Signal transduction of the GPCR Smoothened: a key protein in Hedgehog-regulated morphogenesis and oncogenesis

Rachel E. Woolley



A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2021

Green Templeton College
Division of Structural Biology
University of Oxford

Signal transduction of the GPCR Smoothened: a key protein in Hedgehog-regulated morphogenesis and oncogenesis

Rachel E. Woolley

A thesis submitted for the degree of *Doctor of Philosophy*
Green Templeton College, Division of Structural Biology
University of Oxford, Hilary 2021

ABSTRACT

The Hedgehog signalling pathway is crucial for embryonic development and tissue homeostasis. Misregulation can result in severe developmental defects or oncogenesis, with an estimated 25% of all cancer deaths exhibiting improper Hedgehog signalling. The G-protein coupled receptor Smoothened plays a vital role in this pathway; it is responsible for transducing the Hedgehog signal across the plasma membrane. This makes Smoothened a highly attractive target for anticancer drug development. Despite the importance of Smoothened signalling and the plethora of structural data now available, the molecular mechanism of signal transduction is poorly understood. Here, we explore the mechanism of Smoothened activation and regulation by a range of small molecule ligands. We show that agonist binding at multiple sites can induce conformational changes required for effector binding. We couple these data with mutational analysis to propose distinct roles for each binding site. We also develop and validate a novel small molecule binding assay and adapt it to study allosteric communication within Smoothened. Finally, we present a strategy for use of the novel assay to identify next generation cancer treatments and address drug resistance.

Acknowledgements

Primarily, I would like to thank my supervisor, Professor Christian Siebold, for his expert guidance through the complex maze of Hedgehog signalling. His unwavering enthusiasm and curiosity have been an essential motivator throughout the past 4 years.

I would also like to highlight the contributions of Dr Benjamin Bishop, my unofficial (and somewhat self-appointed) mentor. His practical assistance and advice are second only to his constant support and friendship.

I am lucky to have been part of the Siebold group and would like to thank all the members of this supportive community, both past and present. With particular thanks to the postdocs who shared their invaluable experience; Amalie Rudolf, Tomas Malinauskas, Rebekka Siebold-Schwab, Samuel Griffiths and Eamon Byrne.

I would also like to thank all the members of STRUBI who contribute to making the division an enjoyable and rewarding place to work. In particular, those who keep the lab running smoothly; Margaret Jones, Weixian Lu and Oliver Hayward.

Thank you to the Wellcome Trust for funding of my DPhil studentship and those in the University who work hard to run this programme. I have also benefitted from Cancer Research UK and European Research Council consumables funding. Also thank you to Green Templeton College, whose welcoming environment has provided a calm centre within the Oxford world.

Finally, I would like to thank my friends and family, both within and far away from Oxford, for their constant encouragement and support.

Technical Acknowledgements

I would like to acknowledge all those who have provided technical assistance, advice and expertise to enable the completion of these studies. Specific contributions are highlighted in the relevant sections of this thesis.

Professor Rajat Rohatgi and his laboratory at Stanford University School of Medicine for undertaking important cellular assays and mutational analysis, in addition to providing valuable insights and advice.

Professor Oleg Fedorov and James Bennett at the Centre for Medicines Discovery for sharing their expertise to enable us to develop and run the HTRF assay. I would like to personally thank James Bennett for generously reading numerous plates when Covid-19 restrictions prevented me from doing so.

Dr Loic Carrique, Dr Dimple Karia, Dr Vojta Prazak, Dr Christiane Kowatsch and Claire Coupland for operation of electron microscopes, grid preparation and data processing assistance.

Professor Jan Steyaert and his laboratory at VIB-VUB Center for Structural Biology for identification and provision of nanobody and megabody DNA.

Professor Douglas Covey and his laboratory at Washington University in St. Louis for synthesis of oxysterol-biotin for the HTRF assay, LKM27 and KK174.

Dr Benjamin Bishop for collection of fluorescent cellular images.

Dr Eamon Byrne for initial validation of the HTRF assay.

List of Abbreviations

1D4-beads	Rho1D4 conjugated Sepharose beads
1D4-peptide	Rho1D4 epitope peptide, sequence TETSQVAPA
3C	Human Rhinovirus 3C protease
7-TM	seven transmembrane
ATP	adenosine triphosphate
β_2 AR	β_2 adrenergic receptor
β-OG	n-octyl-β-D-glucopyranoside
BCC	basal cell carcinoma
BLI	bio-layer interferometry
BOC	Brother of CDO
BRIL	thermostabilized apocytochrome b_{562}
CDO	CAM-related/downregulated by oncogenes
cAMP	cyclic adenosine monophosphate
CHS	cholesteryl hemisuccinate
CK1	Casein kinase 1
CMC	critical micelle concentration
CMV	cytomegalovirus
CRD	cysteine rich domain
DDM	n-dodecyl-β-D-maltopyranoside
DHH	Desert hedgehog
DISP	Dispatched
DMEM	Dulbecco's modified eagle's media
DMNG	decyl maltose neopentyl glycol
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DPPC	1,2-dipalmitoyl-sn-glycero-3-phosphocholine
dSMO	Smoothened from *Drosophila melanogaster*
DTT	dithiothreitol
EC50	ligand concentration which gives half-maximal response
ECL	extracellular loop
EVC	Ellis-van Creveld syndrome protein
FBS	fetal bovine serum
FDA	US food & drug administration

FRET	Förster resonance energy transfer
FSEC	fluorescence-detection size exclusion chromatography
GAS1	Growth-arrest specific 1
GDN	glyco-diosgenin
GDP	guanosine diphosphate
GLI	Glioma-associated oncogene family member
GPCR	G-protein coupled receptor
GRK2	GPCR kinase 2
GTP	guanosine triphosphate
HA	Haemagglutinin
Hh	Hedgehog
HHAT	Hedgehog acyltransferase
Hh-C	carboxy-terminal domain of immature Hh
Hh-N	amino-terminal domain of immature Hh
HPE	holoprosencephaly
hSMO	Smoothened from *Homo sapiens*
HTRF	homogeneous time resolved fluorescence
ICD	intracellular domain
ICL	intracellular loop
IHH	Indian hedgehog
IPTG	isopropyl b-D-1-thiogalactopyranoside)
K_D	dissociation constant
kDa	kilodalton
KIF3A	Kinesin-like protein 3A
KK174	novel Hh agonist
LD	linker domain
LKM27	novel Hh agonist
MAP	Mitogen-activated protein
Mb	megabody
MCS	multiple cloning site
mSMO	Smoothened from *Mus musculus*
MSP	membrane scaffold protein
mVs	monovenus
MWCO	molecular weight cut-off
Nb	nanobody
NEAA	non-essential amino acids

Ni-NTA	nickel-nitrilotriacetic acid
OGNG	octyl glucose neopentyl glycol
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEI	polyethyleneimine
PEM	protein expression media
PelB	periplasmic localisation sequence from pectate lyase B
PKA	Protein kinase A
POPC	1-palmitoyl-2-oleoyl-glycero-3-phosphocholine
PTCH	Patched
SAG	Smoothened agonist
SANT	Smoothened antagonist
scFv	single chain variable fragment
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate poly-acrylamide gel electrophoresis
SEC	size exclusion chromatography
SHH	Sonic hedgehog
SMA	styrene maleic acid
SMO	Smoothened (human Smoothened unless specified)
SMOΔC	SMO construct comprising 32-555 (UniProt ID Q99835)
SMO-M2	SMO constitutive activation cancer mutant (W535L)
SSD	sterol-sensing domain
SUFU	Suppressor of Fused
T4L	T4 lysozyme
TEM	transmission electron microscope
TetR	Tet repressor
Tm	melting temperature
TMD	transmembrane domain

Contents

Introduction 1

1.1. Hedgehog signalling 2
1.1.1. Role in embryonic development & disease ____ 4
1.1.2. Role in adult tissue homeostasis & oncogenesis ____ 6
1.1.3. Molecular mechanisms ____ 7

1.2. G-protein coupled receptors 12
1.2.1. Signalling mechanism ____ 12
1.2.2. GPCRs as drug targets ____ 15
1.2.3. Structural divergence and conservation ____ 15
1.2.4. Intracellular signalling cascades ____ 16

1.3. Smoothened 17
1.3.1. Smoothened has a multidomain structure ____ 17
1.3.2. Patched regulates Smoothened via a sterol ligand ____ 19
1.3.3. Multiple small molecules regulate Smoothened ____ 20
1.3.4. Smoothened signals to GLI via an unknown mechanism ____ 22

1.4. Thesis outline 24

Materials & Methods 26

2.1. Expression and purification of membrane proteins 27
2.1.1. Smoothened expression (transient) ____ 27
2.1.2. Smoothened expression (lentiviral) ____ 29
2.1.3. Smoothened small scale expression ____ 31
2.1.4. Smoothened purification ____ 32
2.1.5. Amphipol exchange ____ 34
2.1.6. Nanodisc reconstitution ____ 35

2.2. Expression and purification of soluble proteins 36
2.2.1. G-protein alpha subunit expression ____ 36
2.2.2. G-protein alpha subunit purification ____ 37
2.2.3. Nanobody & megabody bacterial expression ____ 37
2.2.4. Nanobody purification from E. coli ____ 38
2.2.5. Megabody purification ____ 39
2.2.6. Nanobody mammalian expression ____ 39
2.2.7. Nanobody purification from conditioned media ____ 40

2.3. Biophysical techniques 41
2.3.1. Minimal G-protein thermal shift assay ____ 41
2.3.2. FSEC complexation trials ____ 42
2.3.3. SMO pull-down trials ____ 42
2.3.4. Bio-layer interferometry ____ 42

2.4. Structural techniques 44

2.4.1. Electron microscopy studies 44
2.5. Small molecule binding 45
2.5.1. FRET-based CRD binding assay 45
Production of wild type Smoothened 49
3.1. Abstract 50
3.2. Introduction 51
3.2.1. GPCR Expression Systems 51
3.2.2. Solubilisation 54
3.2.3. Purification 56
3.3. Results 58
3.3.1. Stable cell lines produce higher yields of SMO than transient transfection 58
3.3.2. C-terminal tags affect SMO expression levels 60
3.3.3. Purification of increased amounts of monomeric SMOΔC 63
3.3.4. Tandem affinity purification can be used to isolate full-length SMOΔC 65
3.3.5. Detergent-free systems can be successfully used for SMOΔC 67
3.4. Discussion 71
3.4.1. Expression of SMOΔC 71
3.4.2. Purification of SMOΔC 72
3.4.3. SMOΔC in a detergent-free system 73
3.4.4. Summary 75
Development of a novel binding assay 76
4.1. Abstract 77
4.2. Introduction 78
4.2.1. Small molecules bind to and modulate SMO 78
4.2.2. SMO as an anti-cancer drug target 80
4.2.3. SMO small molecule binding assays 82
4.2.4. Concept of a SMO CRD binding assay 84
4.2.5. Assay design 85
4.3. Results 88
4.3.1. Assay validation 89
4.3.2. Optimisation – we can reduce reagent use without sacrificing assay quality 91
4.3.3. Proof of principle – we can identify competition for the CRD site 95
4.3.4. Oxysterols interact with the CRD 99
4.3.5. Allosteric ligands affect CRD site conformation 100
4.4. Discussion 104
4.4.1. Assay sensitivity is of commercial standard 104
4.4.2. SMO CRD FRET assay shows robust specificity 105
4.4.3. Allosteric effects are observed in the wild-type SMO 108
4.4.4. Future directions 110
4.4.5. Summary 111
Smoothened activation 112
5.1. Abstract 113
5.2. Introduction 114

5.2.1. G-protein coupled receptor signal transduction ____ 114
5.2.2. Heterotrimeric G-proteins ____ 116
5.2.3. Smoothened signal transduction ____ 120
5.2.4. Strategies for obtaining active state structures ____ 121

5.3. Results ____ 125
5.3.1. Strategy to obtain an active state SMO structure ____ 125
5.3.2. Adding size and asymmetry for EM ____ 126
5.3.3. Stabilising the SMO active state ____ 129
5.3.4. Design of SMO-specific minimal G-proteins ____ 130
5.3.5. Addition of GDP increases G-protein stability ____ 132
5.3.6. Mini G-proteins interact with SMO ____ 133
5.3.7. SMO activating nanobody ____ 133
5.3.8. Optimising the sample for cryo-EM ____ 135

5.4. Discussion ____ 141
5.4.1. From nanobodies to megabodies ____ 141
5.4.2. Activating Smoothened ____ 142
5.4.3. Lessons from electron microscopy studies ____ 144
5.4.4. Summary ____ 145

Discussion ____ 147

6.1. Smoothened signal transduction ____ 148
6.1.1. Smoothened binding sites show co-operativity ____ 148
6.1.2. Basal Hh signalling relies upon the TMD ____ 150
6.1.3. Cholesterol access to Smoothened ____ 152
6.1.4. Activation of Smoothened ____ 156

6.2. Smoothened downstream signalling ____ 157
6.2.1. Activation resembles class A GPCRs ____ 157
6.2.2. Challenges to studying activation in vitro ____ 158
6.2.3. Fine-tuning Smoothened activation ____ 159

6.3. Overcoming resistance to SMO-targeted anticancer treatments ____ 161
6.3.1. Targeting the vismodegib binding site ____ 161
6.3.2. Existing CRD inhibitors pave the way ____ 162
6.3.3. A strategy to overcome vismodegib resistance ____ 163

6.4. Summary ____ 166

References ____ 167

List of Figures

Figure 1.1. The action of morphogens can be described using the "French Flag Model" *3*

Figure 1.2. Misregulation of Hedgehog signalling results in malformation and oncogenesis. *5*

Figure 1.3. Mammalian Hedgehog signalling *10*

Figure 1.4. GPCRs show conserved structural features *14*

Figure 1.5. Smoothened has a multidomain structure *18*

Figure 1.6. Smoothened interacts with a range of small molecule ligands. *21*

Figure 2.1 Vector map of pHLsec-SMOΔC-1D4 used for transient expression in HEK 293S GnTI$^-$ cells. *28*

Figure 2.2 Vector map of pHR-CMV-TetO2_His-SMOΔC-3C-mVs-1D4 used for creation of stable HEK 293S GnTI$^-$ TetR cell lines *31*

Figure 2.3 Simplified purification pipeline for SMO. *34*

Figure 2.4 FRET assay schematic. *46*

Figure 3.1. Solubilisation of GPCRs *53*

Figure 3.2. Tandem affinity purification of GPCRs *57*

Figure 3.3. Lentiviral expression of SMO *60*

Figure 3.4. C-terminal Venus improves expression of SMOΔC *62*

Figure 3.5. Purification of SMOΔC-3C-mVs-1D4 ..... *64*

Figure 3.6. Tandem affinity purification of SMO *67*

Figure 3.7. SMO in detergent-free systems *69*

Figure 4.1 Multiple small molecule ligand binding sites modulate SMO activity *79*

Figure 4.2 Anticancer drugs target SMO and are susceptible to resistance mutations *81*

Figure 4.3 SMO CRD binding assay was developed using customisable HTRF reagents. *86*

Figure 4.4: Structure guided SMO mutations for ligand binding experiments. 88

Figure 4.5 Z-factor calculations classify assay quality....90

Figure 4.6 Reduction in FRET reagents affects the FRET ratio 93

Figure 4.7 Known CRD-binding ligands can be assessed using our novel FRET assay....96

Figure 4.8 Novel SMO ligands can be characterised using the FRET assay. 101

Figure 5.1. G-protein coupled receptors exhibit complex pharmacology....115

Figure 5.2. G-protein signalling 119

Figure 5.3. Approaches to capturing GPCR active state structures. 123

Figure 5.4. Flowchart showing planned approach to determining the structure of SMO in an active state. 126

Figure 5.5. Megabody production and characterization....127

Figure 5.6. Minimal G-protein production and characterization 131

Figure 5.7. Nb8 binds the active state of SMO 135

Figure 5.8. Optimising cryo-EM for SMOΔC....138

Figure 6.1. Communication between ligand binding sites 149

Figure 6.2 Contribution of cholesterol binding sites to Hh signalling....152

Figure 6.3. Activation of Smoothened 158

Figure 6.4. Overcoming vismodegib resistance 163

Figure 6.5. Possible pathway from assay development to clinical trials 165

Chapter 1

Introduction

1.1. Hedgehog signalling

The fundamental basis of developmental biology is how a near homogeneous fertilised egg emerges into an organism with such complex structures and functions as seen across the natural world. Alan Turing's formative 1952 paper "The chemical basis of morphogenesis" first proposed the concept of such intricate developmental processes being regulated by simple molecular gradients (*Figure 1.1.a*) [1,2]. These molecules, termed morphogens, are secreted by the signalling cell, which diffuse through the surrounding area to receptor cells [1,3,4]. They naturally form a gradient and are able to induce a distinct range of responses in a concentration-dependent manner and hence drive cellular differentiation (*Figure 1.1.b*). Gradient-free models have also been proposed, such as Wolpert's "balancing" model, introduced as a solution to the so-called "French flag problem" [5]. The tricolour flag is commonly used as a tool to visualise and explain differential cell fates. The "balancing" model relies upon self-organisation, where each cell's fate is determined by its neighbours (*Figure 1.1.c*) [6].

The most well-studied of these signalling families are Hedgehog (Hh) and Wnt [7,8]. The Hh pathway was first described in 1980, whereupon the knockout of a *Drosophila melanogaster* gene caused the formation of spiked projections across the embryo (*Figure 1.2.a*). This observation gave the *hedgehog* gene its name [9]. The morphogen ligand was not described until a decade later when three groups

simultaneously showed that the gene product was a secreted peptide, hedgehog [10-12]. This in turn led to the discovery of several Hh variants in vertebrates, including the mammalian homologs Indian (Ihh), Desert (Dhh) and Sonic (Shh) hedgehog [13-15]. The delayed recognition of this vital developmental pathway is in part due to its absence from *Caenorhabditis elegans*, the genome of which contains no identifiable Hh ligand homolog. It does, however, contain a number of hedgehog-related and Patched-related proteins which appear to interact with sterols and control cell growth [16].

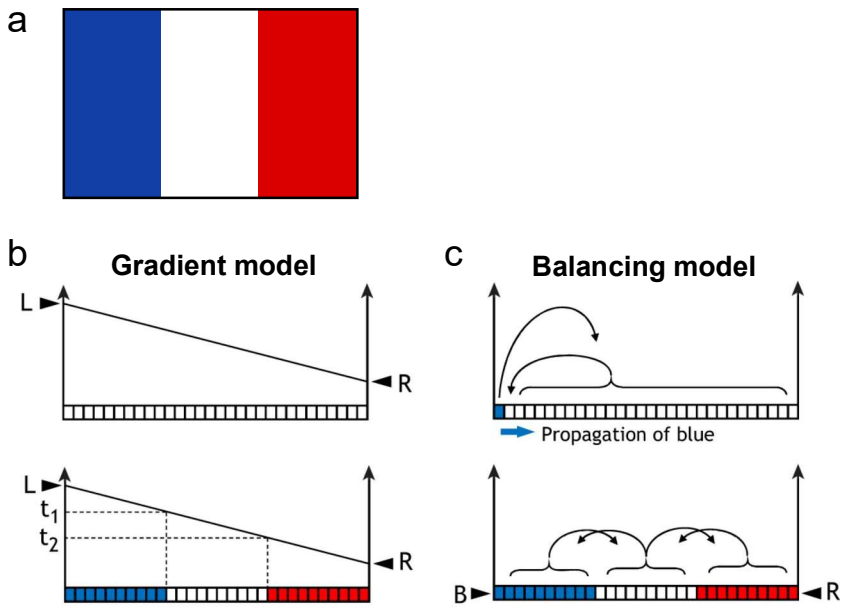


Figure 1.1. The action of morphogens can be described using the "French Flag Model"

a. Depiction of the tricolour French flag where each colour is said to represent a different cellular fate.
b. The gradient model of morphogen action relies upon diffusion from a common source, as proposed by Turing in 1952.
c. Wolpert's balancing model relies on molecules produced by neighbouring cells to determine the fate of a region.

b and c are adapted from Tabata & Takei, 2004, with permission (Licence #1111522-1)

1.1.1. Role in embryonic development & disease

Hh signalling is crucial for correct embryonic development in both *D. melanogaster* and across a range of vertebrates. It regulates organogenesis, body polarisation and patterning, as well as germ cell and bone development [4,13,17]. Any misregulation can result in a range of potentially severe developmental disorders, including holoprosencephaly and Grieg syndrome (causing polydactyly and syndactyly) (*Figure 1.2.b*)[8,18-20]. In mammalian embryonic development the Hh homologs regulate a variety of processes, largely determined by their expression localisation.

Dhh is the homolog which most closely resembles *Drosophila* Hh [21]. In mammals it shows the most restricted expression profile and is limited to the gonads [22-24]. Mouse knockout studies show few effects; the most notable being male infertility due to lack of sperm maturation [22]. Ihh is also expressed in a narrow range of tissues and is largely involved in bone development [25,26]. Mouse knockouts cause a 50% reduction in early embryo viability due to poor yolk-sac development and the surviving embryos show a range of bone defects [25,27]. In humans, homozygous loss of function mutations result in acrocapitofemoral dysplasia, a condition characterised by reduced stature and other bone defects [28].

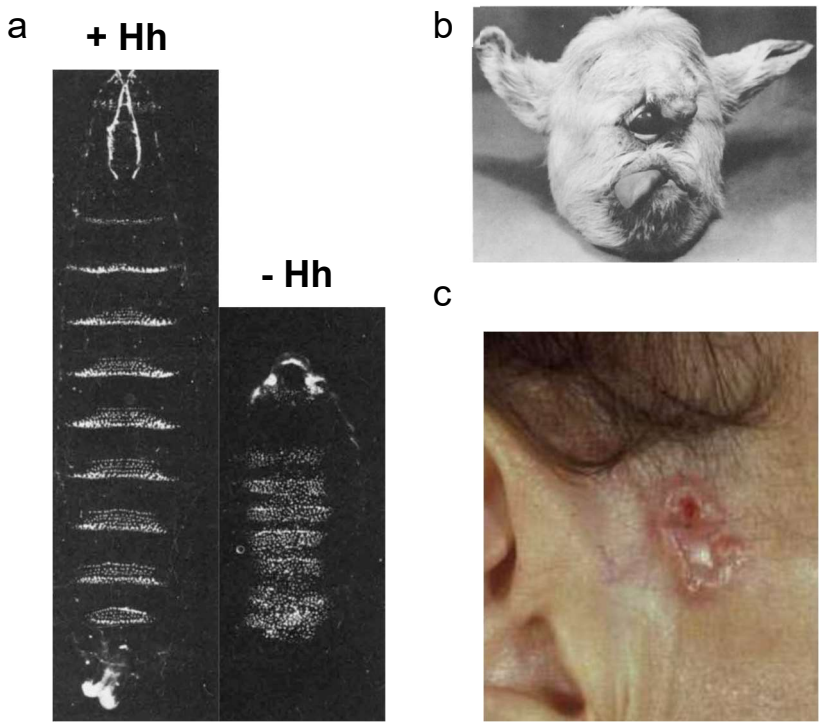


Figure 1.2. Misregulation of Hedgehog signalling results in malformation and oncogenesis.

a. Drosophila melanogaster larvae with and without the Hh gene. The knockout shows stunted growth and loss of body patterning. Reproduced with permission from Nusslein-Volhard & Wieschaus, 1980 (Licence #4987031117070).
b. Cyclopia is a result of improper Hh signalling during embryogenesis. Reproduced with permission from Binns, 1964 (Licence #4987040751370).
c. Basal cell carcinoma is a skin cancer that can be caused by inappropriate Hh signalling in adult tissues. Reproduced with permission from Epstein, 2008 (Licence #4987101134645).

Shh is the most widely expressed mammalian homolog, crucially including the central nervous system. It is both necessary and sufficient to determine ventral identity in developing embryos [29]. Shh knockout mice lack ventral cell types in the brain and show the associated craniofacial abnormalities [8]. These mice show a phenotype which is remarkably similar to holoprosencephaly (HPE) in humans, which often results in part from mutations in the Shh gene [19]. Interestingly, HPE patients present graded phenotypes ranging from severe (alobar HPE) to those

described as clinically normal. This range of conditions can all result from the same Shh mutation, strongly implying that other factors may contribute. HPE has also been linked to abnormalities in the expression of several proteins involved in the biosynthesis and transport of cholesterol, including megalin and Δ-7-sterol reductase [19,30-34].

1.1.2. Role in adult tissue homeostasis & oncogenesis

Like other developmental signalling pathways, Hh plays an important role in the homeostasis of a range of adult tissues. These include the brain, skin, bone, muscle and other tissues [35]. In the adult central nervous system Shh is required for both establishment and maintenance of neural stem cells. This stem cell pool gives rise to transit-amplifying cells which can form new neurons. This role is essential for continued proper brain structure and function [36]. Adult skin is also maintained by Shh, in a similar fashion. The basal epithelial cells depend upon Shh-mediated signalling to continuously produce new skin cells to replace the upper layer of dead cells. However, this mechanism also poses a danger as any hyperactivity caused by Hh pathway mutations can lead to uncontrolled proliferation and basal cell carcinoma (BCC) (*Figure 1.2.c*) [37-40]. These oncogenic mutations implicated in BCC include both Patched (PTCH) loss of function and Smoothened (SMO) gain of function mutations [41,42]. Aberrant hedgehog signalling also accounts for around 25%

of all medulloblastoma cases, a prevalent childhood brain tumour which arises in the cerebellum [43,44]. However, this condition is thought to stem from developmental defects in the neural progenitor pool rather than spontaneous mutations [43].

1.1.3. Molecular mechanisms

For the purposes of this thesis, the following refers to mammalian hedgehog signalling, unless otherwise specified. All hedgehog ligands are expressed as ~45 kDa precursor proteins. They undergo autoproteolytic cleavage resulting in proteasomal degradation of the C-terminal fragment (Hh-C) (*Figure 1.3.a*) [45,46]. This is likely initiated by the recruitment of cholesterol and its subsequent covalent linkage to the C-terminus of the N-terminal fragment (Hh-N) and simultaneous dissociation of Hh-C [45]. HHAT (hedgehog acyltransferase) then further modifies Hh-N by palmitoylating the N-terminus, thus producing dual-lipidated mature Hh, which is around 19 kDa [47,48]. This molecule is then translocated to the plasma membrane for secretion. The mechanism likely involves the cholesterol binding abilities of Dispatched (DISP) and SCUBE2 (*Figure 1.3.a*) [49-51]. This is supported by evidence suggesting cholesterol is responsible for the insolubility of mature Hh, which clearly must be minimised for paracrine signalling to occur. Models suggest the integral membrane protein DISP presents Hh at the extracellular surface, aiding the initial dissociation from the membrane [52]. SCUBE2 can then receive the Hh molecule and increase its solubility by shielding the cholesterol moiety. The

SCUBE2-Hh complex shows increased solubility relative to monomeric dual-lipidated Hh and is therefore more suited to long-range diffusion, a key requirement for morphogen signalling [50].

Hh ligands interact with the 12-pass transmembrane receptor PTCH. This interaction is stabilised by a number of co-receptors; CDO (CAM-related/downregulated by oncogenes), BOC (brother of CDO) and GAS1 (growth-arrest specific 1) (*Figure 1.3.b*) [53]. SHH can interact with two molecules of PTCH simultaneously. Recent structural data have illuminated the binding mechanism and show the involvement of two Hh interfaces [54-57]. The first is a classical protein:protein interface, organised by the Shh-bound metal ions (Ca^{2+} and Zn^{2+}). The second interface shows the importance of the Shh lipid modifications, as both the palmitate and cholesterol moieties bind to PTCH [58]. The palmitate makes use of the N- terminal loop to bind deep within the PTCH molecule, at the sterol-sensing domain [58]. The presence of these two lipid moieties increases the potency of this reaction and PTCH inhibition. This in turn increases stimulation of the downstream Hh pathway [58].

The inhibition of PTCH relieves repression on SMO, the Hh pathway G-protein coupled receptor (GPCR) (*Figure 1.3.c*). It is thought that PTCH regulates SMO by restricting the availability of cholesterol, an endogenous SMO agonist [59-61]. PTCH is thought to function as a cholesterol transporter, supported by evidence of

cholesterol binding and the existence of hydrophobic tunnels in structural models [62]. However, the mechanism by which PTCH-mediated cholesterol transport is able to control SMO access to the sterol is a hotly debated topic within the field [63]. SMO performs a crucial step in the Hh pathway, transduction of the signal across the plasma membrane. The extracellular domain of the GPCR binds to cholesterol, causing a cascade of structural changes to transmit the signal across the bilayer [59-61]. Activated SMO is then able to stimulate downstream effectors (*Figure 1.3.c*). In canonical Hh signalling, this occurs in the primary cilium, a microtubule-based organelle which extends from most vertebrate cells [64]. SMO is phosphorylated and retained in the cilia by β-arrestin and KIF3A [65,66]. This results in the ciliary localisation of the complex composed of the glioma-associated oncogene (GLI) and suppressor of fused (SUFU), its subsequent dissociation and translocation of full-length GLI to the nucleus where it stimulates transcription of Hh target genes [4].

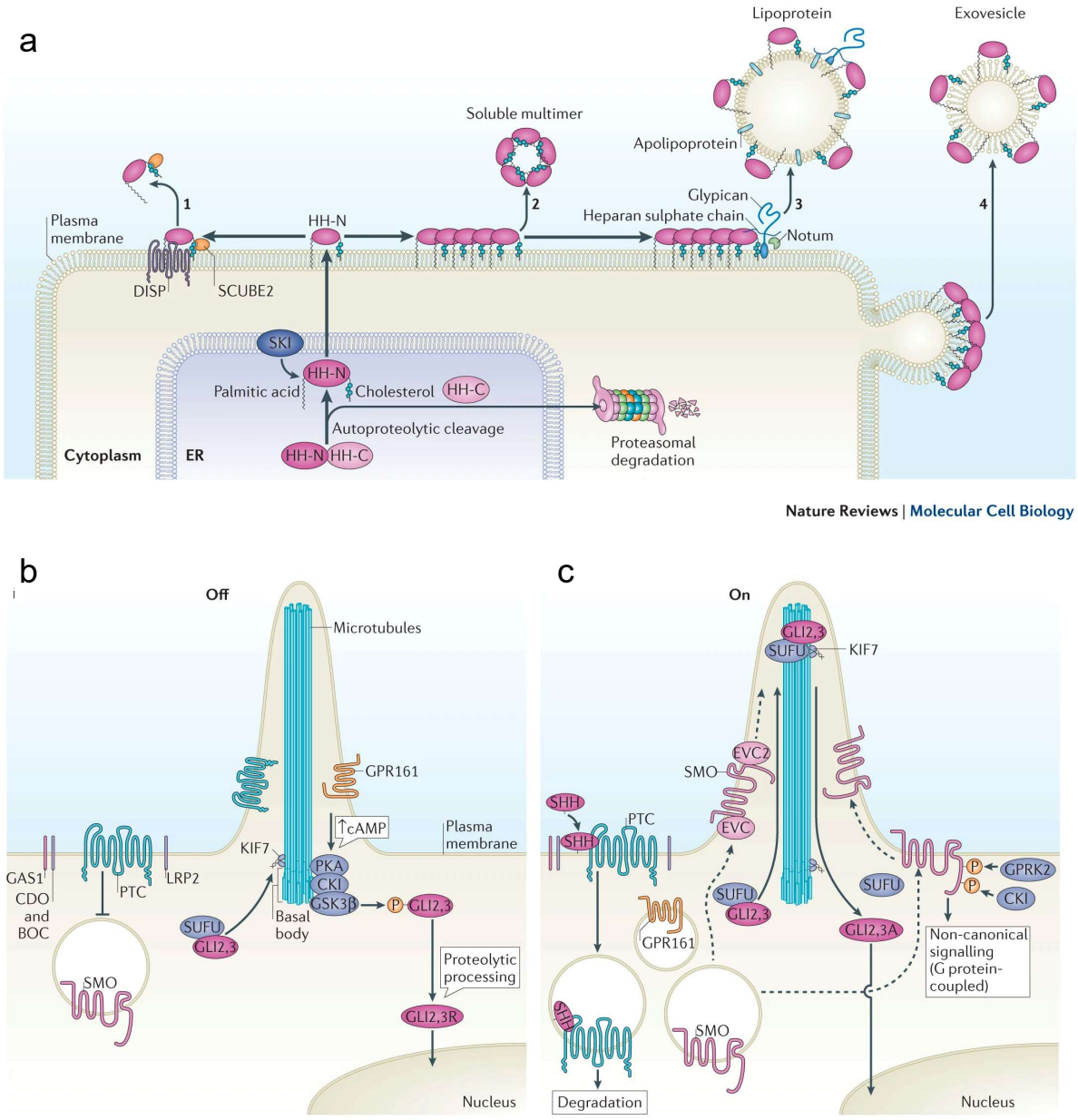


Figure 1.3. Mammalian Hedgehog signalling

a. Hh is produced as a ~45 kDa precursor protein which undergoes autoproteolytic cleavage and degradation of the C-terminal region (HH-C). HH-N is dual lipid modified – cholesterol is added to the C-terminus upon cleavage and the N-terminus is subsequently palmitylated by Hedgehog acyltransferase (HHAT or Skinny Hedgehog SKI). Release from the plasma membrane is not fully understood and several mechanisms have been suggested (1-4).

b. In the resting state at the receiving cell Patched inhibits Smoothened to prevent Hh signalling. Cellular cAMP stimulates PKA to phosphorylate GLI transcription factors and induce degradation.

c. When the Hh ligand is present, Patched is inhibited, allowing SMO to translocate to the primary cilium and induce GLI-mediated transcription of Hh target genes.

Figure reproduced with permission from Briscoe & Therond, 2013 (Licence #4987110616257).

Non-canonical Hh signalling is thought to involve SMO acting as a GPCR to stimulate G-protein cascades. There is evidence to suggest SMO interacts with the Gi/o G-protein class [67-69]. This has been reported to have a supporting role to canonical hedgehog signalling, occurring only in certain cell types [70]. The stimulation of Gi/o proteins may minimise the degradation of GLI and hence increases the amount available to stimulate gene transcription [71]. It is possible that this pathway is specifically activated by certain agonists or even thresholds of activation, though its role has not yet been defined.

1.2. G-protein coupled receptors

G-protein coupled receptors (GPCRs) are a large class of proteins which function to transduce extracellular stimuli across the plasma membrane. Whilst they form one of the biggest and most diverse protein families in mammalian genomes, they are also found in a range of invertebrates, bacteria, yeast & plants [72]. The number of GPCRs in the human genome is estimated to be in excess of 800 [73]. Over 50% of these are olfactory receptors and the remaining proteins respond to a wide range of stimuli, including neurotransmitters, nucleotides and drugs such as caffeine. This wide range of specificities makes GPCRs attractive pharmacological targets, with an estimated 30% of FDA approved drugs binding this receptor class [74].

1.2.1. Signalling mechanism

GPCRs have several conserved features, including an extracellular N-terminus and intracellular C-terminus, linked by seven transmembrane helices (*Figure 1.4.a*). Despite these similarities the stimuli to which they respond are diverse, as are the downstream pathways they regulate. GPCRs respond to light, odorants, hormones, neurotransmitters and a wide range of other molecules [72,75]. They selectively stimulate a subset of nearly 20 different Gα proteins, as well as β-arrestin and non-classical pathways. A general mechanism of activation is seen across GPCR families. It first requires an external stimulus, usually ligand binding at the extracellular side

of the GPCR. This may involve binding sites in the extracellular domain or the extracellular face of the transmembrane bundle. This event is followed by conformational changes which transfer the signal to the intracellular face of the receptor [76]. Specific structural changes are GPCR-specific, though a general pattern is emerging, where Helix VI bends "outwards", possibly breaking a so-called ionic lock and hence facilitating G-protein binding (*Figure 1.4.a*) [77,78]. The GPCR activates G-proteins by initiating guanosine nucleotide exchange and subsequent dissociation of the Gα and Gβγ subunits. These G-protein subunits stimulate the intracellular cascade before arrestin binding attenuates the signal by inducing internalisation of the GPCR [77].

