HhN Cys24Ser mutants result in significantly reduced signalling in mice and abolished signalling in *Drosophila* [44, 72]. Furthermore, mutations in Skinny Hh have been found to lead to developmental defects such as sex development [249] and severe holoprosencephaly with acrania and aghnatia [82]. Skinny Hh catalyses N-palmitoylation of HhN by an unknown mechanism. However, two hypotheses exist, one involving an S-to-N shift, the other involving N-myristoylation-like direct palmitoylation [238]. Skinny Hh is also an important drug target for ligand-dependent Hh cancers, since small-molecule inhibitors of Skinny Hh were found to successfully inhibit the Hh pathway in pancreatic, breast and lung cancer cell lines [158, 199].

Despite significant advancements in our understanding of Skinny Hh over recent years, studying Skinny Hh poses a challenge associated with studying membrane proteins. As a result, many aspects of Skinny Hh, including its structure, molecular mechanism and inhibition remain a mystery. The aim of this chapter was to address these issues by optimising the production and quality of Skinny Hh with the final goal of determining the molecular structure and conducting further drug screening experiments.

We optimised the mammalian expression of hSkinny Hh full-length construct in HEK293S cells. We identified the optimal purification protocol, including

buffers and detergents, to be used to extract and solubilise Skinny Hh to a high yield and purity to allow for crystallisation trials. Furthermore, we identified a set of additives including CYMAL-7-NG, POPC and P-CoA that enhanced the thermostability of Skinny Hh. We tested the effects of inhibitors and identified TLH-45 which significantly increases the stability of Skinny Hh when compared to other inhibitors and additives. In addition, we established that purified Skinny is exceptionally active in palmitoylation assays. We also engineered a number of constructs establishing that removing residues 30-54 of the first luminal loop of Skinny Hh increase the stability, while the addition of Bril at positions 194 and Δ231-233 increased expression. Furthermore, the N-terminal luminal loop of Skinny Hh (residues 30-54) was found to be essential for its activity.

Overall, these results provide a platform for structural studies and drug screening assays. The ability to purify thermostable Skinny Hh in milligram quantities in the presence of additives, substrates and inhibitors provides an opportunity for extensive crystallisation trials. Furthermore, a major setback of *in vitro* Skinny Hh inhibitor screens were low yields and poor stability of purified protein. High yield and superior activity of Skinny Hh construct will allow for the 100,000s of compounds to be screened in an *in vitro* palmitoylation assay using significantly less protein.

Future work will focus on crystallisation trials of Skinny Hh in the presence of substrates Shh peptide and P-CoA, and inhibitor TLH-45. Furthermore, inhibitor screen of 100,000 compounds will be performed in collaboration with Prof. Ed Tate, Imperial College London. Collectively, these findings will contribute to

a better understanding of Skinny Hh, and will contribute to the development of novel inhibitors. Moreover, structural work on Skinny Hh could potentially inform the development of inhibitors of other MBOAT family members. In particular, palmitoyltransferase Porcupine, which is essential for Wnt ligand palmitoylation and proper Wnt signalling, could be exploited as a drug target since activation of Wnt pathway was associated with multiple cancers, similar to Hh signalling.

Chapter 6

Conclusions

The Hh pathway is one of the fundamental morphogen pathways activated during embryonic development. Originally discovered in *Drosophila*, the Hh signalling pathway was later discovered in all metazoans including humans [282]. The Shh morphogen was first discovered in the neural tube and limbs of vertebrates, where it contributed to tissue patterning [9, 10]. Shh, the most widely expressed Hh ligand, and the more specialised Dhh and Ihh, were subsequently found in nearly all tissues where they contributed to patterning, growth and symmetry formation [282]. Consequently, mutations that resulted in defective Hh signalling were found in multiple developmental disorders, most commonly holoprosencephaly [79].

The Hh pathway remains active in adults where it contributes to tissue homeostasis and regeneration [158]. Constitutive activation of Hh signalling was found in multiple cancers including basal cell carcinoma (BCC), medulloblastoma (MB), pancreatic, lung, breast, liver and others [93]. Loss-of-function mutations in the receptor Ptch or gain-of-function mutations in Smo were correlated with BCC and MB, where they supported growth and proliferation of tumour cells in a ligand-independent manner [170]. Two FDA approved inhibitors, Vismodegib and Sonidegib, which target Smo are being used as a treatment for BCC and are being tested for MB [183, 283]. However, patient dropout rates due to side effects and recurrence of disease due to drug resistance have had a negative impact on the clinical use of Smo inhibitors [184]. Hh pathway activation in the absence of mutations was found in a wide range of cancers and was defined as ligand-dependent mechanism [170]. This allowed for other Hh proteins to be explored

as alternative targets. In this thesis, two such targets Shh and Skinny Hh are studied using structural biology approaches.

6.1 Molecular basis for Hh signalling inhibition by Robotnikinin

ShhN has been extensively characterised functionally and structurally. The structure of ShhN revealed a pseudo-active zinc site that is essential for protein stability [128, 131]. Furthermore, structures of ShhN in complex with receptors Cdo and Hhip uncovered a calcium binding site located within the same binding cleft [111, 135, 115]. Interactions with Cdo, Hhip, 5E1 antibody and GAGs identified a 'hotspot' for receptor interactions adjacent to the ion binding cleft, which is comprised of positively charged residues and forms electrostatic interactions with binding partners. The other common feature for interaction with its receptor is the ShhN ion binding cleft. Zinc was shown to be essential for the stability of ShhN and interaction with all receptors, in particular with Hhip, where Hhip-Asp388 coordinates zinc, and the 5E1 antibody, where 5E1-Asp98 interacts with Asp147 [115]. Although calcium is essential for the ShhN-Cdo interaction [111], it has been shown to increase the affinity towards other receptors (Hhip and 5E1) by neutralising a negatively charged ShhN loop and increasing surface complementary [135, 115]. The same binding interface of ShhN, including the charged 'hotspot' and the ion binding cleft is occupied by the 5E1 antibody that

competes with Ptch for ShhN binding [139]. Smaller-ligands like the GAGs and C6S occupy only the charged area 'hotspot'.

Despite the vast amount of information on ShhN-receptor interactions, not much is known about the binding of the small-molecule inhibitor Robotnikinin. However, a plausible binding site would include the receptor interface involving the ion binding cleft or the adjacent charged 'hotspot'. We investigated the interactions between ShhN and Robotnikinin using molecular docking and X-ray crystallography. Our data strongly suggest that Robotnikinin binds within the ion binding cleft, more specifically at the pseudo-active zinc site. The docking studies identified the 131-141 loop to be important in binding along with zinc-coordinating residues His 183 and Asp148, and nearby residues Glu177 and His181. These residues are part of the 5E1-Shh interface and were previously identified as important for binding to Ptch [139, 138]. The crystallographic data revealed binding of partially cleaved Robotnikinin, presumably the macrocyclic portion, within the pseudo-active zinc site. However, establishing the precise molecular binding mechanism would require a complete structure.

From these preliminary data we conclude that Robotnikinin likely binds at the pseudo-active zinc site rather than the positively charged receptor 'hotspot'. Further crystallisation trials are needed to determine a fully bound ShhN-Robotnikinin complex, which would provide insights into precise molecular interactions for the development of Robotnikinin analogues and novel ShhN inhibitors.

6.2 Fragment-based screening for novel ShhN inhibitors

Protein-protein interactions (PPI) are essential for proper function of most cellular functions and are, therefore, implicated in a number of diseases [284]. Structural biology approaches including X-ray crystallography, NMR and biophysical techniques were paramount for understanding PPI and identifying 'hotspot' binding pockets or residues. The nature of interfaces can be obligate or transient, and can range in nature depending on surface complementary (e.g two globular protein-protein epitope, globular protein-peptide or peptide-peptide) [284]. The ShhN interface with binding partners represents a mix between a discontinuous epitope, which includes the charged 'hotspot', and globular protein-peptide interaction, which involves the ion binding cleft of ShhN and secondary structure elements of receptors Cdo and Hhip, and 5E1 antibody [225].

PPIs were for a long time considered undruggable due to their relatively flat nature in contrast to small molecule binding pockets (e.g. enzymatic active sites). Fragment-based lead discovery (FBLD) became a standard method for drug discovery and has been particularly successful in identifying hits for difficult targets such as PPI [285]. In particular, successful hits were identified for bromodomains, HIV LEDGF-integrase and XIAPcaspase9 [286, 285].

Crystallographic fragment-based screen was conducted to identify hits targeting ShhN. For this, the ShhN-SOS complex structure was used where SOS bound to the receptor 'hotspot' adjacent to the ion binding cleft. Fragment-screening

included screening over 1,200 unique fragments resulting in identification of 13 potential hits for ShhN and their high-resolution complex structures. Five of these were found within the ion binding cleft while the remaining 8 were scattered across the ShhN surface. In total, six hits were validated in SPR assay, and the preferred benzenemethanamine derivative was selected based on micromolar affinity binding to the pseudo-active site of ShhN. Initial SAR experiments indicated that the tetrahedral shape as well as the presence of the amino group are essential for binding. The presence of a methoxy group on each phenyl ring and the addition of a carboxyl group to the amino group increased affinity.

In addition to Robotnikinin, this is, to our knowledge, the only other small-molecule inhibitor of ShhN. The primary focus for future work is compound optimisation to achieve higher affinity and ensure specificity for ShhN. The specificity for different Hh ligands ShhN, DhhN and IhhN could be attempted, however, the conserved pseudo-active site is present in all Hh ligands, thus making it challenging. The exact benefit of targeting different ligands in adult cancers is not clear as the most widely expressed HhN is ShhN. Furthermore, pursuing other fragments in particular FRG 48, which bound to the calcium-binding site should be explored.

6.3 Skinny Hh structure and function

Skinny Hh is an essential acyltransferase involved in palmitoylation of Hh ligands [44]. The structure and molecular function of Skinny Hh are not known.

However, recent advances have been made to better understand its topology [263, 262]. Skinny Hh is thought to be composed of 10 transmembrane helices and two re-entrant loops, with two conserved catalytic residues His379 and Asp339, characteristic of the MBOAT superfamily. Furthermore, the second re-entrant loop comprising residues 194-233 was shown to be conserved in MBOATs and a nucleotide-binding Walker A motif was mapped onto the luminal loop between TM 9 and TM 10. We further determined by Skinny Hh construct engineering that removing residues 30-54 from the first luminal loop of Skinny Hh increased its stability but completely abolished its activity. This is contrary to what was previously reported where introducing HA tags within the first luminal loop (positions 47 and 65) reduced the stability of Skinny Hh [262]. Our results might indicate that, in addition to C-terminal catalytic residues, the N-terminal luminal loop is essential for function. Furthermore, we discovered that introducing Bril at positions 194 and 231-233 results in high Skinny Hh expression but reduces its stability, also contrary to what was previously reported [262]. Future studies focusing on the structural investigation of Skinny Hh will be of utmost importance for understanding its catalytic mechanism and molecular function.

Despite low sequence identity, Skinny Hh and ghrelin O-acyltransferase (GOAT) are predicted to have remarkably similar topology [262, 287]. GOAT contains 11 TM domains and contains 2 re-entrant loops and shares approximately 2/3 of topology with Skinny Hh, but has more cytoplasmic loops [287]. GOAT is an enzyme that octanoylates ghrelin, a peptidic hormone involved in regulating appetite and stimulating weight gain in vertebrates [288]. GOAT is essential for

the proper function of ghrelin and targeting GOAT has been of interest as a way to regulate glucose metabolism and weight gain. In fact, a peptide-based inhibitor of ghrelin has been shown to have beneficial effects on weight loss and improved glucose tolerance in mice [289]. Therefore, understanding the structure of Skinny Hh would benefit the understanding of structure and molecular mechanism of GOAT, and might inform future inhibitor development. Porcupine, another MBOAT essential for acylating Wnt morphogens, is a potential target for inhibition of Wnt-related cancers [290, 291]. Porcupine is thought to consist of 11 TM domains and no re-entrant loops. However, it shares the catalytic residues His and Asp as well as the conserved MBOAT region of Skinny Hh (Skinny Hh residues 194-233), which might be important for catalysis [292]. Understanding the catalytic mechanism of Skinny Hh might also contribute to understanding another cancer target MBOAT.

Other groups have identified inhibitors of Skinny Hh called RUSKI and showed that attenuation of Skinny Hh in cancer cell lines leads to the inhibition of Hh signalling and cell growth [198, 199]. We have established that RUSKI inhibitors 41 and 43 are not able to stabilise Skinny Hh. However, RUSKI 101 and 201, and in particular TLH 45 stabilise Skinny Hh significantly and are suitable additives for crystallisation trials. Furthermore, we established a method of obtaining active and thermostable Skinny Hh that we can now use for screening in order to identify novel inhibitors. This has the potential to help validate Skinny Hh as a target in Hh ligand-dependent cancers. Future work will focus on conducting an inhibitor screen and crystallising a complex of Skinny Hh and THL 45, and its

substrates.

In summary, two highly potential targets for Hh inhibition, ShhN and Skinny Hh, were studied in this thesis. Interactions between ShhN and the small molecule inhibitor Robotnikinin indicated that the pseudo-active site of ShhN is the most plausible binding site. Novel hits targeting ShhN were generated in a crystallographic fragment screening, resulting in a hit binding to the ShhN pseudo-active site with micromolar affinity. Skinny Hh acyltrasferase, thermostable and active, was produced in milligram quantities for crystallisation and drug-screening studies. The work in this thesis has advanced our knowledge of alternative targets, and has the basis of Hh pathway inhibition using a wide range of structural biology techniques and provided a basis for future development of Hh pathway inhibitors. The availability of more inhibitors will contribute to studying different aspects of the Hh pathway including the significance Hh pathway inhibition in ligand-dependent Hh cancers.