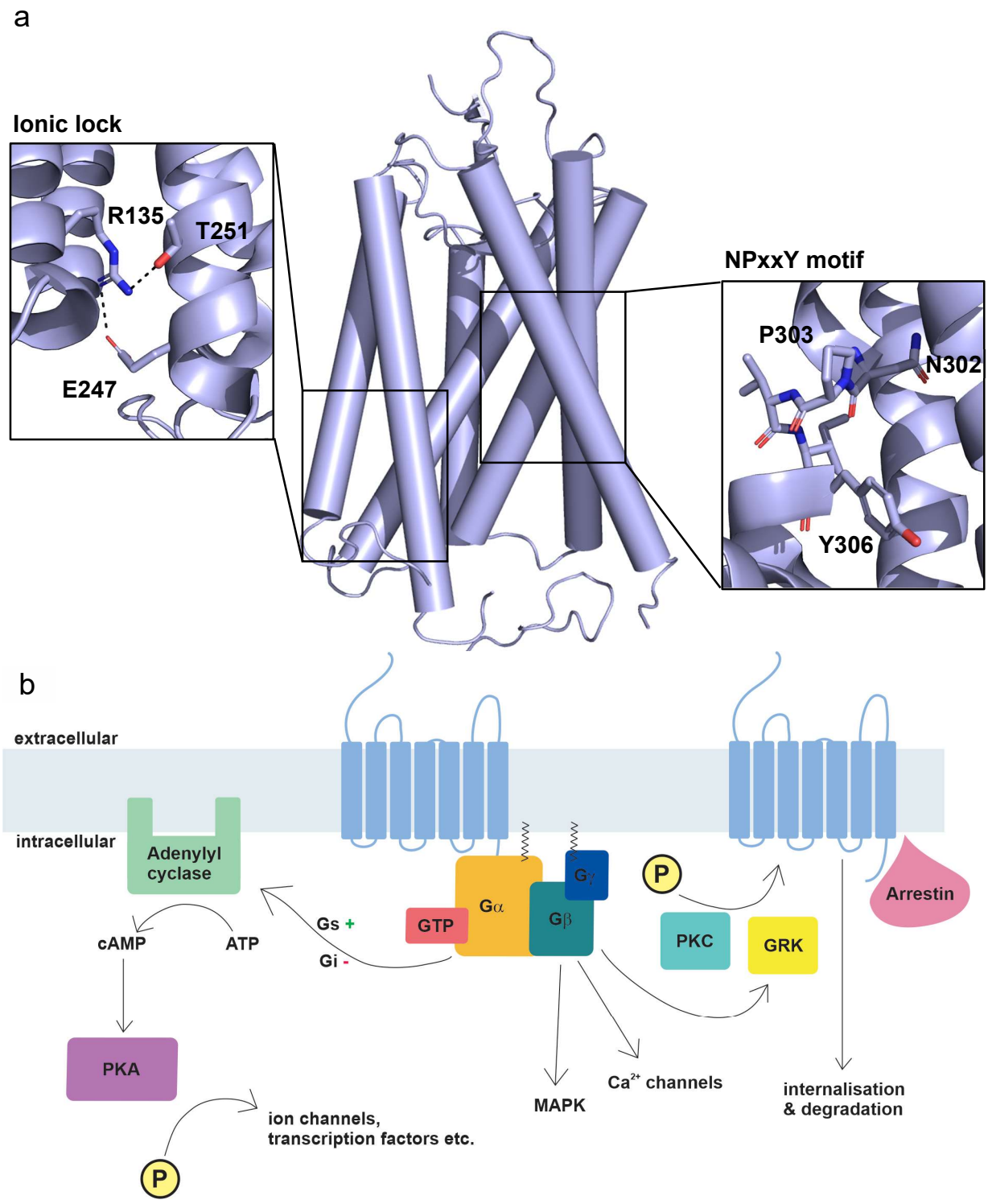


Figure 1.4. GPCRs show conserved structural features.

a. Main image shows a cartoon representation of the GPCR heptahelical bundle from PDB 1F88. Insets show the so-called "ionic lock" and conserved NPxxY motif.

b. GPCR activation of G-proteins stimulates the Gα subunit to regulate adenylyl cyclase production of cAMP and subsequent kinase activation. Gβγ regulates the MAPK pathway, Ca^{2+} channels and recruits GRK to phosphorylate the GPCR. Arrestins bind to the phosphorylated GPCR and induce internalization.

1.2.2. GPCRs as drug targets

One key reason for the widespread interest in GPCRs is their suitability as drug targets. Already, an estimated 35% of US and EU approved drugs target GPCRs, a number which will likely increase as orphan GPCRs are studied further and de-orphanised [75,79]. The diverse range of GPCR stimuli make them very attractive targets, and they are already used in treatment of cancers, cardiac and inflammatory disorders, among others [80].

1.2.3. Structural divergence and conservation

The GPCR superfamily is divided into 5 classes, based on sequence and structural similarity: rhodopsin (class A); secretin (class B1); adhesion (class B2); glutamate (class C); frizzled (class F) [73]. Despite recent advances in GPCR structural biology, 87% of GPCRs remain structurally uncharacterised [81]. Most structural divergence is in the ligand binding sites. In many GPCRs, this is found on the extracellular face of the transmembrane bundle, involving residues from the extracellular loops in ligand stabilisation [76]. The ligand binding pockets are often capable of accommodating several different ligands, which may cause a range of downstream effects or alter their potency. Partial and biased agonism result from this and are instrumental in GPCR signalling (discussed further in Chapter 3) [75,82,83]. At the cytoplasmic face, an ionic lock is observed in some GPCRs (*Figure 1.4.a*). Initially

described as a feature of Class A ligand activated GPCRs, it has recently been highlighted in other classes, including class F [76,84-87]. Another feature is the NPxxY motif, conserved in class A receptors (*Figure 1.4.a*). Minor variation in this region can be found in other GPCRs [59,88,89]. The NPxxY motif is responsible for maintaining a network of organised water molecules thought to enable movement of transmembrane helix VII for activation [88].

1.2.4. Intracellular signalling cascades

Upon agonist binding to a GPCR, conformational changes are transmitted across the transmembrane bundle to the cytoplasmic face of the protein. This new conformation is capable of binding to and activating G-proteins [78]. GPCRs show specificity via the subset of Gα proteins they can interact with, of which there are over 20 isoforms. Heterotrimeric G-proteins bind to GDP (guanosine diphosphate) in the cytoplasm, which stabilises the complex in its inactive state. However, when the GPCR-G-protein complex forms, GDP is rapidly exchanged for GTP (guanosine triphosphate). This stimulates dissociation of the α subunit from the βγ-dimer and the initiation of downstream effector pathways (*Figure 1.4.b*). Gα-GTP regulates adenylyl cyclase and hence the production of cAMP [90,91]. This regulation can be either positive or negative depending on the G protein subtype. The βγ-dimer can stimulate ion channels, such as Ca^{2+} [92]. Signalling ceases when the GTPase domain of Gα hydrolyses GTP to GDP, and the heterotrimer is able to reform.

1.3. Smoothened

1.3.1. Smoothened has a multidomain structure

SMO plays an essential part in the Hh signalling pathway as it transduces the signal across the plasma membrane and into the cell. It is structurally classified as a Class F GPCR. It comprises 4 domains: extracellular cysteine-rich (CRD) and linker (LD) domains, heptahelical transmembrane bundle (TMD) and the intrinsically disordered cytoplasmic domain (ICD) (*Figure 1.5.a*) [93]. This multidomain structure is key for Hh signalling. The deletion of the CRD or mutation of the LD results in constitutive activation of SMO [59,60]. This implies the extracellular region of SMO is involved in stabilising the TMD in an inactive state. From evidence, it has been inferred that this is via extracellular loop 3 and transmembrane helix VI, both of which are thought to be instrumental in the activation of SMO, as well as other GPCRs (*Figure 1.5.b*) [59]. This also suggests a route of structural communication between the CRD and the so-called "ionic lock" at the cytoplasmic face of the TMD [59,84,94].

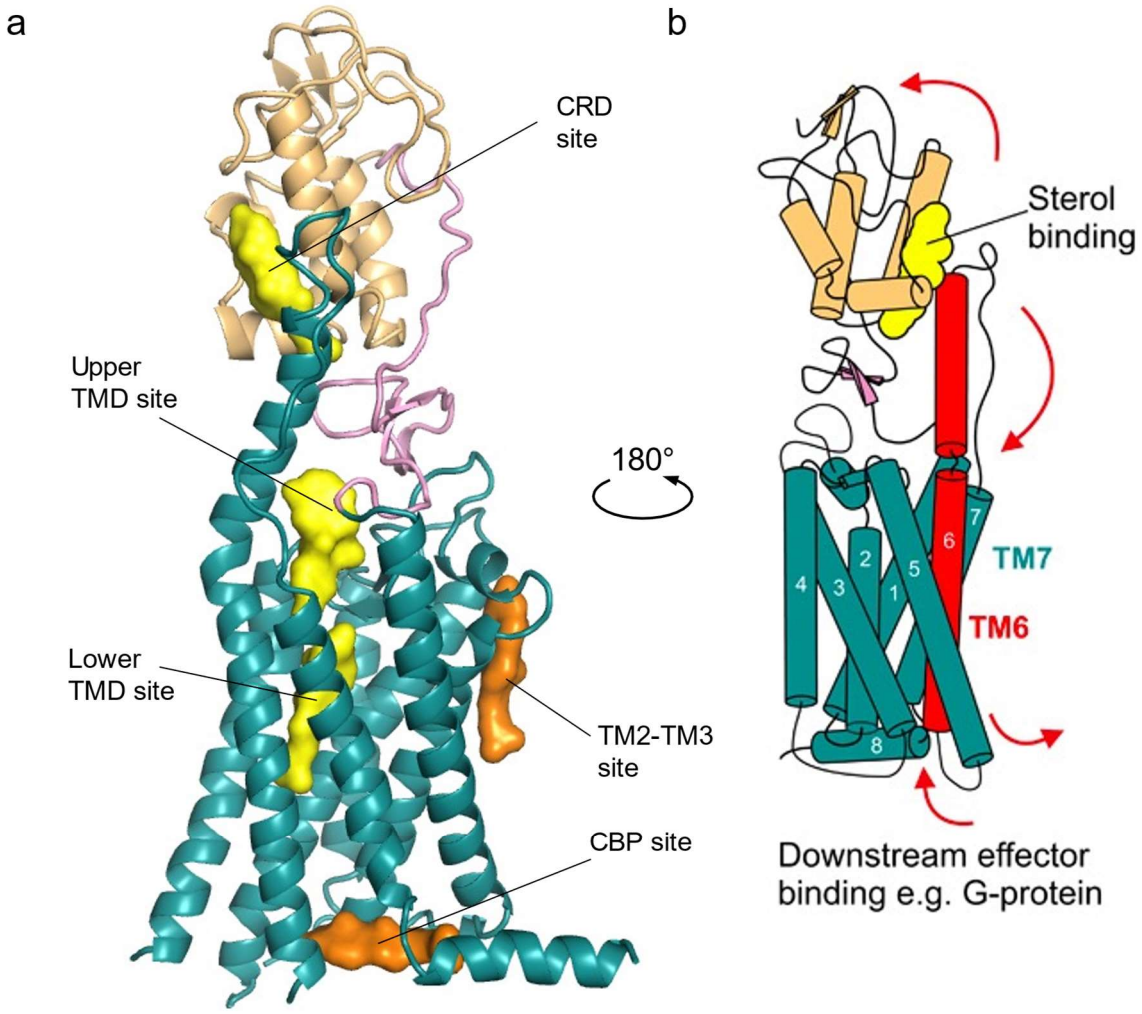


Figure 1.5. Smoothened has a multidomain structure

a. The structure of Smoothened shows the CRD (light orange) stacked above the TMD (teal) with the linker (pink) forming a wedge between the domains. The small molecule binding sites which have been structurally verified are shown in yellow. Additional sterol binding sites have been proposed and are shown in orange (CBP - cytoplasmic binding pocket, TM2-TM3 site). Binding sites are depicted as surface representations of typical ligands.

b. The proposed mechanism of signal transduction from the SMO CRD to intracellular effectors involves a rotation of the CRD which is conveyed via ECL3 to TM6 (red), which is expected to bend outwards (at the cytoplasmic end) upon activation, as seen for other GPCRs.

1.3.2. Patched regulates Smoothened via a sterol ligand

In the absence of Hh ligands, PTCH represses SMO. Most models now predict that PTCH does this by limiting the availability of specific sterols to SMO though the specific mechanism by which this is achieved is currently unknown [63]. There is mounting evidence that PTCH functions as a transmembrane cholesterol transporter and that SMO can bind cholesterol in at least two sites [58-62]. Some recent publications have proposed that SMO can also move cholesterol across the plasma membrane, though there is little evidence or requirement for this [95]. The identification of cholesterol as an endogenous agonist was supported by mutational analysis that showed prevention of cholesterol binding also ablated Shh mediated Hh signalling [60]. However, this also raised questions as the inherent insolubility of cholesterol restricts how it may travel between PTCH and SMO – clearly it cannot simply diffuse through the aqueous cellular environment. Initially, it was suggested that the SMO CRD could bind cholesterol from the plasma membrane. This is an attractive explanation, as PTCH appears to be capable of controlling the relative abundance of cholesterol in each leaflet of the plasma membrane [62]. However, the CRD is around 12Å above the membrane (*Figure 1.5*), so would require cholesterol to completely desorb from the lipid bilayer in order to bind to SMO [96]. This event is highly energetically unfavourable and therefore unlikely to occur spontaneously.

1.3.3. Multiple small molecules regulate Smoothened

Alongside cholesterol, many other natural and synthetic small molecules can modulate SMO activity via a range of binding sites [93]. The CRD has a shallow, hydrophobic groove which when empty interacts with the ECL3 [59,97]. A number of sterols and related molecules interact here, including cholesterol and several side-chain oxysterols (*Figure 1.6*) [60,61,97-101]. Mutational analysis of this site has highlighted that CRD cholesterol binding is both necessary and sufficient for endogenous SMO activation [59,60,102]. The TMD has a classical GPCR ligand binding site, accessed via the extracellular face of the 7-TM bundle [103]. Ligands including the plant steroidal alkaloid cyclopamine and SAG bind here, as well as anticancer therapies such as vismodegib and sonidegib (*Figure 1.6*) [59,103,104]. A lower TMD binding site has also been observed recently, suggested to accommodate a cholesterol molecule and explain the basal levels of SMO signalling (*Figure 1.5.a*) [94]. Several other locations have been identified in molecular dynamics simulations as potential cholesterol binding sites, including the CBP (cytoplasmic binding pocket) and TM2-3 site, though they have yet to be experimentally confirmed [101,105]. They could simply show areas of the membrane-exposed surface of the TMD where membrane cholesterol is most likely to associate.

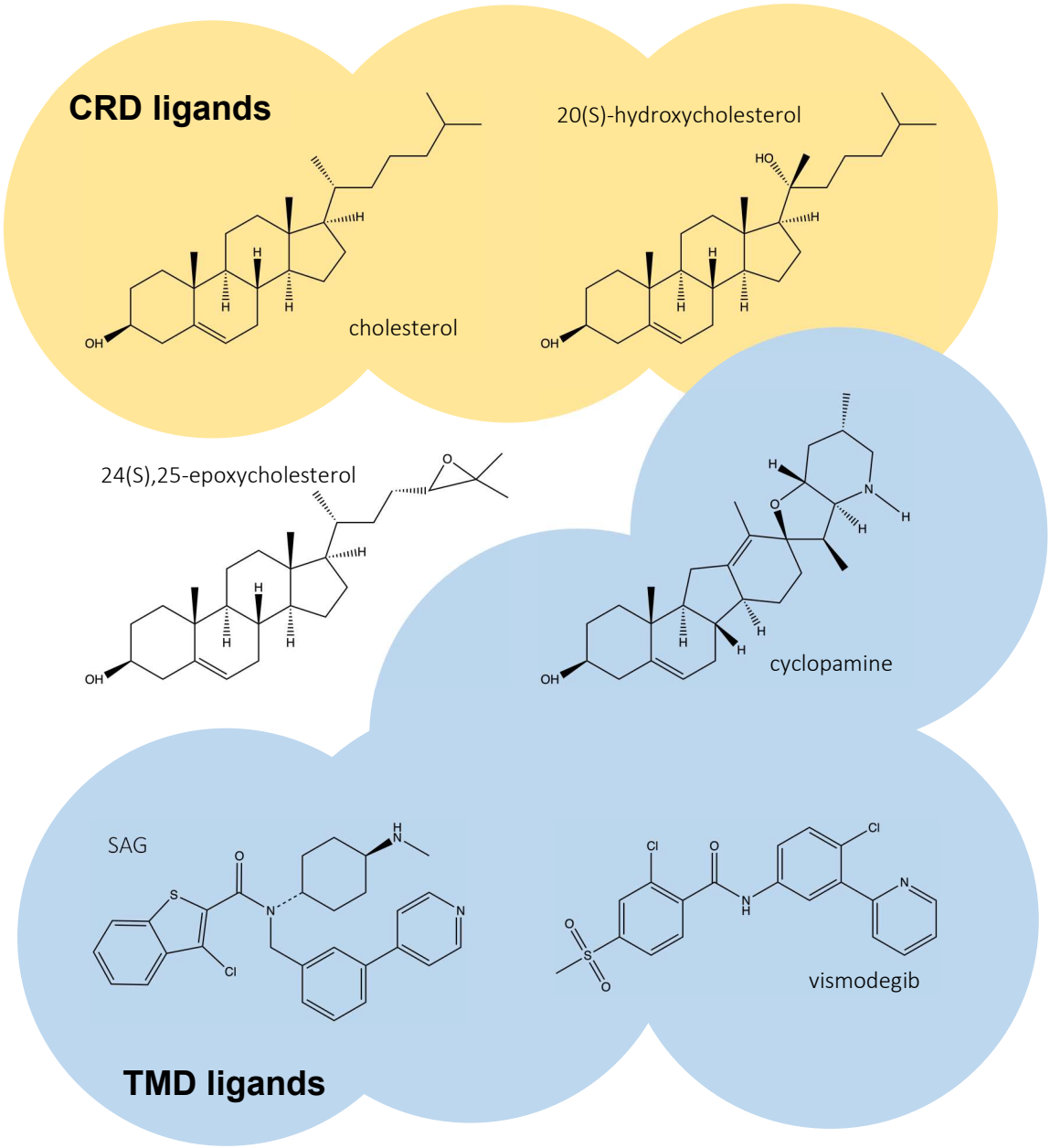


Figure 1.6. Smoothened interacts with a range of small molecule ligands.
Chemical structures of a range of SMO binding small molecules discussed in this thesis. CRD ligands are highlighted in yellow, TMD ligands in blue. 24(S),25-epoxycholesterol has not been definitively linked to one of the two sites as yet. Note – many other compounds are known to interact with SMO, this list is not exhaustive.

1.3.4. Smoothened signals to GLI via an unknown mechanism

Upon SMO activation by a CRD ligand, e.g. cholesterol, the signal is communicated across the molecule to the cytoplasmic face of the molecule. However, the events which unfold as a result of this are less well characterised. It has been suggested that SMO signals by more than one mechanism: the canonical system involving ciliary localisation and the Sufu-GLI complex, and the non-canonical system resembling a typical GPCR signalling cascade [4]. It is possible these systems work synergistically or occur in specific locations or in response to different agonists [106].

In the canonical intracellular signalling pathway, SMO localises to the primary cilium upon activation. This is thought to be as a result of phosphorylation by GRK2 (GPCR kinase 2) and CK1 (casein kinase 1) of the disordered intracellular tail and subsequent interaction with β-arrestin and KIF3A (kinesin-like protein 3A) [66,107,108]. Once in the cilia, SMO forms a ternary complex with EVC (Ellis-van Creveld syndrome protein) and EVC2, which enables signalling [109,110]. The step which links the formation of this SMO-EVC-EVC2 complex to the dissociation of the SuFu-GLI complex is unknown [111,112]. This results in isolated full-length GLI, which can then translocate to the nucleus without being degraded.

The non-canonical pathway describes SMO behaving as a classical GPCR. There is evidence to suggest SMO stimulates the Gi/o inhibitory G-protein classes in a cellular system [67-69,113]. Activation of these heterotrimeric G-proteins causes the inhibition of adenylyl cyclase and a subsequent decrease in cellular levels of cAMP. This in turn reduces activation of PKA (protein kinase A), an enzyme implicated in phosphorylation of GLI transcription factors [114,115]. It is clear that this would enhance Hh signalling, simply due to a lower rate of GLI degradation.

1.4. Thesis outline

SMO is an essential component of the Hh signalling pathway and hence plays a vital role in embryonic development. The pathway is also implicated in a wide range of cancers as both driver and potentiator, making SMO a valuable anti-cancer target. In this thesis I explore the molecular mechanism of SMO, including how small molecule ligands interact at multiple sites to stimulate intramolecular conformational change. I also present a range of approaches to determine structural information about the SMO active state, alongside the initial optimisations for cryo-electron microscopy experiments with this sample.

Materials and Methods are grouped into one chapter (Chapter 2, from page 27).

The results are presented as three chapters; each including a detailed introduction, description of the data and focused discussion. The first of these chapters covers the optimisation of production of 'wild-type' SMO, a key requirement for this thesis (Chapter 3, from page 50). The next chapter focuses upon small molecule and drug binding to SMO. We describe the development of a novel ligand binding assay and show how this will be used to identify next-generation anti-cancer drugs to overcome the emerging problem of drug resistance. We also adapt the assay to explore SMO allostery and the role of a recently discovered sterol binding site (Chapter 4, from page 77). The final results chapter details our approach to

determining an activated structure of SMO. We explore a range of potential binding partners to identify a conformationally selective molecule to stabilise the active state of SMO. We also describe extensive optimisation of cryo-electron microscopy sample preparation and outline the remaining steps towards our structural goal (Chapter 5, from page 112).

This thesis is completed with a discussion chapter which brings our results into the wider context of SMO and Hh signalling and describes our suggestions for future study (Chapter 6, from page 138).

Chapter 2

Materials & Methods

2.1. Expression and purification of membrane proteins

2.1.1. Smoothened expression (transient)

Wild-type human Smoothened was truncated at the N and C termini, to leave the extracellular and transmembrane domains, hereafter referred to as SMOΔC (UniProt ID Q99835; residues 32-555). This construct was obtained from GeneArt (Grand Island, NY) and cloned into the pHLsec vector [116] with the addition of a C-terminal Rho1D4 antibody epitope tag for purification purposes [117]. The pHLsec vector was developed within STRUBI as an effective system for high yield protein production in mammalian cells (*Figure 2.1*). SMOΔC was cloned into the MCS (multiple cloning site) using EcoRI/XhoI restriction sites. Expression is under the control of the CMV (cytomegalovirus) enhancer and chicken β-actin promoter (found to be the strongest combination for a variety of cell lines [118]). pHLsec contains an optimised Kozak sequence and chicken RPTPs secretion signal.

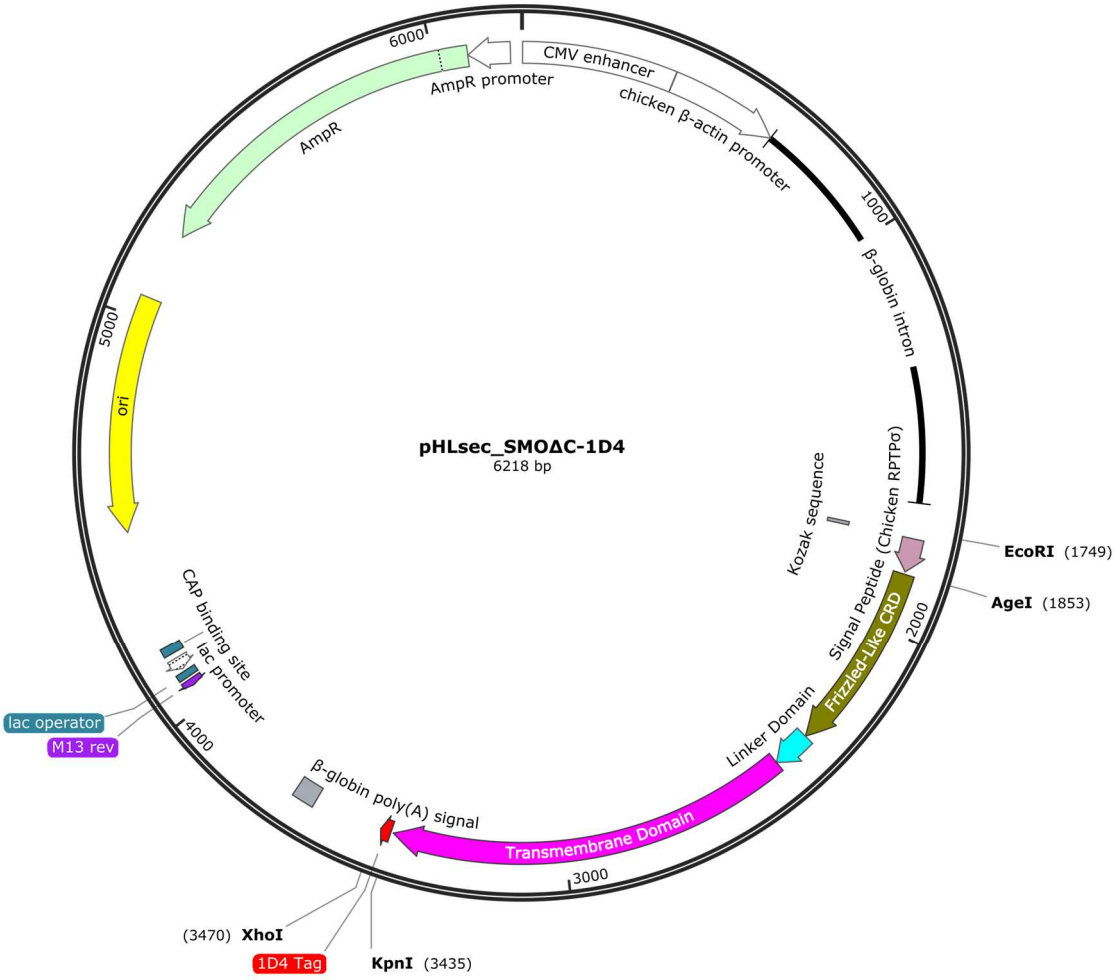


Figure 2.1 Vector map of pHLsec-SMOΔC-1D4 used for transient expression in HEK 293S GnTI⁻ cells.
Gene expression is under the control of the cytomegalovirus (CMV) enhancer/chicken β-actin promoter (white), the optimised Kozak sequence (grey) serves as the translation initiation point. Chicken receptor protein tyrosine phosphatase sigma (RPTPσ) secretion signal peptide (pink) traffics translated protein to the plasma membrane. EcoRI – AgeI – KpnI – XhoI restriction sites allow straightforward subcloning within this vector. AmpR (ampicillin resistance) provides a selection marker for plasmid amplification in *E.coli.* Map produced using SnapGene® software (from GSL Biotech; available at snapgene.com)

SMOΔC was expressed in HEK293S GnTI⁻ (ATCC CRL-3022) cells grown in suspension at 37°C, 8% CO2, 130 rpm in protein expression medium (PEM, Gibco) supplemented with 1% L-Glutamine (Gibco), 1% non-essential amino acids (NEAA, Gibco) and 1% fetal bovine serum (FBS, Gibco). Transient transfection of a typical

batch volume of 9.6 L was performed, essentially as previously described [59] when cells reached densities of 2-3 x10^6 cells per mL. Cells from 0.8 L cultures were collected by centrifugation (1100xg, 7 min) and resuspended in 120 mL Freestyle 293 Expression medium (Gibco) supplemented with 1% L-Glutamine and 1% non-essential amino acids. 1.2 mg PEI Max (Polysciences), 5 mM valproic acid (Sigma-Aldrich) and 0.4 mg plasmid DNA were added, and cultures were incubated for 3 – 5 h at 160 rpm. Cultures were topped up to 0.9 L with recycled PEM, supplemented with 0.5% L-Glutamine, 0.5% non-essential amino acids and 0.5% Penicillin/Streptomycin (Sigma-Aldrich). Cells were harvested 72 h post-transfection by centrifugation (1400xg, 7 min), snap-frozen in liquid nitrogen and stored at -80°C.

2.1.2. Smoothened expression (lentiviral)

The SMOΔC construct used for transient transfection was cloned into pHR-CMV_TetO_2 (Addgene Plasmid #113883, *Figure 2.2*) [119]. The pHR-CMV_TetO_2 transfer plasmid suite was developed within STRUBI to enable the generation of stable cell lines for protein production. When Lenti-X producer cells (Takara/Clontech, cat. no. 632180) are co-transfected with the transfer, envelope (pMD2.G; Addgene Plasmid #12259) and packaging plasmid (psPAX2; Addgene Plasmid #12260) lentiviral particles encoding the transgene are produced and subsequently used to transduce the expression cell line. The transgene, SMOΔC-3C-

mVs-1D4 is expressed under control of a CMV promoter/enhancer and two tet operators. When the HEK293S GnTI$^-$ TetR cell line is used, constitutive intracellular expression of TetR (Tet repressor protein) results in repression of transgene expression by binding to the tet operators. Repression is relieved (and hence expression induced) through addition of tetracycline (or its analogue, doxycycline) which binds to TetR and blocks binding and repression.

After co-transfection of lentiviral producer cells (as above), the supernatant containing lentiviral particles was filtered (0.45 μm), supplemented with 10 μg/mL Polybrene and used to transduce HEK293S GnTI$^-$ TetR cells. Transduced cells were expanded into suspension cultures and grown in Freestyle 293 Expression Medium (Gibco) under the same conditions as cells for transient transfection, with the addition of 2 μg/mL Blasticidin (Thermo Fisher, cat. no. R21001) to all media. When cells reached densities of 3-4x10^6/mL, expression was induced with the addition of 0.1 μg/mL Doxycycline (Merck, cat. no. D3447). Cells were harvested 72 h post-transfection by centrifugation (1500xg, 10 min), snap-frozen in liquid nitrogen and stored at -80°C.

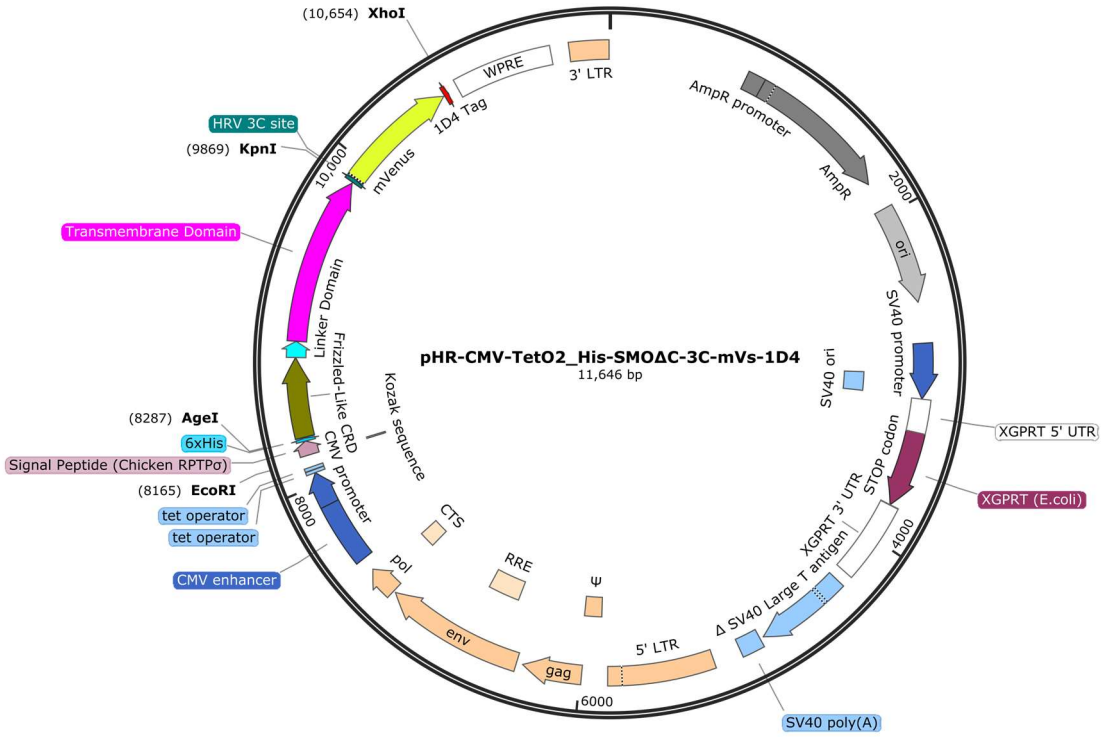


Figure 2.2 Vector map of pHR-CMV-TetO2_His-SMOΔC-3C-mVs-1D4 used for creation of stable HEK 293S GnTI⁻ TetR cell lines.
Genetic elements between the 5′ LTR and 3′ LTR (pale orange) will be packaged in lentiviral particles and stably integrated into the host cell genome as proviral DNA. Tet operators (light blue) can bind Tet repressor protein (TetR) and repress transcription. Gene expression is under the control of the cytomegalovirus (CMV) enhancer/promoter (blue), the optimised Kozak sequence (grey) serves as the translation initiation point. Chicken receptor protein tyrosine phosphatase sigma (RPTPσ) secretion signal peptide (pink) traffics translated protein to the plasma membrane. The SV40 (simian vacuolating virus 40) promoter, ori and Large T antigen together allow episomal replication of the plasmid within mammalian cells, ensuring high copy numbers. EcoRI – AgeI – KpnI – XhoI restriction sites allow straightforward subcloning within this vector and from pHLsec. AmpR (ampicillin resistance) provides a selection marker for plasmid amplification in *E.coli.* Map produced using SnapGene® software (from GSL Biotech; available at snapgene.com)

2.1.3. Smoothened small scale expression

Small scale expression tests were performed by seeding cells of the relevant type

(for lentiviral or transient expression systems) into 6-well plates and growing them

to confluency. Cells were either induced using doxycycline or transiently transfected [120]. After the specified expression time, cells were washed and resuspended in PBS before being harvested by centrifugation (600xg, 2 min). Cells were lysed with 1% (w/v) n-dodecyl-β-D-maltopyranoside (DDM, Anatrace) 5:1 cholesteryl hemisuccinate (CHS, Anatrace) for 90 minutes at 4°C. Supernatant was clarified by centrifugation (15000xg, 15 min, 4°C) before being analysed by either western blotting or FSEC (Fluorescence-detection Size Exclusion Chromatography).

2.1.4. Smoothened purification

Purification of SMO was initially performed as previously described [59]. Optimisation was carried out within this project (purification pipeline in *Figure 2.3*, see Chapter 3 for optimisation).

Thawed cell pellets from lentiviral or transient expression were resuspended in 50 mM HEPES pH 7.5, 300 mM NaCl, 1% protease inhibitor cocktail (Sigma, P8340). This was supplemented with a final concentration of 1% DDM and 0.2% CHS and incubated at 4°C for 90 minutes. Cell lysate was clarified by centrifugation (50,000xg, 30 min, 4°C) and the soluble fraction was retained. 50 mM HEPES pH 7.5, 300 mM NaCl was added to dilute the total detergent concentration to 0.5%, before adding Rho-1D4 antibody (University of British Columbia) coupled CNBr-activated sepharose beads (Cytiva) and incubating for 2 h. Protein-bound beads

were washed with 100x bead volume 50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS and eluted overnight with the addition of either 500 μM 1D4 peptide (TETSQVAPA, GenScript) or 3C protease (produced in-house). If tandem affinity purification was required (Chapter 3.3.4), Ni-NTA Agarose resin (Qiagen) was added and incubated for 1 hour at 4°C before washing sequentially with 1D4 wash buffer and buffer supplemented with 10 mM imidazole. Bound protein was eluted by the addition of 250 mM imidazole. The eluate was concentrated to 250 μL using a Vivaspin 20 PES 100,000 Da MWCO centrifugal concentrator (Sartorius) and loaded onto a Superose 6 Increase 10/300 column (Cytiva), pre-equilibrated in 10 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS. Peak fractions were pooled and concentrated.

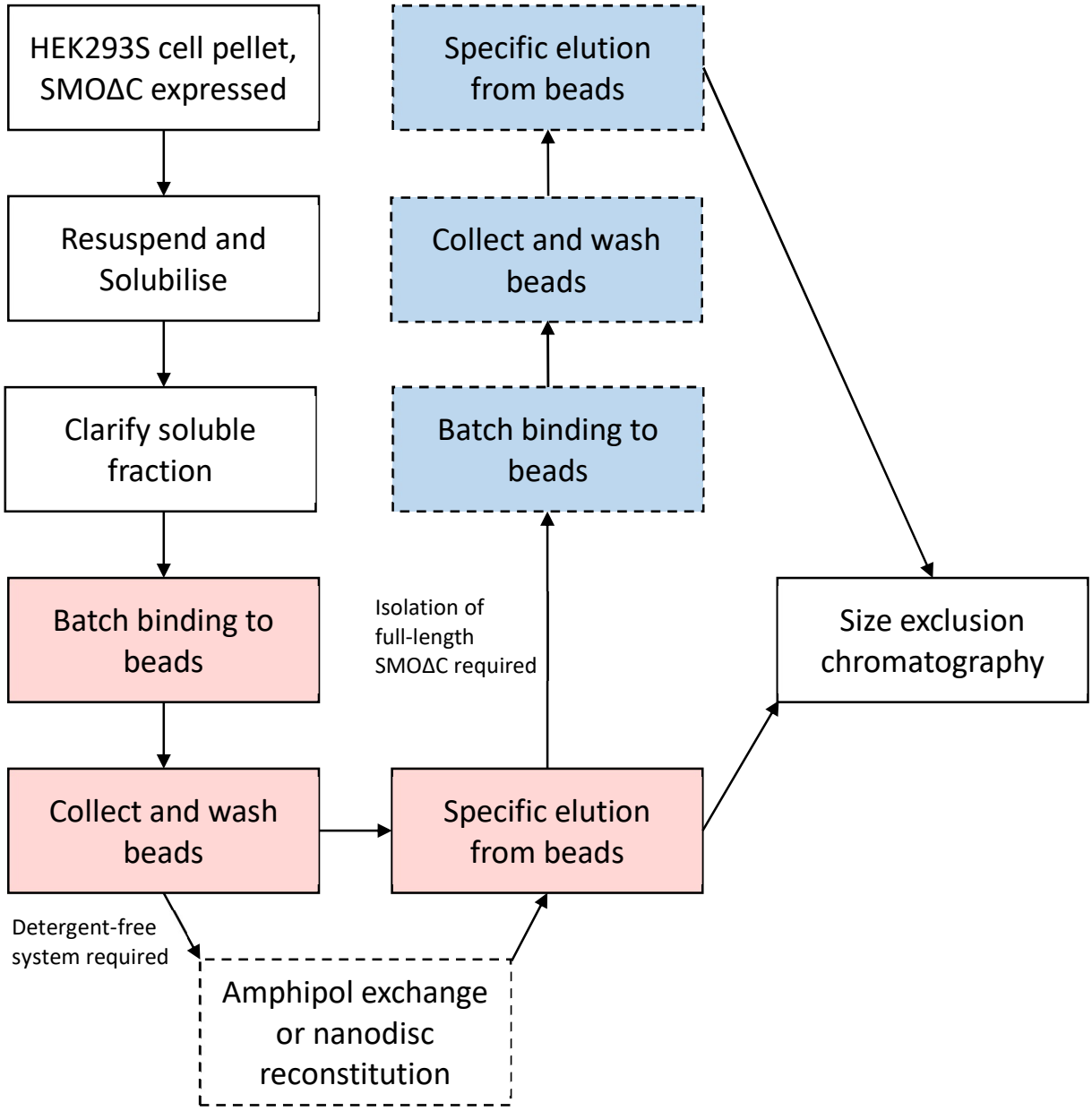


Figure 2.3 Simplified purification pipeline for SMO.
Compulsory procedure shown (solid borders) with optional extra steps (dashed borders) to transfer to a detergent-free system or perform tandem affinity purification.

2.1.5. Amphipol exchange

A8-35 (Anatrace) amphipol stock solutions were used at 1 mg/mL. Amphipols were added to the target protein at an 8:1 ratio (w:w) and diluted to give a final protein concentration of 0.3 mg/mL in 10 mM HEPES pH7.5, 150 mM NaCl, 10% glycerol,

0.03% DDM, 0.006% CHS. This exchange reaction was incubated via rotation for 2-4 h at 4°C before adding 100 mg/mL BioBeads (BioRad), pre-equilibrated in 10 mM HEPES pH 7.5, 150 mM NaCl, and incubating overnight to remove detergent. Amphipol-stabilised protein was removed from the BioBeads and concentrated if necessary.

2.1.6. Nanodisc reconstitution

DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphocholine), DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) and POPC (1-palmitoyl-2-oleoyl-glycero-3-phosphocholine) (Avanti Polar Lipids Inc.) were prepared from chloroform stocks by desiccation. Dried lipids were resuspended in 10 mM HEPES, pH 7.5, 150 mM NaCl, 0.03% DDM, 0.006% CHS, 300 mM Sodium Cholate to give a final lipid concentration of 100 mM. Reconstitution reactions were prepared using 5 mM lipid, 15 mM cholate and varied concentrations of MSP1D1 (expressed and purified in the Siebold lab by C. Kowatsch) and SMOΔC (dependant on lipid:MSP ratio) and incubated for 1 h at 4°C, RT or 37°C (dependent on lipid). 100 mg/mL BioBeads were added and the mixture was incubated overnight at 4°C. Nanodisc-reconstitued protein was then removed from the BioBeads and concentrated if necessary.

2.2. Expression and purification of soluble proteins

2.2.1. G-protein alpha subunit expression

The minimal-Gs construct used was identical to that previously described [121,122]. The mini-Gi1, mini-Gi2 and mini-Go constructs were designed on the basis of the mini-Gs construct. The remaining mini G-protein constructs are as described [123]. For purification purposes, an N-terminal histidine tag (His_6), monovenus (mVs, if applicable) and HRV 3C protease cleavage site (LEVLFQGP) were added. These constructs were obtained from GeneArt (Grand Island, NY) and cloned into the pET22b vector for bacterial expression.

Full length G alpha i1 and G alpha o constructs were obtained from the cDNA Resource Center (www.cdna.org) and cloned into pET22b after an N-terminal histidine tag (His_6) and 3C protease cleavage site.

All G alpha and mini-G proteins were expressed in *E. coli* strain Rosetta (DE3), grown at 37°C in Terrific Broth (TB, Fisher Scientific) supplemented with 0.2% glucose and 5 mM $MgSO_4$ until OD600=0.8. Expression was induced with 200 μM IPTG and the temperature lowered to 20°C. Cells were harvested after 20 h by centrifugation (4000xg, 15 min) and stored at -20°C until purification.

2.2.2. G-protein alpha subunit purification

Thawed cell pellets were resuspended in (40 mM HEPES pH7.5, 500 mM NaCl, 10% glycerol, 5 mM $MgCl_2$, 250 μM guanosine diphosphate (GDP), 500 μM β-mercaptoethanol, EDTA-free protease inhibitor cocktail (Roche) and lysed using a cell disruptor (Constant Systems Ltd). The lysate was clarified by centrifugation (38,000xg, 45 min) and incubated with TALON beads for 1 h at 4°C. Protein-bound beads were washed with 20 mM HEPES pH7.5, 500 mM NaCl, 10% glycerol, 5 mM imidazole, 1 mM $MgCl_2$, 25 μM GDP and bound protein was eluted with 20 mM HEPES pH7.5, 100 mM NaCl, 10% glycerol, 250 mM imidazole, 1 mM $MgCl_2$, 250 μM GDP, 250 μM β-mercaptoethanol. The eluate was loaded onto a Superdex75 or Superdex200 (for full-length G alpha and monovenus tagged mini-G) 16/600 gel filtration column (Cytiva), pre-equilibrated in 10 mM HEPES pH7.5, 300 mM NaCl, 10% glycerol, 1 mM $MgCl_2$, 1 μM GDP, 1 mM dithiothreitol (DTT). Peak fractions were pooled and concentrated. Aliquots were snap-frozen in liquid nitrogen and stored at -80°C.

2.2.3. Nanobody & megabody bacterial expression

Nanobodies were raised in llamas (*Lama glama*) [124] against SMOΔC (UniProt ID Q99835; residues 32-555) containing a BRIL domain (UniProt ID: P0ABE7; residues 23-128) replacing intracellular loop 3 (UniProt ID: Q99835; residues 429-440).

Megabodies were constructed based on these nanobodies by the Steyaert lab (VUB, Brussels). An N-terminal PelB signal sequence (MKYLLPTAAAGLLLLAAQPAMA - for periplasmic expression) and a C-terminal histidine tag (His_6) were added for purification purposes. Megabodies were expressed in *E.coli* strain Wk6Su, grown at 37°C in TB until OD600=0.7. Expression was induced with 1 mM IPTG and the temperature lowered to 25°C. Cells were harvested after 20 h by centrifugation (4000xg, 15 min) and stored at -20°C until purification.

2.2.4. Nanobody purification from *E. coli*

Thawed cell pellets were resuspended in ice-cold TES buffer (200 mM Tris-HCl pH8.0, 20% (w/v) sucrose, EDTA-free protease inhibitor cocktail (Roche)) and incubated for 4 h at 4°C. Further TES buffer was added and diluted before incubating overnight at 4°C. Cell suspension was clarified by centrifugation (10,000 xg, 30 min) and incubated with Ni-NTA resin (Qiagen) for 1 h at 4°C. Protein-bound beads were washed with 50 mM HEPES pH7.5, 500 mM NaCl, followed by 50 mM HEPES pH7.5, 500 mM NaCl, 10 mM imidazole. Protein was eluted in 20 mM HEPES pH7.5, 150 mM NaCl, 250 mM imidazole and loaded onto a Superdex75 16/600 gel filtration column (Cytiva), pre-equilibrated in 10 mM HEPES pH7.5, 150 mM NaCl. Peak fractions were pooled and concentrated to ~30 mg/mL. Aliquots were snap-frozen in liquid nitrogen and stored at -80°C.

2.2.5. Megabody purification

Thawed cell pellets were resuspended in chilled PE buffer (50 mM Tris-HCl pH8.0, 150 mM NaCl, 20% w/v sucrose, 0.4 mg/mL Lysozyme (Sigma) and EDTA-free protease inhibitor cocktail (Roche) and incubated for 30 minutes with shaking at 4°C. Cell suspension was clarified by centrifugation (10,000xg, 30 min) and the supernatant was supplemented with 5 mM $MgCl_2$ and 500 mM NaCl (final concentrations). The supernatant was then filtered through a 0.22 μm filter and incubated with Ni-NTA resin (Qiagen) for 1 h at 4°C. Protein-bound beads were washed with 50 mM Tris-HCl pH7.5, 500 mM NaCl, followed by 50 mM Tris-HCl pH7.5, 500 mM NaCl, 10 mM imidazole. Bound protein was eluted with 50 mM Tris-HCl pH7.5, 100 mM NaCl, 250 mM imidazole. Eluate was loaded onto a Superdex200 16/600 gel filtration column (Cytiva), pre-equilibrated in 10 mM Tris-HCl pH7.5, 150 mM NaCl. Peak fractions were pooled and concentrated. Aliquots were snap-frozen in liquid nitrogen and stored at -80°C.

2.2.6. Nanobody mammalian expression

Nanobodies were cloned into pHR-CMV_$TetO_2$ [119] for lentiviral expression in mammalian cells. C-terminal His_6 or 3C-mVs-His_{12} were added for purification purposes. Lenti-X producer cells were transiently transfected alongside envelope and packaging plasmids, as described (Chapter 2.1.2). Virus-containing media was

subsequently used to transduce HEK293T cells. These stably-transfected cells were then expanded into roller bottles in Dulbecco's Modified Eagle's Media (DMEM, Sigma, cat. no. D6546) supplemented with 1% L-glutamine, 1% non-essential amino acids and 10% FBS. To induce expression, media was switched to 2% FBS and incubated at 37°C, 5% CO_2 for 72 h. Conditioned media was harvested and replaced with fresh complete DMEM supplemented with 2% FBS and incubated for a further 72 h before the final harvest.

2.2.7. Nanobody purification from conditioned media

Cell debris was removed from conditioned media by centrifugation (4000 xg, 10 min, RT) and filtration (0.2 μm). Clarified media was concentrated and dialysed into phosphate buffered saline (PBS) pH 8.0 using a QuixStand™ benchtop system (Cytiva) connected to a 60 cm Xampler Cartridge (Cytiva) with a 10 kDa MWCO. Concentrated media was incubated with TALON beads for 1 h at RT. Protein-bound beads were washed sequentially with PBS pH 8.0 and PBS pH 8.0 supplemented with 5 mM imidazole. Protein was eluted in 10 mM HEPES pH 7.5, 150 mM NaCl, 250 mM imidazole and loaded onto a Superdex75 16/600, pre-equilibrated in 10 mM HEPES pH 7.5, 150 mM NaCl. Peak fractions were pooled, concentrated and aliquots were snap-frozen in liquid nitrogen for storage at -80°C.

2.3. Biophysical techniques

2.3.1. Minimal G-protein thermal shift assay

Thermal shift assays measure the thermal denaturation temperature of proteins under various conditions. This reflects the stability of the protein and can be useful in optimising buffer conditions or construct design. Thermofluor uses fluorescent dyes which bind to hydrophobic residues to measure the protein denaturation. One such dye is SYPRO orange, a fluorophore which is quenched by water [125]. As the temperature increase causes protein unfolding, the hydrophobic core of the protein is exposed. SYPRO orange binds non-specifically to this hydrophobic surface, therefore excluding water and resulting in an increase in fluorescence.

4 μg purified minimal G protein was diluted to 50 μL in 10 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol supplemented with various concentrations of GDP (guanosine diphosphate) and 3x SYPRO orange (Sigma-Aldrich). Samples were heated from 25- 98 °C in 1 °C increments using an Mx3005P qPCR System (Agilent Genomics). Resultant data was fit to a Boltzmann distribution and the Tm values extracted using MATLAB.

2.3.2. FSEC complexation trials

Samples were loaded onto a Superose 6 Increase 3.2/300 column (Cytiva), pre-equilibrated in 10 mM HEPES pH 7.5, 150 mM NaCl, supplemented with 0.03% DDM, 0006% CHS if necessary. Absorbance at 280 nm was monitored and monovenus fluorescence was measured using in-line fluorescence detection (excitation 515 nm, emission 528 nm).

2.3.3. SMO pull-down trials

SMO was incubated overnight at 4°C with nanobody 8 (Nb8)[94] and small molecule ligands, where applicable. Nb8 was immobilised onto TALON beads via a His_6-tag and washed with 10x bead volume chilled 10 mM Hepes pH 7.5, 150 mM NaCl, 0.02% GDN, 0.002% CHS. Nb8 and any bound SMO were eluted in 10 mM Hepes pH 7.5, 150 mM NaCl, 0.02% GDN, 0.002% CHS, 250 mM imidazole and analysed by western blotting.

2.3.4. Bio-layer interferometry

Bio-layer interferometry (BLI) is a label free technique used to measure biomolecular interactions. It measures the interference pattern of white light reflected from a layer of protein on the biosensor tip and an internal reference. Once a ligand has been immobilised on the tip, usually via a specific high affinity

interaction such as biotin/streptavidin, it is presented with an analyte in solution. Any binding of analyte to immobilised ligand causes a change in the optical thickness of the protein layer, measured as a shift in wavelength of reflected light. This technique was developed by ForteBio and is carried out using an Octet® system.

SMOΔC with an N-terminal Avi-tag™ was biotinylated *in vivo* by co-transfection of BirA biotin ligase [126] and then purified and exchanged into amphipols as outlined above (Chapter 2.1.4-5). SMO at ~0.2 μM in 10 mM HEPES pH7.5, 150 mM NaCl, was immobilised onto Dip and Read™ Streptavidin (SA) Biosensors (180 s baseline, 300 s loading, 180 s baseline, at 25°C) using the Octet® RED96e System (ForteBio). After loading, 10-50 μM analyte was applied to the SMO-conjugated biosensors (180 s baseline, 180 s association, 300 s dissociation) as an initial binding test. To subsequently determine a dissociation constant (K_D), a dilution series of the analyte was prepared and sequentially applied (increasing in concentration) to an unused SMO-conjugated biosensor. In parallel, the dilution series was applied to an unloaded biosensor to allow any non-specific binding to be subtracted during analysis. All kinetic analysis was performed using the Octet® Data Analysis software.

2.4. Structural techniques

2.4.1. Electron microscopy studies

The SMO sample was isolated using 1D4 affinity purification, as described. The protein was reconstituted into nanodiscs (if applicable) and mixed with a 2x molar excess of nanobody or megabody, as appropriate. For negative stain analysis 4 μL of sample was applied to glow-discharged Formvar/Carbon 300 mesh copper supports. After a short incubation excess sample was removed by sequential blotting and washing before applying uranyl formate stain. Images were collected on a Tecnai T12 transmission electron microscope (TEM, ThermoFisher) together with Christiane Kowatsch.

For cryo-electron microscopy, 4 μL of purified sample was applied to glow-discharged Quantifoil R1.2/1.3 Cu 300 mesh grids. Excess sample was removed by blotting for 3s with a blot force of -25 under 4°C and 100% humidity conditions. The grids were plunge-frozen in liquid ethane using a Vitrobot™ (ThermoFisher). Cryo-EM images were collected together with Dimple Karia, Vojta Prazak and Loic Carrique using a Glacios cryo-TEM (ThermoFisher).

2.5. Small molecule binding

2.5.1. FRET-based CRD binding assay

We designed a novel FRET-based (Förster resonance energy transfer) assay to study the SMO CRD binding site, based on HTRF (homogeneous time resolved fluorescence) technology for high-throughput compound screening. We made use of a commercially available FRET pair – an anti-HA antibody linked to the FRET donor Eu^{3+} cryptate and streptavidin linked XL665 FRET acceptor (Cisbio). Alongside this we designed target specific components – SMOΔC$_{stable}$-HA (HA-SMOΔC-V329F-BRIL) and a biotinylated oxysterol (20(S)-yne conjugated to biotin via a long linker, prepared by Doug Covey, Washington University School of Medicine, *Figure 2.4.a*). When these components are combined the anti-HA antibody recognises the SMO HA tag, the biotinylated oxysterol binds to the SMOΔC$_{stable}$-HA CRD and is simultaneously recognised by XL665-streptavidin. This brings the FRET pair into close proximity, allowing energy transfer and hence a high FRET signal (*Figure 2.4.b*). The subsequent addition of competing CRD ligands disrupts this by replacing the biotinylated oxysterol in the SMO CRD binding site. This results in a decrease in FRET as a function of competitor concentration, allowing us to measure binding of small molecules to the SMO CRD site.

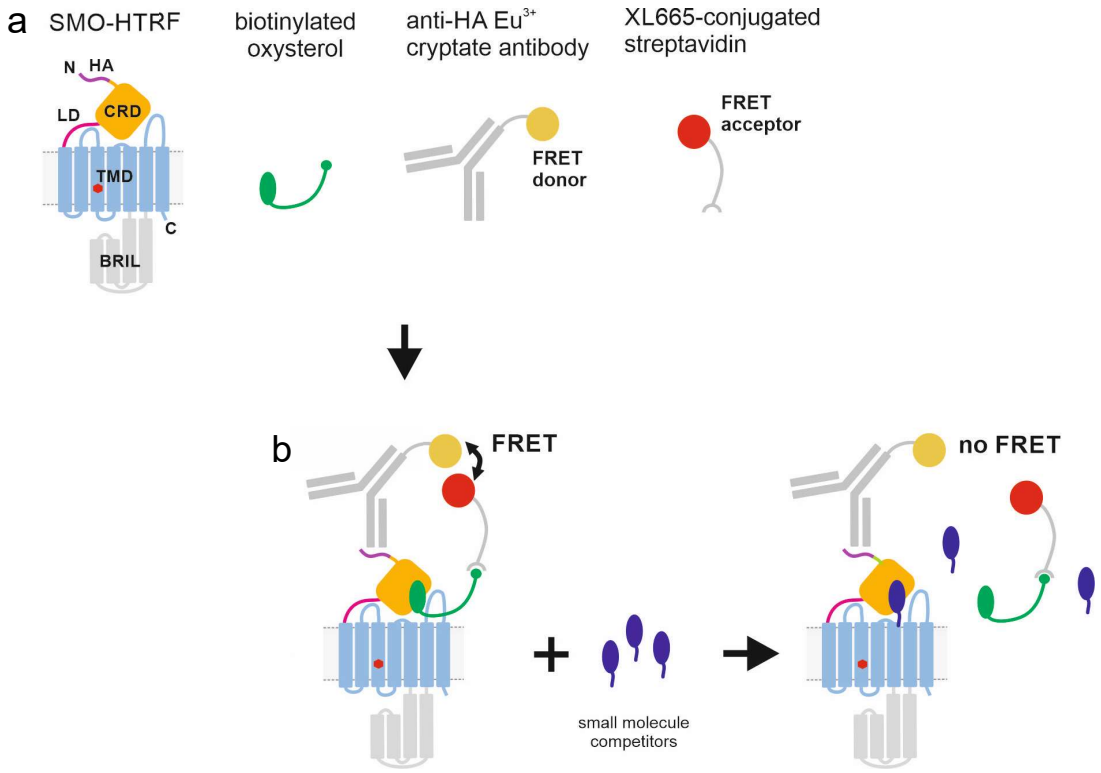


Figure 2.4 FRET assay schematic.

a. SMOΔC$_{stable}$-HA contains an N-terminal hemagglutinin (HA) tag (purple), single point mutation in the TMD, V329F (red), a BRIL domain in place of intracellular loop 3 (grey) and a truncated C-terminus (aa. 555). Biotinylated oxysterol (green) consists of 20(S)-yne conjugated to biotin via a linker, produced by Douglas Covey, Washington University in St. Louis. Anti-HA Eu^{3+} cryptate antibody and XL665-conjugated streptavidin are available from Cisbio (Bagnols sur Cèze, France).

b. SMOΔC$_{stable}$-HA binds to biotinylated oxysterol and forms a complex which produces a FRET signal. Small molecule competitors replace the biotinylated oxysterol and FRET ceases to occur.

Potential small-molecule competitors were initially dissolved in 50% DMSO and dispensed as an 11-point dilution series at 10x final concentration in 10 mM HEPES pH 7.5, 150 mM NaCl, 0.03% DDM, 0.006% CHS. A competitor free reference was included in each row and a DMSO-only dilution series was run alongside each experiment. The SMO-specific assay components were combined to give final in-assay concentrations of 50 nM SMOΔC$_{stable}$-HA, 150 nM biotinylated oxysterol in 10

mM HEPES pH 7.5, 150 mM NaCl, 0.03% DDM, 0.006% CHS. The FRET components were combined to give final in-assay concentrations of 18.75 nM XL665-Streptavidin, 0.4 nM Anti-HA Eu^{3+} cryptate in PPI Europium Detection buffer (Cisbio). The SMO-specific and FRET component mixtures were then combined and transferred into the 96-well plate containing the competitor dilution series. Each condition was mixed thoroughly and any air bubbles were removed by centrifugation (1000xg, 2 min, RT). Each condition was then divided into four 20 µL replicates in the assay plate (ProxiPlate-384 Plus, white (PerkinElmer)). The assay was read using a PHERAstar FS microplate reader (BMG LABTECH) with an excitation wavelength of 337 nm and emission detection at 620 and 665 nm. Experimental data are presented as a FRET ratio:

$$FRET\ ratio = \left(\frac{Emission665}{Emission620}\right) x10000$$

Data were processed using GraphPad Prism version 8.0.0 for Mac (GraphPad Software, San Diego, California USA, www.graphpad.com). Data were normalised to 0% and 100% bound biotinylated oxysterol, using experimentally determined minimum and maximum values. A minimum value of 568.9 was determined as the mean from 96 replicates of $SMO\Delta C_{stable}^{D95A}$-HA (a mutant that cannot bind oxysterols in the CRD), our negative control. The maximum value for each experiment was taken from the competitor free reference in each dataset, to reflect

experimental variation. EC50 values were calculated for each independent experiment using a four-parameter dose response binding curve in GraphPad Prism. Cyclopamine data were instead fitted to a bell-shaped dose response binding curve. Average EC50 values for small molecule competitors were determined by calculating the mean and 95% confidence interval of the logEC50 [127].

Chapter 3

Production of wild type Smoothened

3.1. Abstract

As with many structural biology projects, producing sufficient amounts of pure protein was not trivial. Integral membrane proteins are especially challenging to overexpress and a wide range of variables must be considered. The subsequent solubilisation and purification of such proteins must also be carefully established. For this study, the production of wild-type SMO was required, thereby limiting our use of recognised protein engineering techniques often employed to improve the expression and stability of G-protein coupled receptors. We therefore adopted and optimised a novel lentiviral method of creating stable cell lines for protein expression. Alongside this I improved our existing purification protocol for SMO in order to maximise yields and remove degradation product contamination. I also designed and led the optimisation of detergent-free systems for use with SMO in a range of biophysical and structural studies. Our efforts in this chapter resulted in a near 8-fold increase in protein yield, making wild-type SMO a viable target for structural studies.

3.2. Introduction

As with many structural biology projects, a major hurdle is the production of milligram quantities of pure, stable protein. GPCRs are more challenging than some, as they are endogenously expressed at low levels, meaning over-production can often result in aggregation [128]. As integral membrane proteins, they also require the addition of a molecule (most often detergent) to stabilise them outside of the cellular membrane environment, to which they are native [129]. Further complications arise when studying wildtype or active constructs of GPCRs as these states are inherently transient, flexible and hence unstable. Nevertheless, there have been a number of recent advances in the field which have enabled a range of GPCRs to be produced in high enough amounts for structural studies [130].

3.2.1. GPCR Expression Systems

Several well-characterised protein expression systems have been used successfully for production of human GPCRs, though each has distinct advantages and disadvantages. *E. coli* and yeast (most commonly *Pichia pastoris*) have been used successfully and provide the advantage of rapid cell-growth to high densities in inexpensive media as well as ease of manipulation of genetic material [130,131]. However, the membrane environment of both yeast and bacteria differs significantly to that present in mammalian plasma membranes; most notably in the

amount of cholesterol present. As cholesterol is known to bind a number of GPCRs, it is thought to be an important factor in keeping GPCRs functional [130]. These systems also fail to add N-linked glycosylation to the extracellular domains of these membrane proteins, as would be in their native cell-type.

The most commonly used expression system for GPCRs is insect cells, namely *Spodoptera frugiperda Sf9* or *Trichoplusia ni* High Five cells. Insect cell systems are capable of producing milligram quantities of GPCRs per litre of culture. They are also able to perform most post-translational modifications observed in mammalian cells, however some heterogeneous glycosylation has been shown to prevent complete maturation of some GPCRs [131]. Insect cell membranes also differ in lipid composition from mammalian cells, again lacking in high amounts of cholesterol as insects cannot synthesise the sterol [132,133]. However, this can be mitigated by adding lipids to the growth media either as a mixture [134], or a specific lipid such as cholesterol [135]. It has also been suggested that the large quantities of protein produced include a substantial misfolded population, likely due to the impairment of the secretory pathway resulting from the baculovirus infection [136].

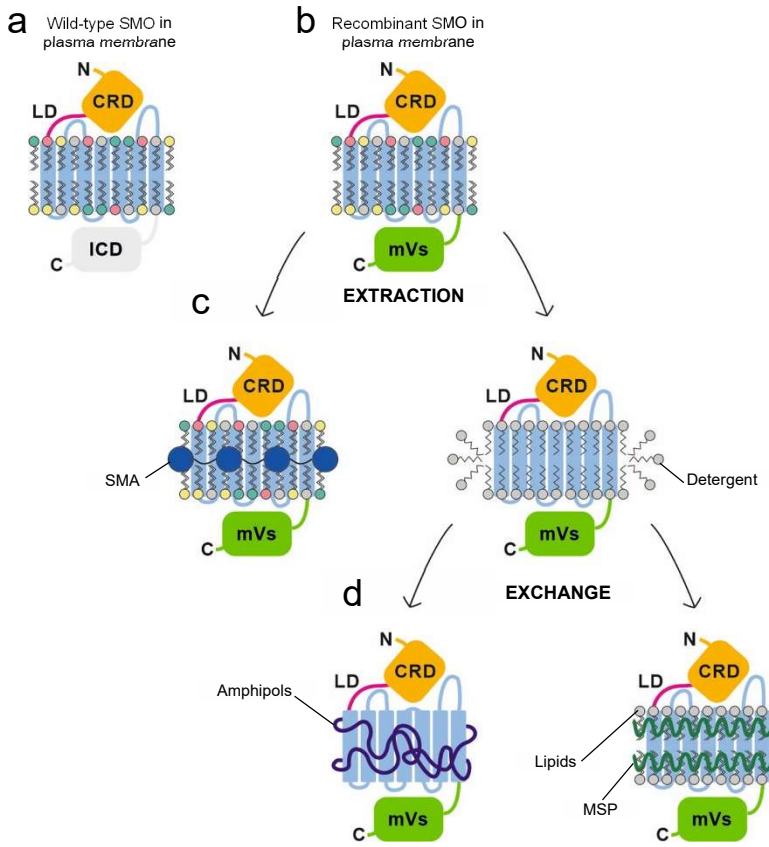


Figure 3.1. Solubilisation of GPCRs

a. Wild type SMO in the plasma membrane. A range of lipids in the membrane surround the TMD of SMO.
b. Recombinant SMOΔC-3C-mVs-1D4 in the plasma membrane after over-expression.
c. Extraction from the plasma membrane can be performed using SMALPs or detergent.
d. Detergent-solubilised SMO can be exchanged into amphipols or nanodiscs.

Perhaps the most obvious expression system for human GPCRs is mammalian cells. There are numerous advantages to using a mammalian expression system, including the native membrane environment (*Figure 3.1.a.*), full range of post-translational modifications, and glycosylation [131]. However, culturing mammalian cells is expensive, time-consuming and requires specialist equipment. There are two methods to produce protein; transient or stable transfection. Transient transfection requires a cationic compound (e.g. PEI, polyethylenimine) to introduce DNA into

the cell, while stable transfection can be achieved in a number of ways, including a non-replicative virus to insert the DNA into genome. The lentiviral method was compared to alternative approaches to high yield protein expression, such as BacMam and piggyBac [137-140]. The main advantages of the lentiviral method were speed, increased protein yields and the ability to store cell lines for re-expansion [119]. Transient transfection is often simpler, yet generating stable cell lines gives a greater amount of control over expression and higher yields [141].

3.2.2. Solubilisation

After determining a suitable expression system for production of the GPCR, the next challenge is removing the protein from its native bilayer, ideally whilst retaining its fold and function (*Figure 3.1.b.*). This is most commonly achieved using detergents (*Figure 3.1.c.*); amphiphilic molecules which are capable of solubilising membrane proteins by satisfying both hydrophobic transmembrane regions and hydrophilic loops [130]. They can be classified as ionic (e.g. sodium dodecyl sulphate, SDS), zwitterionic (e.g. 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate, CHAPS) and non-ionic (e.g. DDM) based on the charge of the polar head group. Non-ionic detergents are the mildest class and the most commonly used for solubilisation of membrane proteins for structural studies [129]. However, no detergent for the solubilisation of functional GPCRs has been proven to be universally superior, meaning the optimal detergent must be determined

experimentally [142]. Currently, the only alternative to using detergents to extract membrane proteins is styrene maleic acid (SMA, *Figure 3.1.c.*). In the presence of biological membranes SMA spontaneously polymerises surrounding a portion of the bilayer and any membrane proteins contained within to form a SMA lipid particle (SMALP) [130].

After detergent extraction from the bilayer, membrane proteins can be exchanged into amphipathic polymers (amphipols) or reconstituted into rafts of phospholipid bilayer, stabilised by an amphipathic helical protein (nanodiscs, *Figure 3.1.d.*). This circumvents any issues caused by the presence of free detergent micelles in the sample [143]. Amphipols bind exclusively to the hydrophobic transmembrane regions of membrane proteins in order to make them water-soluble. Amphipols have been shown to thermostabilise bacteriorhodopsin (BR) by up to 16°C in comparison to the detergent-solubilised protein. It has also been shown that amphipols can enable refolding of BR after denaturation by SDS [144]. Membrane proteins can be reconstituted into nanodiscs by adding selected lipids and membrane scaffold protein (MSP) and gradually removing detergent [145]. An advantage of nanodiscs is the customisability – the MSP can be varied to control the diameter of the nanodisc formed, and the lipids can be selected to vary the composition of the bilayer [146]. Whilst empty nanodiscs will exist alongside protein-containing discs (as empty detergent micelles are found alongside detergent-solubilised proteins), these can be

easily removed via affinity purification. These approaches are invaluable for studies benefitting from a detergent-free system.

3.2.3. Purification

GPCRs are purified using similar techniques to soluble proteins. Initially, the soluble fraction of cell lysate is usually subjected to affinity purification. The common hexa-histidine tag may be used, but often more specialised affinity purifications are chosen; including FLAG and 1D4 [131]. These techniques both require specific antibodies, which can be immobilised on sepharose beads. The N-terminal FLAG epitope (DYKDDDD(A/K)) is recognised by the FLAG M1 antibody and has been used to purify a number of GPCRs for structural studies [78,131]. The 1D4 epitope consists of the C-terminal residues of rhodopsin (TETSQVAPA) and the Rho1D4 monoclonal antibody is highly specific for this sequence, often resulting in extremely pure eluate [147,148]. It is also a common strategy to include two different affinity tags, one located at each of the N- and C-termini of the protein (*Figure 3.2*.). This enables tandem affinity purification and ensures the final sample includes exclusively full-length GPCR molecules [131].

Another important step in purifying GPCRs can be proteolytic cleavage. It may be desirable to remove the affinity tags or even the native N- or C-termini, as these regions may be flexible and hence hinder structural studies. Extensive studies have

been performed to test the efficiency of the most commonly used proteases in a wide range of detergents, including HRV 3C, TEV and thrombin [131,149].

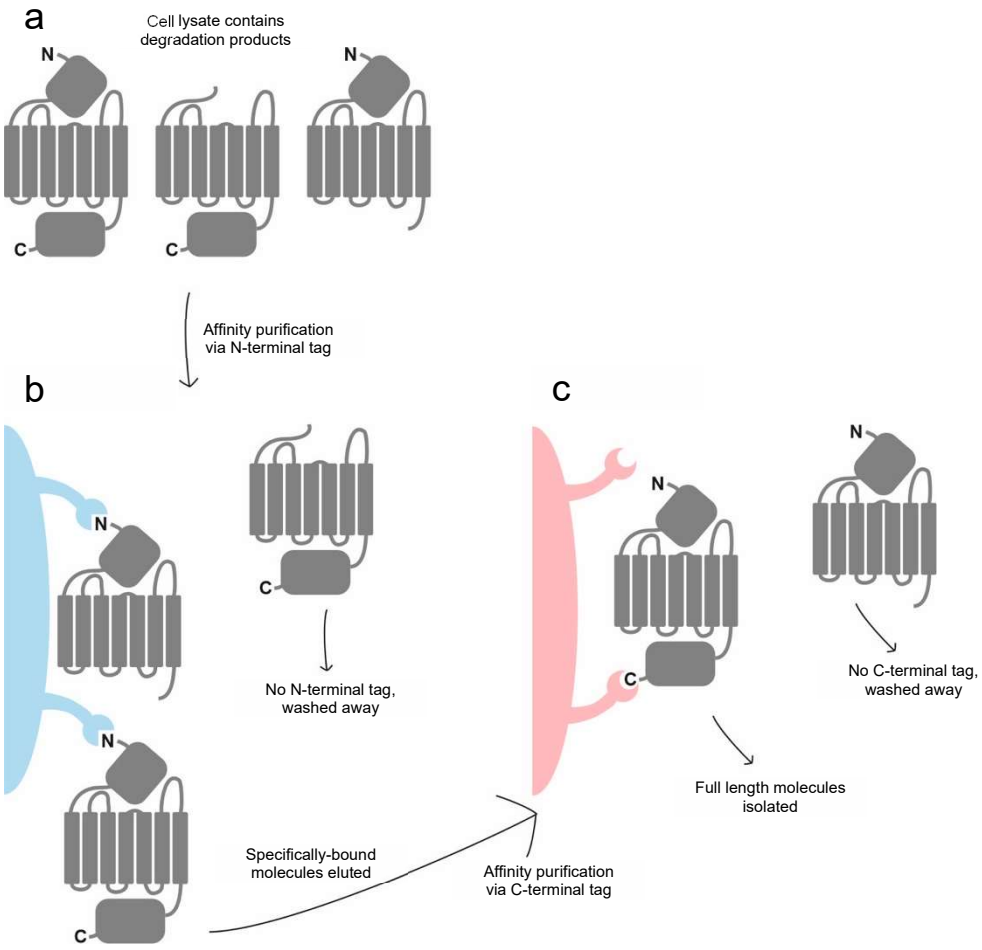


Figure 3.2. Tandem affinity purification of GPCRs

a. Cell lysate contains full-length GPCR and degradation products.
b. Initial purification via N-terminal tag removes GPCR with degraded N-terminus.
c. Specifically-eluted GPCR then purified via C-terminal tag. This removes GPCR with degraded C-terminus, leaving only full-length GPCR.

3.3. Results

3.3.1. Stable cell lines produce higher yields of SMO than transient transfection

In previous SMO studies in our research group, large-scale transient transfection using PEImax was employed to produce milligram quantities of protein. Recent advances within both our and neighbouring groups have enabled the use of a novel method of generating stable cell lines using a lentiviral system utilising the pHR plasmid [119]. This method has been designed for straightforward transfer from our previously used system; the pHLsec expression plasmid [116]. The lentiviral system allows rapid production of non-replicative virus particles for safe and efficient transduction of a variety of cell types. Any construct in the pHLsec plasmid can be transferred using EcoRI/XhoI cloning to the pHR-CMV-$TetO_2$ plasmid, which is then used (in combination with two other plasmids) to transfect virus-producing cells (*Figure 3.3.a.*). For all large-scale preparations of SMO, HEK293S GnTI$^-$ TetR cells were used, adapted for suspension culturing, as they ensure homogeneous non-complex glycosylation and allow inducible expression of SMO [119,150].

After creation of stably-transfected cell lines, transduction efficiency can be estimated either using FACS or a fluorescence microscope, providing the construct includes a fluorescent protein tag (*Figure 3.3.b.*). This can be compared to transient

transfection efficiency, which is estimated to be <30% in suspension cultures (general observation, *Fig 3.3.c*). The transduction efficiency of SMO genes into HEK293S GnTI$^-$ TetR cells was estimated to be 87.5% [119], though subsequent optimisation has increased this further (general observation).