Chapter 7

Materials and Methods

7.1 Methods for Chapters 2-4

7.1.1 Molecular Biology

7.1.1.1 Construct design and expression vectors

Constructs of human ShhN (Genbank ID: L38518.1), DhhN (Genbank ID: NP_066382, 39-194) and IhhN (Genbank ID: NG_016741.1, 44-194) were cloned into bacterial expression vector pET22B (Invitrogen) that contains a C-terminal 6xHis-tag and an additional N-terminal methionine residue. Constructs used were ShhN24II (residues A23-A193; double mutant A23I, C24I), ShhNΔ39 (residues L39-A193) and DhhNΔ39 (L40-V194).

7.1.1.2 Transformation

E. coli Rosetta^{TM}2(DE3) Singles^{TM} competent cells (Novagen) were incubated with 1ug of plasmid DNA on ice for 2 min. The cells were heat-shocked at 42 °C for 45 s and then placed on ice for 2 min before being plated on LB-agar plates, supplemented with 100 ug/ml ampicillin (Sigma).

7.1.2 Protein Production

7.1.2.1 Protein expression

Inoculated *E. coli* Rosetta^{TM}2(DE3) Singles^{TM} (Novagen) cell cultures were grown in LB medium (Sigma) supplemented with 100 ug/ml ampicillin (Sigma) until the optical absorbance (A600) reached 0.6-0.8. Protein expression was induced

with 0.2 mM Isopropyl -D-1-thiogalactopyranoside (IPTG, Sigma) and the cultures were incubated overnight at 18 °C. Cells were harvested by centrifugation at 8,000×g for 15 min at 4 °C and the pellets were stored at -20 °C.

7.1.2.2 Protein purification

His-tagged constructs were purified using affinity chromatography followed by size exclusion chromatography. Cell pellets were resuspended in buffer: PBS, 360 mM NaCl, 10 mM β-ME, supplemented with bacterial protease inhibitors (Sigma). The lysate was sonicated for 4 min at 30 % amplitude (10 s on, 30 s off) on ice and centrifuged at 30,000×g at 4 °C for 60 min. The supernatant was diluted two-fold with buffer and incubated with 8 ml of equilibrated TALON beads (50 % slurry, Clontech) for 1 h at RT. The beads were collected in a medium-sized glass chromatography column (Bio-rad) and were washed with 50 CV of buffer (wash 1) followed by 50 CV wash with 10 mM imidazole buffer (wash 2). Proteins were eluted in 250 mM imidazole buffer and further purified by size-exclusion chromatography (Superdex S75 16/60, GE Healthcare) in buffer 10 mM Hepes 7.5, 150 mM NaCl. Proteins were concentrated using 5 kDa cut-off concentrators (Millipore), flash-frozen in liquid nitrogen and stored at -80 °C.

7.1.2.3 Protein crystallisation

Crystallisation screens were set up in 96-well Greiner or MRC Swissci plates using a Cartesian Technologies robot that dispenses 100 nL protein solution plus 100 nL reservoir solution at 3 different ratios 1:1, 2:1 and 1:2 in sitting drop

vapor diffusion format [293]. Crystallisation plates were maintained at 20.5 °C in a TAP Homebase storage vault and imaged via a Veeco visualization system [294]. Shh co-crystallisation attempts with Robotnikinin are described in detail in Chapter 2. For fragment screening, Shh24II was crystallised at 1.8 mg/ml or 3 mg/ml in the presence of 3 mM $CaCl_2$ and 10 mM sucrose octasulphate (SOS) (Sigma). ShhN24II-SOS initially crystallised in multiple conditions from commercial screens: Morpheous screen (conditions C1, C4, H8), Hampton Crystal screen 1 and 2 (conditions B3, C4) and Emerald Wizard 1&2 (condition A10). The optimised conditions used for the fragment-screening were the following: 48% Morpheous C1 condition, 1:1 drops (0.9 M NPS* (*0.3 M sodium nitrate, 0.3 M sodium phosphate dibasic, 0.3 M ammonium sulphate), 0.1 M imidazole, MES pH 6.5, 20% v/v PEG 500* MME; 10% w/v PEG 20000) and 28% block 1 C4, drop 1:2 (PEG 8000 30% w/v, sodium acetate trihydrate 0.2M, sodium cacodylate pH 6.5 0.1 M). Crystals appeared within 1-2 days but were left to grow for 5-7 days before soaking.

7.1.3 Crystallographic fragment-screen

7.1.3.1 Fragment Libraries

Two fragment libraries for screening were provided by the XChem Group at Diamond Light Source (DLS) beamline I04-1: Diamond and SGC poised library (DSPL) [221] and the commercially available Maybridge 1000 fragment library (Maybridge). Both libraries contained poised fragments that were identified based

on possessing a simple follow-up chemistry strategy. Poised fragments have the ability to be broken down into at least two synthons that can readily be elaborated into analogue libraries using parallel chemistry and can undergo common chemical reactions. These reactions include amide coupling, N-arylation, ester synthesis, ether synthesis, n-alkylation via Sn2, reductive amination, sonogashira coupling, sulfonamide coupling, souzuki coupling and urease/thiourease as previously described [221]. DSPL and Maybridge libraries were dissolved in DMSO or ethylene glycol at concentrations of 200 or 500 mM and accounted for $\sim$ 1000 screened fragments. A tailored library of 50 zinc-specific compounds was assembled by collaborators Prof Angela Russell and Dr Graham Wynne (Department of Chemistry, University of Oxford) and was supplied at 10 mM concentration in DMSO and ethylene glycol. A library of additional 50 compounds was assembled by collaborators from the OxXChem group and supplied at 200 mM in DMSO.

7.1.3.2 Crystal soaking

Crystals were imaged 1-3 days before fragment screening using a Veeco visualization system [294] and images were sorted using the TexRank software [295] that allowed for drops to be selected based on crystal size and quality. Fragment libraries were dispensed using an Echo acoustic liquid handler [296] to a final concentration of 30% DMSO (v/v) concentration and the crystals were left to soak at room temperature for 1h. Ethylene glycol was used as a cryoprotectant at 20% (v/v) before cryocooling the crystals with liquid nitrogen.

7.1.3.3 Crystallographic data collection, processing and refinement

All data were collected using the automated data collection pipeline at DLS beamline I04-1 [297] and were further analysed by an automatic software pipeline that included data processing using fast_dp [297] and Xia2 (CCP4) [298, 299]. Difference maps were calculated using ShhN-SOS structure as input for DIMPLE (part of the CCP4 suit [300]) within ISPyB available at the beamline. All data were viewed and managed through the XChem explorer, a graphical workflow tool for large-scale data management and structure determination [301]. Structures were viewed and real-space refined in Coot [302]. Subsequent rounds of refinement were done using program Refmac [303].

7.1.3.4 PanDDAs hit identification

The pan-dataset density analysis (PanDDA) software [224] was used to analyse all datasets after refinement derived form the Dimple pipeline. A common problem in crystallographic fragment screening is the inability to detect bound ligands due to their low occupancy. This is because traditional SigmaA-weighted 2Fo-Fc electron density maps present an average of two states (bound and unbound), which resembles the high-occupancy or major state (unbound) more than it does the minor (bound). Furthermore, the difference maps (Fo-Fc) fail to present the minor (bound) state correctly (Figure 7.1, a and b).

PanDDA produces an average ground state map by aligning 40-200 related isomorpous SigmaA-weighted 2Fo-Fc maps and calculating an average. The ground

state map is adjusted by the Background Density Correction (BDC) factor before being subtracted from each individual data set to produce the PanDDa or event map (Figure 7.1, c).

$$[\text{event map}] = [\text{dataset map}] - \text{BDC} \times [\text{ground state map}] \tag{7.1}$$

The event map more accurately represents the minor state, and allows for low-occupancy ligands and protein conformation to come into view. Once the model is built, PanDDA calculates a number of statistical parameters to validate the quality of the model after each round of refinement. PanDDa assesses the model by calculating how well it correlates with observed density by producing

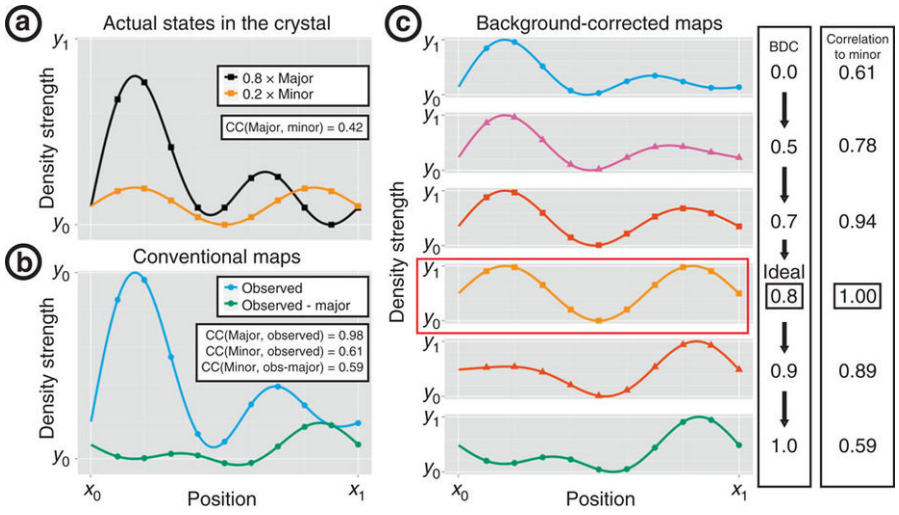


Figure 7.1: **Principle behind the PanDDa software.** a) The crystal may contain two diverse states, a major (black) present in 80 % of the crystal and a minor (orange) present in only 20 %. b) The traditional SigmaA-weighted electron density maps (blue) present an average state of the two. Because the traditional map resembles the major state, difference Fo-Fc maps (green) present the subtraction of the major state from the observed dataset. As a result, the minor state is not presented correctly with the exception of regions where the minor state has higher density than the major state. c) Background density correction (BDC) factor ranges from 0.0 (observed dataset) to 1.0 (Fo-Fc difference map). Only when applied at 0.8, the event map will correlate with the minor state (orange). Adapted from [224].

real-space correlation coefficient (RSCC, > 0.7). It further calculates the ratio between ligand B-factor and the surrounding protein (B-factor ratio, ~ 1), the movement of a ligand throughout refinement (RMSD, < 1 Å) and the difference density around the binding site as real-space Z difference score (RSZD, < 3).

The Z difference score is a statistical score that resembles the standard deviation and is calculated at every point in each dataset to show how much a particular point deviates from the ground state average maps. High Z scores at the 'event' site indicate high significance of density, while low Z scores for the area around the 'event' (RSZD) should indicate low noise levels. The Z difference score can conveniently be converted into a Z map, which is used for model building alongside PanDDa maps (Equation (7.1)) and as guidance during refinement.

7.1.4 Hit validation assays

7.1.4.1 Commercially available compounds

All compounds were kept as 10 mM stocks in anhydrous DMSO (Sigma) and stored at -20 °C. The list of all the purchased compounds can be found in Appendix Section 1.2.

7.1.4.2 Surface Plasmon Resonance (SPR)

SPR is a technique used to measure the binding between proteins and another binding partner like a small-molecule, antibody or another protein, and determine the affinity and kinetics of this interaction. The protein is used as a 'ligand' and

is immobilised onto an SPR sensor chip, while the binding partners or 'analytes' are injected over the sensor chip at a concentration gradient. SPR sensor chip contains a microfluidics system of 4 parallel flow-cells built on a layer of glass. Each flow cell contains a thin layer of gold and a layer of carboxymethyl-dextran to which the protein is coupled to. A beam of polarised light is internally reflected by the glass surface, however, a fraction of energy is absorbed by the free electrons of the gold layer (surface plasmon) and lost from the reflected light. The highest loss of energy happens at an angle called the resonance angle or θ_{spr}. This angle is dependent on the refractive indices of the two layers (protein and glass) neighbouring the gold layer. As the analytes bind to the protein, the refractive index changes and the the change in θ_{spr} is recorded.

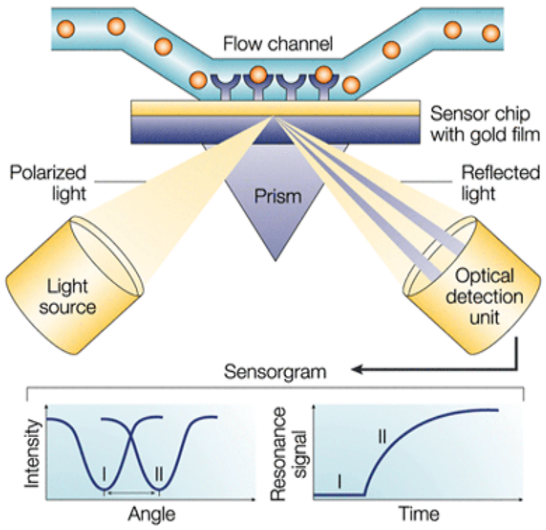


Figure 7.2: **Schematic illustration of the principles behind surface plasmon resonance.** Protein is immobilised onto a sensor chip surface coated with a layer of gold, while analytes are flown over. The resonance angle or θ_{spr} is correlated to the refractive index of the layers surrounding gold. Upon analyte binding, the refractive index changes resulting in a change in the resonance angle. Adapted from [304].