An important step in changing to a stable expression system is to re-determine the optimum conditions for protein production. This was performed using small-scale (6-well plate format) induction tests and subjecting the solubilised cell lysate to FSEC. This allows rapid comparison of a wide range of conditions. As a minimum, it is recommended to test temperature, concentration of doxycycline (the inducing agent) and post-induction time before harvest, though media additives can also be screened in this way. I have shown that SMO expression is more effective at 37°C in comparison to 30°C (*Figure 3.3.b.*). The yields of monomeric SMO also increased continually over a 72 h period, meaning it was not necessary to harvest early to prevent degradation or aggregation. The level of SMO expression was highest when induced with a 10-100x lower concentration of doxycycline than recommended in the publication (*Figure 3.3.b.*). This may indicate that high concentrations of doxycycline produce SMO more quickly and hence fail to reach maturity.

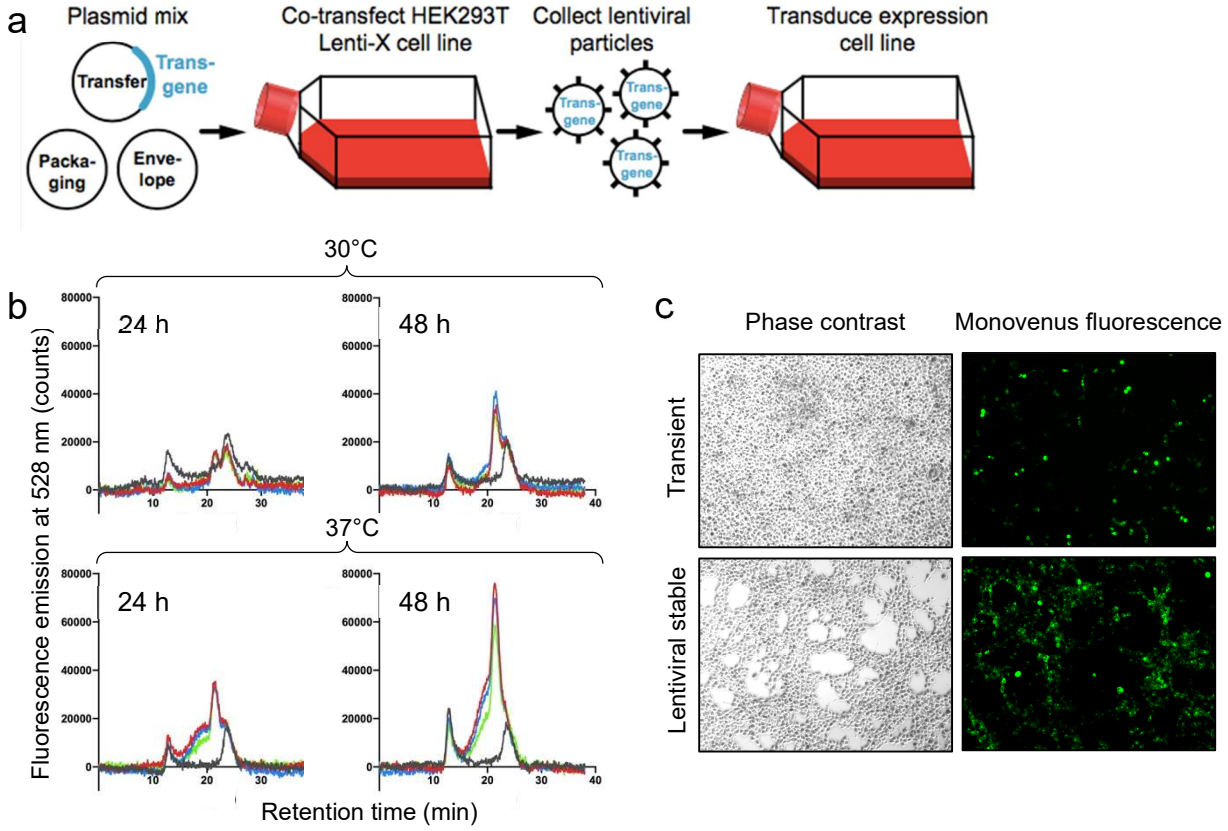


Figure 3.3. Lentiviral expression of SMO

a. Production of a stable cell line using the lentiviral method. Figure reproduced with permission from Elegheert *et al.*, 2018.
b. Optimisation of SMOΔC-3C-mVs-1D4 expression from stable cell line testing temperature, time of induction and doxycycline concentration (0µg/mL - grey, 0.1µg/mL - red, 1µg/mL - blue, 10µg/mL - green).
c. Stable and transiently-transfected HEK293S TetR cells overexpressing SMOΔC-3C-mVs-1D4. Images were taken using a Nikon Eclipse TE2000U Inverted microscope with the assistance of Dr Benjamin Bishop.

3.3.2. C-terminal tags affect SMO expression levels

One of the major challenges for this project was to improve the yields of SMOΔC to a level where structural studies would be possible. This construct comprises amino acids 32-555 of the human Smoothened protein (UniProt ID Q99835), spanning from

the CRD at the N-terminus to the end of Helix VIII at the C-terminus. As a crucial part of this project was to study the wild-type protein (or as close as possible), we were unable to use any stabilising point mutations or fusions that had been employed for GPCR structure determination in the past. Previous SMO studies in our group made use of the Val329Phe point mutation, which greatly improves expression levels [59,60,101]. Unfortunately, this mutation also locks SMO in an inactive conformation, making it unsuitable for studies of wild-type SMO. Another established strategy in the GPCR field is to insert either T4 lysozyme (T4L) or a thermostabilised apocytochrome b562RIL (BRIL) domain in the place of intracellular loop 3 (ICL3) [151,152]. This often stabilises the transmembrane bundle and hence improves yields and crystallisability [131].

As these recognised strategies were unsuitable for this project, I tested fusions and tags that were not expected to alter the function of SMO. This included a range of affinity tags, the AviTag™, and mono-venus fluorescent protein (mVs) (*Figure 3.4.a.*). Initially these features were added at the C-terminus of the SMOΔC construct, followed by a 1D4 tag. It has previously been observed in our research group that the addition of an AviTag™ to constructs improves expression (T. Malinauskas, *unpublished*) and that is the case for SMOΔC. Western blotting was used to determine relative expression levels for several differentially tagged constructs. It appears that addition of a C-terminal AviTag™ increases expression

of SMOΔC though the inclusion of mVs at the C-terminus of SMOΔC is even more effective (*Figure 3.4.b.*).

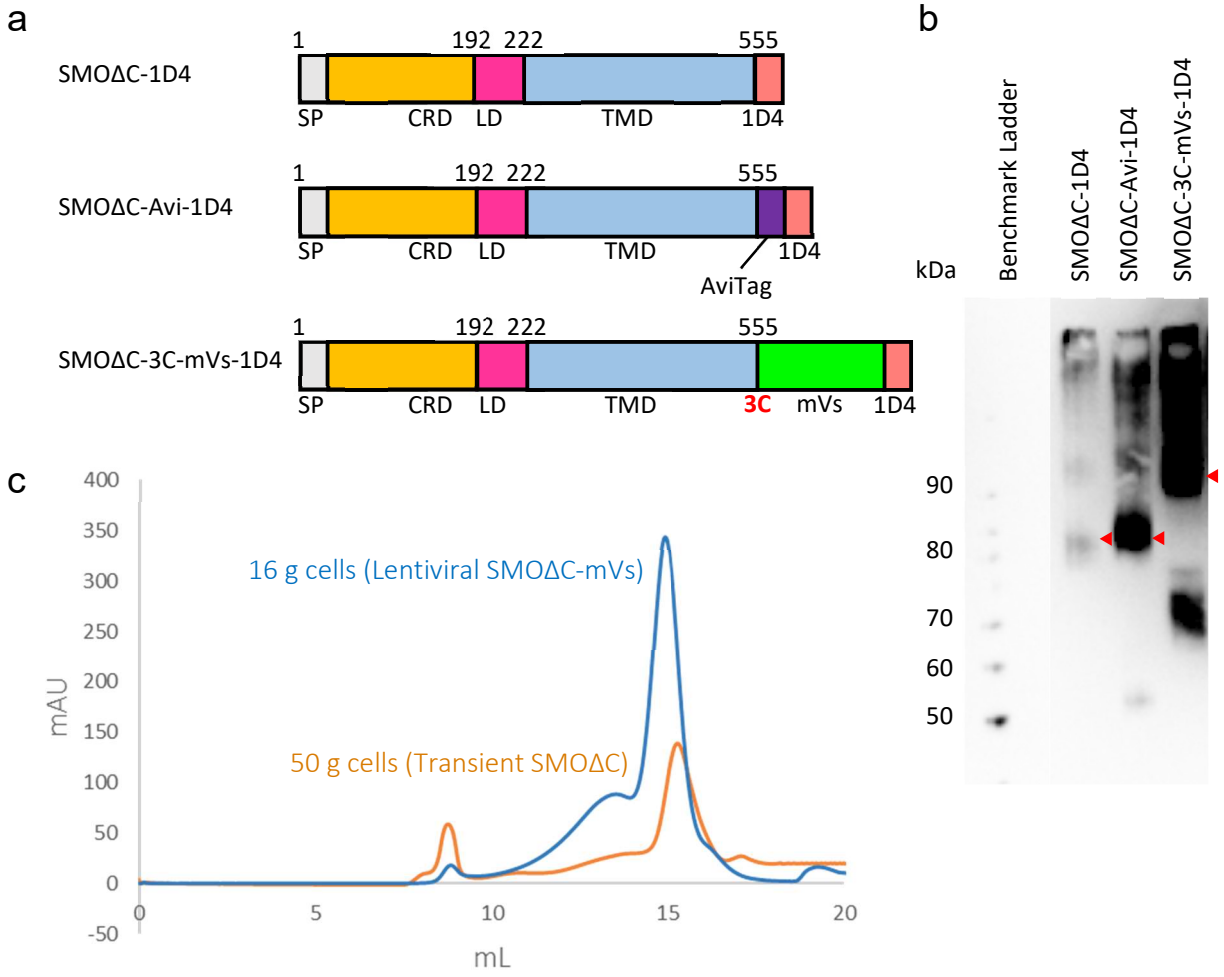


Figure 3.4. C-terminal Venus improves expression of SMOΔC

a. Comparison of constructs tested. SP– signal peptide, CRD – cysteine rich domain, LD – linker domain, TMD – transmembrane domain, 1D4 – Rho-1D4 tag, mVs – monovenus.
b. 1D4 western blot. Red arrows highlight the full-length SMO band for each construct.
c. Comparison of Superose 6 10/300 traces showing expression of SMOΔC expressed using transient transfection and SMOΔC-3C-mVs-1D4 expressed in a stable cell line.

We combined the novel lentiviral system for generating stable cell lines with the optimised conditions and construct. The overall effect of these changes increased

yields by a factor of approximately 8 (*Figure 3.4.c.*). Most importantly, this has been achieved without having to move away from a wild-type SMO construct.

3.3.3. Purification of increased amounts of monomeric SMOΔC

Another potential strategy to increase the yield of SMOΔC is to improve the purification technique used. Previous work in our research group has shown that solubilisation of SMO from cell membranes is efficient in a range of detergents including OGNG (octyl glucose neopentyl glycol), DMNG (decyl maltose neopentyl glycol), β-OG (n-octyl-β-D-glucopyranoside) and DDM [153]. However, it was shown that DDM with 10:1 CHS provides additional thermostability to SMO, in comparison to detergents such as OGNG [153]. We explored whether additional compounds could improve this solubilisation process using a 96-condition additive screen, designed in-house [153]. This revealed that we could subtly improve the solubilisation of SMOΔC through addition of low millimolar concentrations of dithiothreitol (DTT, Sigma). Adding this resulted in an increase in the yield of monomeric SMOΔC relative to oligomeric SMOΔC (*Figure 3.5.c.*), suggesting that SMOΔC oligomerises (at least in part) due to cysteine residues forming intermolecular disulphide bonds.

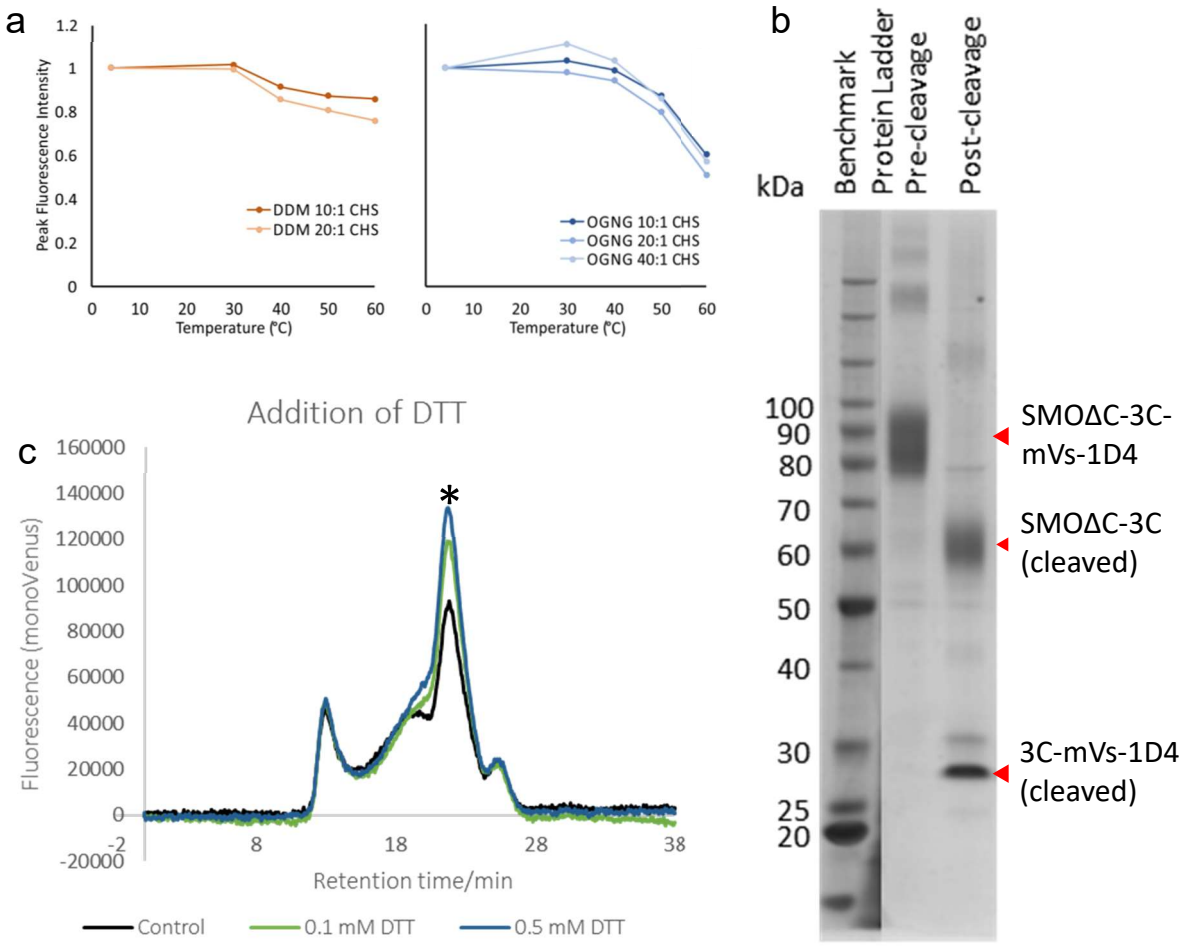


Figure 3.5. Purification of SMOΔC-3C-mVs-1D4

a. DDM provides greater thermostability for SMO, in comparison with OGNG (experiments performed and analysed by E.F.X. Byrne)
b. SDS-PAGE gel showing efficiency of 3C cleavage of SMOΔC-3C-mVs-1D4.
c. Addition of DTT marginally improves the amount of monomeric SMO (*) relative to oligomeric aggregate.

Alongside solubilisation, we also focused upon the affinity purification procedure as a potential area for improvement. Traditionally, elution from Rho1D4 conjugated beads is achieved by outcompeting the protein's 1D4 tag with the 1D4 peptide (amino acid sequence TETSQVAPA). As the protein tag and the free peptide are of identical sequence, they have a very similar affinity for the Rho1D4 antibody. This means competition-based elution is inherently inefficient as an equilibrium is

reached with significant quantities of the protein bound to the Rho1D4 beads via the 1D4 tag. I therefore decided to investigate other means of elution. The best expressing SMOΔC construct is SMOΔC-3C-mVs-1D4 so I decided to make use of the 3C cleavage site. Elution in this manner will remove the mVs-1D4 tag from the protein, but this is often necessary anyway. The efficiency of HRV 3C protease (produced in house by S.C. Griffiths, R. Woolley) cleavage of SMOΔC-3C-mVs-1D4 in solution was tested and subsequently analysed via SDS-PAGE. Cleavage is shown to be nearly 100% efficient for this SMO construct (*Figure 3.5.b.*). As a result, the cleavage method of elution was adopted for all purifications, unless the C-terminal mVs-1D4 was required for subsequent analyses.

3.3.4. Tandem affinity purification can be used to isolate full-length SMOΔC

Upon closer examination of SMOΔC purification we observed an additional species in SDS-PAGE and FSEC analysis. This appeared to be a degradation product of SMOΔC as it could be detected via anti-1D4 western blotting and FSEC. This led to the conclusion that the N-terminus of SMOΔC must be cleaved in this sub-species (*Figure 3.6.a.*). We first attempted to prevent this via a range of approaches including: adding L-arginine to the expression medium; decreasing expression time; adding additional protease inhibitors during solubilisation; and ensuring every step was carried out on ice (where possible). However, these strategies were

unsuccessful. We therefore decided to circumvent the problem by isolating the full-length protein from the mixture. To do this I designed and cloned several constructs with dual tags to enable a tandem affinity purification, purifying first via a C-terminal tag and then purifying via an N-terminal tag, for example (*Figure 3.6.b.*). We decided to keep the C-terminal 1D4 tag, as this system has been optimised for use with SMO and has the advantage of a highly pure elution. The best expressing construct tested had an N-terminal hexa-histidine (His_6) tag, making the full construct His_6-SMOΔC-3C-mVs-1D4. We have shown that tandem affinity purification of this construct results in a single species; full-length His_6-SMOΔC-3C-mVs-1D4 (*Figure 3.6.c.*).

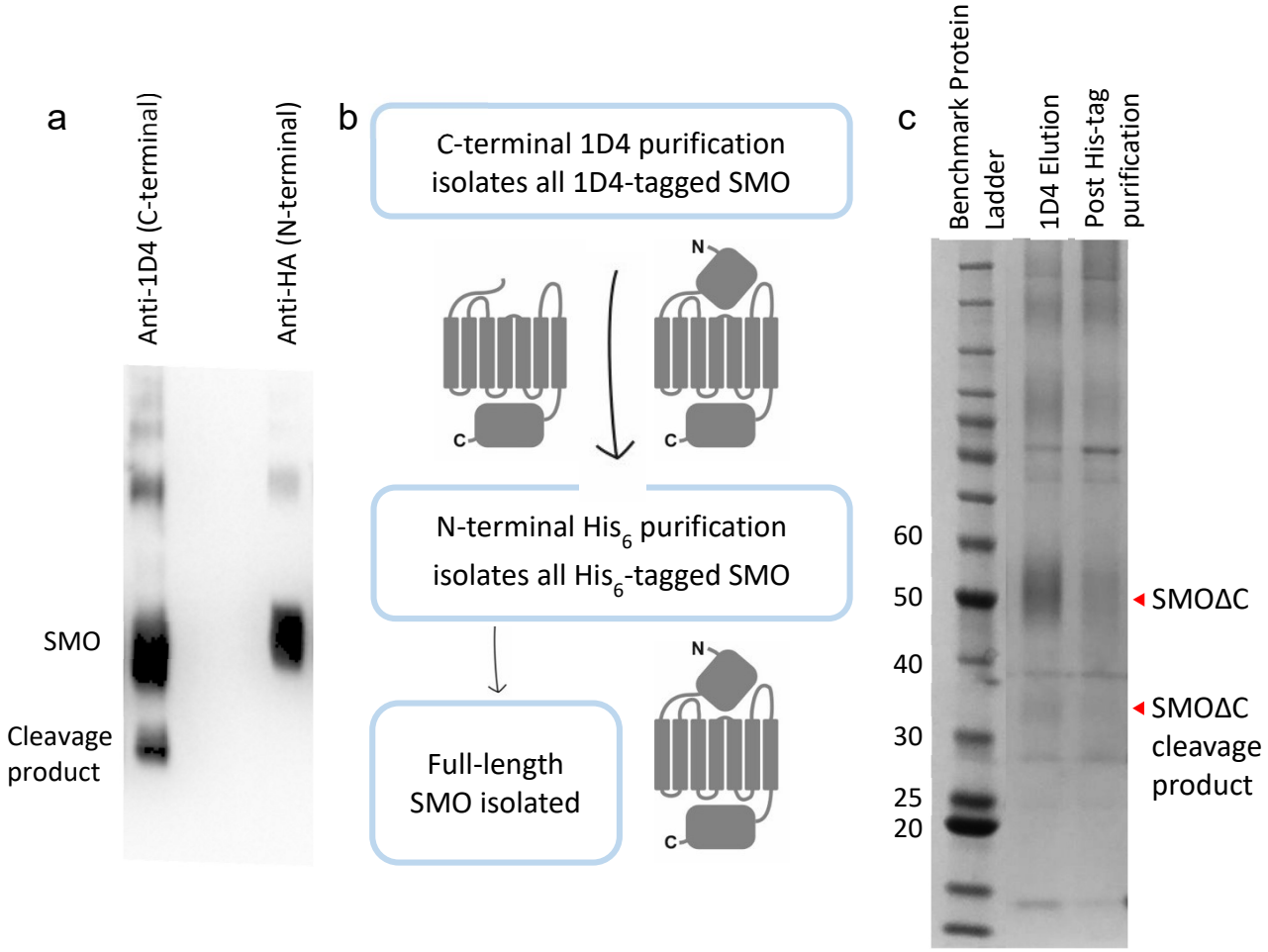


Figure 3.6. Tandem affinity purification of SMO

a. C-terminal 1D4 western blot shows two species, N-terminal western blot shows one of the species is lacking the N-terminus.
b. Schematic outlining the SMO tandem affinity purification.
c. SDS-PAGE gel showing isolation of full-length SMO only.

3.3.5. Detergent-free systems can be successfully used for SMOΔC

Detergent-free systems can be useful for a range of experiments involving membrane proteins. Detergent micelles (those containing protein and empty micelles) can cause problems for both biophysical and structural studies. Detergent-solubilised proteins may be restricted to a particular conformational state due to

detergent stabilisation [154]. This, along with steric clashes with the detergent micelle, can prevent protein ligand binding to the membrane protein. Empty detergent micelles can also be problematic, especially as they are often enriched alongside the solubilised protein during centrifugal concentration. This becomes an issue in studies such as electron microscopy, where empty micelles are easily mistaken for protein-containing 'particles'.

To combat this, we have optimised the exchange of SMOΔC into amphipols. An excess of amphipols (A8-35) is mixed with detergent-solubilised membrane protein and then (after an incubation period) the detergent molecules are gradually removed. I have also optimised amphipol exchange to be carried out during the initial purification to maximise yields. This results in a size exclusion profile resembling that of SMOΔC purified in detergent, suggesting the majority of solubilised protein has been successfully exchanged into amphipols (*Figure 3.7.a.*). Amphipol-exchanged protein was used for binding and complexation studies.

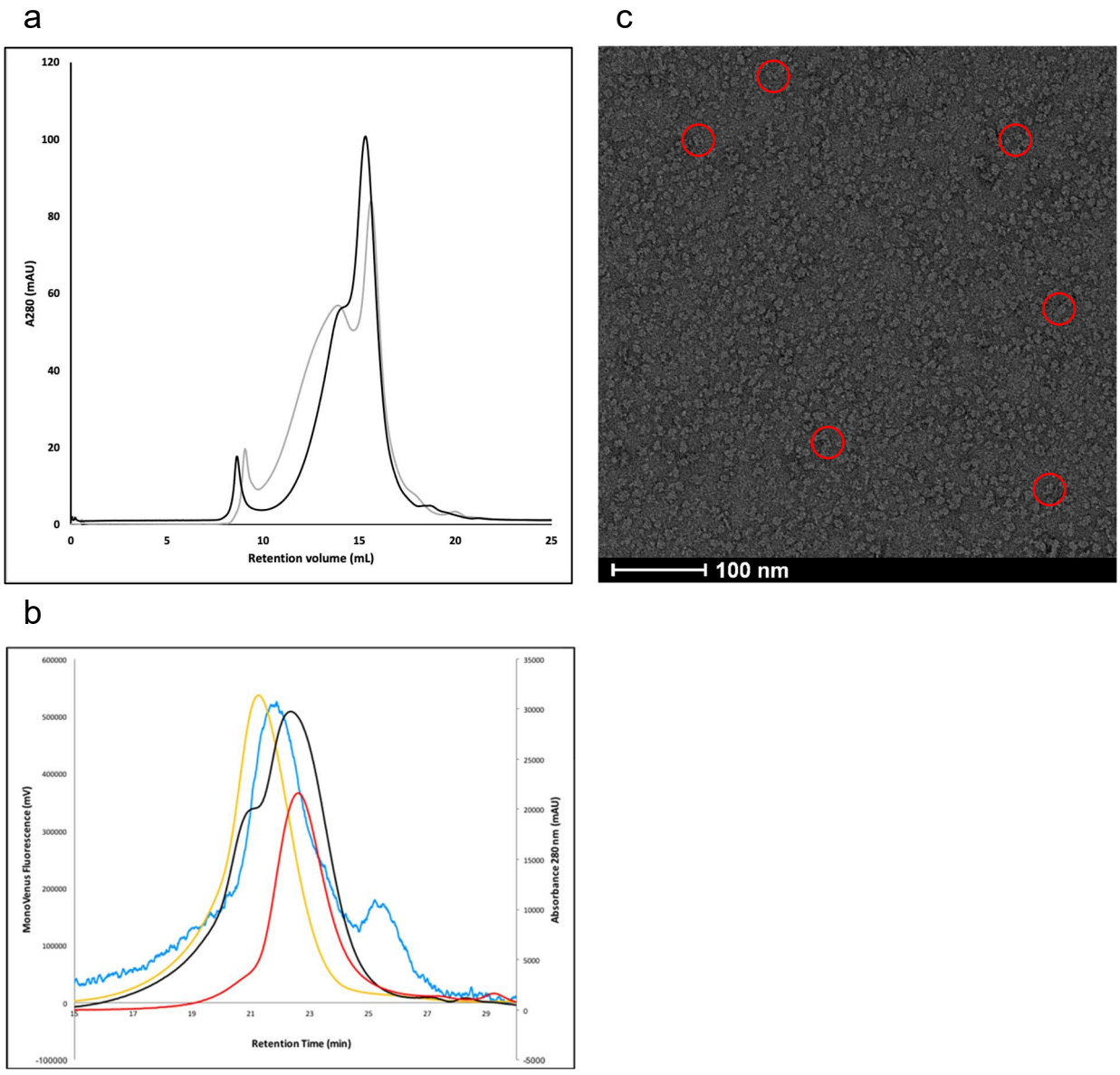


Figure 3.7. SMO in detergent-free systems

a. SMO can be exchanged in amphipols during purification. Superose 6 Increase 10/300 traces of SMO in detergent (black) and amphipols (grey).
b. SMO can be reconstituted into nanodiscs. Red – empty nanodiscs, blue - SMOΔC-3C-mVs-1D4 in detergent, black – SMOΔC-3C-mVs-1D4 reconstituted in nanodiscs A_{280nm}, yellow - SMOΔC-3C-mVs-1D4 reconstituted in nanodiscs mVs fluorescence.
c. Negative stain TEM image of SMO in nanodiscs. Nanodisc stacking highlighted in red.

I have also carried out extensive trials to explore nanodisc reconstitution of SMOΔC. Nanodisc reconstitution is performed in a similar way to amphipol exchange, by adding the lipids and MSP to detergent-solubilised membrane protein, before

gradually removing the detergent. I tested a range of lipids and MSP constructs to find the optimum combination for SMOΔC. I used the fluorescent SMOΔC construct to assess nanodisc reconstitution via FSEC. This analysis showed that SMOΔC was successfully incorporated into the nanodisc as the fluorescent peak increased in molecular weight in comparison to either SMOΔC alone or empty nanodiscs (*Figure 3.7.b.*). We aimed to maximise the yields of SMOΔC and so used an excess of lipids and MSP. This led to a population of empty nanodiscs being produced. These nanodiscs would cause the same problems as empty detergent micelles during electron microscopy studies, as they would be difficult to differentiate from protein-filled nanodiscs. We have also observed nanodisc 'stacking' in negative stain TEM, which can mask protein particles (*Figure 3.7.c.*). However, unlike detergent micelles, we can remove empty nanodiscs. The crucial difference between these solubilising molecules is that detergent must be maintained above the CMC in the solution to prevent micelles dissociating. There is no such issue with nanodiscs, meaning these samples can be washed extensively in non-nanodisc-containing buffer to remove excess discs. I chose to do this whilst SMOΔC in nanodiscs was bound to 1D4 beads. Any empty discs should not be bound to these beads and any that are bound non-specifically dissociate after a mild salt wash (500 mM NaCl). This results in a specific elution of SMOΔC in nanodiscs.

3.4. Discussion

3.4.1. Expression of SMOΔC

The lentiviral system of producing stable cell lines has many advantages over transient transfection. Perhaps the most obvious is the improved yield of SMOΔC, as discussed here and widely noted as one of the key factors for structure determination [155]. The efficiency of transduction is much higher than that of transient transfection, meaning a higher number of cells will produce the chosen construct [119]. Ease of use is also a benefit of stable cell lines [156]. When sufficient cell mass is reached, stable cell lines are induced. In our lentiviral system, this consists of adding the desired concentration of doxycycline. However, a large-scale transient transfection of HEK293S cells adapted for suspension is a much larger undertaking – a process that requires several hours work and gives far less control over expression.

An important step in moving to using a stable expression system is the optimisation. The lentiviral system discussed here was designed to be customisable and includes many variants not covered here. This includes a range of plasmids with different promoters to control expression levels and plasmids with an internal ribosome entry site (IRES) and fluorescent protein encoded. This provides a method of judging transduction efficiency if the protein of interest cannot be tagged [119].

However, the optimisation screen shown here (*Figure 3.3.b.*) provides a quick and simple way to pinpoint the ideal expression conditions. Other variables could also be tested in this way including cell density or passage number upon induction and a range of common additives, such as valproic acid or sodium butyrate [157].

Construct design is a crucial part of obtaining milligram quantities of any protein. Due to our aim of using a wild-type construct, we were limited in what could be altered. The discovery that C-terminal tags have a substantial effect on expression levels of SMOΔC was key. Through adding a cleavable mVs tag, we were able to convincingly improve expression whilst retaining the wild-type SMOΔC sequence. It is not clear why the fluorescent protein tag improves expression of SMOΔC – possible mechanisms include aiding folding of the entire construct or protection from aggregation, which are thought to be characteristics of some fusion proteins [158]. Overall, through optimising the expression of SMOΔC we have increased the yields nearly 8-fold, making SMOΔC a viable target for structural studies.

3.4.2. Purification of SMOΔC

Purification of membrane proteins is not a trivial undertaking. GPCRs are intrinsically flexible and unstable and there is no universally reliable method for isolating these proteins [130,131]. Previously, our group has successfully purified a thermostabilised SMO construct for crystallisation [59]. I optimised this protocol for

the wild-type construct SMOΔC by maximising the yield of protein that we could obtain from the limited amount expressed. We traditionally used the 1D4 affinity purification system, which, due to high affinity binding of tag to antibody, gives a highly pure elution. However, this high affinity interaction has the inherent feature of being unlikely to dissociate [148]. During purification, this can hamper elution. I decided to harness the highly specific nature of the Rho1D4 system but elute using 3C protease cleavage, thereby taking advantage of the efficiency of the protease.

After working with the wild-type construct, we observed a sub-species and theorised that it was SMOΔC lacking the N-terminus. I developed a tandem affinity purification to circumvent this issue and isolate full-length SMOΔC only. This is an established technique in the field of GPCR purification, resulting in a highly pure sample by removing degradation products as well as other contaminants [159-163]. This method proved to be an effective way to purify full-length SMOΔC and will be essential for future structural studies exploring the interdomain communication across SMO.

3.4.3. SMOΔC in a detergent-free system

Having the ability to study SMOΔC in detergent-free systems is likely to be beneficial for a range of techniques. We have so far made use of this for binding and electron microscopy studies. Amphipol-exchanged SMOΔC behaves in a similar

manner to detergent-solubilised SMOΔC in size exclusion chromatography. It has been useful for bio-layer interferometry (BLI) studies as it removes the necessity for detergent-containing buffer (see Chapter 5). It is also possible that large detergent micelles can occlude protein-protein binding.

The reconstitution of SMOΔC into phospholipid nanodiscs is also highly useful for a range of studies. Nanodiscs are an attractive tool for electron microscopy studies, as they provide an environment with greater similarity to the cellular membrane than that of a detergent micelle. They also remove the need for detergent-containing buffers. An advantage of nanodiscs is the ability to adjust the lipid composition of the artificial bilayer. For a number of GPCRs, including SMO, native membrane lipids are thought to bind and even affect function. One such lipid is cholesterol, which interacts extensively with the TMD of SMOΔC. Not only does cholesterol bind to the outer face of the transmembrane bundle, but also to the hydrophobic groove of the CRD, and a buried ligand binding pocket within the TMD [59-61,94]. Studies have shown that supplementing the nanodisc lipid mixture with cholesterol is sufficient for SMO-dependent G-protein activation [68]. The potential manipulation of the activation state of SMOΔC through nanodisc lipid composition is an interesting and relevant field to explore.

3.4.4. Summary

In conclusion, I have successfully improved the yield and quality of purified monomeric SMOΔC. This was achieved through a combination of techniques to increase protein expression, including adoption of the lentiviral method of stable transfection and construct manipulation. I have also adapted the purification of SMOΔC to increase the population of monomeric SMO and the proportion of this which can be isolated from cells. Finally, I have optimised protocols for amphipol exchange and nanodisc reconstitution of SMOΔC, to enable future studies in detergent-free systems.

Chapter 4

Development of a novel binding assay

4.1. Abstract

The Hedgehog (Hh) signalling pathway is commonly activated in oncogenesis and can act as both driver and potentiator of tumour growth. As such, it is an attractive target for anticancer drug development efforts. There are already several Smoothened (SMO) inhibitors in clinical use against various cancers, however, drug resistance is an emerging obstacle. Here, we build on previous studies of SMO small molecule binding and allosteric behaviour to suggest an alternative target site for anticancer drugs and combat resistance. We describe a novel fluorescence-based assay to specifically identify novel binders for this site. We evaluate the sensitivity and specificity of the new assay and demonstrate its scope by characterising novel SMO binders. We conclude that our assay is ideally suited for use in high-throughput screening and future drug development.

4.2. Introduction

4.2.1. Small molecules bind to and modulate SMO

SMO has multiple distinct small molecule binding sites with the capacity to affect signalling, found in different regions of the protein (*Figure 4.1a*) [59-61,69,93,100,101,103,164-168]. This includes the well-characterised cysteine-rich domain (CRD) hydrophobic groove and the transmembrane domain (TMD) classical GPCR ligand binding site. Recently, several further sterol binding sites have been suggested [94,101,105]. Structural studies reveal a cholesterol binding site buried deep within the TMD, allowing simultaneous occupancy with molecules such as SAG [94]. Computational docking studies have also proposed other cholesterol binding sites on the outer face of the TMD bundle – likely sites where cholesterol from the plasma membrane may associate with SMO [105]. These sites have currently not been shown to modulate Hh pathway activity. In contrast, the CRD and main TMD site have been studied extensively using structural information to design specific SMO mutants. These experiments have suggested an intriguing allosteric relationship; however, the precise molecular mechanism is still not understood. It is known that each binding site is capable of interacting with both agonists and antagonists, yet how they modulate the activation of SMO is less well-established. The first multi-domain structure of SMO revealed a stacked arrangement of the CRD and TMD, bridged by the LD and ECL3. This suggested a route of communication via ECL3, which

interacts with the CRD hydrophobic groove and is capable of transducing an extracellular signal across the plasma membrane via TM helix VI, from which it stems (*Figure 4.1.b*) [59]. This proposed signal transduction mechanism supports the hypothesis that the CRD site is vital for endogenous signalling.

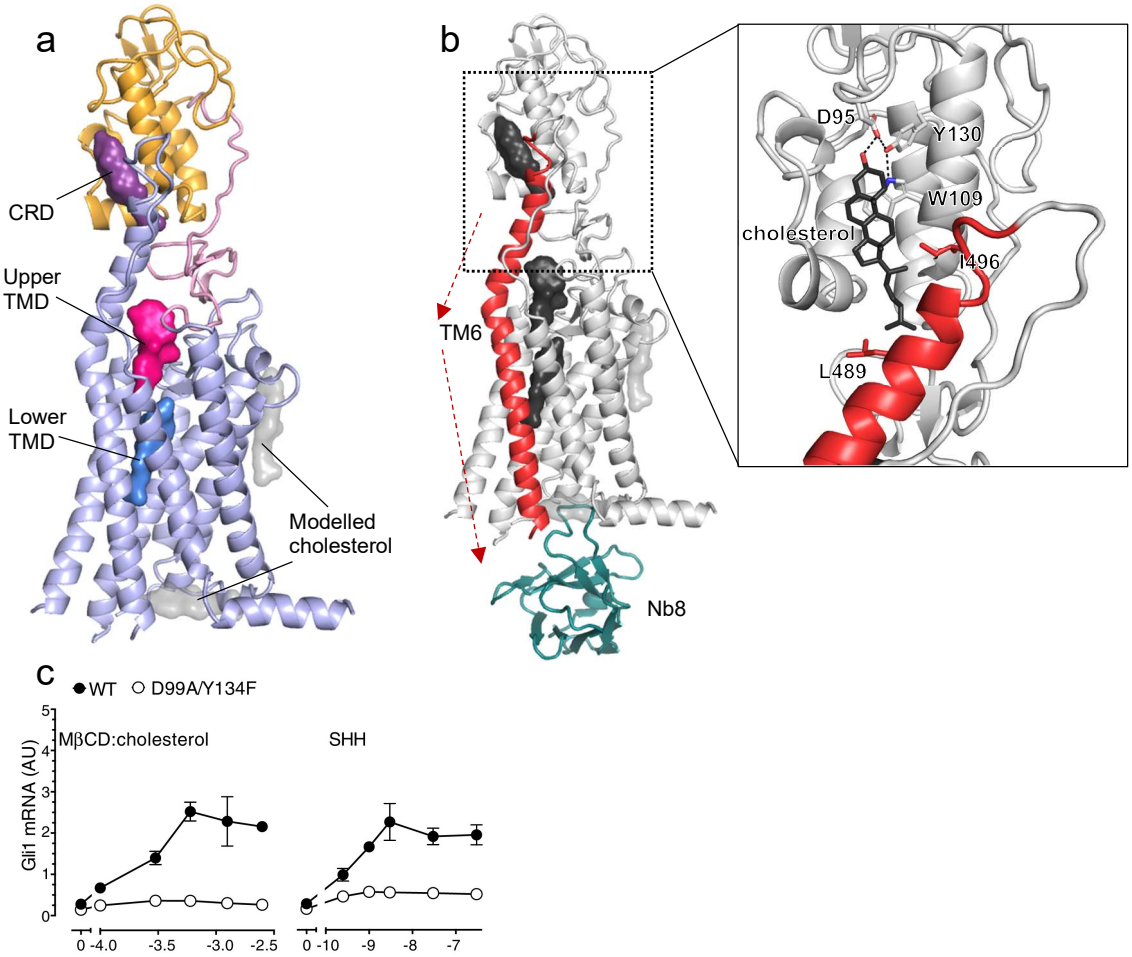


Figure 4.1 Multiple small molecule ligand binding sites modulate SMO activity

a. 3-dimensional structure of SMO with multiple ligand binding sites shown. SMO CRD – orange, SMO LD – pink, SMO TMD – grey-blue. Structurally characterized ligand binding sites and modelled cholesterol interaction sites are labelled.

b. Proposed route of signal transduction across SMO from the CRD site via extracellular loop 3 and transmembrane helix 6 (highlighted in red) to intracellular binding partners e.g.Nb8 (teal). Inset shows CRD cholesterol binding and the contribution of ECL3, residues involved are labelled.

c. GLI1 mRNA assay for Hh pathway activation shows the importance of D99 and Y134 – note these experiments were performed using mSMO, these residues are equivalent to D95 and Y130 in hSMO. Data from Luchetti *et al.*, 2016, reproduced under creative commons license Attribution 4.0 International.

It is predicted that cholesterol is the endogenous SMO agonist, and this has been reinforced by recent proposals that Patched is a cholesterol transporter [56-58,62,63]. The study of CRD binding site mutants has been instrumental in reaching this conclusion. The SMO D95A mutant has been shown to prevent binding of sterols to the CRD hydrophobic groove by disrupting the network of hydrogen bonds that mediates the interaction between SMO and the 3β-hydroxyl [59]. This mutant is not activated by either cholesterol or oxysterols yet does respond to SAG. Crucially it is insensitive to SHH-mediated activation, suggesting this binding site is necessary for endogenous signalling [60,61]. Furthermore, the G111F mutation prevents oxysterol binding yet retains cholesterol binding. This SMO mutant does respond to SHH with a similar magnitude as wild type SMO [93]. Taken together, these studies strongly suggest that cholesterol interaction with the CRD is the endogenous signalling mechanism.

4.2.2. SMO as an anti-cancer drug target

The Hh pathway is an attractive target for cancer drugs due to the prevalence of cancers displaying aberrant Hh signalling. Hh mutations can either initiate oncogenesis or potentiate existing cancers, causing increased tumour aggressiveness and likelihood of metastasis [169,170]. It has been estimated that around 25% of all cancer deaths show misregulated Hh activation, making the pathway a crucial target [171,172].

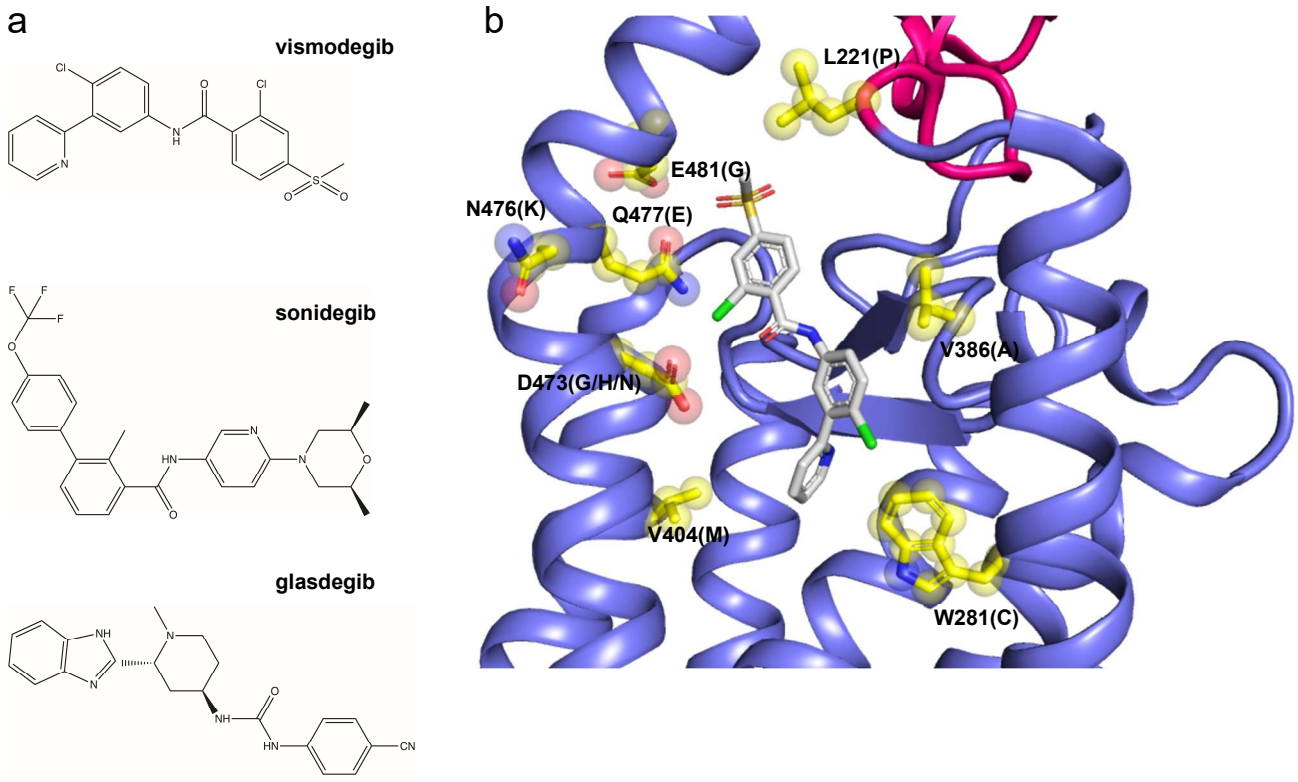


Figure 4.2 Anticancer drugs target SMO and are susceptible to resistance mutations.

a. Chemical structures of the three SMO inhibitors currently approved for clinical use against cancers: vismodegib (Roche), sonidegib (Novartis) and glasdegib (Pfizer).
b. Vismodegib (white) interacts with the SMO TMD binding pocket. Commonly mutated residues observed in cancer patients following vismodegib treatment are shown in yellow (wild type residue pictured, common mutants shown in brackets).

Despite the importance of the CRD site for endogenous signalling, the TMD site has been the target of all current SMO inhibitors developed for clinical use as anticancer drugs. This includes vismodegib (Erivedge, Roche); the first Hh inhibitor approved for clinical use, sonidegib (Odomzo, Novartis); which is also used against advanced basal cell carcinoma and glasdegib (Daurismo, Pfizer); which is used to treat acute myeloid leukaemia (*Figure 4.2.a*) [173-175]. Vismodegib treatment has been shown to result in dramatic tumour regression when used against metastatic medulloblastoma, a rare but aggressive form of brain tumour [176]. It is also used

clinically against locally advanced or metastatic basal cell carcinoma, as these cancers are often inoperable [177]. Unfortunately, these improvements are often transient, leading to reoccurrence of tumours and further metastasis due to acquired vismodegib resistance [171,176,178,179]. Of all vismodegib resistant tumours it was found that around 50% can be attributed to SMO mutations, most of which prevent vismodegib binding but remain responsive to SHH-mediated activation (*Figure 4.2.b*) [178,179]. This includes the common D473H vismodegib resistance mutant, which prevents vismodegib inhibition, but is still inhibited by Patched. In contrast, the constitutively active cancer mutant SMO-M2 (W535L) shows much higher basal activity levels than wild type SMO and is much less sensitive to Patched inhibition [179,180]. Remarkably, it was revealed that both D473H and SMO-M2 are sensitive to inhibition by CRD binding antagonists [97]. However, current SMO CRD binding inhibitors have low affinity and poor pharmacological attributes (including high hydrophobicity), hence an inhibitor with more favourable drug-like qualities would be of great pharmaceutical interest [97]. This provides a promising alternative approach to the development of next-generation SMO inhibitors.

4.2.3. SMO small molecule binding assays

There are currently several published SMO small molecule binding assays. They all rely exclusively on monitoring cyclopamine competition to detect binding of other small molecules [104,166,181,182]. The assay developed first makes use of BODIPY-

cyclopamine, a synthetic form of cyclopamine conjugated to a fluorescent dye, BODIPY FL (boron-dipyrromethene fluorescein). SMO transfected COS-1 cells are incubated with BODIPY-cyclopamine and various concentrations of potential competing small molecules. The cells are analysed for green fluorescence using FACS (fluorescence-activated cell sorting) to quantify BODIPY-cyclopamine binding [166,181]. In a similar manner, an assay using a radioactive form of cyclopamine was later developed [104]. In this study, ^{3}H-cyclopamine is added to SMO-containing membranes alongside various concentrations of potential competing ligands. Membranes are thoroughly washed before being prepared and analysed for radioactivity. Both methods can produce a concentration dependent competition curve. Recently, this type of assay was advanced, by combining the use of BODIPY-cyclopamine with a second fluorophore, nanoluciferase, conjugated to SMO [182]. When nanoluciferase is excited, BRET (bioluminescence resonance energy transfer) can occur – hence transferring energy to a BRET partner within 10 nm. When BODIPY-cyclopamine is bound to SMO it is within this distance so accepts the energy and emits it as a fluorescent signal. As a BRET signal can be measured when BODIPY-cyclopamine is bound to SMO, addition of competing ligands can be used (as above) to cause BODIPY-cyclopamine dissociation and hence a reduction in the signal [182]. One of the main advantages of this is reduced non-specific signal as the measure of BODIPY-cyclopamine fluorescence alone is sensitive to background fluorescence from non-transfected cells. However, all described assays share a common issue. The use of cyclopamine as the tagged ligand is problematic as it is

known to bind SMO at two distinct sites in the CRD and TMD [97,181,183]. Firstly, this introduces complexity as either site may be responsible for any competition measured. Secondly, cyclopamine binds to the TMD with higher affinity hence biasing results towards this site. This may partly explain why most recently discovered SMO inhibitors are TMD binders [170].

4.2.4. Concept of a SMO CRD binding assay

The observation that CRD antagonists can inhibit signalling of vismodegib-resistant and constitutively active SMO mutants is of great relevance to anti-cancer drug design [97]. This raises the possibility of designing SMO inhibitors that will bypass the TMD site, and therefore any resistance mutations, but still prevent Hh signalling. Additionally, considering the hypothesis that the CRD is the endogenous binding site, one may reasonably predict that any mutations conferring resistance to CRD binding drugs might also prevent SHH-mediated activation. This makes the CRD an extremely interesting and potentially 'druggable' target site. Unfortunately, the lack of an available assay for detection of SMO CRD binding is an obstacle to finding new drug candidates. This dearth of an appropriate technique is what we sought to resolve.

4.2.5. Assay design

It is accepted that measuring small molecule binding to SMO directly is difficult. This is why all current binding assays measure binding indirectly, through competition with a ligand that is labelled to make it detectable. We also adopted this strategy, the major alteration in our assay design being the labelled ligand. As discussed, all current assays use cyclopamine which is not optimal for discovering CRD binding ligands. We chose to use a high affinity CRD binding oxysterol (20(S)-yne) as the base for our detectable ligand. We then needed a method to measure the amount of ligand bound to SMO in the presence of various concentrations of competing ligands. We decided to use Förster resonance energy transfer (FRET) to characterise this interaction. FRET is the transfer of energy from one fluorescent molecule to another and its efficiency is crucially dependent on proximity of the FRET pair [184,185]. For this assay, we chose to use a FRET-based technique called HTRF (homogeneous time resolved fluorescence) and associated reagents [186]. HTRF makes use of rare-earth lanthanides as donor fluorophores; they have longer emission half-lives than other fluorophores, making their signal stable over time and therefore ideally suited to high-throughput screening [187,188].

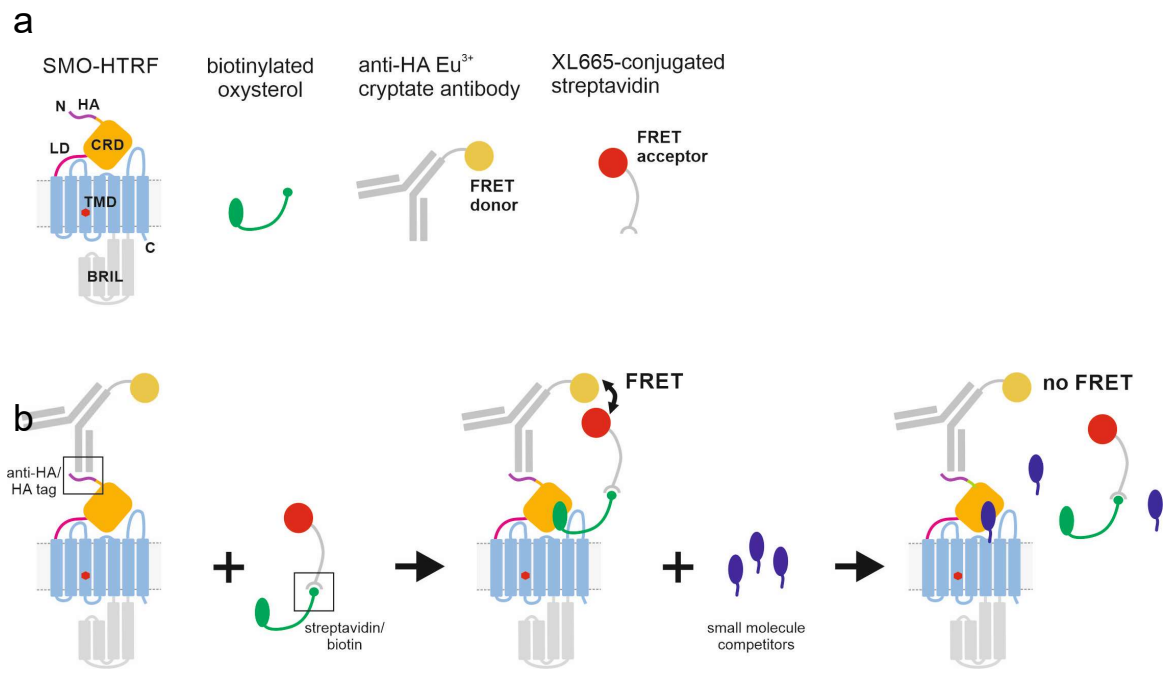


Figure 4.3 SMO CRD binding assay was developed using customisable HTRF reagents.

a. Components required for the SMO CRD binding assay: SMO with an N-terminal HA tag, biotinylated oxysterol (green), anti-HA mAb conjugated FRET donor (yellow), streptavidin linked FRET acceptor (red)
b. To perform the assay, all components are combined to form a complex which brings the donor and acceptor into proximity of each other and allows energy transfer. Potential competitors of the biotinylated oxysterol are added, and any displacement is measured as a decrease in FRET emission.

Cisbio (Bagnols-Sur-Cèze, France) produce a number of HTRF partners, each conjugated to an antibody or other specific recognition sequence. We decided to use Europium (Eu^{3+}) cryptate as our FRET donor and XL665 as the FRET acceptor. To adapt this to our system we chose Eu^{3+} cryptate conjugated to an anti-HA monoclonal antibody which recognises a HA (hemagglutinin) tag engineered onto the N-terminus of SMO (*Figure 4.3.a*). For simplicity, we performed initial experiments using a thermostabilised form of SMO ($SMO\Delta C_{stable}$) which incorporates a BRIL domain and V329F mutation. Our collaborators at Washington University in St. Louis (Douglas Covey laboratory) synthesised a biotinylated form

of 20(S)-yne for this assay, so we chose XL665 conjugated to a streptavidin molecule (*Figure 4.3.a*). When combined these four components form a complex defined by two very high affinity interactions – HA epitope antibody recognition and biotin-streptavidin. However, the interaction between SMO and the biotinylated ligand is weaker and hence may be broken by competing molecules thereby distancing the FRET pair and reducing the signal (*Figure 4.3.b*). Therefore, we can monitor complex dissociation due to competitor binding via the FRET signal.

4.3. Results

For all initial experiments SMOΔC$_{stable}$-HA solubilised in DDM/CHS was used. This construct is a modified version of hSMO, consisting of UniProt ID Q99835; residues 32-555 with an N-terminal HA tag. "Stable" refers to the V329F mutation and BRIL domain inserted into ICL3. All other mutations utilised in this chapter are labelled and are shown in *Figure 4.4*.

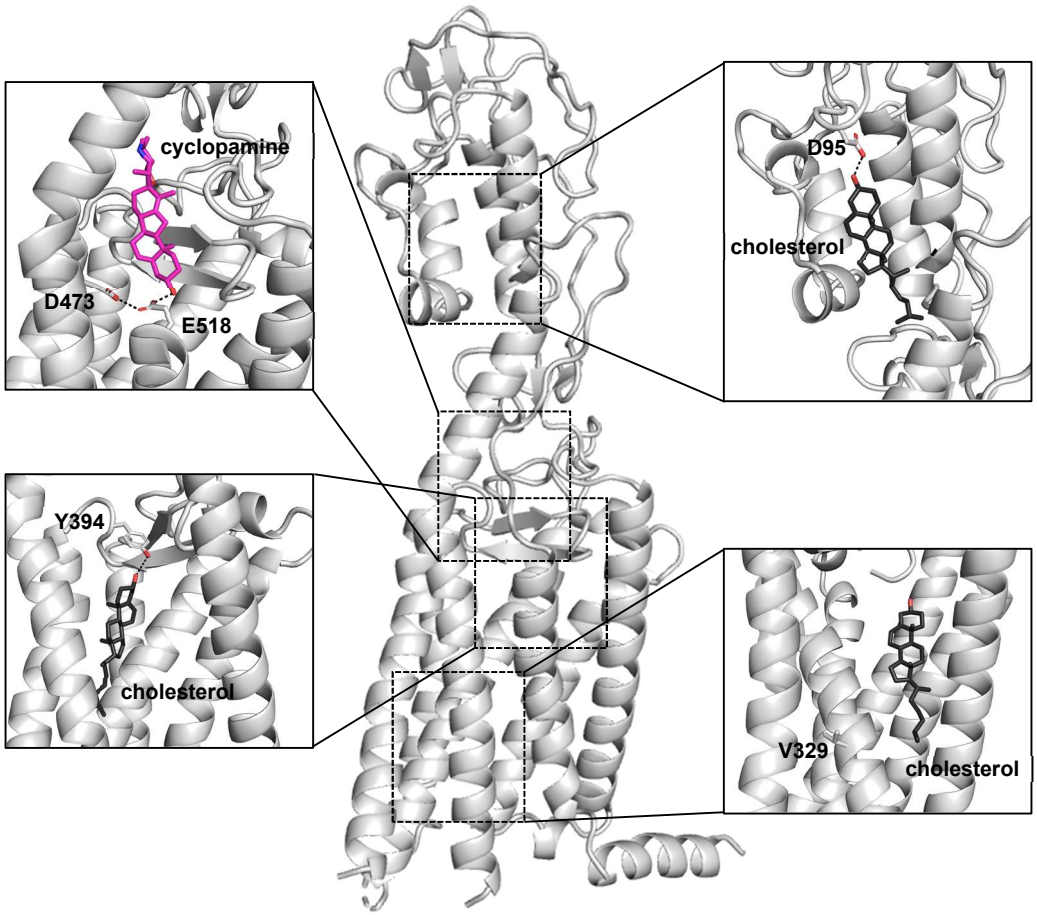


Figure 4.4: Structure guided SMO mutations for ligand binding experiments.
SMOΔC is shown as a cartoon. All mutations employed in our HTRF assay are depicted in insets, alongside any relevant ligands. Cholesterol is shown in black, cyclopamine in pink.

4.3.1. Assay validation

As with any new method, our binding assay must first be validated. To be able to use such a technique to accurately identify novel compounds it must have the required suitability and sensitivity. Whilst statistical parameters such as signal-to-background (S/B) and signal-to-noise (S/N) are often used, they are imperfect. The S/B is simply the ratio between the mean of the positive and negative controls, with no consideration for the variation in each dataset. S/N additionally considers deviation from the mean, but just for the negative dataset. We chose to use the more relevant Z-factor [189]. This measure calculates the separation between signal (s) and control (c) datasets as the ratio of total data variability ($3\sigma_s + 3\sigma_c$) over the separation of the two means ($\mu_s - \mu_c$) (*Figure 4.5.a*). The Z-factor is calculated as follows:

$$Z = 1 - \frac{3\sigma_s + 3\sigma_c}{|\mu_s - \mu_c|} \quad (4.1)$$

A Z-factor of 1 represents the ideal assay and is only achievable with no variation within either signal or control datasets. Z-factors between 1 and 0.5 are classed as an excellent assay, whilst scores from 0.5 to 0 are classed as marginal. If an assay achieves a Z-factor of less than 0 it means the variability of the data is greater than the separation of positive and negative results (signal and control) and hence the assay is unsuitable [189].

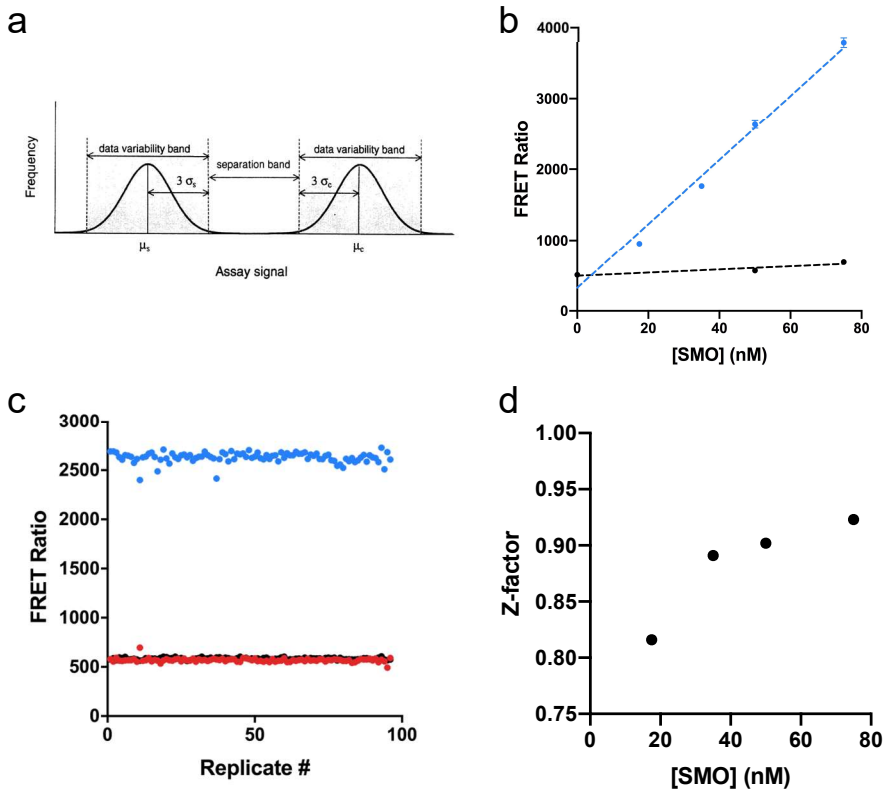


Figure 4.5 Z-factor calculations classify assay quality

a. Z-factor calculations assess the separation between positive and negative signals, also factoring in the expected variance in each data set. Reproduced from Zhang et al., 1999 under 'Gratis Reuse'.
b. Concentration of SMOΔC$_{stable}$-HA (blue) has a noticeable effect on FRET ratio. Concentration of SMOΔC$_{stable}$-HA^{D95A} (black) does not.
c. Experimental data for 96 replicates for SMOΔC$_{stable}$-HA (blue), SMOΔC$_{stable}$-HA^{D95A} (red) and no SMO (black), used to calculate the Z-factor.
d. Z-factor scores vary with SMOΔC$_{stable}$-HA concentration.

An important consideration for the calculation of the Z-factor is the signal and control datasets. These should be the upper and lower limits of the assay output. For the SMO FRET assay, the upper limit (signal dataset) is observed when the FRET components are combined in the absence of any competitor. For the control dataset (lower limit) we tested two systems, one with no SMO protein added and another with SMOΔC$_{stable}^{D95A}$-HA. The D95A mutation prevents oxysterol binding

to the CRD site and is shown to mirror the behaviour of the no SMO protein system [59]. We present the effect of concentration of SMOΔC$_{stable}$-HA and SMOΔC$_{stable}^{D95A}$-HA on the FRET ratio. Our results show a strong correlation between increasing concentration of SMOΔC$_{stable}$-HA and FRET signal, yet no such effect from increasing concentration of SMOΔC$_{stable}^{D95A}$-HA (*Figure 4.5.b*). This demonstrates the specific nature of our assay and highlights that we are exclusively detecting ligand bound at the CRD site. We carried out 96 repeats of each condition in a 384-well plate and used Equation 4.1 to calculate the Z-factor (*Figure 4.5.c*). We determined the Z-factor at a range of SMOΔC$_{stable}$-HA concentrations between 17.5 and 75 nM and found them all to be between 0.5 and 1, suggesting an excellent assay (*Figure 4.5.d*)[189].

4.3.2. Optimisation – we can reduce reagent use without sacrificing assay quality

A key step in the development of a high throughput screen is maximising the number of conditions that can be screened per plate or set quantity of reagents, whilst maintaining assay quality. Initially we examined the concentration of oxysterol-biotin required for maximum response. We titrated this reagent against constant SMO, XL665-streptavidin and anti-HA concentrations. We tested two different XL665-streptavidin concentrations as the high affinity binding partner of the oxysterol-biotin. Upon initial examination of the data, we noted minimal

response at low concentrations of oxysterol-biotin, until a threshold is reached where the response peaks and subsequently begins to decline at higher concentrations (*Figure 4.6.a*). We were able to identify this threshold as when the concentration of oxysterol-biotin is at an 8x molar excess to that of XL665-streptavidin. We observed a higher FRET signal when 62.5 nM XL665-streptavidin was used, however this was accompanied by higher background signal. To achieve the maximum response with this high concentration of XL665-streptavidin, a higher concentration of oxysterol-biotin must also be used. However, if the concentration of SMOΔC$_{stable}$-HA is not increased by the same factor, the ratio of oxysterol-biotin:SMOΔC$_{stable}$-HA will increase. We hypothesised that if this ratio is too high, free oxysterol-biotin (not bound to SMOΔC$_{stable}$-HA) will bind to XL665-streptavidin, distant from SMOΔC$_{stable}$-HA and anti-HA Eu^{3+} cryptate, effectively removing a proportion of XL665-streptavidin from the reaction and hence decreasing the possible FRET response. This can be seen as a decrease in signal after the threshold maximum and is described as the hook effect, a common problem in immunoassays (*Figure 4.6.a*) [190-192]. It is also possible that excess oxysterol-biotin may decrease the sensitivity to competing ligands.

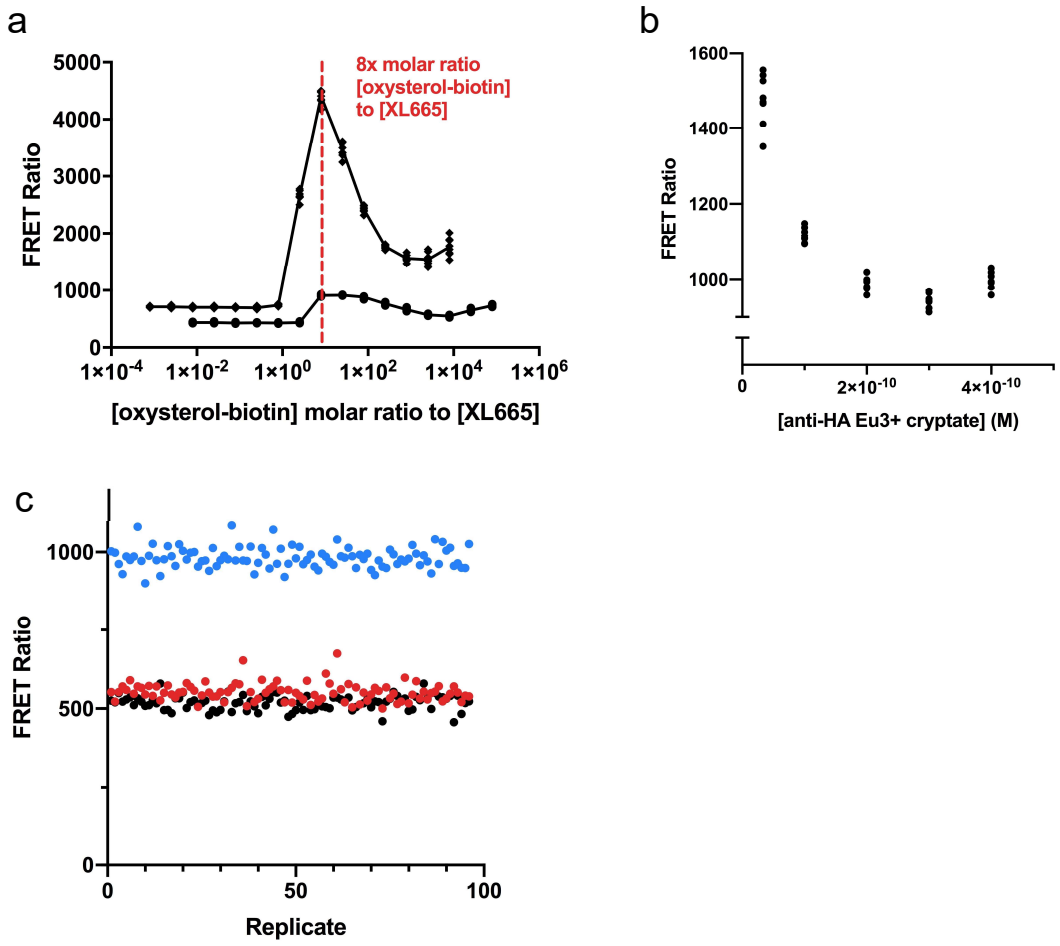


Figure 4.6 Reduction in FRET reagents affects the FRET ratio

a. Oxysterol-biotin titration, alongside XL665-streptavidin at 6.25 nM (circles) and 62.5 nM (diamonds). Concentration of oxysterol-biotin presented as a molar ratio to XL665.
b. Titration of anti-HA Eu^{3+} cryptate antibody and its effect on the FRET ratio.
c. Z-factor assessment when reducing multiple components simultaneously. Experimental data for 96 replicates for $SMO\Delta C_{stable}$-HA (blue), $SMO\Delta C_{stable}$-HA^{D95A} (red) and no SMO (black).

In addition, we tested whether the concentration of anti-HA Eu^{3+} cryptate antibody could be reduced without decreasing assay quality. We titrated this reagent against constant $SMO\Delta C_{stable}$-HA, oxysterol-biotin and XL665-streptavidin. We observed little difference in FRET ratio when the concentration of anti-HA Eu3+ cryptate antibody is 0.1 nM or higher (*Figure 4.6.b).* Below this point we noted increased

FRET ratios, due to disproportionate decreases in emission at 665 nm relative to 620 nm. We hypothesise that this is due to the very low concentrations of antibody. The signal at 665 nm is the emission from XL665-streptavidin, reliant upon its proximity to anti-HA Eu^{3+} cryptate and hence inherently lower than the signal at 620 nm. Due to the low concentration, the photons emitted by excited donor fluorophore (anti-HA Eu^{3+} cryptate) will also be very low. We postulate that FRET is still occurring, but the low level of photons exciting XL665 leads to a very low number of photons emitted, hence approaching the minimum limit of the detector. This creates an artificial minimum plateau in the 665 nm signal, whilst the 620 nm signal continues to decrease. This causes the FRET ratio to increase, for reasons unrelated to the biological system.

We combined our findings from the above experiments and recalculated the Z-factor. For the recalculation we used 50 nM SMOΔC_{stable}-HA and a 1:1 molar ratio of oxysterol-biotin. We kept an 8-fold molar excess of oxysterol-biotin relative to XL665-streptavidin, as this gave the maximum FRET ratio (*Figure 4.6.a*). We also reduced the anti-HA Eu^{3+} cryptate antibody to 0.1 nM, as this had little effect upon the FRET ratio (*Figure 4.6.b*). We performed 96 replicates of this condition and of each control using SMOΔC_{stable}^{D95A}-HA and no SMO. This dataset gave a Z-factor of 0.568 (*Figure 4.6.c*). This result is classified as an "excellent" assay as it is above 0.5, however it is markedly lower than the previous calculated Z-factor for this SMO concentration, 0.902 [189]. This appears to be due to the lower concentrations of

oxysterol-biotin and XL665-streptavidin, as the reduction of anti-HA Eu^{3+} cryptate to 0.1 nM has little effect upon the FRET ratio (*Figure 4.6.b*). We concluded that the Z-factor with minimal reagent use still represented a robust assay and these conditions are therefore sufficient for initial compound screens. When titration experiments are required it is possible to use higher reagent concentrations to increase the data range and simplify analysis. For the dose-response curves presented in this chapter, we used 50 nM SMO, 50 nM oxysterol-biotin, 6.25 nM XL665-streptavidin and 0.4 nM anti-HA Eu^{3+} cryptate.

4.3.3. Proof of principle – we can identify competition for the CRD site

After determining that our assay was robust and having identified the optimal conditions for compound screening and characterisation, we needed to show that small molecules could out-compete oxysterol-biotin for the SMO CRD site. We chose to use a set of pre-characterised compounds, including known SMO CRD binders and negative controls. We dispensed an 11-point serial dilution and DMSO-only reference for each potential competitor, before adding the FRET reagents from a pre-made master-mix. Each condition was divided into four replicates and transferred to a 384-well assay plate. The FRET ratio for each well was measured and the quadruplicate mean was used for further analysis. Data were normalised and fit Equation 4.2 using GraphPad Prism (for full details, see Chapter 2.5.1). Data

are presented as concentration curves for each of SMOΔC$_{stable}$-HA and SMOΔC-HA (*Figure 4.7.a,b*).