SPR experiments were performed on a Biacore S200 machine (GE Healthcare) developed especially for fragment screening. Running buffer (10 mM Hepes 7.5, 150 mM NaCl, 0.05% Tween 20, 5% DMSO) was prepared fresh with anhydrous DMSO (Sigma). The ShhN24II construct, used as ligand, was covalently immobilized onto Biacore CM5 S series chips (GE Healthcare) via amine coupling. Purified protein was diluted in sodium acetate, pH 5.5 to final concentration of 80 ug/ml. Both reference and active flow cells were treated with N-hydroxysuccinimide (NHS) and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) 1:1 mixture for 900s at 5 µl/min. The active flow cell was then saturated with protein to 3000-7000 response units (RU) at 10 µl/min. When the RU level was achieved both reference and active flow cells were treated with ethanolamine for 900s at 5 µl/min to block any remaining activated sites.

Small-molecule compounds were used as analytes and were dissolved in anhydrous DMSO (Sigma). Stock solutions were diluted in buffer to reach DMSO concentration of 5% and were further prepared in a dilution series using freshly prepared running buffer. Solvent correction series were calculated from 8 different buffer solutions with varying DMSO concentrations in the range of 4.5-5.8 %. Analytes were injected at either 80 µl/min for 120 s (120 s dissocitaion time) or at 50 µl/min for 40 s (80 s dissociation time). Robotnikinin was used as a positive control and Rmax control. The Biacore S200 evaluation software was used to prepare and evaluate data. All data sets were double blanked by subtracting reference flow cell and negative control sensorgrams. Data were also solvent corrected for DMSO discrepancies as described before [305]. The results

were evaluated using the affinity data evaluation module. The sensorgrams were exported and plotted in GraphPad Prism 7 (GraphPad Software).

7.1.4.3 Isothermal titration calorimetry (ITC)

ITC is a biophysical technique used to measure thermodynamics, stoichiometry and affinity of biomolecular interactions. Microcalorimeter contains two cell, a reference and a sample cell, which are maintained at the exact same temperature. The reference cell contains water, while the sample cell contains one of the binding partners. The other binding partner is injected in small aliquots from a syringe until binding saturation is reached. If the two molecules interact, the change in temperature in the sample cell is detected upon every injection and energy is used to bring the temperature back to equilibrium between the two cells. The amount of energy used is plotted against the molar ratio of the interacting ligands and the resulting isotherm is fitted to a binding model, from which relevant parameters are calculated.

ITC experiments were performed at 25 °C using MicroCal VP-ITC (Malvern). Protein solution was dialysed into PBS buffer overnight, and the compound was dissolved using the dialysis buffer. ShhN was placed in the sample cell at 20 µM, while FRG 155 was injected 26× at 200s intervals at 200 µM concentration. The NanoAnalyzer software (TA Instruments) was used to calculate stoichiometry (N) and binding affinity constant (Kd).

7.1.4.4 Gli-reporter assay

Gli-reporter assay is a common assay used to measure the activity of Hh signalling in cells. In short, the NIH 3T3 cells are starved to induce the formation of primary cilia and Hh responsiveness, and are treated with ShhN24II and respective compounds for 24h. The cells are harvested and all RNA is reverse-transcribed in RT-PCR reaction. Next, Gli1, Hh pathway specific transcriptional activator, and GAPDH, glyceraldehyde 3-phosphate dehydrogenase, are amplified in the presence of SYBR green fluorescent dye and detected using QPCR.

3T3 NIH cells (ATCC) were grown in DMEM media (Sigma) supplemented with 1 % L-glutamine, 1 % nonessential amino acids and 10% fetal bovine serum (Gibco by Life Technologies). Cells were plated onto 96-well flat-bottom plates at 20 x 10^3 cells per well and were grown for 24 h at 37 °C, 8% CO_2. The cells were starved for 24 h in low-serum media supplemented with 0.5% newborn calf serum (NCB) and treated with 50 nM Shh24II and the respective compounds at a final concentration of 0.5% DMSO. Cells were then incubated for 24 h at 37 °C, 8% CO_2. The QPCR assay was done using Power SYBR green Cell-to-Ct kit (Thermofisher Scientific) as per manufacturer's instructions. The cells were washed with PBS, resuspended in lysis solution an incubated with 1 unit of DNase I for 2 min, after which stop solution was added. The lysate was mixed with reverse transcriptase (RT) master mix and incubated at 37 °C for 1 h followed by 5 min at 95 °C. The QPCR reaction was prepared by combining SYBR green solution, 2 uM of either Gli primers (5'-CCA AGC CAA CTT TAT

GTC AGG G-3', 5'-AGC CCG CTT CTT TGT TAA TTT GA-3') or GAPDH primers (5'-AGT GGC AAA GTG GAG ATT-3', 5'-GTG GAG TCA TAC TGG AAC A-3') with RT samples. The QPCR cycle was set at 95 °C for 10 min, followed by 40 cycles of: 15 s at 95 °C, 60 s at 60 °C, and 5 s at 65 °C; and 5 s at 95 °C.

All samples were done in 2 or 3 biological replicates and 2 technical replicates, and 'no RT enzyme' and 'no template' QPCR samples were used as controls. The quantified cycles (Cq) were baseline subtracted, and an average Cq was calculated for each sample, along with standard deviation (SD). Both parameters were normalised relative to the housekeeping gene GAPDH and the blank sample.

7.1.5 *In silico* docking and molecular dynamics

7.1.5.1 Molecular structures and modelling of missing residues

The following structures of ShhN and DhhN were used for simulations and docking (pdb IDs): 1VHH, 2WFX, 2WFR, 2WFQ, 2WG3, 2WG4, 3N1R, 3HO5, 3D1M, 3M1N, 3MWX, 4C4M, 4C4N. Missing residues were modelled using the homology modelling software Modeller 9.9 [306].

7.1.5.2 Molecular dynamics simulations

The molecular dynamics of ShhN (pdb: 3N1R) was simulated in Gromacs 4.5.4. [307] using the OPLS-AA [308] and AMBER99SB-ILDN [309] force fields. Protein was dissolved in a cubic box with the SPC-E water model with 1.0 nm distance

from the box edge, and neutralised using Na^+ and Cl^- ions. Energy minimisation was carried out using the steepest descent method, and the Particle Mesh Ewald (PME) method was used for electrostatic interaction calculations. ShhN was equilibrated under the NVT ensemble for 200 ps and the NPT ensemble for 1000 ps. During the equilibration, the temperature was stabilised at 300 K and the pressure at 1 atm using the Berendsen thermostat and barostat [310], respectively. The molecular dynamics simulations was run for 100 ns with 2 fs timestep in both OPLS-AA and AMBER99SB-ILDN as a control. For the simulations, the V-rescale thermostat [311] and Parrinello-Rahman barostat [312] were used.

7.1.5.3 Molecular docking

The structure of Robotnikinin was drawn and minimised in MarvinSketch-6.3.0 (ChemAxon http://www.chemaxon.com), and docked in Gold using Gold Wizard [128]. All of the HhN structures were prepared by removing the solvent and any ligands, leaving only the HhN, and zinc and calcium ions. Hydrogens were added as part of the structure preparation process in the Gold Wizard. The binding site was selected by listing protein residues as per Gold manual and all the other settings were left as default. The ChemPLP scoring method [313] was used and every docking run was repeated three times.

7.1.5.4 Lead generation and consensus docking

Lead compounds were generated using the Xpoise software at DLS XChem (not published). Xpoise was developed as *in silico* follow up to xpoise library screens

developed by the DLS and SGC labs [221]. The software is designed to break down each fragment into multiple synthons, assign a reaction filter (from predefined common chemical reactions) and then to screen commercially available fragments to create novel compound libraries.

Consensus docking was done simultaneously, as described previously, using Autodock 4.2.6. [314] and Autodock vina 1.1.2 [315]. Each fragment library was docked to the protein structure where original fragment hit was obtained. The protein was prepared in Python Molecule Viewer [316] using Autodock tools [314] by removing waters and ligands, and adding hydrogens and Gasteiger charges. Ligands were converted into pdbqt format on the fly using Autodock 4 scripts. The box size that defined the search space was set to 15x15x15 Å with the centre of the original fragment hit as the centre of the search box. All of the remaining docking settings were left as default. After docking was completed, the RMSD was calculated between all Autodock and Vina poses. For RMSD calculation, a lower bound RMSD (lb) calculation was used (adapted from Vina source code by Teresa Paramo, SBCB) that takes into account atom types and overcomes challenges with symmetric and flexible ligands. Where more than 1 consensus pair was found, the one with the lowest energy was picked. Binding energies of consensus pairs were normalised by dividing the score with the cube root of heavy atoms [236], and were ranked from lowest to highest.

7.1.5.5 Molecular analysis

Molecular dynamics trajectories were analysed in VMD [317]. The docking results were analysed in Maestro [318]. Some aspects of structure analysis and some figures were made in PyMOL [212]. The graphs were made in Gnuplot [319].

7.2 Methods for Chapter 5

7.2.1 Molecular Biology

7.2.1.1 Construct design

Constructs of Skinny Hh from human (Genbank ID: NM_001122834.3, natural variant VAR_024743), fruit fly (Genbank ID: AF393157.1) and zebrafish (Genbank ID: NM_001114890.2) were cloned into pHLsec vector for mammalian expression lacking the secretion signal sequence [271]. The vector contains N-terminal Kozac sequence and a C-terminal 1D4 tag (TETSQVAPA) or a 3C protease recognition sequence (LEVLFQGPG) followed by a mVenus and a 1D4 tags [95].

7.2.1.2 Restriction digests, ligations and transformations

The fly and zebra fish genes were synthesised and codon-optimised for mammalian expression by LifeTechnologies. The genes were sub-cloned into pHL-sec vectors using EcoRI and KpnI and ligated using QuickLigase (NEB) enzymes as per manufacturer instructions. For cloning Skinny Hh constructs containing loop

deletions and Bril inserts, overlapping PCR strategy was used.

Gene fragments were amplified in a PCR reaction containing pyrobest buffer 1x (Takara), 250 µM dNTPs, 1 µM PCR templates, 10 ng DNA, 1 U Pyrobest DNA polymerase in 50 µl reactions. The PCR setting were: 95 °C 3 min, 34 cycles: 95 °C 30 s, 55-66 °C 30 s, 72 °C 90 s; 72 °C 5 min. The primers were added only after the first 5 cycles. The amplified DNA was digested using EcoRI and KpnI (NEB) and ligated into pHLsec vectors using QuickLigase (NEB), as per manufacturer's instructions. Transformation was performed as described in Section 7.1.1.2. but with DH5α^{TM} competent cells (Life Technologies).

7.2.2 Protein production and analysis

7.2.2.1 Small-scale expression

HEK293T cells were seeded in 6 well plates at 1 x 10^5 cells/ml and grown to approximately 90% confluency in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Sigma), 1% non-essential amino acids (Gibco) and 1% L-Glutamine (Gibco). The cells were transiently transfected by combining 8 µl of Lipofectamine transfection reagent (LifeTechnologies) in 250 µl of serum-free DMEM mixture with 4 µg of DNA in 250 µl serum-free DMEM. The Lipofectamine-DNA mixture was incubated for 20 min at RT and added to cells. Cells were harvested by centrifugation 48 h post-transfection. Cell pellets were stored at -20 °C.

7.2.2.2 Small-scale purification

Samples were kept at 4 °C at all times. The buffers used were: membrane buffer (MB) 20 mM HEPES pH 7.2, 300 mM NaCl; 10 % OGNG + 40:1 CHS; wash buffer (WB): MB + 0.15 % (3 CMC) OGNG 40:1 CHS; elution buffer (EB): wash + 500 uM 1D4 peptide. Cell pellets were placed on ice and lysed in 400 µl of MB with 1:100 mammalian protease inhibitors (Sigma) and 1.5 % OGNG, and rotated for 90 min. The samples were centrifuged at 16,000×g for 10 min at 4 °C to separate detergent-solubilised protein. The supernatant was diluted with MB to reach final concentration of detergent of approximately 1% and was incubated with 60 µl 50% slurry of Anti-Rho-1D4-beads (University of British Columbia) for 2 h rotating. The sample was centrifuged at 34×g for 30 s and the supernatant was carefully removed. The beads were washed twice with 20 CV of WB, followed by a centrifugation step at 34×g for 30 s. The beads were incubated with 3CV of EB rotating overnight and were centrifuged at 34×g for 30 s to remove eluate. The elution step was repeated with 1 h incubation time at 4 °C.

7.2.2.3 DNA purification

Large-scale protein expression in mammalian cells requires a significant amount of DNA. *E. coli* DH5α^{TM} competent cells were transformed with the appropriate plasmid and grown as 3L cultures overnight at 37 °C shaking. Cells were harvested in the morning and DNA was purified using Endotoxin-Free Plasmid

High Speed Giga Kit (Qiagen) as per manufacturer's instructions. One DNA Giga-Prep resulted in approximately 12 mg of purified plasmid DNA, sufficient for 24 litre transfection.

7.2.2.4 Large-scale expression

HEK293S cells were grown in 6-9 L suspension cultures to a density of 2-3 x 10^6 cells/ml of culture. 0.4 mg of DNA and 1.2 mg of PEI were individually pre-mixed with 20 ml Freestyle media (supplemented with 1% NEAA, 1% L-Glu), and then combined and incubated for 10 min. Cells from 800 ml of culture were harvested by centrifuged at 1000×g for 7 min. Cells were resuspended in DNA/PEI/Freestyle mixture and were added to 80 ml of Freestyle supplemented with 5 mM valproic acid (VPA) in a 2 L flasks. The cells were incubated in a shaking incubator for 3-6 h at 37 °C, 160 rpm, 8% CO2. The cells were topped up with PEM (supplemented with 0.5 % NEAA, 0.5 % L-Glu, 0.5 % Penicillin-Streptomycin) to make up 750 ml of culture and were grown for 2 days at 37 °C, 130 rpm, 8 % CO_2. The cells were harvested by centrifugation at 1400×g for 12 min, flash frozen in liquid nitrogen and stored at -80 °C.

7.2.2.5 Large-scale purification

The sample was kept at 4 °C at all times. The buffers used were: i) membrane buffer (MB) 20 mM HEPES pH 7.2, 300 mM NaCl, 10% glycerol, 10 mM L-arginine; ii) 10% DDM/DMNG/OGNG (w/v) + 40:1 CHS dissolved in water; iii) wash buffer I (WBI): MB + 10×CMC DDM/DMNG/OGNG 40:1

CHS; wash buffer II (WBII): MB + 3×CMC DDM/DMNG/OGNG 40:1 CHS; elution buffer (EB): WBII + 500 µM 1D4 peptide and SEC buffer: 2×CMC DMM/DMNG/OGNG 40:1 CHS, 10 mM HEPES pH 7.2, 300 mM NaCl, 5% glycerol. Flash-frozen pellets were thawed on ice and dissolved in 4 ml of MB per 1 mg of cell pellet. The detergent OGNG 40:1 CHS was added to final concentration of 1.5% and the sample was incubated for 90 min shaking or stirring at 4 °C. The sample was centrifuged at 16,000×g for 12 min at 4 °C. The supernatant was pooled, diluted to 1% of OGNGN 40:1 CHS and incubated with approximately 100 µl of 50% Anti-Rho1D4-beads per 1 g of pellet for 2 h. The sample was passed in a column, washed with 25 CVs of WB I and 25 CVs of WB II supplemented with DDM 40:1 CHS. Beads were incubated with 1 CV of EB supplemented with DDM 40:1 CHS, rotating overnight (ON). In the morning, the eluate was collected and the beads were incubated with another 1 CV of EB for 1 h. The eluate was collected and mixed with additives 0.05 % CYMAL-7-NG, 0.01 mg/ml POPC, Shh peptide (1:1.5 M protein:peptide) and non-hydrolysable PCoA 16:0 (1:1.5 M protein:PCoA) before being concentrated to 500 µl using 100-kDa cut-off vivaspin concentrators (Sartorius). The sample was gel filtrated on a Superdex 200 Increase 10/300 column using SEC buffer supplemented with DDM. The same additives were added again before protein was concentrated, flash frozen and stored at -80 °C.

7.2.2.6 Western Blot

20 µl sample was combined with 5 µl reducing protein loading dye and ran on an SDS-PAGE in MES buffer (Life Technologies) for 30 min at 180 V. The protein was transferred onto a nitrocellulose membrane in transfer buffer (Life Technologies) at 20 V for 30 min. The membrane was blocked with 5% milk PBS solution at RT and incubated with Rho purified Monoclonal Antibody at 1:10,000 or mouse Anti-5xHis antibody at 1:1,000 ratio, dissolved in 5% milk PBS solution, for 20 min at RT. The membrane was washed with PBS, 0.1% Tween 3x2 min. The membrane was incubated with Goat-alpha mouse IgG - Peroxidase antibody (Sigma A0168) at 1:2,000 dilution in 5% milk PBS for 20 min at RT. The membrane was washed with PBS, 0.1% Tween 3x2 min. The bands were visualised using a Clarity ECL Western Blotting Substrate (BioRad).

7.2.2.7 Thermostability and additive screens

Samples were prepared by combining 0.1 mg/ml Skinny Hh containing a Rho-1D4 tag with additives followed by incubation at a range of temperatures (30-80 °C) in a thermocycler for 1 h. The samples were analysed on a Shimadzu HPLC machine by injecting 10 µl sample on a Superose 6 3.2/300 in SEC buffer, detecting fluorescence and/or UV depending on the constructs. For the MPAS screen, additives were diluted 10x in protein solution before incubation. The membrane protein additive screen was assembled in-house comprising of: detergents, pH scree (with and without salt), salts, sugars, lipids, amino acids, alcohols, non-detergent

sulfobetaines, and denaturants and reducing agents. The full list of MPAS additives is provided in Appendix Section 1.3. The additional Skinny Hh specific additives used were: palmitoylated and non-palimtoylated Shh peptide (JPT peptide), Palmitoyl-CoA (Sigma), Guanosine 5'-[β,γ-imido]triphosphate (Sigma), Adenosine 5'-[γ-thio]triphosphate (Sigma), Phosphatidylcholin (Sigma), Phosphatidylethanolamine (Sigma), 16:0 Ether Coenzyme A (Avanti polar lipids). RUSKI compounds and TLH 45 compounds were obtained from Prof Ed Tate (Imperial College London).

7.2.3 Crystallisation

7.2.3.1 Lipidic cubic phase

The monoolein was melted at 42 °C and mixed with Skinny Hh containing a rho-1D4 tag at 60:40 lipid:protein ratio. A total of 57 µl of this mixture containing 1 mM P-CoA at 2.9 mg/ml was placed in two glass syringes connected with a coupler. The two were mixed gradually with a plunger until the mixture went clear. The drops were set up on siliconised glass sandwich plates using a Gryphon Modular Microdispenser System where one drop contained 50 nl of protein + 800 µl of mother liquor. The screens tested were MemMeso (Molecular Dimensions) and homemade screens provided by collaborator Prof Simon Newstead, University of Oxford.

7.2.3.2 Vapour diffusion

Screens were set up as described in section 7.1.2.3 in MRC Swiss-sci plates. Skinny Hh was crystallised at 1.4 mg/ml at three different drop ratios (1:1, 2:1 and 1:2). The screens tested were Morpheus, MemGold, MemGold2 (Molecular Dimensions) and plates were placed at 4 °C.

Appendices

1.1 Skinny Hh sequence alignment

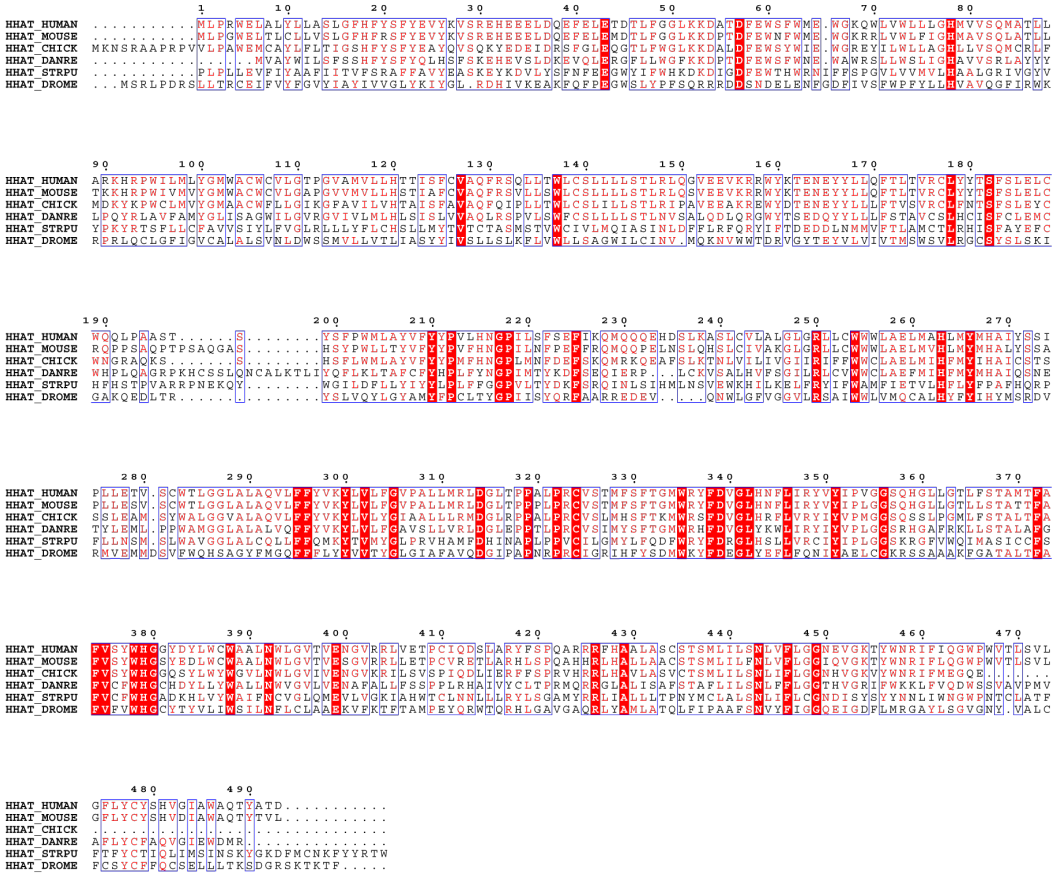


Figure 3: **Alignment of Skinny Hh protein sequenced.** Human, mouse, chick, zebra fish (DANRE), sea urchin (STRPU) and fly (DROME) Skinny Hh sequences were aligned using Multalin [133]. Conserved resides are shown in red.

1.2 List of fragments and compounds

Compound name	CAS number	Supplier	Catalogue number
Fragments			
FRG 17	864068-98-4	Sigma	CDS017420
FRG 48	87-52-5	Sigma	G10806
FRG 155	2538-34-3	Sigma	EN300-12273
FRG 159	203436-48-0	Sigma	ENA508627114
FRG 185	32087-05-1	Sigma	COM448686003
FRG 303	341021-30-5	Enamine	Z274555794
FRG 505	934-32-7	Sigma	171778
FRG 668	328288-82-0	Key Organics	MS-2014
Follow-up comopunds			
C1	4099-51-8	Sigma	D6071-10MG
C2	21749-83-7	Sigma	KEY1288004100
C3	131224-22-1	Enamine	EN300-57901
C7	1306603-44-0	Sigma	ENA321764431
C20	19293-62-0	Sigma	CDS000086
C21	91-00-9	Sigma	A53605
C22	574-66-3	Acros Organics - Fisher	342330050
C23	525-03-1	Spectrum Chemicals	TCA0622
C24	2157-52-0	Sigma	S375004
C25	40718-29-4	Sigma	PH010798
C26	3060-50-2	Acros Organics - Fisher	171720050
C27	3963-62-0	Acros Organics - Fisher	117250050
C30	1883-32-5	Acros Organics - Fisher	188220020
C31	720-44-5	Spectrum Chemicals	TCI M1278
C32	611-94-9	Sigma	M10301
C33	728-87-0	Sigma	162582
C34	119-61-9	Sigma	427551
Robotnikinin	1132653-79-2	Glentham Life Sciences	GL2680

Figure 4: **List of purchased fragments and compounds.**

1.3 Membrane protein additive screen

MPAS - COMPLETED SCREEN

Location	Condition
A1	MilliQ H_2O (blank)
A2	0.50% (w/v) LMNG
A3	0.33% (w/v) digitonin
A4	0.50% (w/v) CYMAL-6 NG
A5	0.50% (w/v) CYMAL-7 NG
A6	100 mM $CaCl_2$
A7	100 mM $MgCl_2$
A8	100 mM $MnCl_2$
A9	100 mM $ZnCl_2$
A10	(100 mM) $FeCl_3$
A11	500 mM KCl
A12	500 mM LiCl
B1	MilliQ H_2O (blank)
B2	100 mM $NiSO_4$
B3	0.1 mg/mL POPC in 0.2% (w/v) DMNG
B4	0.05 mg/mL POPE in 0.2% (w/v) DMNG
B5	0.1 mg/mL POPG in 0.2% (w/v) DMNG
B6	0.085 mg/mL POPC:POPE:POPG (2:1:2) in 0.2% (w/v) DMNG
B7	0.1 mg/mL POPC in 0.2% (w/v) DDM
B8	0.05 mg/mL POPE in 0.2% (w/v) DDM
B9	0.1 mg/mL POPG in 0.2% (w/v) DDM
B10	0.085 mg/mL POPC:POPE:POPG (2:1:2) in 0.2% (w/v) DDM
B11	0.1 mg/mL POPC in 1.0% (w/v) βOG
B12	0.1 mg/mL POPE in 1.0% (w/v) βOG
C1	MilliQ H_2O (blank)
C2	0.1 mg/mL POPG in 1.0% (w/v) βOG
C3	0.085 mg/mL POPC:POPE:POPG (2:1:2) in 1.0% (w/v) βOG
C4	100 mM arginine
C5	100 mM glutamate
C6	100 mM glycine
C7	1M NDSB-256
C8	1M NDSB-195
C9	10% (v/v) glycerol
C10	10% (v/v) ethylene glycol
C11	1M urea
C12	500 mM sodium acetate, pH4.0
D1	MilliQ H_2O (blank)
D2	500 mM sodium acetate, pH4.4
D3	500 mM sodium citrate, pH5.0
D4	500 mM sodium citrate, pH5.4
D5	500 mM sodium cacodylate, pH6.0
D6	500 mM sodium cacodylate, pH6.4
D7	500 mM HEPES, pH7.0
D8	500 mM HEPES, pH7.4
D9	500 mM Tris, pH8.0
D10	500 mM Tris, pH8.4
D11	500 mM CAPSO, pH9.0
D12	500 mM CAPSO, pH9.4