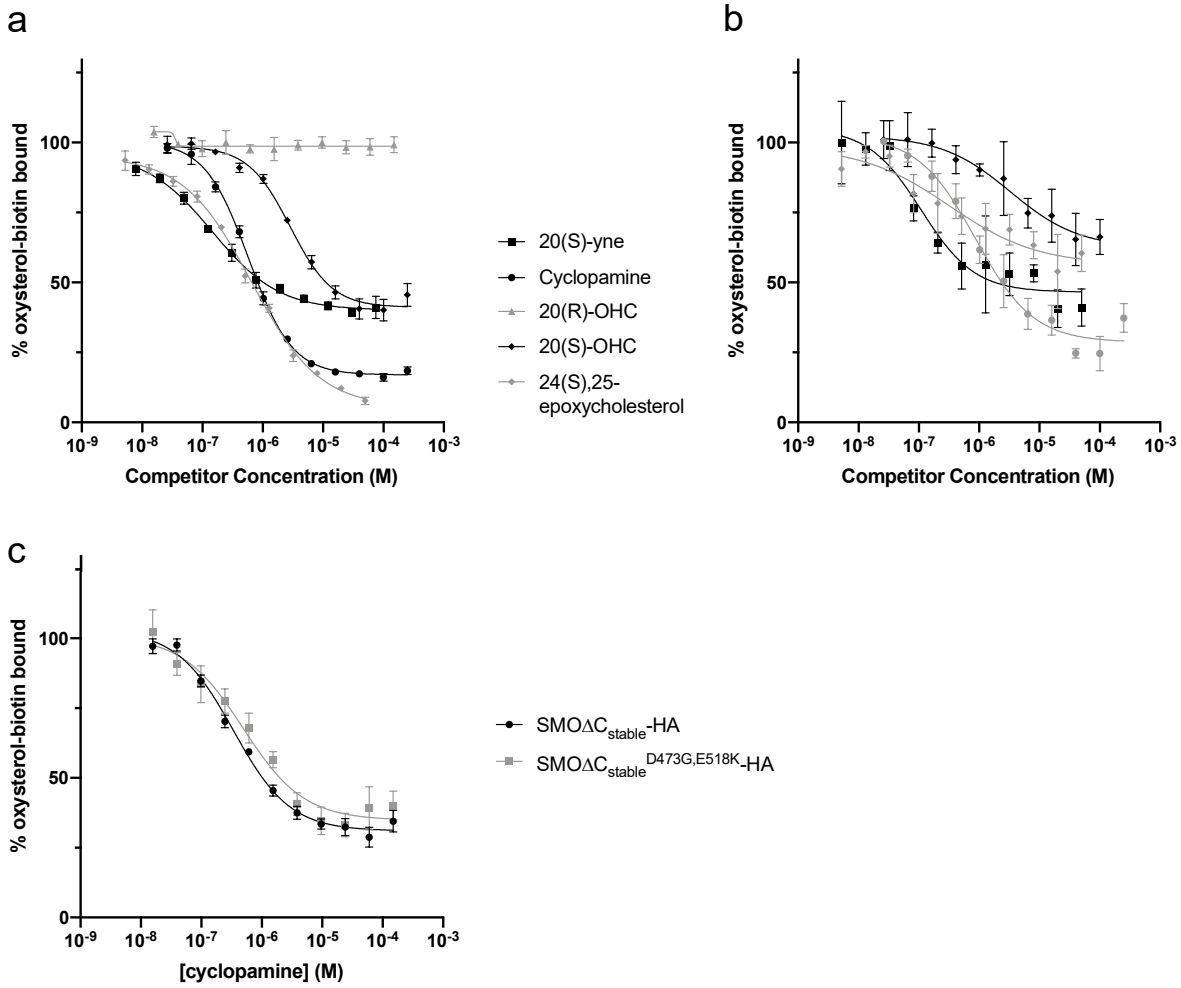


Figure 4.7 Known CRD-binding ligands can be assessed using our novel FRET assay

a. Concentration curves for known SMO ligands binding to SMOΔC$_{stable}$-HA. Legend shown for clarity; 20(S)-yne (black squares), cyclopamine (black circles), 20(R)-OHC (grey triangles), 20(S)-OHC (black diamonds), 24(S),25-epoxycholesterol (grey diamonds).
b. Concentration curves for known SMO ligands binding to SMOΔC-HA. Data represented as above.
c. Dose-response curve for cyclopamine binding to SMOΔC$_{stable}$-HA (black circles) and SMOΔC$_{stable}^{D473G,E518K}$-HA (grey squares).

Competitor concentration dependent dissociation of oxysterol-biotin is clearly observed for the known SMO CRD binders. Increasing concentrations of known

CRD binders, such as 20(S)-OHC, show the expected outcome for a high affinity CRD binding molecule; a sigmoidal dose-dependent response. However, the minimum plateau varies between 10-50% bound oxysterol-biotin. This is likely due to relative binding affinities of each ligand. Interestingly, it is 20(S)-OHC and 20(S)-yne which only displace around 50% of the oxysterol-biotin. This observation is likely due to a non-biased equilibrium being reached between SMO bound to the oxysterol-biotin, and these competing oxysterols, which are near-identical. 20(R)-OHC is an enantiomer of 20(S)-OHC, which has not been shown to bind SMO [99]. We observe no binding in our assay, and hence use 20(R)-OHC as a relevant negative control.

$$\%\ oxysterol\text{-}biotin\ bound = min + [competitor]^H \left(\frac{max - min}{[competitor]^H + EC50^H}\right) \quad (4.2)$$

Fitting the data to Equation 4.2 enabled us to calculate EC50 values (the concentration at which half-maximal response is observed) using the maximum and minimum values for each data set (*max* and *min*, respectively), molar competitor concentration ([*competitor*]) and the hill slope (*H*). If we compare these data to published EC50 values we see a range of correlations (*Figure 4.7.b*, Table 4.1). For the positive controls, 20(S)-yne and 20(S)-OHC, the EC50 values calculated in our FRET assay are within the same order of magnitude as previously published values.

Table 4.1: Comparison of EC50 values

Competitor	FRET assay EC50 SMOΔC$_{stable}$-HA (μM)	FRET assay EC50 SMOΔC-HA (μM)	EC50 published (μM)
20(S)-yne	0.118 (0.084 – 0.165)	0.107 (0.073 – 0.156)	0.309 [99]
20(S)-OHC	3.10 (1.22 – 7.85)	2.73 (0.056 – 13.2)	3.00 [99]
20(R)-OHC	-	-	-
Cyclopamine	0.459 (0.340 – 0.620)	1.01 (0.853 – 1.20)	0.300 [181]
24(S),25-epoxycholesterol	1.00 (0.392 – 2.57)	0.348 (0.264 – 0.459)	-
LKM27	0.178 (0.103 – 0.307)	0.409 (0.264 – 0.634)	-

Calculated and published EC50 values for a range of SMO ligands. FRET assay EC50 values shown are the mean of three independent replicates, where each data point is measured in quadruplicate. Range shown in parentheses is the 95% confidence interval of the average EC50.

We calculated the HTRF EC50 for cyclopamine as similar to that previously published, despite the latter incorporating cyclopamine binding to the SMO TMD (Table 4.1). We explored this further by testing a SMO mutant (D473G, E518K) which cannot bind to cyclopamine via the TMD [193] (*Figure 4.7.c*). The EC50 value calculated for SMOΔC$_{stable}^{D473G,E518K}$-HA was 0.378 μM, in comparison to 0.459 μM for SMOΔC$_{stable}$-HA. The similarity in these EC50 values suggests they are not affected by cyclopamine binding to the SMO TMD. This demonstrates that the assay exclusively measures CRD binding.

4.3.4. Oxysterols interact with the CRD

Whilst the most obvious application for our SMO binding assay is to discover previously unidentified SMO CRD binders, it is also useful in providing additional information about known ligands. There are several ligands known to bind SMO, but without conclusive evidence showing where they interact with the receptor. For example, 24(S),25-epoxycholesterol has been suggested to bind SMO at the CRD, TMD and an additional (as yet unidentified) site [69,101]. We are now able to add our data concerning 24(S),25-epoxycholesterol binding to SMO. Using our FRET assay, we show CRD binding of 24(S),25-epoxycholesterol to SMOΔC$_{stable}$-HA, with an EC50 of 1.00 μM (*Table 4.1*). For comparison, 20(S)-OHC, a known CRD binder, has an EC50 of 3.10 μM. These data support the hypothesis that 24(S),25-epoxycholesterol binds to the SMO CRD site, however, due to the specificity of our assay it does not disprove additional interaction sites.

We also characterised a novel ligand developed by collaborators (Doug Covey, Washington University) known as LKM27 (or 6-azi-20(S)-yne)(*Figure 4.8.a*). Mutational analysis of Hh pathway activation performed by our collaborators (Rajat Rohatgi, Stanford University) showed that LKM27 Hh agonism was dependent on mSMO residue Y134 (hSMO equivalent Y130) (*Figure 4.8.b*). This residue has previously been shown to be essential for oxysterol binding to the CRD of SMO and so this data strongly implies LKM27 is a CRD binder [97]. We observed

competition for the CRD site with an EC50 of under 200 nM (Table 4.1, *Figure 4.8.c*). This binding efficiency is comparable with 20(S)-yne and hence clearly supports the hypothesis that LKM27 interacts with the CRD site of SMO.

4.3.5. Allosteric ligands affect CRD site conformation

Alongside testing novel compounds and optimising our assay for next-generation cancer drug development, we aimed to study SMO signal transduction. Currently, the molecular mechanism of SMO activation is unknown. This includes transduction of the extracellular Hh signal across the plasma membrane and intra-molecular communication between the SMO small molecule binding sites. Previously, synergy has been observed between agonists binding the CRD and upper TMD sites [99]. Recent structural data has highlighted an intriguing lower TMD cholesterol binding site [94]. Together with collaborators (Rajat Rohatgi, Stanford University), we sought to explore how cholesterol binding at this site may cooperate with agonist binding at other sites and ultimately contribute to Hh signalling.

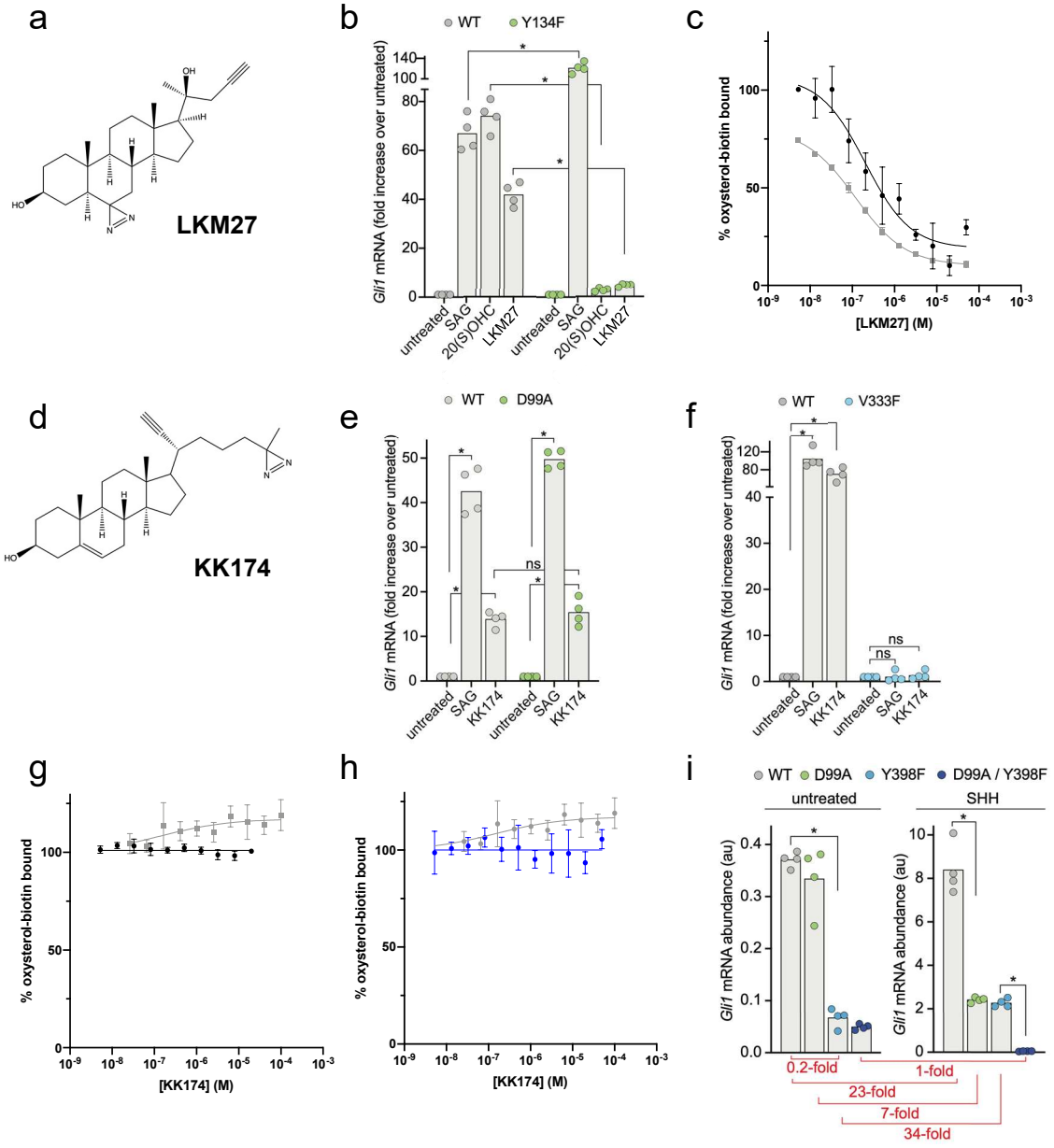


Figure 4.8 Novel SMO ligands can be characterised using the FRET assay.

a. Chemical structure of novel ligand LKM27.
b. Hh activation assays show agonism of LKM27 is dependent on Y134 (hSMO Y130).
c. Dose-response curve for LKM27; SMOΔC$_{stable}$-HA (black circles) and SMOΔC-HA (grey squares).
d. Chemical structure of novel ligand KK174.
e. Hh activation assays show KK174 agonism is not dependent upon the CRD site.
f. Hh activation assays show KK174 agonism is abrogated by V333F (hSMO V329F).
g. Dose-response curve for KK174; SMOΔC$_{stable}$-HA (black circles) and SMOΔC-HA (grey squares).
h. Dose-response curve for KK174; SMOΔC-HA (grey) and SMOΔC^{Y394F}-HA (blue).
i. Comparison of basal and SHH-activated Hh pathway signalling, measured via GLI1 mRNA abundance.

Hh activation assays shown here were performed by Maia Kinnebrew and Raj Rohatgi.

To explore the allosteric effects on CRD binding, I chose to use the SMOΔC-HA construct, as the nearest to wild-type SMO that we can produce that is compatible with the assay. I also used a novel cholesterol-like agonist KK174, synthesised by collaborators at Washington University (Douglas Covey) (*Figure 4.8.d*). Mutational analysis and Hh pathway activation assays, performed by our collaborators (Rajat Rohatgi, Stanford University) suggested that KK174 interacts exclusively with the deep TMD sterol binding site (*Figure 4.8.e,f*). We initially compared the KK174 interaction with SMOΔC$_{stable}$-HA to SMOΔC-HA and found only the wildtype to be responsive (*Figure 4.8.g*). We observed a unique increase in oxysterol-biotin binding to SMOΔC-HA. This strongly supports the hypothesis that KK174 does not bind to the CRD site, though is able to modulate its binding properties via allosteric communication. We postulated that in the SMOΔC$_{stable}$-HA construct the V329F mutation either prevents binding, or the stabilised TMD is incapable of undergoing the required conformational change for this allosteric communication to occur. The data from Hh activation assays supports this as KK174 is unable to activate the V333F mSMO mutant (equivalent to hSMO V329F (*Figure 4.8.f*)). Our collaborators at Stanford also identified a point mutation Y398F (hSMO Y394F) which abrogates this interaction and hence KK174-mediated SMO activation. By comparing the wild-type SMO and the Y394F mutant, we can identify allosteric communication with the CRD. Addition of KK174 causes an increase in oxysterol-biotin binding to SMOΔC-HA, but not SMOΔC^{Y394F}-HA (*Figure 4.8.h*). The dependency of KK174 binding on Y394 suggests an allosteric effect on the CRD. Interestingly, KK174

causes an increase in the population of SMO molecules that bind oxysterol-biotin at the CRD site. This is indicative of positive co-operativity between these two spatially separate sites.

4.4. Discussion

4.4.1. Assay sensitivity is of commercial standard

One of the most important factors in developing a new technique is ensuring that the results will be useful. When considering high-throughput assays like the SMO CRD FRET assay, we needed to be confident that a positive and negative result can be easily distinguished. As we designed this assay to produce dose response curves, we also required the range within the assay to be wide enough to measure this. To assess the quality of our assay we calculated the Z-factor for a range of workable SMOΔC$_{stable}$-HA concentrations between 17.5 and 75 nM. We found all assay conditions to give a Z-factor score greater than 0.8, which is classed as an 'excellent assay' (*Figure 4.5.d*).

We assessed the Z-factor relative to published examples of two comparable commercially available assays currently in use. The Z'-LYTE™ Assay (Invitrogen) is a popular tool using FRET to study kinase activity [194]. FRET is disrupted upon phosphorylation of the Z'-LYTE™ peptide, allowing efficient study of kinase inhibitors for possible pharmacological application. A recent study used Z'-LYTE™ to explore the effect of molecular chaperones on the activity of AMP-activated kinase (AMPK) [195,196]. The optimised conditions used for this system gave a Z-factor of 0.66. AlphaScreen™ (PerkinElmer) is a popular customisable system for carrying

out high throughput protein interaction studies [197]. Much like the SMO CRD FRET assay described here, the FRET donor and acceptor beads are conjugated to either streptavidin or antibodies, making the system highly adaptable. For example, it has been used to identify inhibitors of Heat shock protein 90 (Hsp90), a key player in the folding and maturation of oncogenic proteins and hence an attractive anticancer target [198]. The study describes a large-scale compound screen with conditions optimised to give a Z-factor of 0.77.

The SMO CRD FRET assay Z-factor of >0.8 compares favourably with these published examples and indicates suitability for use in high-throughput screening. We also observed an improved Z-factor with increased SMOΔC$_{stable}$-HA concentration, likely due to the higher signal from the positive control. This provides scope to adapt the assay to prioritise either measurement range or minimal reagent use. This will be of use across the different stages of large-scale lead-like compound screening for drug development.

4.4.2. SMO CRD FRET assay shows robust specificity

When developing the SMO CRD FRET assay, it was essential to determine the degree of specificity displayed. We took several approaches to ensure our assay would only detect changes to CRD binding. Initially, we needed to determine the

suitability of the FRET-dependent CRD-binding compound (oxysterol-biotin). We made use of the synthetic oxysterol 20(S)-yne, as it combines the high affinity CRD binding ability of 20(S)-OHC and a terminal alkyne group for "click chemistry' reactions, making it ideally suited for conjugation to biotin. As SMO contains multiple sterol binding sites, it was crucial to determine that the oxysterol-biotin exclusively targets the CRD site of interest. Our studies of SMOΔC$_{stable}^{D95A}$-HA show complete abrogation of SMO-dependent FRET as a result of this single point mutation, which has been shown to prevent sterol binding to the SMO CRD through activation and structural studies [59-61].

To further explore this, we chose to study the plant alkaloid and potent SMO antagonist, cyclopamine. Cyclopamine is unusual as it can bind to both the SMO CRD and TMD sites simultaneously. Despite this, our initial studies using cyclopamine as the competitor ligand were successful and no indication of any TMD binding effect was apparent. However, to guarantee the robustness of our assay we decided to employ a SMO double mutant (SMOΔC$_{stable}^{D473G,E518K}$-HA) which cannot bind cyclopamine in the TMD. We performed parallel cyclopamine dose-response experiments using SMOΔC$_{stable}$-HA and SMOΔC$_{stable}^{D473G,E518K}$-HA (*Figure 4.7.c*). Resulting EC50 values were within an order of magnitude of each other, supporting the specificity of the assay. To ensure the V329F mutation in SMOΔC$_{stable}$-HA was not preventing an allosteric effect of cyclopamine TMD binding upon the CRD, I repeated the same experiment using SMOΔC-HA.

Interestingly, the EC50 values for cyclopamine binding to SMOΔC-HA were also similar to those observed for SMOΔC$_{stable}$-HA. This suggests cyclopamine binding to the SMO TMD does not have a measurable effect on the CRD site.

It has previously been shown that cyclopamine can non-competitively inhibit oxysterol-mediated SMO activation [99]. Our data suggests this inhibition is not due to a conformational change in the CRD site and hence a lowered binding affinity for the oxysterol, as previously thought. If this was the case, when cyclopamine is titrated against SMOΔC-HA (the sole construct with the ability to bind cyclopamine at the TMD and the capacity for inter-domain conformational changes) the decreased affinity of the CRD site for oxysterol-biotin would be reflected in an apparent lower EC50 for cyclopamine competition for the CRD. Our data shows no such effect, hence implying that cyclopamine binding to the TMD has no structural effect upon the CRD binding site. We postulate that cyclopamine binding to the TMD does not affect the conformation of the CRD site, but rather prevents the transmission of an activating conformational change from the CRD across the TMD. This model could describe how SMO antagonists, including anticancer drug vismodegib, are capable of overriding SHH-mediated SMO activation.

4.4.3. Allosteric effects are observed in the wild-type SMO

When we began developing the SMO CRD FRET assay, we used the SMOΔC$_{stable}$-HA construct, which is based on a thermostabilised form of SMO containing the V329F mutation and a BRIL domain. This construct was ideal for initial validation and optimisation due to improved yields over the wild-type protein, making it easier and cheaper to produce. Previous studies have also demonstrated that despite its modifications, CRD ligand binding is unaffected [59,60,153]. However, a more ambitious aim for this assay was to use the SMO wild-type construct, the optimisation of which is described in Chapter 3. We predicted the wild-type construct might allow us to observe allosteric effects upon ligand binding to SMO, as conformational changes would not be hindered by non-physiological stabilisation. To explore this, we used SMOΔC$_{stable}$-HA as a pseudo-negative control, as the TMD is locked in an inactive state and so unable to change conformation. Our studies of SMOΔC-HA produced similar EC50 values to SMOΔC$_{stable}$-HA for ligands characterised to exclusively bind to the CRD.

However, the novel compound KK174, predicted to bind at the lower TMD sterol binding site, showed distinct binding effects towards the two constructs. SMOΔC$_{stable}$-HA appeared insensitive to KK174, whilst SMOΔC-HA showed a positive correlation between oxysterol binding and KK174 concentration. We

hypothesised that this was due to KK174 binding to the lower TMD site (as predicted) and causing a conformational change in the CRD site. This effect would not be observed in the SMOΔC$_{stable}$-HA construct due to its stabilisation. The allosteric effect of KK174 binding to this site was confirmed using the SMOΔC^{Y394F}-HA construct, which specifically abolishes binding to sterols in the lower TMD site. This mutant showed insensitivity to KK174 which suggests it disrupts binding to SMO (*Figure 4.8.h*).

The observation that sterol binding in the lower TMD site can modulate CRD ligand binding has important implications for the mechanism by which SMO transduces the Hh signal. Data produced by collaborators (Rajat Rohatgi, Stanford University) suggests this sterol binding site is crucial for Hh signalling (*Figure 4.8.i*). They show SHH-mediated pathway activation is reduced by approximately 75% when this site is disrupted, an effect also observed for the CRD site. Interestingly, the two sites appear to contribute differently to basal SMO activity. Mutations blocking sterol binding to the CRD have a negligible effect on basal Hh signalling, however the disruption of sterol binding in the lower TMD site causes a 5-fold reduction. This implies that the lower TMD site is essential for basal Hh signalling, but also 'primes' SMO for full activation.

4.4.4. Future directions

As demonstrated here, the SMO CRD FRET assay is ideally suited for use in high throughput screening to identify novel binders. Our aim when designing this assay was to provide a tool to detect CRD-specific SMO inhibitors and with possible anticancer drug development potential. Due to the highly sensitive and specific nature of our assay, we are confident that it can be used to identify and characterise such molecules. Our decision to use HTRF components means that the assay signal will be stable over an extended period of time, and the excellent Z-factor score also gives scope for assay miniaturisation. Together, these features make our assay optimal for high-throughput screening. To that end, we are currently involved in an industrial collaboration screening over 250,000 lead-like compounds using this assay.

A key result from the experiments presented here is the ability to characterise allosteric effects using this assay. We plan to continue this study to provide valuable insights into the mechanism of SMO signal transduction. We will combine this novel assay with the SMO-interacting proteins described in Chapter 5 to assess how they affect the conformation of SMO. We hypothesise that molecules such as Nb8 and G-proteins (which bind the active state of SMO) will have a similar effect as KK174 and stabilise SMO in a conformation more suited to oxysterol binding.

4.4.5. Summary

In this chapter I have described the conceptualisation, development and application of a novel SMO CRD binding assay. We have shown the assay to be robust, specific and sensitive. I have demonstrated the assay's potential to identify novel CRD binders and characterise their binding efficacy. Crucially I have also observed allosteric co-operative binding effects, which provide evidence for a two-site mode of SMO activation. We recommend this assay for use in large-scale compound screening and plan to explore the mechanism of SMO signal transduction further using this technique.

Chapter 5

Smoothened activation

5.1. Abstract

Smoothened (SMO) is a key player in Hedgehog (Hh) signalling with the crucial role of transducing the signal across the plasma membrane. Though the structure and function of SMO have been well studied, it is unclear how the extracellular Hh signal is translated to the intracellular response. In this chapter, we explore a range of approaches with the aim of elucidating how SMO activation is communicated through the molecule. We describe our intended strategy for structural solution and present data for each stage of optimisation. We show SMO binding to a conformationally selective SMO nanobody in the presence of a novel range of agonists. These data have key implications for our understanding of SMO signalling and provide a robust foundation to achieve our structural goal.

5.2. Introduction

5.2.1. G-protein coupled receptor signal transduction

G-protein coupled receptors (GPCRs) are the largest class of membrane proteins in the human genome and account for over a third of targets of all FDA-approved drugs [75]. They are structurally characterised by their heptahelical transmembrane bundle, flanked by variable extracellular amino and intracellular carboxyl termini. Over half of all GPCRs are olfactory receptors, which recognise a range of odorant molecules within the sensory nervous system. The remainder of GPCRs respond to a wide variety of stimuli, including adenosine, opioids, glutamate and in the case of rhodopsin, light [77]. These ligands are capable of inducing conformational changes that are transmitted across the receptor to the cytoplasmic region, stimulating binding to a range of effector molecules, including G-proteins and arrestins. If stabilised in an active conformation, this enables G-protein binding and GTP exchange (*Figure 5.1.a.*).

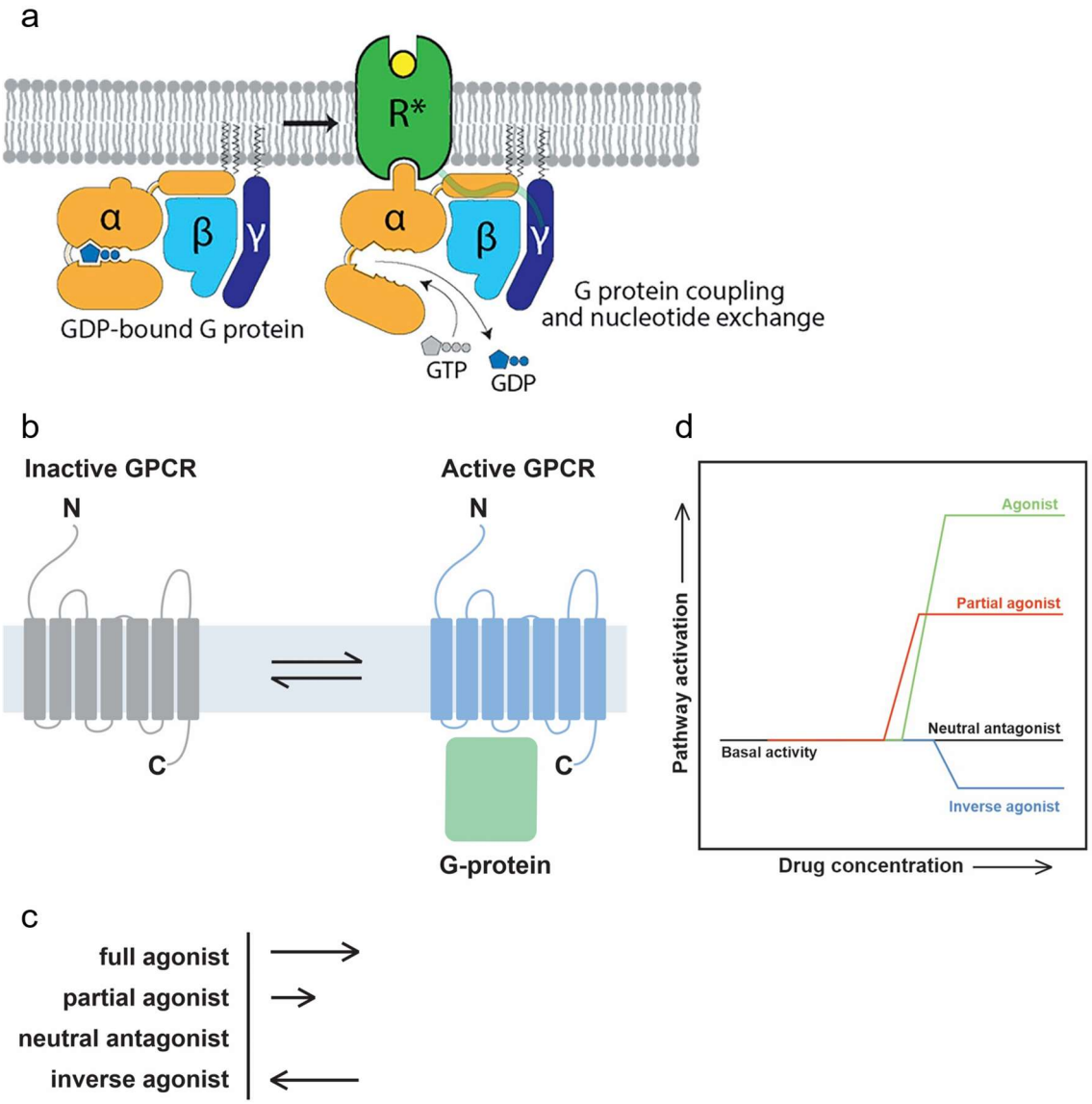


Figure 5.1. G-protein coupled receptors exhibit complex pharmacology

a. GDP-bound G-protein interacts with an activated GPCR molecule. Nucleotide exchange is initiated and the GTP-bound G-protein dissociates to stimulate downstream signalling. Figure reproduced with permission from Komolov, 2017 (Licence # 5035450611116).

b. Simplified schematic showing how GPCRs are described as an equilibrium of states, a subset of which can bind effector molecules.

c. Different ligands bias the GPCR equilibrium towards a subset of states. Typical effects of some ligand classes are shown. Arrow length represents extent of response.

d. Schematic showing how the ligand classes affect signalling. This is the outcome of equilibrium bias shown in c.

GPCRs often display a low level of basal activity, due to a small proportion of molecules being in an active conformation [92]. It is thought that GPCRs exist in an equilibrium between several conformational states, each with differing propensity to bind to and activate G-proteins (or other binding partners) (*Figure 5.1.b*). Ligand binding simply pushes the equilibrium to favour a specific activation state by decreasing the free energy required to adopt this conformation (*Figure 5.1.c*) [76]. Due to this mechanism, there are a range of ligand classes, rather than simply activators and inhibitors. GPCRs often respond to both full and partial agonists, which shift the equilibrium towards active conformations to a greater or lesser extent [83,92]. Inverse agonists have the opposite effect, causing the equilibrium to favour inactive states and hence decreasing activity below basal levels (*Figure 5.1.d*). Finally, neutral antagonists bind to the receptor but do not alter the equilibrium, hence maintaining basal activity levels and preventing other ligands binding to the same site [76,77]. Ligand bias has also been observed in GPCRs as some ligands can stimulate both G-protein and arrestin binding to varying degrees or favour a specific G-protein subtype or isoform [76,92].

5.2.2. Heterotrimeric G-proteins

Heterotrimeric G-proteins consist of an alpha (α), beta (β) and gamma (γ) subunit (*Figure 5.2.a*). The complex is stabilised by GDP (guanosine diphosphate) binding and dissociation is induced by GPCR-dependent GDP to GTP exchange [199]. The α-

subunit has an alpha-helical domain and GTPase domain, which contribute to stabilisation and hydrolysis of GTP, respectively [200]. The GTPase domain is lipid-modified with myristoyl and/or palmitoyl groups attached to the N-terminal helix, which encourage membrane association [201]. The GTPase domain forms extensive contacts with the GPCR when complex formation occurs, which may account for up to 100% of G-protein-GPCR contacts [78]. This domain also interacts with the Gβγ-complex, which exists as a stable dimer. The β-subunit has a β-propeller domain and N-terminal helical region which forms a coiled-coil with the N-terminus of the γ-subunit [200]. The Gβγ-complex is also membrane associated via C-terminal prenylation of the γ-subunit [201].

The different isoforms of each subunit can associate in a number of different combinations and are generally classified into subtypes, based on sequence homology of the α-subunit [202]. The classes are G_s, G_i, G_q and $G_{12/13}$. Each class has a different effect on the downstream signalling cascade. G_s and G_i have opposite effects on the same pathway. G_s α-subunits stimulate adenylyl cyclase to convert adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP), whilst G_i inhibit this enzyme. cAMP then activates cAMP-dependent protein kinase A (PKA), which phosphorylates a wide range of proteins, including ion channels and transcription factors (*Figure 5.2.b*). Some G_i α-subunits are also able to activate phosphodiesterase (PDE) proteins, which down regulate cAMP production [199]. The Gβγ-complex also contributes to intracellular signalling by activating cascades such

as mitogen-activated protein (MAP) kinase signalling and recruiting G-protein coupled receptor kinase (GRK) to phosphorylate the GPCR [203,204]. This enables arrestin binding and internalisation of the GPCR [205].

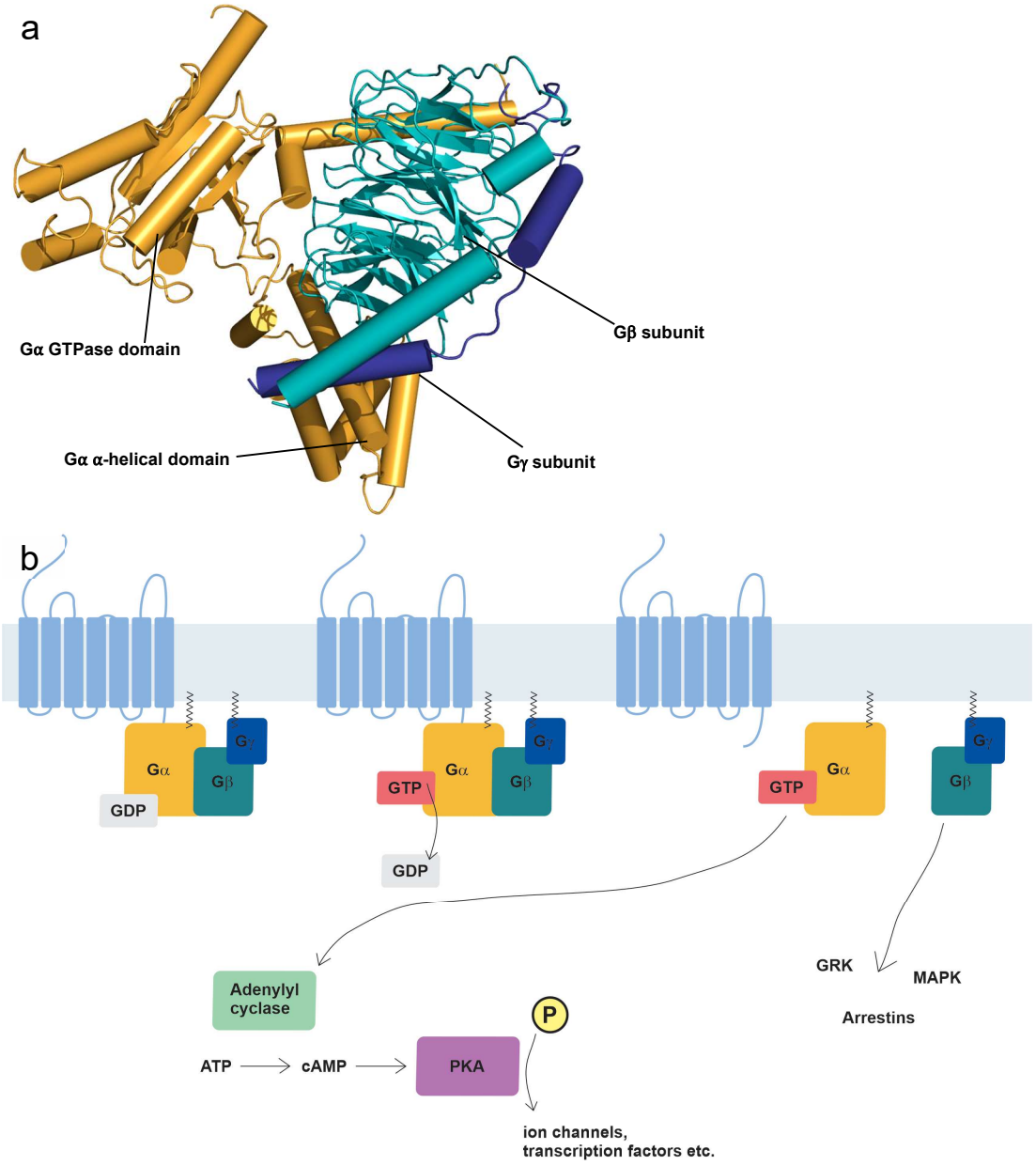


Figure 5.2. G-protein signalling

a. Typical structure of a heterotrimeric G-protein. α-subunit – orange, β-subunit – teal, γ-subunit – dark blue. PDB 6OT0 used.

b. Schematic showing the cascade of events when an active GPCR interacts with a G-protein. GDP to GTP exchange is initiated and causes dissociation of the Gα and Gβγ-subunits. Gα subunits regulate adenylyl cyclase production of cyclic-AMP (Gs isoforms stimulate and Gi inhibits). Increased concentrations of cellular cAMP induce PKA-mediated phosphorylation of a range of targets. The Gβγ-subunit can stimulate the MAP kinase cascade and recruit GRK to phosphorylate the GPCR. This enables arrestin binding and ultimately GPCR internalization.