Location	Condition
E1	MilliQ H_2O (blank)
E2	200 mM sodium thiocyanate
E3	50% (v/v) glycerol
E4	50% (v/v) ethylene glycol
E5	10% (v/v) tert-butanol
E6	100 mM DMSO
E7	1M guanidinium hydrochloride
E8	50% (v/v) PEG 4,000
E9	50% (v/v) PEG 400
E10	20% (w/v) D(+)-glucose
E11	20% (w/v) sucrose
E12	100 mM sucrose octasulphate (SOS)
F1	MilliQ H_2O (blank)
F2	10 mM heparin
F3	50% (v/v) (+/-)-2-methyl-2,4-pentanediol (MPD)
F4	100 mM $CoCl_2$
F5	20% (v/v) ethanol
F6	1 mM DTT
F7	5 mM DTT
F8	100 mM alanine
F9	100 mM methionine
F10	100 mM serine
F11	100 mM L-proline
F12	100 mM phenol
G1	MilliQ H_2O (blank)
G2	2M sodium malonate pH7.0
G3	1M NH_4Cl
G4	1M $MgSO_4$
G5	50% (v/v) polypropylene glycol P 400
G6	20% (v/v) isopropanol
G7	50% (v/v) tacsimate pH7.0
G8	20% (w/v) benzamidine hydrochloride
G9	100 mM EDTA
G10	10% (v/v) 1,4-dioxane
G11	300 mM Gly-Gly-Gly (triglycine)
G12	500 mM sodium acetate, pH4.0, 4M NaCl
H1	MilliQ H_2O (blank)
H2	500 mM sodium acetate, pH4.4, 4M NaCl
H3	500 mM sodium citrate, pH5.0, 4M NaCl
H4	500 mM sodium citrate, pH5.4, 4M NaCl
H5	500 mM sodium cacodylate, pH6.0, 4M NaCl
H6	500 mM sodium cacodylate, pH6.4, 4M NaCl
H7	500 mM HEPES, pH7.0, 4M NaCl
H8	500 mM HEPES, pH7.4, 4M NaCl
H9	500 mM Tris, pH8.0, 4M NaCl
H10	500 mM Tris, pH8.4, 4M NaCl
H11	500 mM CAPSO, pH9.0, 4M NaCl
H12	500 mM CAPSO, pH9.4, 2M NaCl

Figure 5: **List of conditions used for the Membrane protein additive screen (MPAS).**

Bibliography

[1] Nusslein-Volhard, C. *et al.* (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature*, **287**: 795–801.

[2] Mohler, J. (1988) Requirements for Hedgehog, a Segmental Polarity Gene, in Patterning Larval and Adult Cuticle of *Drosophila*. *Genetics*, **120**: 1061–1072.

[3] Ingham, P. W. *et al.* (1991) Role of the *Drosophila patched* gene in positional signalling. *Nature*, **353**: 184–187.

[4] Lee, J. *et al.* (1992) Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell*, **71**: 33–50.

[5] Mohler, J. and Vani, K. (1992) Molecular organization and embryonic expression of the hedgehog gene involved in cell-cell communication in segmental patterning of *Drosophila*. *Development*, **115**: 957–971.

[6] Tabata, T. *et al.* (1992) The Drosophila hedgehog gene is expressed specifically in posterior compartment cells and is a target of engrailed regulation. *Genes Dev*, **6**(12b): 2635–2645.

[7] Tashiro, S. *et al.* (1993) Structure and expression of hedgehog, a Drosophila segment-polarity gene required for cell-cell communication. *Gene*, **124**(2): 183–189.

[8] 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**: 1417–1430.

[9] Krauss, S. *et al.* (1993) A funactionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in Zebrafish embryos. *Cell*, **75**: 1431–1444.

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

[11] Chang, D. T. *et al.* (1994) Products, genetic linkage and limb patterning activity of a murin hedgehog gene. *Development*, **120**: 3339–3353.

[12] 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**: 761–775.

[13] Wolpert, L. (1969) Positional information and the spatial pattern of cellular differentiation. *J of Theor Biol*, **25**(1): 1–47.

[14] Heberlein, U. *et al.* (1995) Growth and differentiation in the *Drosophila* eye coordinated by hedgehog. *Nature*, **373**: 709–711.

[15] Dominguez, M. (1999) Dual role for Hedgehog in the regulation of the proneural gene atonal during ommatidia development. *Development*, **126**: 2345–2353.

[16] Forbes, A. *et al.* (1996) Hedgehog is required for the proliferation and specification of somatic cells during egg chamber assembly in *Drosophila* oogenesis. *Development*, **122**: 1125–1135.

[17] Zhang, Y. and Kalderon, D. (2000) Regulation of cell proliferation and patterning in *Drosophila* oogenesis by Hedgehog signaling. *Development*, **127**: 2165–2176.

[18] Struhl, G. *et al.* (1997) Hedgehog organizes the pattern and polarity of epidermal cells in the *Drosophila* abdomen. *Development*, **124**: 2143–2154.

[19] Pankratz, M. and Hoch, M. (1995) Control of epithelial morphogenesis by cell signaling and integrin molecules in the *Drosophila* foregut. *Development*, **121**: 1885–1898.

[20] Huang, Z. and Kunes, S. (1996) Hedgehog, transmitted along retinal axons, triggers neurogenesis in the developing visual centers of the *Drosophila* brain. *Cell*, **86**: 411–422.

[21] Huang, Z. and Kunes, S. (1998) Signals transmitted along retinal axons in *Drosophila*: Hedgehog signal reception and the cell circuitry of lamina cartridge assembly. *Development*, **125**: 3753–3764.

[22] Glazer, L. and Shilo, B. (2001) Hedgehog signaling patterns the tracheal branches. *Development*, **128**: 1599–1606.

[23] Deshpande, G. *et al.* (2001) Hedgehog signaling in germ cell migration. *Cell*, **106**: 759–769.

[24] Bitgood, M. *et al.* (1996) Sertoli-cell signaling by Desert hedgehog regulates the male germline. *Curr. Biol.*, **6**: 298–304.

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

[26] Lanske, B. *et al.* (1996) PTH/PTHrP receptor in early development and Indian hedgehog-regulated bone growth. *Science*, **273**: 663–66.

[27] Vortkamp, A. *et al.* (1996) Indian hedgehog and parathyroid hormone-related protein regulate the rate of cartilage differentiation. *Science*, **273**: 613–622.

[28] Dyer, M. *et al.* (2001) Indian hedgehog activates hematopoiesis and vasculogenesis and can respecify prospective neurectodermal cell fate in the mouse embryo. *Development*, **128**: 1717–1730.

[29] R&D systems. Sonic Hedgehog: a morphogen involved in axon guidance. *www.rndsystems.com.*

[30] Zeller, R. *et al.* (2009) Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat Rev Genet*, **10**(12): 845–858.

[31] Ingham, P. W. and McMahon, A. P. (2001) Hedgehog signaling in animal development: paradigms and principles. *Genes Dev*, **15**(23): 3059–87.

[32] Levin, M. *et al.* (1995) A molecular pathway determining leftright asymmetry in chick embryogenesis. *Cell*, **82**: 803–814.

[33] Zhang, X. *et al.* (2001) Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R asymmetry by the mouse node. *Cell*, **105**: 781–792.

[34] Dahmane, N. and Ruiz-i Altaba, A. (1999) Sonic hedgehog regulates the growth and patterning of the cerebellum. *Development*, **126**: 3089–3100.

[35] Johnson, R. *et al.* (1994) Ectopic expression of Sonic hedgehog alters dorsalventral patterning of somites. *Cell*, **79**: 1165–1173.

[36] Bellusci, S. *et al.* (1997) Involvement of Sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. *Development*, **124**: 53–63.

[37] Roberts, D. *et al.* (1995) Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development*, **121**: 3163–3174.

[38] Outram, S. *et al.* (2000) Hedgehog signaling regulates differentiation from double-negative to double-positive thymocyte. *Immunity*, **13**: 187–197.

[39] Fan, C.-M. and Tessier-Lavigne, M. (1994) Patterning of mammalian somites by the surface ectoderm and the notochord: Evidence for sclеretome induction by Sonic hedgehog/Vhh- 1. *Cell*, **79**: 1175–1186.

[40] Porter, J. *et al.* (1996) Cholesterol modification of hedgehog signaling proteins in animal development. *Science*, **274**: 255–259.

[41] Porter, J. A. *et al.* (1996) Hedgehog Patterning Activity: Role of a Lipophilic Modification Mediated by the Carboxy-Terminal Autoprocessing Domain. *Cell*, **86**(1): 21–34.

[42] Perler, F. B. (1998) Protein Splicing of Inteins and Hedgehog Autoproteolysis: Structure, Function, and Evolution. *Cell*, **92**(1): 1–4.

[43] Hall, T. M. T. *et al.* (1997) Crystal Structure of a Hedgehog Autoprocessing Domain: Homology between Hedgehog and Self-Splicing Proteins. *Cell*, **91**: 85–97.

[44] Chamoun, Z. *et al.* (2001) Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. *Science*, **293**(5537): 2080–4.

[45] Pepinsky, R. B. *et al.* (1998) Identification of a Palmitic Acid-modified Form of Human Sonic hedgehog. *J Biol Chem*, **273**(22): 14037–14045.

[46] Buglino, J. A. and Resh, M. D. (2008) Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *J Biol Chem*, **283**(32): 22076–88.

[47] Rietveld, A. *et al.* (1999) Association of sterol- and glycosylphosphatidylinositol-linked proteins with Drosophila raft lipid microdomains. *J Biol Chem*, **274**(17): 12049–54.

[48] Peters, C. *et al.* (2004) The cholesterol membrane anchor of the Hedgehog protein confers stable membrane association to lipid-modified proteins. *Proc Natl Acad Sci U S A*, **101**(23): 8531–6.

[49] 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 Dev*, **18**(6): 641–59.

[50] Briscoe, J. *et al.* (2001) A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol Cell*, **7**(6): 1279–91.

[51] 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): 803–15.

[52] Tian, H. *et al.* (2005) Mouse Disp1 is required in sonic hedgehog-expressing cells for paracrine activity of the cholesterol-modified ligand. *Development*, **132**(1): 133–42.

[53] Amanai K, J. J. (2001) Distinct roles of Central missing and Dispatched in sending the Hedgehog signal. *Development*, **128**: 5119–5127.

[54] Tukachinsky, H. *et al.* (2012) Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand. *Cell Rep*, **2**(2): 308–20.

[55] Creanga, A. *et al.* (2012) Scube/You activity mediates release of dually lipid-modified Hedgehog signal in soluble form. *Genes Dev*, **26**(12): 1312–25.

[56] Callejo, A. *et al.* (2011) Dispatched mediates Hedgehog basolateral release to form the long-range morphogenetic gradient in the Drosophila wing disk epithelium. *Proc Natl Acad Sci U S A*, **108**(31): 12591–8.

[57] Zeng, X. *et al.* (2001) A freely diffusible form of Sonic hedgehog mediates long-range signalling. *Nature*, **411**(6838): 716–20.

[58] Jakobs, P. *et al.* (2014) Scube2 enhances proteolytic Shh processing from the surface of Shh-producing cells. *J Cell Sci*, **127**(Pt 8): 1726–37.

[59] Gallet, A. *et al.* (2006) Cholesterol modification is necessary for controlled planar long-range activity of Hedgehog in *Drosophila* epithelia. *Development*, **133**(3): 407–18.

[60] Kohtz1, J. D. *et al.* (2001) N-terminal fatty-acylation of sonic hedgehog enhance the induction of rodent ventral forebrain neurons. *Development*, **128**: 2351–2363.

[61] Gallet, A. *et al.* (2003) Cholesterol modification of hedgehog is required for trafficking and movement, revealing an asymmetric cellular response to hedgehog. *Dev Cell*, **4**(2): 191–204.

[62] Goetz, J. A. *et al.* (2006) A highly conserved amino-terminal region of sonic hedgehog is required for the formation of its freely diffusible multimeric form. *J Biol Chem*, **281**(7): 4087–93.

[63] Feng, J. *et al.* (2004) Synergistic and antagonistic roles of the Sonic hedgehog N- and C-terminal lipids. *Development*, **131**(17): 4357–70.

[64] Briscoe, J. and Therond, P. P. (2013) The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat Rev Mol Cell Biol*, **14**(7): 416–29.

[65] 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. *Dev Cell*, **18**(4): 605–20.

[66] Gallet, A. *et al.* (2008) Cellular trafficking of the glypican Dally-like is required for full-strength Hedgehog signaling and wingless transcytosis. *Dev Cell*, **14**(5): 712–25.

[67] Ortmann, C. *et al.* (2015) Sonic hedgehog processing and release are regulated by glypican heparan sulfate proteoglycans. *J Cell Sci*, **128**(12): 2374–2385.

[68] Panakova, D. *et al.* (2005) Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature*, **435**(7038): 58–65.

[69] Eugster, C. *et al.* (2007) Lipoprotein-heparan sulfate interactions in the Hh pathway. *Dev Cell*, **13**(1): 57–71.

[70] Tanaka, Y. *et al.* (2005) FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. *Nature*, **435**(7039): 172–177.

[71] Vyas, N. *et al.* (2014) Vertebrate Hedgehog is secreted on two types of extracellular vesicles with different signaling properties. *Sci Rep*, **4**: 7357.

[72] 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 Dev*, **18**(6): 641–59.

[73] Huang, X. *et al.* (2007) Region-specific requirement for cholesterol modification of sonic hedgehog in patterning the telencephalon and spinal cord. *Development*, **134**(11): 2095–105.

[74] Li, Y. *et al.* (2006) Cholesterol modification restricts the spread of Shh gradient in the limb bud. *Proc Natl Acad Sci U S A*, **103**(17): 6548–53.

[75] Burke, R. *et al.* (1999) Dispatched, a Novel Sterol-Sensing Domain Protein Dedicated to the Release of Cholesterol-Modeified Hedgehog from Signaling Cell. *Cell*, **99**: 803–815.

[76] Kawakami, T. *et al.* (2002) Mouse dispatched mutants fail to distribute hedgehog proteins and are defective in hedgehog signaling. *Development*, **129**(24): 5753–5765.

[77] Tokhunts, R. *et al.* (2010) The Full-length Unprocessed Hedgehog Protein Is an Active Signaling Molecule. *J Biol Chem*, **285**(4): 2562–2568.

[78] Maity, T. *et al.* (2005) Molecular mechanisms of Sonic hedgehog mutant effects in holoprosencephaly. *Proc Natl Acad Sci U S A*, **102**(47): 17026–17031.

[79] Roessler, E. *et al.* (2009) The mutational spectrum of holoprosencephaly-associated changes within the SHH gene in humans predicts loss-of-function through either key structural alterations of the ligand or its altered synthesis. *Human Mutation*, **30**(10): E921–E935.

[80] Lee, J. D. and Treisman, J. E. (2001) Sightless has homology to transmembrane acyltransferase and is required to generate active Hedgehog protein. *Curr Biol*, **11**: 1147–1152.

[81] Dawber, R. J. *et al.* (2005) Differential range and activity of various forms of the Hedgehog protein. *BMC Dev Biol*, **5**: 21.

[82] Dennis, J. F. *et al.* (2012) Mutations in Hedgehog acyltransferase (Hhat) perturb Hedgehog signaling, resulting in severe acrania-holoprosencephaly-agnathia craniofacial defects. *PLoS Genet*, **8**(10): e1002927.

[83] Rohatgi, R. *et al.* (2007) Patched1 Regulates Hedgehog Signaling at the Primary Cilium. *Science*, **317**(5836): 372–376.

[84] Han, Y.-G. *et al.* (2008) Hedgehog signaling and primary cilia are required for the formation of adult neural stem cells. *Nat Neurosci*, **11**(3): 277–284.

[85] Carpenter, D. *et al.* (1998) Characterization of two patched receptors for the vertebrate hedgehog protein family. *Proc Natl Acad Sci U S A*, **95**(23): 13630–13634.

[86] Nakano, Y. *et al.* (1989) A protein with several possible membrane-spanning domains encoded by the Drosophila segment polarity gene patched. *Nature*, **341**(6242): 508–513.

[87] Stone, D. M. *et al.* (1996) The tumour-supressor gene patched encodes a candidate receptor for Sonic hegdehog. *Nature*, **384**: 129–133.

[88] Marigo, V. *et al.* (1996) Biochemical evidence that Patched is the Hedgehog receptor. *Nature*, **384**: 176–179.

[89] Taipale, J. *et al.* (2002) Patched acts catalytically to suppress the activity of Smoothened. *Nature*, **418**(6900): 892–896.

[90] Fuse, N. *et al.* (1999) Sonic hedgehog protien signals not as a hydrolytic enzyme but as an apparent ligand for Patched. *Proc Natl Acad Sci U S A*, **96**: 10992–10999.

[91] Milenkovic, L. *et al.* (2009) Lateral transport of Smoothened from the plasma membrane to the membrane of the cilium. *J Cell Biol*, **187**(3): 365–374.

[92] Chen, Y. and Struhl, G. (1996) Dual Roles for Patched in Sequestering and Transducing Hedgehog. *Cell*, **87**: 553–563.

[93] Wu, F. *et al.* (2017) Hedgehog Signaling: From Basic Biology to Cancer Therapy. *Cell Chem Biol*, **24**(3): 252 – 280.

[94] Denef, N. *et al.* (2000) Hedgehog induces opposite changes in turnover and subcellular localization of Patched and Smothened. *Cell*, **102**: 521–531.

[95] Byrne, E. F. X. *et al.* (2016) Structural basis of Smoothened regulation by its extracellular domains. *Nature*, **535**(7613): 517–522.

[96] Wang, C. *et al.* (2014) Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nat Commun*, **5**: 4355.

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

[98] Nachtergaele, S. *et al.* (2013) Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *Elife*, **2**: e01340.

[99] Whalen, D. M. *et al.* (2013) Structural insights into proteoglycan-shaped Hedgehog signaling. *Proc Natl Acad Sci U S A*, **110**(41): 16420–5.

[100] Nedelcu, D. *et al.* (2013) Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nat Chem Biol*, **9**(9): 557–64.

[101] Nachtergaele, S. *et al.* (2012) Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nat Chem Biol*, **8**(2): 211–20.

[102] Luchetti, G. *et al.* (2016) Cholesterol activates the G-protein coupled receptor Smoothened to promote Hedgehog signaling. *eLife*, **5**: e20304.

[103] DeBerardinis, A. M. *et al.* (2013) Identification of Vitamin D3-Based Hedgehog Pathway Inhibitors That Incorporate an Aromatic A-Ring Isostere. *ACS Med Chem Lett*, **4**(7): 590–595.

[104] Bidet, M. *et al.* (2011) The Hedgehog Receptor Patched is Involved in Cholesterol Transport. *PLoS One*, **6**: e23834.

[105] Bijlsma, M. *et al.* (2006) Repression of Smoothened by Patched-Dependent (Pro-)Vitamin D3 Secretion. *PLoS Biology*, **4**: e232.

[106] Tseng, T. *et al.* (1999) The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J Mol Microbiol Biotechnol*, **1**: 107–125.

[107] Eaton, S. (2008) Multiple roles for lipids in the Hedgehog signalling pathway. *Nat Rev Mol Cell Biol*, **9**(6): 437–45.

[108] Allen, B. L. *et al.* (2007) The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development. *Genes Dev*, **21**(10): 1244–57.

[109] 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. *Dev Cell*, **10**(5): 647–56.

[110] Okada, A. *et al.* (2006) Boc is a receptor for sonic hedgehog in the guidance of commissural axons. *Nature*, **444**(7117): 369–373.

[111] McLellan, J. S. *et al.* (2008) The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. *Nature*, **455**(7215): 979–83.

[112] Izzi, L. *et al.* (2011) Boc and Gas1 Each Form Distinct Shh Receptor Complexes with Ptch1 and Are Required for Shh-Mediated Cell Proliferation. *Dev Cell*, **20**(6): 788 – 801.

[113] Allen, B. *et al.* (2011) Overlapping Roles and Collective Requirement for the Coreceptors GAS1, CDO, and BOC in SHH Pathway Function. *Dev Cell*, **20**(6): 775 – 787.

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

[115] Bishop, B. *et al.* (2009) Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. *Nat Struct Mol Biol*, **16**(7): 698–703.

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

[117] Capurro, M. I. *et al.* (2008) Glypican-3 Inhibits Hedgehog Signaling during Development by Competing with Patched for Hedgehog Binding. *Dev Cell*, **14**(5): 700 – 711.

[118] Chong, Y. C. *et al.* (2015) Bifurcating action of Smoothened in Hedgehog signaling is mediated by Dlg5. *Genes Dev*, **29**(3): 262–276.

[119] Dorn, K. V. *et al.* (2012) A Smoothened-Evc2 Complex Transduces the Hedgehog Signal at Primary Cilia. *Dev Cell*, **23**: 823–835.

[120] Hui, C. C. and Angers, S. (2011) Gli proteins in development and disease. *Annu Rev Cell Dev Biol*, **27**: 513–37.

[121] Zhang, Y. *et al.* (2013) Structural insight into the mutual recognition and regulation between Suppressor of Fused and Gli/Ci. *Nat Commun*, **4**: 2608.

[122] Tukachinsky, H. *et al.* (2010) A mechanism for vertebrate Hedgehog signaling: recruitment to cili and dissociation of SuFu-GLI protein complexes. *JCB*, **191**: 415–428.

[123] Tuson, M. *et al.* (2011) Protein kinase A acts at the basal body of the primary cilium to prevent Gli2 activation and ventralization of the mouse neural tube. *Development*, **138**(22): 4921–4930.

[124] Chen, Y. *et al.* (2011) Sonic Hedgehog Dependent Phosphorylation by CK1 and GRK2 Is Required for Ciliary Accumulation and Activation of Smoothened. *PLOS Biology*, **9**(6): 1–16.

[125] He, M. *et al.* (2014) The kinesin-4 protein Kif7 regulates mammalian Hedgehog signalling by organizing the cilium tip compartment. *Nat Cell Biol*, **16**(7): 663–672.

[126] Liu, A. *et al.* (2005) Mouse intraflagellar transport proteins regulate both the activator and repressor functions of Gli transcription factors. *Development*, **132**(13): 3103–3111.

[127] Tukachinsky, H. *et al.* (2010) A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu–Gli protein complexes. *J Cell Biol*, **191**(2): 415–428.

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

[129] Holmes, M. and Matthews, B. (1981) Binding of hydroxamic acid inhibitors to crystalline thermolysin suggests a pentacoordinate zinc intermediate in catalysis. *Biochemistry*, **20**: 6912–6920.

[130] Farshi, P. *et al.* (2011) Dual Roles of the Cardin-Weintraub Motif in Multimeric Sonic Hedgehog. *Journal of Biological Chemistry*, **286**(26): 23608–23619.

[131] Day, E. S. *et al.* (1999) Zinc-dependent structural stability of human sonic hegdehog. *Biochemistry*, **38**: 14868–14880.

[132] 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. *J Biol Chem*, **285**(32): 24584–24590.

[133] Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl Acids Res*, **16**(22): 10881–10890.

[134] Chuang, P. *et al.* (2003) Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes Dev*, **17**(3): 342–7.

[135] 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. *Nat Struct Mol Biol*, **16**(7): 691–7.

[136] Tukachinsky, H. *et al.* (2016) Mechanism of inhibition of the tumor suppressor Patched by Sonic Hedgehog. *Proc Natl Acad Sci U S A*, **113**(40): E5866–E5875.

[137] Taylor, F. R. *et al.* (2001) Enhanced Potency of Human Sonic Hedgehog by Hydrophobic Modification. *Biochemistry*, **40**(14): 4359–4371.

[138] Pepinsky, R. B. *et al.* (2000) Mapping Sonic Hedgehog-Receptor Interactions by Steric Interference. *J Biol Chem*, **275**(15): 10995–11001.

[139] Maun, H. R. *et al.* (2010) Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. *J Biol Chem*, **285**(34): 26570–26580.

[140] Cardin, A. D. and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **9**(1): 21–32.

[141] Rubin, J. B. *et al.* (2002) Cerebellar proteoglycans regulate sonic hedgehog responses during development. *Development*, **129**(9): 2223–2232.

[142] Williams, E. H. *et al.* (2010) Dally-like core protein and its mammalian homologues mediate stimulatory and inhibitory effects on Hedgehog signal response. *Proc Natl Acad Sci U S A*, **107**(13): 5869–5874.

[143] Filmus, J. and Capurro, M. (2014) The role of glypicans in Hedgehog signaling. *Matrix Biology*, **35**(Supplement C): 248 – 252.

[144] Chan, J. A. *et al.* (2009) Proteoglycan interactions with Sonic Hedgehog specify mitogenic responses. *Nat Neurosci*, **12**(4): 409–417.

[145] 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): 3933–3941.

[146] Roessler, E. *et al.* (1996) Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Gen*, **14**: 357–360.

[147] Reiter, J. F. and Leroux, M. R. (2017) Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol*, **18**(9): 533–547.

[148] Kang, S. *et al.* (1997) GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet*, **15**(3): 266–268.

[149] Kelley, R. I. *et al.* (1996) Holoprosencephaly in RSH/Smith-Lemli-Opitz syndrome: Does abnormal cholesterol metabolism affect the function of sonic hedgehog? *ACS Med Gen*, **66**(4): 478–484.

[150] Ma, G. *et al.* (2011) Indian hedgehog mutations causing brachydactyly type A1 impair Hedgehog signal transduction at multiple levels. *Cell Res*, **21**(9): 1343–1357.

[151] Salama, G. *et al.* (2015) Cyclopia: A Rare Condition with Unusual Presentation A Case Report. *Clin Med Insights Pediatr*, **9**: 19–23.

[152] Pineda-Alvarez, D. E. *et al.* (2011) A broad range of ophthalmologic anomalies is part of the holoprosencephaly spectrum. *ACS Med Gen Part A*, **155**(11): 2713–2720.

[153] Dubourg, C. *et al.* (2004) Molecular screening of SHH, ZIC2, SIX3, and TGIF genes in patients with features of holoprosencephaly spectrum: Mutation review and genotypephenotype correlations. *Hum Mutat*, **24**(1): 43–51.

[154] Heussler, H. S. *et al.* (2002) Extreme variability of expression of a Sonic Hedgehog mutation: attention difficulties and holoprosencephaly. *Arch Dis Child*, **86**(4): 293–296.

[155] Orioli, I. e. a. (2001) Identification of novel mutations in SHH and ZIC2 in a South American (ECLAMC) population with holoprosencephaly. *Hum Genet*, **109**: 1–6.

[156] Gao, B. *et al.* (2001) Mutations in IHH, encoding Indian hedgehog, cause brachydactyly type A-1. *Nat Genet*, **28**(4): 386–388.