5.2.3. Smoothened signal transduction

Smoothened is classed as a GPCR, due to its heptahelical transmembrane bundle and sequence homology to other class F GPCRs, namely Frizzled, the Wnt pathway GPCR [206,207]. It has also been shown that SMO binds both G_i class G-proteins and arrestins [66,67]. It is thought that SMO can interact with multiple isoforms within the G_i class, including G_{i1}, G_{i2} and G_o [67]. These Gα-subunits are capable of downregulating adenylyl cyclase-mediated cAMP production (*Figure 5.2.b*). This prevents PKA activation and phosphorylation of the GLI transcription factors [115,208]. This means GLI is not targeted for degradation and can instead enter the nucleus to stimulate transcription of Hh target genes [209].

The G$\beta\gamma$-complex also interacts with SMO and recruits GRK2 to phosphorylate the SMO C-terminal tail [65,108]. This enables β-arrestin 2 binding which causes the translocation of activated SMO to the primary cilia, an essential event in Hh signalling [66,210-212]. β-arrestin may mediate the interaction of SMO with KIF3A (kinesin-like protein 3A), though it has also been suggested that it is involved in removal of GPR161 from the cilia, a GPCR which negatively regulates Hh signalling [213]. SMO interaction with KIF3A ultimately leads to the dissociation of the GLI-SuFu (suppressor of fused) complex. This enables GLI to enter the nucleus and stimulate transcription of Hh target genes[214]. Whilst some of the intermediate steps are still

unclear, the two cascades from SMO binding to a heterotrimeric G-protein provide a logical mechanism for GLI-mediated transcription of hedgehog target genes.

5.2.4. Strategies for obtaining active state structures

A number of structures of GPCRs in their active states have been determined over the last decade. A common feature of these structures is the necessity of an intracellular binding partner to stabilise the active conformation. Agonist binding is not sufficient to maintain this state for structure determination [215]. A range of molecules have been used for this purpose, including heterotrimeric G-proteins (or derivatives), arrestins and nanobodies (Nb) (*Figure 5.3*).

As heterotrimeric G-proteins are a physiological binding partner of active GPCRs, they are a clear and rational choice of molecule to stabilise this state. The first GPCR-G-protein structure was determined by lipidic cubic phase X-ray crystallography in 2011, depicting the β_2 adrenergic receptor (β_2AR) bound to the G_s heterotrimer [78]. It was followed by two more G_s-GPCR complexes [216,217] and several $G_{i/o}$-GPCR complex structures, determined by cryo-electron microscopy [218-221]. Many of these structures include crystallisation aids, including GPCR-stabilising fusion proteins (covered in Chapter 3.2) and G-protein heterotrimer stabilisers. The initial β_2AR-G_s structure required both a T4-lysozyme fusion at the N-terminus of the GPCR and

Nb35, which stabilises the G_s αβ interface [78]. The structure of the μ-opioid receptor in complex with heterotrimeric G_i was stabilised by a single chain variable fragment (scFv16) which was raised to specifically stabilise G_i [221,222]. The need for extra stabilisation of trimeric G-proteins is due to the transient nature of the GPCR-G-protein complex. In cells, formation of this complex enables GDP to GTP exchange and dissociation of the G-protein heterotrimer. This means the complex must be artificially stabilised in order to make structural studies possible.

Alternately, G-protein derivatives have also been used to trap GPCRs in an active conformation. The structure of β_2AR-G_s showed that the receptor exclusively interacted with the Gα-subunit, with over 85% of the contacts involving residues of the $G\alpha_s$ C-terminal α5 helix [78]. This made possible the determination of GPCR-GαCT peptide complexes for both rhodopsin and β_2AR [223-226]. The importance of the Gα-subunit also prompted the concept of "minimal G-proteins" – engineered Gα-subunits, designed to be stable crystallisation aids for a range of GPCRs [121-123]. The suite of mini G-proteins began with G_s only, but was extended to include mini $G_{i/o}$ proteins. Unique chimeric mini G-proteins were also developed, with the stable scaffold and Nb35-binding ability of G_s combined with $G_{i/o}$ specificity (*Figure 5.3*) [123,227].

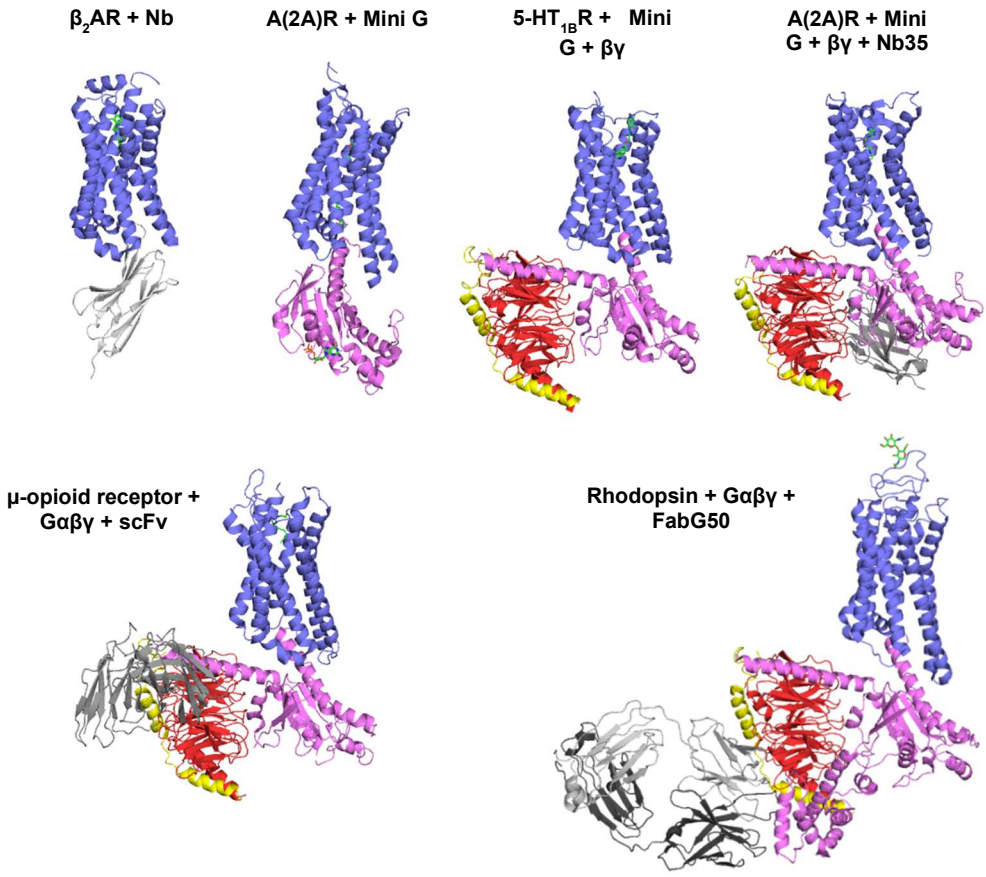


Figure 5.3. Approaches to capturing GPCR active state structures.
β2AR activated with nanobody (PDB:3P0G). **A(2A)R** activated with mini G-protein (PDB:5G53). **5-HT1BR** activated with heterotrimeric G-protein (PDB:6G79). **A(2A)R** activated with mini heterotrimeric G-protein, stabilised by nanobody 35 (PDB:6GDG). **μ-opioid receptor** activated with heterotrimeric G-protein, stabilised with single-chain variable fragment (scFv) (PDB:6DDE). **Rhodopsin** activated with heterotrimeric G-protein, stabilised with FabG50 (PDB:6CMO).

The other physiological GPCR binding partners – arrestins – have also been used as active state stabilisers for structural studies. It was shown that rhodopsin adopts a highly similar conformation when bound to both arrestins and the $G\alpha_s$CT peptide [228]. This interaction is less well characterised than GPCR-G-protein complex formation, partially due to the requirement for the presence and phosphorylation

of the C-terminal intracellular tail of the GPCR [229]. This region is often truncated for high yield protein production, as it often leads to degradation of the GPCR.

Lastly, specifically raised active state stabilising nanobodies have been used for structure determination [94,230-234]. This may be achieved by using an activated form of the GPCR, either agonist-bound or constitutively active mutant, to immunise nanobody-producing camelids [124,235]. The nanobody library must then be reverse screened to isolate those that favour the active conformation over the inactive. The resulting nanobodies may be able to confer enough stability to the GPCR for crystallisation to be successful [234,235].

5.3. Results

5.3.1. Strategy to obtain an active state SMO structure

The pursuit of any active state GPCR structure is not trivial. When embarking on this project we had multiple components to consider and optimise (*Figure 5.4*). The crucial requirement was a high-quality SMO sample with the ability to adopt the active state. In other words, we could not adopt many of the stabilisation strategies commonly used for GPCRs. This is described in Chapter 3. We selected cryo-EM (electron microscopy) as the method of structure solution with the aim of avoiding crystal artefacts and non-physiological contacts. Cryo-EM is also more suited to large multi-protein complexes with predicted flexible regions. This technique also enables the use of detergent-free systems such as lipid nanodiscs, which can be used to study the effect of membrane composition on proteins such as SMO. Lastly, we needed a method of "trapping" SMO in an active conformation, which is expected to be inherently transient and dynamic.

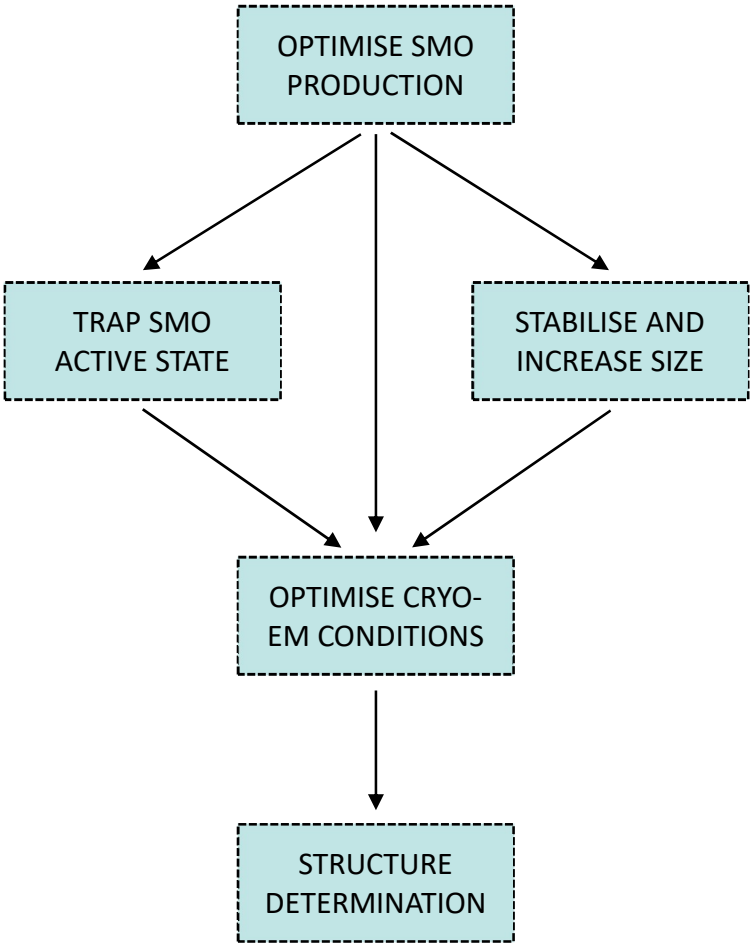


Figure 5.4. Flowchart showing planned approach to determining the structure of SMO in an active state.

5.3.2. Adding size and asymmetry for EM

To optimise our SMO sample for cryo-EM we predicted an increased particle size would be beneficial. We also aimed to make the shape of the particle of interest more asymmetrical. To study SMO alone would be inherently difficult due to its shape and properties. The TMD makes up the majority of the protein and must be surrounded by either a detergent micelle or nanodisc, both of which are roughly symmetrical and featureless, which makes averaging and alignment difficult. We

explored using a range of SMO-binding proteins to increase the size and provide a 'feature' for alignment.

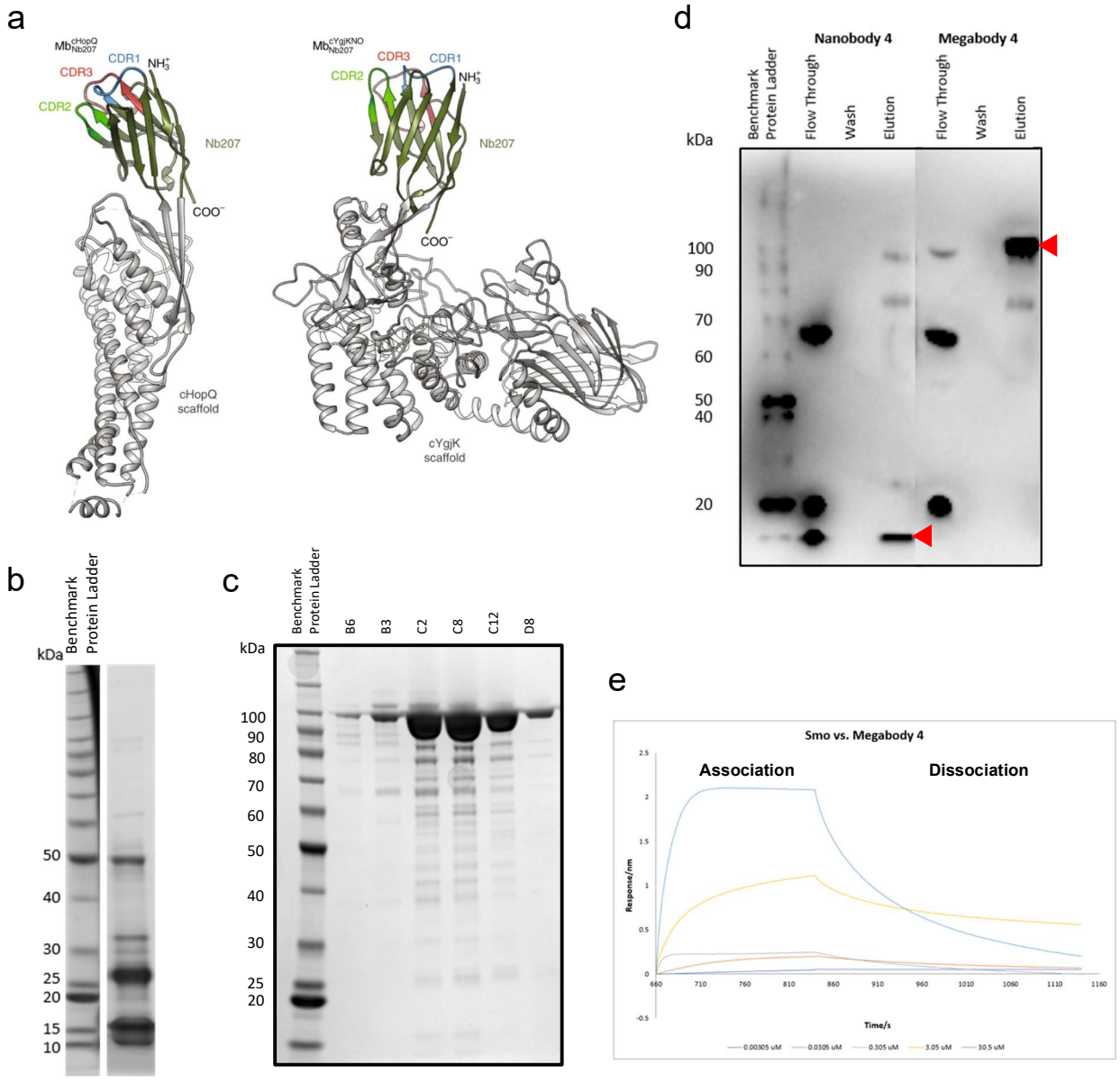


Figure 5.5. Megabody production and characterization.

a. Typical megabody structures showing scaffold protein and nanobody fusion. Reprinted with permission from Uchański, 2021 (Licence # 5044750105969).
b. SDS-PAGE analysis of purified 56 kDa megabody.
c. SDS-PAGE analysis of 100 kDa megabody size-exclusion fractions.
d. Western blot analysis of SMO pull down of nanobody or megabody 4.
e. BLI analysis of SMO - Megabody 4 binding where 660 to 840 seconds show association and 840 to 1140 show dissociation.

Nanobodies have proved a highly valuable tool for structural biology. In addition to those which favour a specific desired conformation, nanobodies confer stability to disordered proteins or transient complexes and are highly successful as a crystallisation aid [78,124,230,234-236]. To complement the recent expansion of cryo-electron microscopy, the pioneers of nanobodies for structural biology developed megabodies (*Figure 5.5.a*) [237]. They are engineered by inserting the nanobody sequence into that of a larger scaffold to produce a stable and well-folded protein. The major advantages of megabodies for cryo-EM are increased particle size and the disruption of preferential orientation. This makes them especially promising for small or poorly distributed samples, including membrane proteins [237].

We previously obtained a range of nanobodies which specifically recognised SMO from the Steyaert Lab and identified two promising constructs through binding studies [153]. We mapped the recognised epitope to the extracellular region of SMO using fluorescence microscopy and determined that they did not affect Hh signalling (experiments performed by R. Siebold-Schwab) [153]. Due to their binding location and lack of interference with signalling, these nanobodies could be used simultaneously with any intracellular SMO binding partner. We selected these for development into two classes of megabodies, DNA of which was kindly provided by the Steyaert Lab (VIB, Brussels). We produced each megabody as recommended in WK6 *E. coli*, using a periplasmic localisation sequence for simple purification [237]. The periplasmic fraction was isolated and subjected to His-tag affinity and size

exclusion chromatography. We compared the 56 and 100 kDa megabody scaffolds and found the latter to be more amenable to large-scale production. We observed significant degradation of the 56 kDa megabodies during the purification process and so focussed upon the 100 kDa scaffold (*Figure 5.5.b,c*). We used pull-down assays and bio-layer interferometry to show that megabody 4 specifically interacts with SMO in a concentration dependent manner, as the corresponding nanobody does (*Figure 5.5.d,e*). In BLI studies we also observed a significant amount of dissociation, suggesting a weak interaction (*Figure 5.5.e*). This was reflected in the K_D values where 5μM was calculated for Mb4 in comparison to 14 nM for Nb4 [153].

5.3.3. Stabilising the SMO active state

Crucially, when pursuing an active state structure, we needed to ensure that the desired state was captured. This may be possible using small molecule agonists, however the pharmacology of GPCRs suggests such a mechanism of activation would be transient and conformationally dynamic. To increase the likelihood of capturing a stable active state, a conformation-selective binding protein can be used. This can be an endogenous effector or an engineered molecule, for example a minimal G-protein or nanobody.

5.3.4. Design of SMO-specific minimal G-proteins

Initially, we chose to adapt the minimal G-protein strategy used for the adenosine receptor ($A_{2A}R$) for SMO [121]. $A_{2A}R$ recognises and activates G_s-proteins, contrasting SMO which contacts $G_{i/o}$-proteins [67,121,122]. When this project began the only minimal G-protein that had been designed and optimised was mini-G_s. We chose the G-protein isoforms G_{i1}, G_{i2} and G_o due to their potent and specific activation by SMO [67]. We also produced mini-G_s to act as a control throughout. We followed the principles used for the design of mini-G_s and structurally verified their suitability for each of the chosen isoforms [122].

Minimal G-proteins are based solely on the Gα subunit, as this region was shown to exclusively contact the GPCR [78]. To minimise the protein, three sections of Gα were deleted to give a more compact molecule (*Figure 5.6.a*). The N-terminal α-helix is reduced by 16-17 amino acids, removing the region which contacts the G$\beta\gamma$-subunits. The most dynamic region of G_s in the GPCR-G-protein complex structure was the Gα alpha helical domain (GαAH) [78], which is removed and replaced by a short glycine-serine linker. The switch III loop is also deleted, constraining β4 and α3 in a more stable arrangement. Mutations were used to stabilise the mini-G at the nucleotide binding pocket and P-loop residues were also mutated to constrain switch II and to improve packing of the protein core with α1 and α5 helices (*Figure 5.6.a*). One further mutation was noted to improve diffraction for the $A_{2A}R$-mini-G_s

complex crystals [122]. After publication, we also added a selection of the mini-$G_{i/o}$ proteins designed by the Tate laboratory [123].

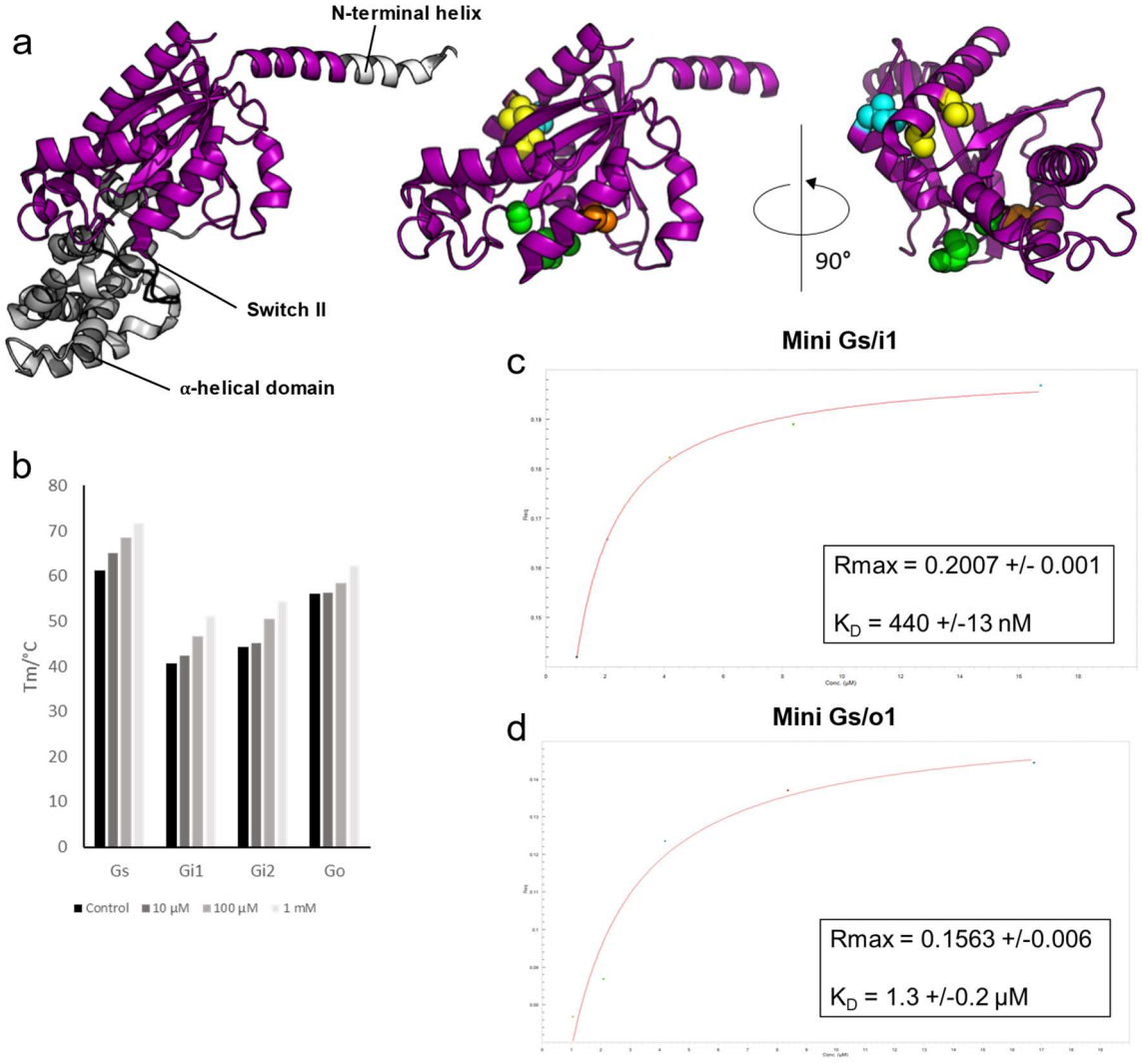


Figure 5.6. Minimal G-protein production and characterization

a. Design of minimal G-proteins. Deleted regions of Gα are labelled and shown in grey. Mutated residues are highlighted as follows; nucleotide binding pocket – green, α1 and α5 packing – yellow, switch II stabilizing – orange, crystal packing – cyan.
b. Effect of GDP addition on thermostability of a range of minimal G-proteins.
c. BLI analysis of SMOΔC binding to mini Gs/i1 in the presence of SAG.
d. BLI analysis of SMOΔC binding to mini Gs/o1 in the presence of SAG.

5.3.5. Addition of GDP increases G-protein stability

To produce minimal G-proteins, Rosetta (DE3) *E. coli* were transformed with pET22b containing the gene for each mini-G construct. Cells were grown to optimum density and expression of the lac operator-controlled gene was induced using IPTG. Cell lysate was subjected to His-tag affinity purification and size exclusion to isolate pure, monomeric minimal G-proteins. Yields varied from 3-17mg purified protein per litre of culture.

We tested the concentration dependent effect of GDP on minimal G-proteins, as this nucleotide is known to stabilise the G-protein heterotrimer. Though the minimal G-protein nucleotide binding pocket was engineered to be stable when empty, we observed a clear increase in melting temperature (Tm) as a result of increased GDP concentration (*Figure 5.6.b*). As expected, the different minimal G-protein isoforms varied in stability; however, the presence of GDP increased the melting temperature ubiquitously by up to 10°C. We subsequently included GDP in all purification buffers to aid stabilisation of the minimal G-proteins. We noted that the nucleotide would need to be removed or replaced with GTP before GPCR binding, though this has previously been achieved via addition of apyrase.

5.3.6. Mini G-proteins interact with SMO

To determine which minimal G-protein isoforms might be successful we carried out binding studies. We used bio-layer interferometry (BLI) to identify which proteins would form a complex with SMO that could possibly be used for structure solution. We immobilised amphipol solubilised SMOΔC via a biotinylated N-terminal AviTag™ onto streptavidin coated BLI tips (Pall Biotech) and presented each minimal G-protein as a separate experiment. From initial tests we identified 2 potential binders; mini-Gs/i1(48) and mini-Gs/o1(long) (*Figure 5.6.c,d*) [123]. We then performed a serial dilution of the G-proteins to produce a binding curve and calculate dissociation constants. We found mini-Gs/i1(48) to bind SMO with a higher affinity and K_D=440 nM. Mini-Gs/o1(long) also showed binding to SMO with K_D=1.3 μM (*Figure 5.6.c,d*).

5.3.7. SMO activating nanobody

As for several other GPCRs, a nanobody was identified which selectively binds an activated conformation of SMO (*Figure 5.7.a*) [94,230-232]. Nanobody 8 (Nb8) interacts with the intracellular side of the SMO transmembrane domain, when it is activated by either SAG or cholesterol [94]. We decided to produce a stock of this protein to trap SMO in the active state. I generated a HEK 293T cell line stably transfected to produce the construct Nb8-3C-mVs-His_{12}. I decided to add a 3C cleavable

monovenus fluorescent tag to the nanobody construct for future use in FSEC complexation studies. Incidentally, this also increased expression substantially (data not shown). This protein was purified from conditioned media using His-tag affinity and size exclusion chromatography. From 2 sequential harvests a final yield of 20 mg pure monomeric nanobody was achieved per litre of conditioned media.

To assess binding, I performed a series of pull-down experiments to identify conditions under which Nb8 can bind to SMO. After combining SMOΔC-HA with Nb8 I added small molecules, as applicable. The reaction was incubated overnight at 4°C before addition of pre-equilibrated TALON beads, which will interact with Nb8 via a His-tag. After incubation, the beads were washed, and the imidazole elution was subjected to SDS-PAGE and western blot analysis to identify bound species (*Figure 5.7.b*). I chose an anti-HA antibody (Sigma Aldrich) to specifically probe SMOΔC-HA and captured a chemiluminescent image using the ChemiDoc system (BioRad). Results showed that Nb8 binds minimal amounts of SMO when no ligand is added. This also occurs when known SMO antagonists cyclopamine or vismodegib are present. However, when 20(S)-OHC, SAG or KK174 are added, SMO is clearly pulled down by Nb8 (*Figure 5.7.b*). This suggests Nb8 binding is dependent on SMO conformation and that addition of agonists 20(S)-OHC, SAG or KK174 cause a conformational change at the intracellular side of SMO. Interestingly, each of these ligands binds to a different previously identified SMO ligand binding site, showing they can all induce the conformational change

required for Nb8 binding. Further studies will be needed to reveal how these diverse binding events result in a common signalling output, namely Nb8 binding.

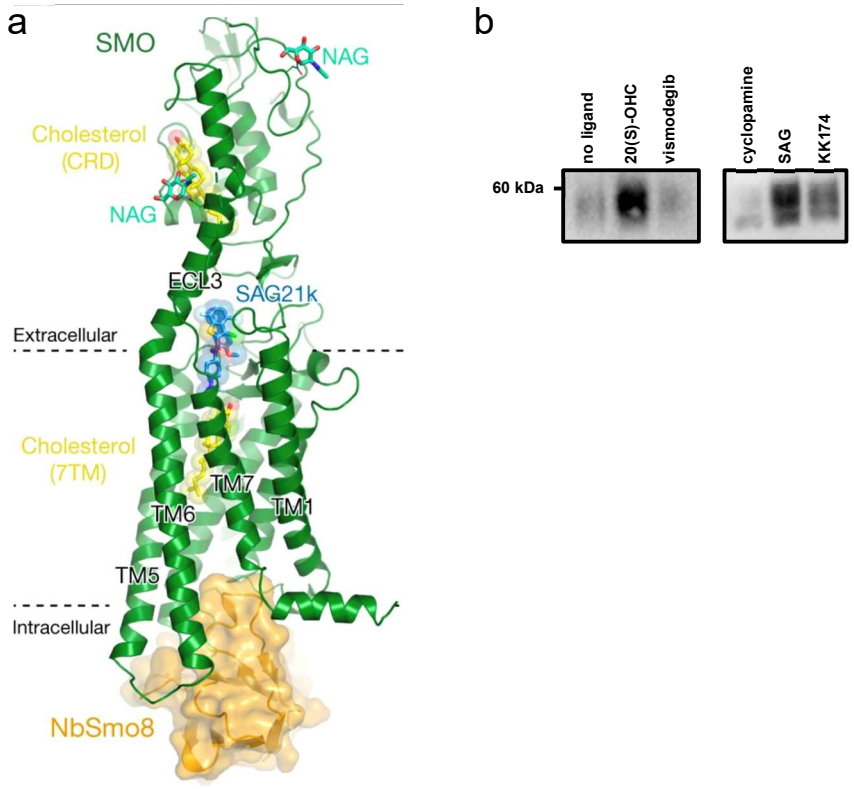


Figure 5.7. Nb8 binds the active state of SMO

a. Structure of mSMO-SAG bound to Nb8 revealed two bound cholesterol molecules. PDB 6O3C reprinted with permission from Deshpande, 2019 (Licence # 5044840789967)
b. Anti HA tag western blot analysis of Nb8 pull down of SMOΔC in the presence of various ligands, where applicable.

5.3.8. Optimising the sample for cryo-EM

Alongside ensuring the protein complex was suitable for our aim, we also needed to optimise the sample for EM studies. When this project began, there was no data available to show how SMO behaved on EM grids. We needed to explore not only

the grid type and plunging conditions but the sample properties and additives too. This included testing detergent-solubilised SMO and detergent-free systems, as well as alignment aids and optional buffer components.

Initially, we compared DDM-solubilised SMO and SMO reconstituted into lipid nanodiscs (*Figure 5.8.a*). There are theoretical advantages for each system; producing detergent-solubilised protein is considerably more straightforward, though nanodisc reconstitution provides a more physiological environment and enables study of membrane-lipid effects. However, detergent-solubilised protein samples are inevitably contaminated with empty detergent micelles and some detergents can restrain transmembrane proteins. Nanodisc reconstitution introduces multiple new variables including scaffold protein and lipid composition, which requires extensive optimisation. Empty nanodiscs can also be a problem if not successfully removed.

We used negative staining to assess sample quality, aggregation and distribution (*Figure 5.8.a,* top row). We compared DDM-solubilised SMO to a range of nanodiscs. We used FSEC to analyse different combinations of membrane scaffold protein and lipids and select the optimal combination (see Chapter 3). The MSP1D1 POPC nanodiscs showed a good distribution of homogenous isolated particles. The DDM-solubilised SMO also showed a good distribution with mostly isolated particles alongside some aggregation. When these samples were analysed under cryo-

conditions the distinction was clearer. Whilst the detergent sample remained consistent, the nanodisc reconstituted SMO was not visible on the grids (*Figure 5.8.a*, bottom row). This could be due to the blotting conditions or the nanodiscs degrading due to crystalline ice or interaction with the air-water interface during plunging.

We also needed to assess which binding partners would be appropriate for cryo-EM. We have previously characterised the SMO-binding nanobodies provided to us by the Steyaert lab (VIB, Brussels) and shown them to bind with high affinity to the extracellular region of SMO with no noticeable effect on Hh signalling. We initially analysed the SMO-Nb4 complex to determine whether the nanobody added enough extra density to enable satisfactory alignment. 2D classes initially looked promising, with a clear protrusion to one side of the detergent micelle (*Figure 5.8.b*). However, alignment was unsuccessful, and the resulting volume showed no clear density for the extracellular domain. We therefore decided to focus on megabodies from this point onwards.

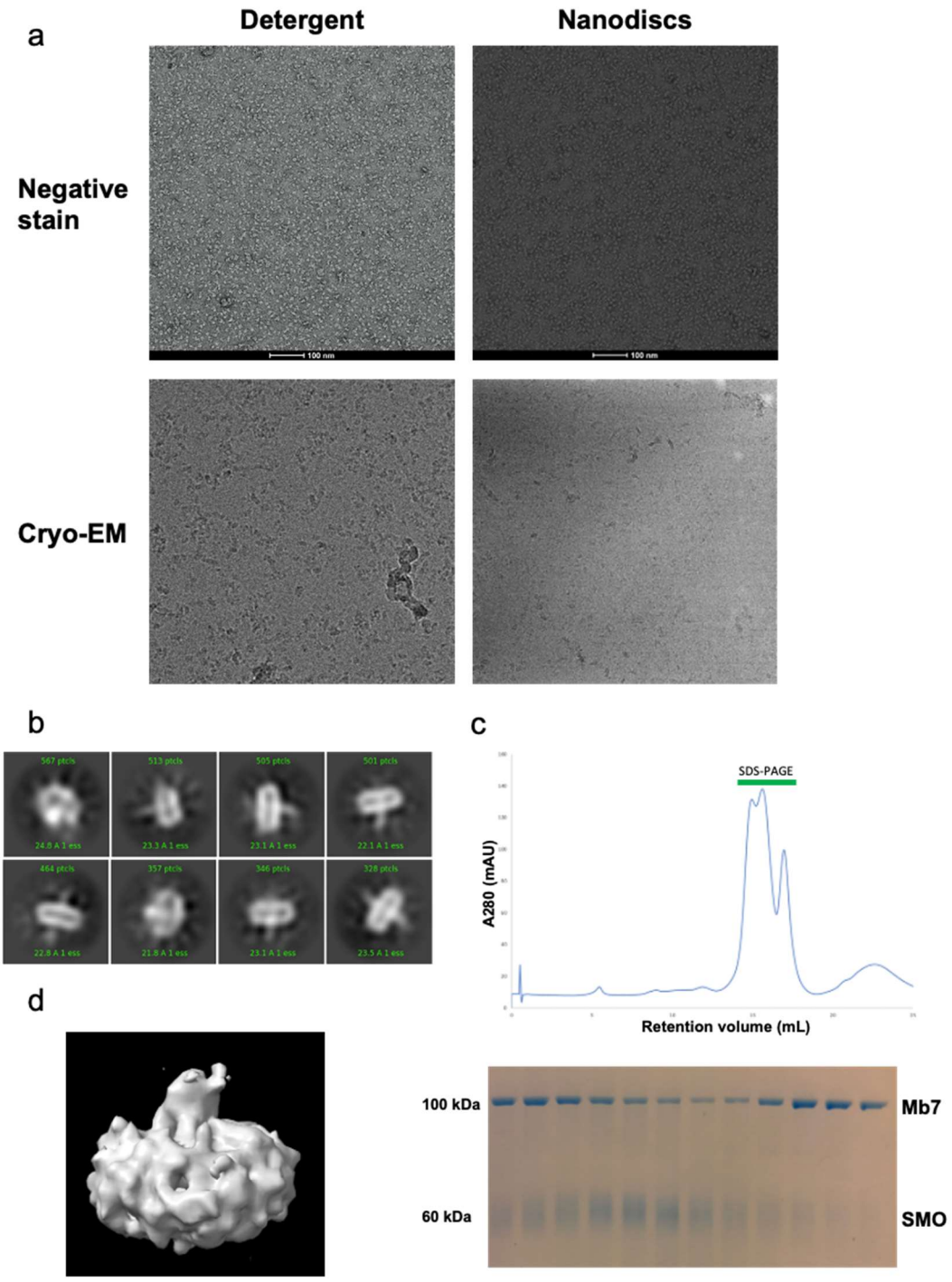


Figure 5.8. Optimising cryo-EM for SMOΔC

a. Micrographs comparing detergent solubilized and nanodisc reconstituted SMO in negative stain analysis and cryo-EM. Cryo-EM data collected at 130,000x magnification.
b. Representative 2D classes of SMOΔC-Nb4 complex generated from 32,000 particles picked and processed in cryoSPARC.
c. Size exclusion chromatography of SMOΔC-Mb7 complexation and SDS-PAGE analysis of fractions.
d. Ab-initio reconstitution of SMOΔC-Mb7 complex generated from 42,000 particles processed in cryoSPARC.