[157] Villavicencio EH, I. P., Walterhouse DO (2000) The Sonic Hedgehog-PatchedGli Pathway in Human Development and Disease. *ACS J Hum Genet*, **67**: 1047–1054.

[158] Petrova, R. and Joyner, A. L. (2014) Roles for Hedgehog signaling in adult organ homeostasis and repair. *Development*, **141**(18): 3445–3457.

[159] Han, Y.-G. *et al.* (2008) Hedgehog signaling and primary cilia are required for the formation of adult neural stem cells. *Nat Neurosci*, **11**(3): 277–284.

[160] Machold, R. *et al.* Sonic Hedgehog Is Required for Progenitor Cell Maintenance in Telencephalic Stem Cell Niches. *Neuron*, **39**(6): 937–950.

[161] Ahn, S. and Joyner, A. L. (2005) In vivo analysis of quiescent adult neural stem cells responding to Sonic hedgehog. *Nature*, **437**(7060): 894–897.

[162] Lai, K. *et al.* (2003) Sonic hedgehog regulates adult neural progenitor proliferation in vitro and in vivo. *Nat Neurosci*, **6**(1): 21–27.

[163] Bambakidis, N. C. *et al.* (2003) Sonic hedgehoginduced neural precursor proliferation after adult rodent spinal cord injury. *J Neurosurg Spine*, **99**(1): 70–75.

[164] Garcia, A. D. R. *et al.* (2010) Sonic Hedgehog Regulates Discrete Populations of Astrocytes in the Adult Mouse Forebrain. *J Neurosci*, **30**(41): 13597–13608.

[165] Patel, S. S. *et al.* (2017) Targeting sonic hedgehog signaling in neurological disorders. *Neurosci Biobehav Rev*, **74**(Pt A): 76–97.

[166] Das, I. *et al.* (2013) Hedgehog Agonist Therapy Corrects Structural and Cognitive Deficits in a Down Syndrome Mouse Model. *Sci Transl Med*, **5**(201): 201ra120.

[167] Bragina, O. *et al.* (2010) Smoothened agonist augments proliferation and survival of neural cells. *Neurosci Lett*, **482**(2): 81 – 85.

[168] Peterson, R. and Turnbull, J. (2012) Sonic Hedgehog is Cytoprotective against Oxidative Challenge in a Cellular Model of Amyotrophic Lateral Sclerosis. *J Mol Neurosci*, **47**(1): 31–41.

[169] Yuan, F. *et al.* (2015) Efficient generation of region-specific forebrain neurons from human pluripotent stem cells under highly defined condition. *Sci Rep*, **5**: 18550.

[170] Amakye, D. *et al.* (2013) Unraveling the therapeutic potential of the Hedgehog pathway in cancer. *Nat Med*, **19**(11): 1410–1422.

[171] Hanna, A. and Shevde, L. A. (2016) Hedgehog signaling: modulation of cancer properies and tumor microenvironment. *Mol Cancer*, **15**: 24.

[172] Xie, J. *et al.* (1998) Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature*, **391**(6662): 90–92.

[173] Berman, D. M. *et al.* (2002) Medulloblastoma Growth Inhibition by Hedgehog Pathway Blockade. *Science*, **297**(5586): 1559–1561.

[174] Hahn, H. *et al.* (1998) Rhabdomyosarcomas and radiation hypersensitivity in a mouse model of Gorlin syndrome. *Nat Med*, **4**(5): 619–622.

[175] Hahn, H. *et al.* (1996) Mutations of the Human Homolog of Drosophila patched in the Nevoid Basal Cell Carcinoma Syndrome. *Cell*, **85**(6): 841 – 851.

[176] Gorlin, R. (1995) Nevoid basal cell carcinoma syndrome. *Dermatol Clin*, **13**: 113–125.

[177] Gailani, M. R. *et al.* (1996) The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas. *Nat Genet*, **14**(1): 78–81.

[178] Vorechovsky, I. e. a. (1997) Somatic mutations in the human homologue of Drosophila patched in primitive neuroectodermal tumours. *Oncogene*, **15**: 361–366.

[179] Barreto, D. *et al.* (2000) PTCH Gene Mutations in Odontogenic Keratocysts. *Journal of Dental Research*, **79**(6): 1418–1422.

[180] Brugires, L. *et al.* (2012) High Frequency of Germline SUFU Mutations in Children With Desmoplastic/Nodular Medulloblastoma Younger Than 3 Years of Age. *Journal of Clinical Oncology*, **30**(17): 2087–2093.

[181] Reifenberger, J. *et al.* (1998) Missense Mutations in SMOH in Sporadic Basal Cell Carcinomas of the Skin and Primitive Neuroectodermal Tumors of the Central Nervous System. *Cancer Research*, **58**(9): 1798–1803.

[182] Dlugosz, A. *et al.* (2012) Vismodegib. *Nat Rev Drug Discov*, **11**(6): 437–438.

[183] Burness, C. B. (2015) Sonidegib: First Global Approval. *Drugs*, **75**(13): 1559–1566.

[184] Sekulic, A. *et al.* (2012) Efficacy and Safety of Vismodegib in Advanced Basal-Cell Carcinoma. *New England Journal of Medicine*, **366**(23): 2171–2179.

[185] Tang, J. Y. *et al.* (2016) Inhibition of the hedgehog pathway in patients with basal-cell nevus syndrome: final results from the multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet Oncology*, **17**(12): 1720 – 1731.

[186] Mukherjee, S. *et al.* (2006) Hedgehog Signaling and Response to Cyclopamine Differ in Epithelial and Stromal Cells in Benign Breast and Breast Cancer. *Cancer biology & therapy*, **5**(6): 674–683.

[187] Watkins, D. N. *et al.* (2003) Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature*, **422**(6929): 313–317.

[188] Karhadkar, S. S. *et al.* (2004) Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature*, **431**(7009): 707–712.

[189] Berman, D. M. *et al.* (2003) Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature*, **425**(6960): 846–851.

[190] Varnat, F. *et al.* (2009) Human colon cancer epithelial cells harbour active HEDGEHOG-GLI signalling that is essential for tumour growth, recurrence, metastasis and stem cell survival and expansion. *EMBO Molecular Medicine*, **1**(6-7): 338–351.

[191] Theunissen, J. W. and de Sauvage, F. J. (2009) Paracrine Hedgehog signaling in cancer. *Cancer Res*, **69**(15): 6007–6010.

[192] Yauch, R. L. *et al.* (2008) A paracrine requirement for hedgehog signalling in cancer. *Nature*, **455**(7211): 406–410.

[193] Li, X. *et al.* (2014) Sonic Hedgehog Paracrine Signaling Activates Stromal Cells to Promote Perineural Invasion in Pancreatic Cancer. *Clinical Cancer Research*, **20**(16): 4326–4338.

[194] Clevers, H. (2011) The cancer stem cell: premises, promises and challenges. *Nat Med*, **17**: 313–319.

[195] Bailey, J. M. *et al.* (2009) Sonic hedgehog paracrine signaling regulates metastasis and lymphangiogenesis in pancreatic cancer. *Oncogene*, **28**(40): 3513–3525.

[196] Huang, S. *et al.* (2006) Activation of the hedgehog pathway in human hepatocellular carcinomas. *Carcinogenesis*, **27**(7): 1334–1340.

[197] Stanton, B. Z. *et al.* (2009) A small molecule that binds Hedgehog and blocks its signaling in human cells. *Nat Chem Biol*, **5**(3): 154–156.

[198] Petrova, E. *et al.* (2013) Inhibitors of Hedgehog acyltransferase block Sonic Hedgehog signaling. *Nat Chem Biol*, **9**(4): 247–9.

[199] Rodgers, U. R. *et al.* (2016) Characterization of Hedgehog Acyltransferase Inhibitors Identifies a Small Molecule Probe for Hedgehog Signaling by Cancer Cells. *ACS J Chem Biol*, **11**(12): 3256–3262.

[200] Rodriguez-Blanco, J. *et al.* (2013) The hedgehog processing pathway is required for NSCLC growth and survival. *Oncogene*, **32**(18): 2335–45.

[201] Konitsiotis, A. D. *et al.* (2014) Attenuation of hedgehog acyltransferase-catalyzed sonic Hedgehog palmitoylation causes reduced signaling, proliferation and invasiveness of human carcinoma cells. *PLoS One*, **9**(3): e89899.

[202] Carney, T. J. and Ingham, P. W. (2013) Drugging Hedgehog: signaling the pathway to translation. *BMC Biology*, **11**(13): 37.

[203] Ng, J. M. and Curran, T. (2011) The Hedgehog's tale: developing strategies for targeting cancer. *Nat Rev Cancer*, **11**(7): 493–501.

[204] Epstein, E. H. (2008) Basal cell carcinomas: attack of the hedgehog. *Nat Rev Cancer*, **8**(10): 743–754.

[205] Raffel, C. *et al.* (1997) Sporadic Medulloblastomas Contain PTCH Mutations. *Cancer Research*, **57**(5): 842–845.

[206] Sanchez, P. *et al.* (2004) Inhibition of prostate cancer proliferation by interference with Sonic Hedgehog-Gli1 signaling. *PNAS*, **101**(34): 12561–12566.

[207] Peng, L. F. *et al.* (2009) Syntheses of aminoalcohol-derived macrocycles leading to a small-molecule binder to and inhibitor of Sonic Hedgehog. *Bioorg Med Chem Lett*, **19**(22): 6319 – 6325.

[208] Dockendorff, C. *et al.* (2012) Macrocyclic Hedgehog Pathway Inhibitors: Optimization of Cellular Activity and Mode of Action Studies. *ACS Med Chem Lett*, **3**(10): 808–813.

[209] Hwang, S. *et al.* (2013) Molecular dynamics simulations of sonic hedgehog-receptor and inhibitor complexes and their applications for potential anti-cancer agent discovery. *PLoS One*, **8**(7): e68271.

[210] Berman, H. M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**(1): 235–242.

[211] Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*, **267**(3): 727 – 748.

[212] The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

[213] Paul, S. M. *et al.* (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*, **9**(3): 203–14.

[214] Rees, D. C. *et al.* (2004) Fragment-based lead discovery. *Nat Rev Drug Discov*, **3**(8): 660–72.

[215] Murray, C. W. and Rees, D. C. (2009) The rise of fragment-based drug discovery. *Nat Chem*, **1**(3): 187–92.

[216] Lipinski, C. A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, **23**(1): 3–25.

[217] Lipinski, C. A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharm Tox Methods*, **44**(1): 235–249.

[218] Congreve, M. *et al.* (2003) 'Rule of Three' for fragment-based lead discovery. *Drug Discov. Today*, **8**: 876–877.

[219] Gleeson, M. P. (2008) Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. *J Med Chem*, **51**(4): 817–834.

[220] Patel, D. *et al.* (2014) Advantages of crystallographic fragment screening: functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog Biophys Mol Biol*, **116**: 92–100.

[221] Cox, O. B. *et al.* (2016) A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem Sci*, **7**(3): 2322–2330.

[222] Hassell, A. M. *et al.* (2007) Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, **63**(Pt 1): 72–9.

[223] Müller, I. (2017) Guidelines for the successful generation of protein–ligand complex crystals. *Acta Cryst D*, **73**(2): 79–92.

[224] Pearce, N. M. *et al.* (2017) A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat Commun*, **8**: 15123.

[225] Beachy, P. A. *et al.* (2010) Interactions between Hedgehog proteins and their binding partners come into view. *Genes Dev*, **24**(18): 2001–12.

[226] War, A. R. *et al.* (2012) Mechanisms of plant defense against insect herbivores. *Plant Signal Behav*, **7**(10): 1306–20.

[227] Zhou, J. (2006) Norepinephrine transporter inhibitors and their therapeutic potential. *Drugs Future*, **29**(12): 1235–1244.

[228] Sun, Y. *et al.* (2013) Identification of adiponectin receptor agonist utilizing a fluorescence polarization based high throughput assay. *PLoS One*, **8**(5): e63354.

[229] Irwin, J. J. *et al.* (2012) ZINC: A Free Tool to Discover Chemistry for Biology. *J Chem Info Model*, **52**(7): 1757–1768.

[230] SciFinder Scholar, version 2004.2; Chemical Abstracts Service: Columbus, OH, 2004; RN 133876-92-3.

[231] S., K. *et al.* (2016) PubChem Substance and Compound databases. *Nucleic Acids Research*, **44**: D1202–D1213.

[232] Friesner, R. A. *et al.* (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem*, **47**(7): 1739–1749.

[233] Hartenfeller, M. *et al.* (2012) DOGS: Reaction-Driven de novo Design of Bioactive Compounds. *PLOS Comp Biol*, **8**(2): 1–12.

[234] Houston, D. R. and Walkinshaw, M. D. (2013) Consensus docking: improving the reliability of docking in a virtual screening context. *J Chem Inf Model*, **53**(2): 384–90.

[235] Tuccinardi, T. *et al.* (2014) Extensive consensus docking evaluation for ligand pose prediction and virtual screening studies. *J Chem Inf Model*, **54**(10): 2980–6.

[236] Jacobsson, M. and Karlen, A. (2006) Ligand Bias of Scoring Functions in Structure-Based Virtual Screening. *J Chem Inf Model*, **46**: 1334–1343.

[237] Wallach, I. *et al.* (2011) Normalizing Molecular Docking Rankings using Virtually Generated Decoys. *J Chem Info Model*, **51**(8): 1817–1830.

[238] Buglino, J. A. and Resh, M. D. (2012) Palmitoylation of Hedgehog Proteins. *Vitam Horm*, **88**: 229–252.

[239] Chang, C. *et al.* (2014) Membrane-bound O-acyltransferases (MBOATs). *Front Biol*, **6**(3): 177–182.

[240] Hofmann, K. (2000) A superfamily of membrane-bound O-acyltransferases with implications for wnt signaling. *Trends Biochem Sci*, **25**: 111–112.

[241] Bosson, R. *et al.* (2006) GUP1 of Saccharomyces cerevisiae encodes an O-acyltransferase involved in remodeling of the GPI anchor. *Mol Biol Cell*, **17**: 2636–2645.

[242] Guo, Z. Y. *et al.* (2005) The active site His-460 of human acyl-coenzyme A:cholesterol acyltransferase 1 resides in a hitherto undisclosed transmembrane domain. *J Biol Chem*, **280**: 37814–37826.

[243] Lin, S. *et al.* (2003) Human acyl-coenzyme A: cholesterol acyltransferase expressed in chinese hamster ovary cells: Membrane topology and active site location. *Mol Biol Cell*, **14**: 2447–2460.

[244] Shindou, H. *et al.* (2009) Identification of membrane O-acyltransferase family motifs. *Biochem Biophys Res Commun*, **383**: 320–325.

[245] Yang, J. *et al.* (2008) Identification of the acyltransferase that octanoylates ghrelin, an appetite-stimulating peptide hormone. *Cell*, **132**: 387–396.

[246] Zhai L, C. S., Chaturvedi D (2004) Drosophila wnt-1 undergoes a hydrophobic modification and is targeted to lipid rafts, a process that requires porcupine. *J Biol Chem*, **279**: 33220–33227.

[247] Micchelli, C. *et al.* (2002) Rasp, a putative transmembrane acyltransferase, is required for Hedgehog signaling. *Development*, **129**: 843–851.

[248] Lee, J. and Treisman, J. (2001) Sightless has homology to transmembrane acyltransferases and is required to generate active Hedgehog protein. *Curr Biol*, **11**: 1147–1152.

[249] Callier, P. *et al.* (2014) Loss of function mutation in the palitoyl-trasferase HHAT to syndromic 46, XY disorder of sex development by impending hedgehog protien palmitoylation and signaling. *PLoS Genet*, **10**(5): e1004340.

[250] Liang, X. *et al.* (2001) Heterogeneous fatty acylation of Src family kinases with polyunsaturated fatty acids regulates raft localization and signal transduction. *J Biol Chem*, **276**: 30987–30994.

[251] Magee, A. and Courtneidge, S. (1985) Two classes of fatty acids acylated proteins exist in eukaryotic cells. *EMBO*, **4**(5): 1137–1144.

[252] Hancock JF, C. J. M. C., Magee AI (1989) All ras proteins are polyisoprenylated but only some are palmitoylated. *Cell*, **57**: 1167–1177.

[253] Miura, G. *et al.* (2006) Palmitoylation of the EGFR ligand Spitz by Rasp increases Spitz activity by restricting its diffusion. *Dev Cell*, **10**: 167–176.

[254] Hardy, R. Y. and Resh, M. D. (2012) Identification of N-terminal residues of Sonic Hedgehog important for palmitoylation by Hedgehog acyltransferase. *J Biol Chem*, **287**(51): 42881–9.

[255] Loisel, T. *et al.* (1999) Activation of the beta(2)-adrenergic receptor-Galpha(s) complex leads to rapid depalmitoylation and inhibition of repalmitoylation of both the receptor and Galpha(s). *J Biol Chem*, **274**: 31014–31019.

[256] Bhatnagar, R. S. *et al.* (1999) The structure of myristoyl-CoA:protein N-myristoyltransferase. *BBA - Mol Cell Biol Lip*, **1441**(2): 162 – 172.

[257] Mann, R. K. and Beachy, P. A. (2004) Novel lipid modifications of secreted protein signals. *Annu Rev Biochem*, **73**: 891–923.

[258] Farazi, T. *et al.* (2001) The biology and enzymology of protein N-myristoylation. *J Biol Chem*, **276**: 39501–39504.

[259] Rudnick, D. A. *et al.* (1991) Kinetic and structural evidence for a sequential ordered Bi Bi mechanism of catalysis by Saccharomyces cerevisiae myristoyl-CoA:protein N-myristoyltransferase. *J Biol Chem*, **266**(15): 9732–9.

[260] Lanyon-Hogg, T. *et al.* (2017) Microfluidic Mobility Shift Assay for Real-Time Analysis of peptide N-palmitoylation. *SLAS Discovery*, **22**: 418–424.

[261] Long, J. *et al.* (2015) Identification of a Family of Fatty-Acid-Speciated Sonic Hedgehog Proteins, Whose Members Display Differential Biological Properties. *Cell Reports*, **10**(8): 1280 – 1287.

[262] Matevossian, A. and Resh, M. D. (2015) Membrane Topology of Hedgehog Acyltransferase. *J Biol Chem*, **290**: 2235–43.

[263] Konitsiotis, A. D. *et al.* (2015) Topological analysis of Hedgehog acyltransferase, a multi-palmitoylated transmembrane protein. *J Biol Chem*, **290**: 3293–307.

[264] Buglino, J. A. and Resh, M. D. (2010) Identification of conserved regions and residues within Hedgehog acyltransferase critical for palmitoylation of Sonic Hedgehog. *PLoS One*, **5**(6): e11195.

[265] Consortium, T. U. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, **45**: D158–D169.

[266] Ramakrishnan, C. *et al.* (2002) A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins. *Protein Eng*, **15**: 783–98.

[267] Petrova, E. *et al.* (2015) Hedgehog acyltransferase as a target in pancreatic ductal adenocarcinoma. *Oncogene*, **34**: 263–8.

[268] Lanyon-Hogg, T. *et al.* (2016) Synthesis and characterisation of 5-acyl-6,7-dihydrothieno[3,2-c]pyridine inhibitors of Hedgehog acyltransferase. *Data Brief*, **7**: 257–81.

[269] Lanyon-Hogg, T. *et al.* (2015) Modulation of Amide Bond Rotamers in 5-Acyl-6,7-dihydrothieno[3,2-c]pyridines. *J Org Chem*, **80**(9): 4370–7.

[270] Lanyon-Hogg, T. *et al.* (2015) Click chemistry armed enzyme-linked immunosorbent assay to measure palmitoylation by hedgehog acyltransferase. *Anal Biochem*, **490**: 66–72.

[271] Aricescu, A. R. *et al.* (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr*, **62**(Pt 10): 1243–50.

[272] Deller, M. *et al.* (2016) Protein stability: a crystallographers perspective. *Acta Crystal F, Struct Biol Commun*, **72**: 72–95.

[273] Meer, G. v. *et al.* (2008) Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol*, **9**: 112–124.

[274] Ratha, A. *et al.* (2009) Detergent binding exmplains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A*, **106**(6): 1760–1765.

[275] Backliwal, G. *et al.* (2008) Valproic acid: a viable alternative to sodium butyrate for enhancing protein expression in mammalian cell cultures. *Biotechnol Bioeng*, **101**: 182–189.

[276] Davie, J. R. (2003) Inhibition of Histone Deacetylase Activity by Butyrate. *J Nutr*, **133**(7): 2485S–2493S.

[277] Gttlicher, M. *et al.* (2001) Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO*, **20**: 6969–6978.

[278] Parker, J. and Newstead, S. (2016) Membrane protein crystallization: Current trends and future perspectives. *Adv Exp Med Biol*, **922**: 61–72.

[279] Parker, J. and Newstead, S. (2012) Current trends in -helical membrane protein crystallization: An update. *Prot Sci*, **21**: 1358–1365.

[280] Misquitta, Y. and Caffrey, M. (2003) Detergents Destabilize the Cubic Phase of Monoolein: Implications for Membrane Protein Crystallization. *Biophys J*, **85**(5): 3084–3096.

[281] Chun, E. *et al.* (2012) Fusion Partner Toolchest for the stabilization and crystallization of G Protein-couples Receptors. *Structure*, **20**: 967–976.

[282] Ingham, P. W. *et al.* (2011) Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat Rev Genet*, **12**(6): 393–406.

[283] Dessinioti, C. *et al.* (2014) Vismodegib for the treatment of basal cell carcinoma: results and implications of the ERIVANCE BCC trial. *Future Oncol*, **10**: 927–36.

[284] Scott, D. E. *et al.* (2016) Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat Rev Drug Discov*, **15**(8): 533–550.

[285] Turnbull, A. *et al.* (2014) Fragment-based drug discovery and proteinprotein interactions. *Res Rep Biochem*, **4**: 13–26.

[286] Peat, T. S. *et al.* (2012) Small Molecule Inhibitors of the LEDGF Site of Human Immunodeficiency Virus Integrase Identified by Fragment Screening and Structure Based Design. *PLOS ONE*, **7**(7): 1–11.

[287] Taylor, M. S. *et al.* (2013) Architectural Organization of the Metabolic Regulatory Enzyme Ghrelin O-Acyltransferase. *Journal of Biological Chemistry*, **288**(45): 32211–32228.

[288] Kirchner, H. *et al.* (2009) GOAT links dietary lipids with the endocrine control of energy balance. *Nat Med*, **15**(7): 741–745.

[289] Barnett, B. P. *et al.* (2010) Glucose and Weight Control in Mice with a Designed Ghrelin O-Acyltransferase Inhibitor. *Science*, **330**(6011): 1689–1692.

[290] Liu, J. *et al.* (2013) Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proceedings of the National Academy of Sciences*, **110**(50): 20224–20229.

[291] Ho, S. Y. and Keller, T. H. (2015) The use of porcupine inhibitors to target Wnt-driven cancers. *Bioorg Medic Chem Lett*, **25**(23): 5472 – 5476.

[292] Rios-Esteves, J. *et al.* (2014) Identification of Key Residues and Regions Important for Porcupine-mediated Wnt Acylation. *Journal of Biological Chemistry*, **289**(24): 17009–17019.

[293] Walter, T. *et al.* (2003) A procedure for setting up high-throughput nanolitre crystallization experiments. I. Protocol design and validation. *J Appl Crystal*, **36**(2): 308–314.

[294] Mayo, C. J. *et al.* (2005) Benefits of Automated Crystallization Plate Tracking, Imaging, and Analysis. *Structure*, **13**(2): 175–182.

[295] Ng, J. T. *et al.* (2014) Using textons to rank crystallization droplets by the likely presence of crystals. *Acta Crystallogr D Biol Crystallogr*, **70**(10): 2702–18.

[296] Collins, P. M. *et al.* (2017) Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallogr D Struct Biol*, **73**(3): 246–255.

[297] Winter, G. and McAuley, K. E. (2011) Automated data collection for macromolecular crystallography. *Methods*, **55**(1): 81–93.

[298] Winter, G. *et al.* (2013) Decision making in xia2. *Acta Crystallogr D Biol Crystallogr*, **69**(Pt 7): 1260–73.

[299] Winter, G. (2010) *xia2*: an expert system for macromolecular crystallography data reduction. *J Appl Crystal*, **43**(1): 186–190.

[300] Winn. and et al., M. D. (2011) Overview of the CCP4 suite and current developments. *Acta Cryst*, **D67**: 235–242.

[301] Krojer, T. *et al.* (2017) The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallogr D Struct Biol*, **73**(Pt 3): 267–278.

[302] Emsley, P. and Cowtan, K. (2004) *Coot*: model-building tools for molecular graphics. *Acta Crystallographica Section D*, **60**(12 Part 1): 2126–2132.

[303] Murshudov, G. N. *et al.* (2011) *REFMAC*5 for the refinement of macromolecular crystal structures. *Acta Crystal D*, **67**(4): 355–367.

[304] Cooper, M. A. (2002) Optical biosensors in drug discovery. *Nat Rev Drug Discov*, **1**(7): 515–528.

[305] Karlsson, R. *et al.* (2000) Biosensor Analysis of DrugTarget Interactions: Direct and Competitive Binding Assays for Investigation of Interactions between Thrombin and Thrombin Inhibitors. *Anal Biochem*, **278**(1): 1–13.

[306] Eswar, N. *et al.* (2006) Comparative protein structure modeling with MODELLER. *Curr Prot Bioinfo*, **15**: 561–563.

[307] Hess, B. *et al.* (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *JCTC*, **4**: 435–447.

[308] Kaminski, G. A. and Friesner, R. A. (2001) Evaluation and Reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical claculations on peptides. *J Phys Chem*, **105**: 6474–6487.

[309] Lindorff-Larsen, K. *et al.* (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**(8): 1950–8.

[310] Berendsen, H. J. C. *et al.* (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys*, **81**(8): 3684.

[311] Bussi, G. *et al.* (2007) Canonical sampling through velocity rescaling. *J Chem Phys*, **126**(1): 014101.

[312] Parrinello, M. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*, **52**(12): 7182.

[313] Korb, O. *et al.* (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model*, **49**: 84–96.

[314] Morris, G. M. *et al.* (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*, **30**(16): 2785–91.

[315] Trott, O. and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, **31**(2): 455–61.

[316] Sanner, M. F. (1999) Python: A Programming Language for Software Integration and Development. *J Mol Graphics Mod*, **17**: 57–61.

[317] Humphrey, W. *et al.* (1996) VMD: Visual Molecular Dynamics. *J Mol Graph*, **14**: 33–38.

[318] Schrdinger Release 2014-3: Maestro, version 9.9, Schrdinger, LLC, New York, NY, 2014.

[319] Williams, T. and Kelley, C. (2011) Gnuplot 4.5: an interactive plotting program. URL hhtp://gnuplot.info.