The megabodies available were based on the known binders Nb4 and Nb7 [153]. They were previously analysed by BLI and cell staining assays. These experiments calculated a K_D of 14 nM for Nb4. Fluorescence microscopy was used to show that both Nb4 and Nb7 interact with the extracellular domain of SMO. It was also confirmed that these nanobodies do not have an effect on Hh signalling [153]. We obtained megabodies constructed from the sequences of Nb4 and Nb7 from the Steyaert lab (VIB, Brussels). We initially needed to ensure they bound SMO as their corresponding nanobodies do. We tested this using BLI and analysed complex formation using size exclusion chromatography, showing that both megabodies interact with SMO.

Initially, we pursued the SMO-Mb4 complex for EM studies. We attempted to purify this complex via size exclusion, but no protein was eluted. We concluded the complex was most likely aggregating and not entering the column matrix. Based on anecdotal evidence (personal communication, Loic Carrique) we chose to combine SMO and Mb4 directly before applying to EM grids and plunge-freezing. It has previously been a successful strategy used to prevent aggregation-prone samples from doing so. However, subsequent data collection suggested we were unsuccessful as all isolated particles that could be processed were of either SMO or Mb4, with the complex appearing to occur only in the aggregated regions. We hence focused on the SMO-Mb7 complex, which was successfully purified via size exclusion, indicating there was no significant aggregation. We observed a mixed

population, with a complex peak alongside each component alone after size exclusion (*Figure 5.8.c*). We proceeded with the complex peak and applied it to EM grids. Analysis of this sample unfortunately showed little complex either, with most particles showing either SMO or Mb7 alone. 3D reconstitution from this data set showed a micelle with a small protrusion, likely the SMO CRD (*Figure 5.8.d*). We concluded that the complex must have dissociated, either spontaneously or due to vitrification.

5.4. Discussion

5.4.1. From nanobodies to megabodies

Since their introduction, nanobodies have proven to be a valuable tool for structural biology as crystallisation chaperones [58,124,230,238]. They have been shown to stabilise flexible complexes and increase crystallisation chances, as well as trapping specific conformational states [78,230]. However, recent advances in the field of cryo-electron microscopy have led to different properties being required of a protein structure solution aid. In addition to stabilisation and conformational selectivity, cryo-EM also benefits from increased size of the particle of interest, as well as asymmetrical features to aid alignment for averaging. To address this, megabodies were developed by the Steyaert lab (VIB, Brussels) (*Figure 5.5.a*) [237].

When analysing the SMO-specific nanobodies and megabodies at our disposal, we observed specific binding of Nb4 and Nb7. This was also reflected by the corresponding megabodies tested. However, we encountered degradation of the megabodies during production and aggregation of the SMO-Mb complexes (*Figure 5.5.b*). We also failed to capture the complex on cryo-EM grids (*Figure 5.8*). Overall, whilst the megabodies are an attractive tool for cryo-EM studies, complex formation and subsequent structure determination requires optimisation. As the SMO

megabody range is very limited, we struggled to overcome these issues. For future studies, we suggest a wider range of megabodies should be trialled.

5.4.2. Activating Smoothened

A range of approaches were taken to stabilise SMO in an active conformation. As described here, we designed, produced and characterised a suite of minimal G-proteins as potential SMO binding partners. We showed that addition of GDP enhanced the stability across all G-protein isoforms, despite targeted mutations to stabilise the nucleotide binding site when empty [122]. However, the primary structure of a GPCR and minimal G-protein complex showed a GDP molecule bound in this site [121]. This implies the mutations have decoupled GPCR binding from GDP dissociation, rather than stabilised the apo-state.

Despite our initial success in designing a range of minimal G-proteins as potential SMO binders, subsequent interaction studies highlighted unexpected obstacles. Whilst we observed binding of Gi and Go isoforms to SMO in BLI experiments, we struggled to form this complex for structural studies (*Figure 5.6.c,d*). There are many possible reasons for this interaction being difficult to capture in a stable complex. Firstly, we know from GPCR biology that the endogenous interaction between a GPCR and G-protein is transient – once the GPCR stimulates the exchange of GDP to GTP, the G-protein dissociates. Though the minimal G-proteins are unable to

catalyse this exchange, the transient nature of the binding event suggests a relatively weak interaction. Secondly, the importance of the lipid membrane environment for GPCR - G-protein interactions has been highlighted recently. In a neurotensin receptor cryo-EM structure, the G-protein forms an extensive network of interactions with the GPCR and the lipid nanodisc used for solubilisation [239]. By using detergent micelles for solubilisation, we are removing lipids that are potentially important for complex formation. This could explain why we cannot reliably observe the SMO – G-protein interaction in a detergent environment. Finally, solubilising membrane proteins with detergents can also remove endogenous cofactors. As discussed in Chapter 4, a recently highlighted cholesterol binding site is thought to play a role in SMO activation and downstream signalling [94]. It is possible that over-expression, detergent solubilisation and purification disrupts this site, and the cholesterol is displaced. Though this is difficult to test, we are exploring whether different detergents have an effect on complex formation, including SMO – minimal G-protein.

In contrast to the minimal G-proteins, Nb8 has been shown to bind SMO in the presence of a number of different agonists (*Figure 5.7.b*). It was initially raised against SAG-bound SMO, which revealed a cholesterol molecule bound in the lower TMD site and also the CRD site [94]. Our data shows that only one of these sites must be occupied for Nb8 binding, as 20(S)-OHC, SAG and KK174 can all stimulate Nb8 binding individually. The observation that the agonists can bind and activate

SMO either independently or in parallel has implications for SMO signal transduction. From these data alone we cannot conclude whether each agonist favours the same conformation, or whether a subset of SMO conformational states is able to bind Nb8. It is possible that the ligands bind cooperatively, with each subsequent agonist biasing the SMO equilibrium further towards an active effector-bound state and increasing the SMO affinity for Nb8 or a G-protein. The relationship between the SMO binding sites is an intriguing area and requires further exploration.

5.4.3. Lessons from electron microscopy studies

Though we have not yet obtained a structure of SMO in an active state, we have explored a range of conditions and samples for subsequent EM studies. Our initial experiments using lipid nanodiscs were promising, though require optimisation for cryo-EM study. SMO is amenable to reconstitution into a nanodisc environment and negative stain analysis shows minimal aggregation of the sample (*Figure 5.8.a*). The use of lipid nanodiscs in cryo-EM studies of GPCRs is still minimal, though this technology is likely to grow in popularity as researchers study complex membrane proteins and their interactions. For SMO, nanodiscs enable several interesting areas of study. Initially, the possibility of membrane lipids contributing to complex formation could be crucial to characterise SMO downstream signalling, whether that is G-protein dependent or not [239]. The role of cholesterol in Hh signalling could

also be explored using lipid nanodiscs, as sterols can be incorporated into these discs, allowing physiological-like interactions [193]. This could also be a promising approach to studying the effect of membrane composition on SMO, including the ciliary membrane environment [240].

We also explored suitable binding partners for SMO cryo-EM studies. We concluded nanobodies were not large enough to act as alignment aids and megabodies would be more appropriate. Unfortunately, the limited megabodies at our disposal introduced further problems including complex dissociation and aggregation. To continue this approach a different set of megabodies are necessary, which is now possible. We aim to create a megabody using Nb8 and the recently published method for constructing novel megabodies [237]. We will make use of scaffold plasmids kindly gifted by the Steyaert lab (VIB, Brussels).

5.4.4. Summary

In this chapter we have explored the necessary steps for determining an active-state SMO structure. We have explored a range of binding partners for activation or stabilisation. We have shown SMO Nb8 to bind a range of agonist-induced SMO conformations and highlighted some key considerations for recreating endogenous complexes *in vitro*. We have produced and characterised nanobodies and megabodies to use as structure solution aids and enable alignment and averaging

of EM data. We have also explored using both detergent and lipid nanodiscs to solubilise SMO and despite difficulties, we believe lipid nanodiscs are a promising approach for this project as they provide the most physiological environment.

Chapter 6

Discussion

6.1. Smoothened signal transduction

6.1.1. Smoothened binding sites show co-operativity

The mere existence of multiple small molecule binding sites in SMO raises the question of inter-site communication [99]. Previous studies have focused upon the relationship between the sterol binding groove in the CRD and the drug binding site in the TMD [59,99]. It has been shown that CRD agonists and TMD agonists can act synergistically to increase the amplitude of SMO activation [99]. When 20(S)-OHC is added, the EC50 of SAG activation decreases. The same is true for the reciprocal experiment, where SAG addition causes a drop in the EC50 of 20(S)-OHC [99]. This shows synergy between the two sites and confirms there is communication between the CRD and TMD (*Figure 6.1.a*)[59]. Also tested by the authors was the combination of a CRD agonist and TMD antagonist. The IC50 of cyclopamine was not affected despite increasing concentrations of 20(S)-OHC, suggesting non-competitive inhibition (*Figure 6.1.b*). However, the maximum level of Hh activation from saturating 20(S)-OHC concentrations is reduced by cyclopamine, which shows cyclopamine has a 'limiting' effect on SMO signalling. In contrast, the IC50 of SANT-1 is affected by increasing concentrations of 20(S)-OHC [99]. This implies that the lower binding position of SANT-1 makes it susceptible to regulation by the CRD.

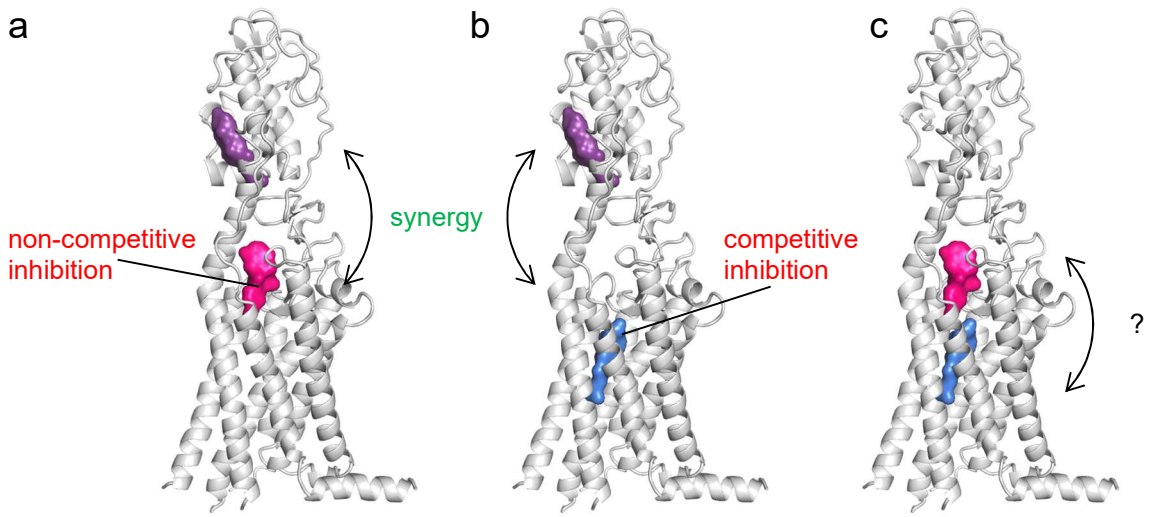


Figure 6.1. Communication between ligand binding sites

a. CRD agonists (purple) synergise with upper TMD agonists (pink), e.g. 20(S)-OHC and SAG. Upper TMD antagonists non-competitively inhibit CRD agonism, e.g. cyclopamine.
b. CRD agonists positively co-operate with lower TMD agonists, (blue) e.g. cholesterol and cholesterol. Lower TMD antagonists competitively inhibit CRD agonism, e.g. SANT-1.
c. The relationship between the upper and lower TMD sites is as yet unknown. SAG and KK174 could be used to explore synergy, whilst cyclopamine could be used to test inhibition.

Since the revelation of the lower TMD cholesterol binding site, we asked what relationship this site has with the others. Through our study of KK174 using our novel CRD binding assay, we have shown positive cooperation between the lower TMD and the CRD sterol site (*Figure 4.8.g*). It is not yet known whether the upper and lower TMD sites act synergistically or communicate in any way (*Figure 6.1.c*). Due to the varied binding positions of upper TMD ligands, it is possible some may clash with a sterol bound in the lower TMD site. This will be a fascinating area to study in the future.

6.1.2. Basal Hh signalling relies upon the TMD

The discovery of the lower TMD site has revealed valuable insights about how SMO is activated [69,94,95]. It has also raised a number of additional questions, namely, why are there two binding sites for the same ligand and what is each of their roles? Mutational data collected by our collaborators shows that disruption of either site reduced SHH-mediated activation of the pathway by approximately 75%, whereas disrupting both sites abolishes it (*Figure 4.8.i*). This supports our hypothesis of positive cooperation between the sites, as the effect of dual site activation is greater than the sum of the two sites separately (*Figure 4.8.1, 6.2.a*).

We have further explored the roles of each site by using SMOΔCRD, a construct lacking the CRD which results in constitutive activation of SMO (*Figure 6.2.b*) [59,193]. The increase in Hh activation upon deletion of the CRD suggests it plays a regulatory role in SMO signalling; when the site is empty it appears to prevent activation, though agonist binding induces it [59-61,101]. One mechanism by which this could occur is via ECL3 and TM6. It has been shown that the empty CRD binding groove interacts with ECL3, and this has been suggested as a structural route of signal transduction [59,69,94,95,168]. This supports the idea that the CRD regulates SMO by stabilising this region in an inactive confirmation and could partly explain why cholesterol binding at the lower site only (as mimicked by the D99A mutant) is only capable of partially activating SMO (*Figure 4.8.1*). Another interesting observation

from our studies on SMOΔCRD is that the Y398F mutation ablates both basal and SHH-mediated signalling (*Figure 6.2.b*). This tyrosine residue is ideally placed to form a stabilising hydrogen bond to the 3β-hydroxyl group of cholesterol in the lower TMD site [94,95]. This strongly suggests that the basal constitutive activity displayed by SMOΔCRD is dependent upon the lower cholesterol binding site. This concept is repeated in wild type SMO, where we see that basal activity is dependent on the lower TMD site, not the CRD site. From these data, we propose that the lower TMD site is responsible for the majority of basal Hh activity, when the SHH ligand is not present (*Figure 6.2.a*). When SHH is present, both sites appear to be important, which agrees with our hypothesis of positive cooperation. This two-site mechanism of activation may be a way of making Hh signalling threshold-dependent, as required for gradient-based morphogen signalling.

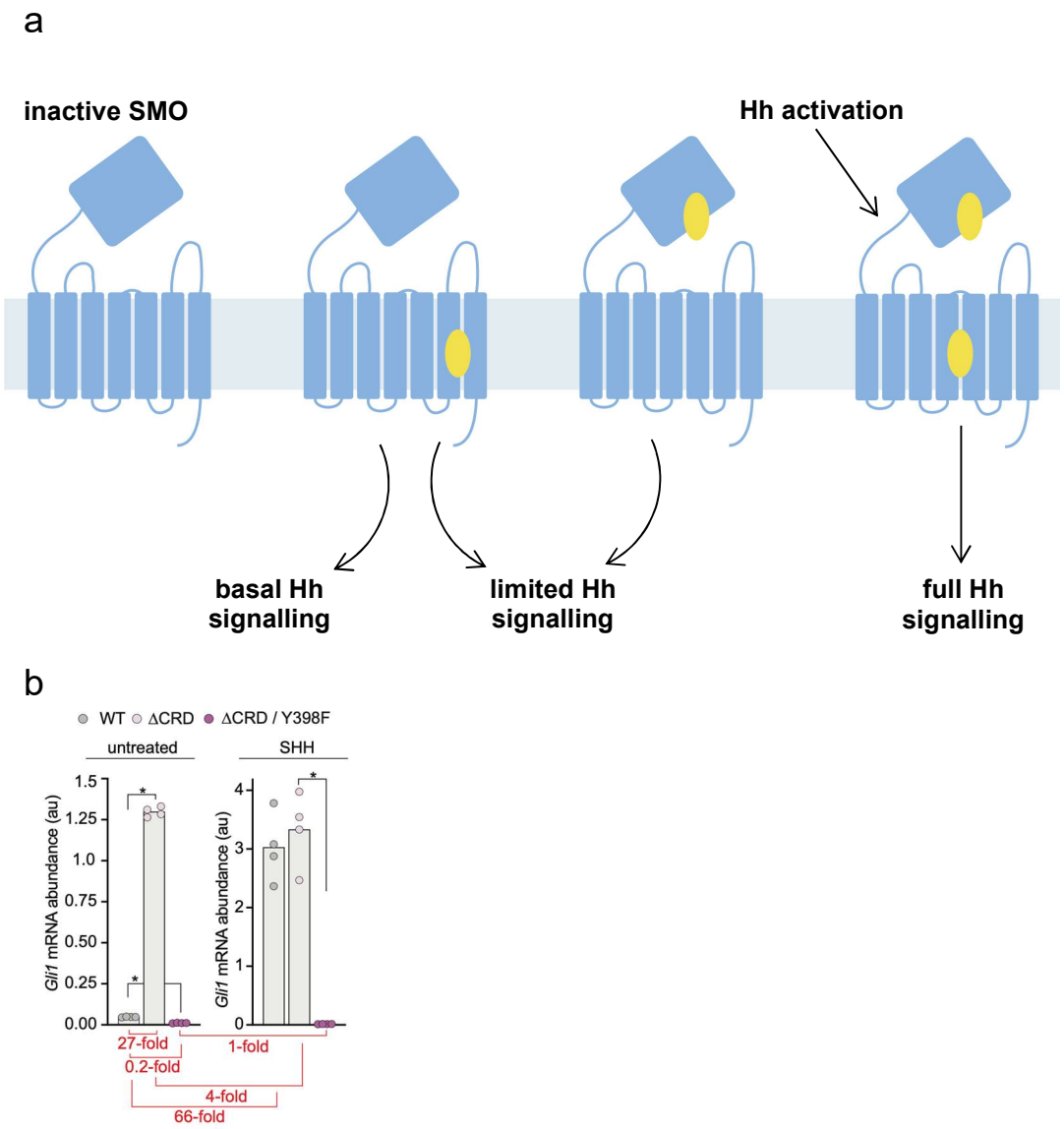


Figure 6.2 Contribution of cholesterol binding sites to Hh signalling

a. Cholesterol binding in the lower TMD site is responsible for basal Hh signalling. Cholesterol binding to the CRD site alone also induces limited Hh activation. Dual-site occupancy is required for full Hh pathway activation.
b. Constitutive activity of SMOΔCRD is abolished by mutation of the lower TMD site.

6.1.3. Cholesterol access to Smoothened

The first observations of membrane cholesterol interactions with transmembrane receptors were made over 30 years ago [241]. Since this discovery it has become clear

that cholesterol is involved in the stabilisation of a range of GPCRs through transient and dynamic interactions with the outer transmembrane helices [151,242-245]. For this reason, it (or a derivative such as cholesterol hemisuccinate) is often added during purification or crystallisation. However, SMO is unique in that cholesterol binds away from the membrane-exposed outer helices, in the core of the transmembrane bundle and in the extracellular CRD [59-61,94,95,168]. Cholesterol binding at these sites is responsible for endogenous activation of both the receptor and the Hh pathway, a clear distinction from the general stabilisation seen for other GPCRs [60,61]. An important unanswered question in Hh signalling is how the availability of such an abundant cellular component is controlled to the extent that it can act as an agonist for such a tightly regulated pathway.

When considering the regulation of cholesterol availability, we must also address the presence of two SMO binding sites for the same ligand. As we have shown these sites have differing roles, we must explore how each is regulated separately. It should not be possible to rely directly upon ligand availability – as this would be the same for each site. However, SMOs multidomain structure provides distinct environments for each binding site. The CRD site is a shallow hydrophobic groove, ideally suited to sterol binding [59-61,97,100,165]. However, it sits approximately 12Å above the membrane [59]. In contrast, the lower TMD site is buried within the heptahelical bundle of SMO, surrounded by the hydrophobic environment of the plasma membrane, also a potential cholesterol source. One could assume that basal Hh

signalling is simply a result of membrane cholesterol interacting with the lower TMD site. The fact that this site is only capable of activating Hh to a low level may explain how SMO can exist in a cholesterol-rich membrane environment without displaying increased levels of Hh signalling. However, this does not answer the question of how cholesterol accesses the CRD site, or how the lower TMD site contributes to SHH-mediated activation of Hh (albeit to a lesser degree than the CRD site). To answer these questions, we must consider both the mechanism of action of Patched (PTCH) and a central feature of the Hh pathway; the primary cilium.

The inhibition of SMO by PTCH is a key step in the Hh pathway as it regulates the transduction of the signal across the plasma membrane. Despite its importance it is not well understood. It has long been predicted that PTCH regulates SMO by restricting access to a small molecule ligand, largely due to the lack of a physical interaction between the two proteins and the non-stoichiometric mode of inhibition [183,246]. Multiple studies have now shown that cholesterol activates SMO and the Hh pathway and it is hence hypothesised to be the endogenous agonist [59-61,168]. Recent studies of PTCH have supported this theory and highlighted the possibility that PTCH functions as a cholesterol transporter [54-58,62,63]. Structural data shows a putative tunnel from the outer membrane leaflet to the extracellular domain, providing a potential route for cholesterol transport between the extracellular region and the membrane [247]. Despite this, it is unclear whether PTCH simply

depletes membrane cholesterol or adjusts the abundance between the inner and outer leaflets – neither of which provide a clear method of SMO regulation [240].

To fully explore this question, we must also consider the role of the primary cilium. In Hh signalling, several steps occur within the primary cilium; in the resting state PTCH is localised to the membrane region at the base of the cilium and the ciliary pocket, upon inhibition and internalisation of PTCH, active SMO moves to the cilia membrane [4,66,114,211,246]. It is possible that the role of the cilia is simply to bring the necessary components for Hh signalling into proximity. However, it has been suggested that the nature of the ciliary membrane may also play a role [68,101,240]. Unfortunately, the exact composition of the membrane is not clear, though the cholesterol abundance is thought to be lower than the plasma membrane. We also know that the cilia membrane contains "lipid rafts"; cholesterol-based lipid microdomains regulated by gangliosides and sphingolipids [248,249]. It was recently shown that cilia membranes have a higher proportion of these "lipid rafts" than the plasma membrane, and that this sequestration leads to lower cholesterol accessibility [240]. The combination of PTCH-driven cholesterol efflux and sequestration of any remaining leads to a membrane environment with very low accessible cholesterol. This is likely a key feature of the primary cilium which prevents inappropriate activation and makes it suitable for Hh signalling. This ultra-low cholesterol environment would therefore be highly sensitive to PTCH

inhibition, leading to a significant change in the local concentration of accessible cholesterol.

6.1.4. Activation of Smoothened

Whilst we cannot clarify the mechanism by which PTCH activates SMO, we have added valuable information about the role of the lower TMD binding site. We also highlight the regulatory role of the CRD in limiting basal Hh signalling. The two-site model and the positive cooperation displayed suggest a "double-lock" mechanism of SMO activation (*Figure 6.2.a*). With a single cholesterol molecule bound in either site, modest activation of the pathway is seen, however dual site binding causes a substantial increase and full activation of the pathway. This system prevents inappropriate activation of the pathway, which is crucial to avoid outcomes such as tumorigenesis.

6.2. Smoothened downstream signalling

6.2.1. Activation resembles class A GPCRs

A number of active SMO structures have been made available recently, showing SMO binding to a G_i-protein and activating nanobody [69,94,95]. From these structures we can see the classic GPCR activation signature – the outward bend of TM6, though a smaller movement is observed in comparison to some class A GPCRs (*Figure 6.3*) [69]. The structures have also questioned the so-called "ionic lock" as it appears to be unbroken in some active structures. This may prove to be a key difference between SMO and class A GPCRs. At the time of writing, there is no active structure with an empty upper TMD site. As this site has not been shown to play a role in endogenous signalling, it would be interesting to compare structures with and without a ligand bound here. Particularly interesting would be the dual-cholesterol bound state of SMO, as we believe this would give a snapshot closest to that of endogenously activated SMO.

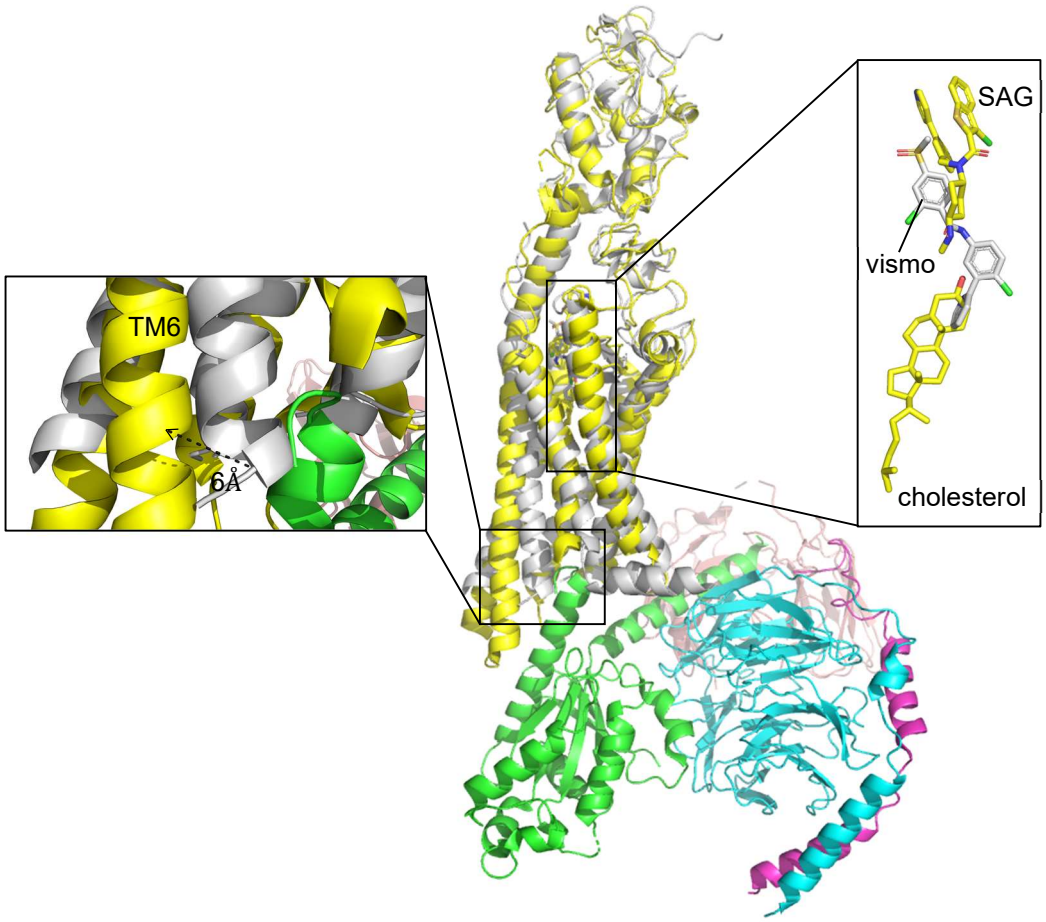


Figure 6.3. Activation of Smoothened
Structure of inactive hSMO bound to vismodegib (grey) from PDB 5L7I superimposed with active hSMO bound to SAG and cholesterol (yellow) and Gi-protein α-aubunit (green), β-subunit (cyan) and γ-subunit (magenta). Insets show movement of TM6 indicative of GPCR activation and the relative positions of ligands in each structure.

6.2.2. Challenges to studying activation *in vitro*

We have also noted some possible challenges of *in vitro* SMO studies. The delayed revelation of the lower TMD cholesterol binding site raises questions about its occupancy in previous studies. The first structures of SMO were obtained using SMO constructs which were locked in an inactive conformation, using the V329F mutation or a BRIL domain [59,103,104,250]. We know from recent structural data and

mutational analysis that this mutant would not bind cholesterol in the TMD due to steric clashes (*Figure 4.8.f*)[94,95]. However, the first active structure of SMO used non-mutated SMO and still did not observe this cholesterol molecule [69]. In the paper, the authors suggest the ligand added (24(S),25-epoxycholesterol) is bound across the upper and lower TMD sites, though the density is quite poor and inconclusive and the structure was only determined to medium to low resolution (~5 Å resolution in the SMO helical bundle and no density for the CRD domain) [69]. This could indicate partial occupancy of the two ligands, if they compete for the sterol site. This contributes extra complexity to studies of SMO activation, as we must consider whether cholesterol is bound at the lower TMD site, a variable which is not trivial to measure.

6.2.3. Fine-tuning Smoothened activation

Another area of interest within Hh signalling is how the signal is "tuned", that is to say, how full and partial activation are selectively achieved. As discussed above, there is evidence to suggest that single-site binding induces partial activation of the pathway, whereas dual-site binding is necessary to unlock the SMO "double-lock" and stimulate full activation (*Figure 4.8.i, 6.2.a*). We must now explore what differences in each SMO state can be translated to potency of downstream activation. Do these binding events induce different conformations of SMO? If so, are these conformations biased toward different intracellular SMO binders? Do the

binding events induce the same conformation with differing stability or longevity? We explored this through studies of nanobody 8 binding to SMO in the presence of site selective ligands (*Figure 5.7.b*). This nanobody was identified as binding to SMO when activated by SAG and cholesterol in both the CRD and lower TMD site [94]. We have highlighted that nanobody 8 also binds to SMO when activated via each of these three sites independently (*Figure 5.7.b*). These data suggest that agonist binding at each site does not induce distinct SMO conformations, though this may be confirmed through future structural studies. We must also consider the presence of an endogenous ligand, such as cholesterol bound in the lower TMD site, as discussed above. The capacity for KK174 induced binding suggests that cholesterol is not bound in this purified SMO sample (*Figure 5.7.b*). However, we may speculate that the lower amount of bound SMO in the presence of KK174 in comparison to SAG or 20(S)-OHC indicates a partial activation of SMO. It is possible that a population of SMO bound to cholesterol in the lower TMD site exists, decreasing the efficacy of KK174 and enhancing that of SAG or 20(S)-OHC. However, due to the non-quantitative nature of this experiment we cannot formally conclude as such. This will be a future area to explore, perhaps making use of site-specific ligands and point mutations.

6.3. Overcoming resistance to SMO-targeted anticancer treatments

6.3.1. Targeting the vismodegib binding site

Basal cell carcinoma (BCC) is the most common form of skin cancer and the most frequently occurring of all cancer types. It is predominantly driven by hyperactive Hh signalling caused by loss of function PTCH mutations or occasionally constitutively activating SMO mutation [251]. Vismodegib was approved by the FDA in 2012 for use against locally advanced or metastatic BCC and proved to be effective in approximately 50% of cases [178]. However, within a year of treatment, drug resistance develops in 20% of patients, leading to tumour regrowth (*Figure 6.4*)[104,178,251]. Interestingly, this acquired resistance is not due to pathway switching, as often seen in tumours, but specific mutations at the vismodegib binding site of SMO (*Figure 4.2.b*) [59]. The pattern of these mutations shows residues across the vismodegib binding site undergoing mutation to confer resistance [59]. A key conclusion from this phenomenon is that the integrity of this site cannot be necessary for Hh activation, as the observed cancer mutations that prevent vismodegib binding still retain responsiveness to normal Hh signal activation. If it was required, the extensive mutations seen would almost certainly be detrimental to Hh activation, and hence tumour growth. Unfortunately, this means that the upper TMD site targeted by vismodegib is highly susceptible to mutations that do

not hinder SMO signalling, hence failing to prevent the hyperactive signalling driving tumour growth.

6.3.2. Existing CRD inhibitors pave the way

Our strategy to target the SMO CRD site stemmed partially from our observations of mutations around the TMD site and our prediction that the endogenous binding site would be less susceptible to this (*Figure 6.4*). However, we were also encouraged by previous data showing the behaviour of CRD-binding SMO antagonists. Notable small molecules are 20(R)-yne, 20-keto-yne, ALLO-2 and budesonide [97,165,252,253]. It has been shown that these antagonists are able to inhibit not just wild-type SMO, but vismodegib resistant and constitutively active mutant forms of SMO [97,253]. This is highly encouraging for the prospect of developing SMO CRD inhibitors. So why aren't these molecules being used as anti-cancer treatments? The main issue with these molecules is that they are all sterol or sterol-like molecules [99,165,252]. This is unsurprising as they target a binding site designed for cholesterol. However, this means that they are largely hydrophobic and insoluble, which is not ideal for a drug candidate. Another problem is that they are relatively weak binders – when studying 20(R)-yne and 20-keto-yne, a 250-fold higher concentration than vismodegib was necessary to achieve comparable inhibition of SMO wild-type [97]. Taken together, these suggest a novel lead compound is necessary for drug development.

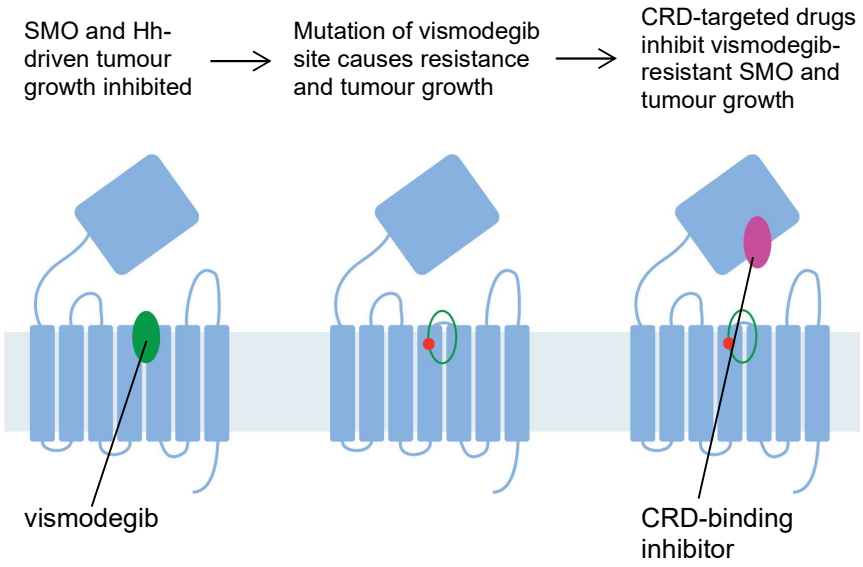


Figure 6.4. Overcoming vismodegib resistance
Treatment of Hh-potentiated cancers with vismodegib is often successful initially. Resistance can develop through mutation of residues in the SMO vismodegib binding site, allowing tumours to grow and spread. Our strategy uses CRD-targeted inhibitors to bypass the vismodegib site and inactivate Hh signalling.

6.3.3. A strategy to overcome vismodegib resistance

We have developed and validated a novel FRET-based assay to detect SMO CRD binders, ideally suited to high-throughput screening (see Chapter 4). We propose that this assay could be used to identify and develop SMO inhibitors as next-generation drugs to tackle vismodegib resistant tumour growth (*Figure 6.4*). There are several steps that would be necessary to achieve this (*Figure 6.5*). In its current form, our assay is run in a 384-well format, though due to its robust nature shown by an excellent Z-factor score, we believe it would be amenable to miniaturisation (*Figure 4.5*) [189]. This would enable a higher number of compounds to be tested whilst using less material. While a classic randomised compound screen would be

valuable, we also suggest a targeted approach would be beneficial. Due to the unique nature of the SMO CRD binding groove and the pool of known agonists and antagonists, a structure-activity relationship analysis could be worthwhile. This may enable us to detect the differences between similar agonists and antagonists to identify key binding moieties. Using SMO structural data, we can also dock potential molecules into the CRD site and compare this to inactive and active SMO conformations. Once the HTRF assay has been used to identify competing molecules, we must analyse them in terms of activity. This assay is purely able to detect binding, not provide information on the effect upon signalling. This should be explored using a GLI mRNA Hh activation assay to study potential hits. If inhibitory molecules are detected in this study, they should then be analysed to estimate binding affinity and suitability for *in vivo* use. It is likely they would need to be modified to increase the affinity for SMO or enhance their solubility. Once these steps have been completed, it is possible that we may identify a molecule with the potential for use against vismodegib-resistant cancers.

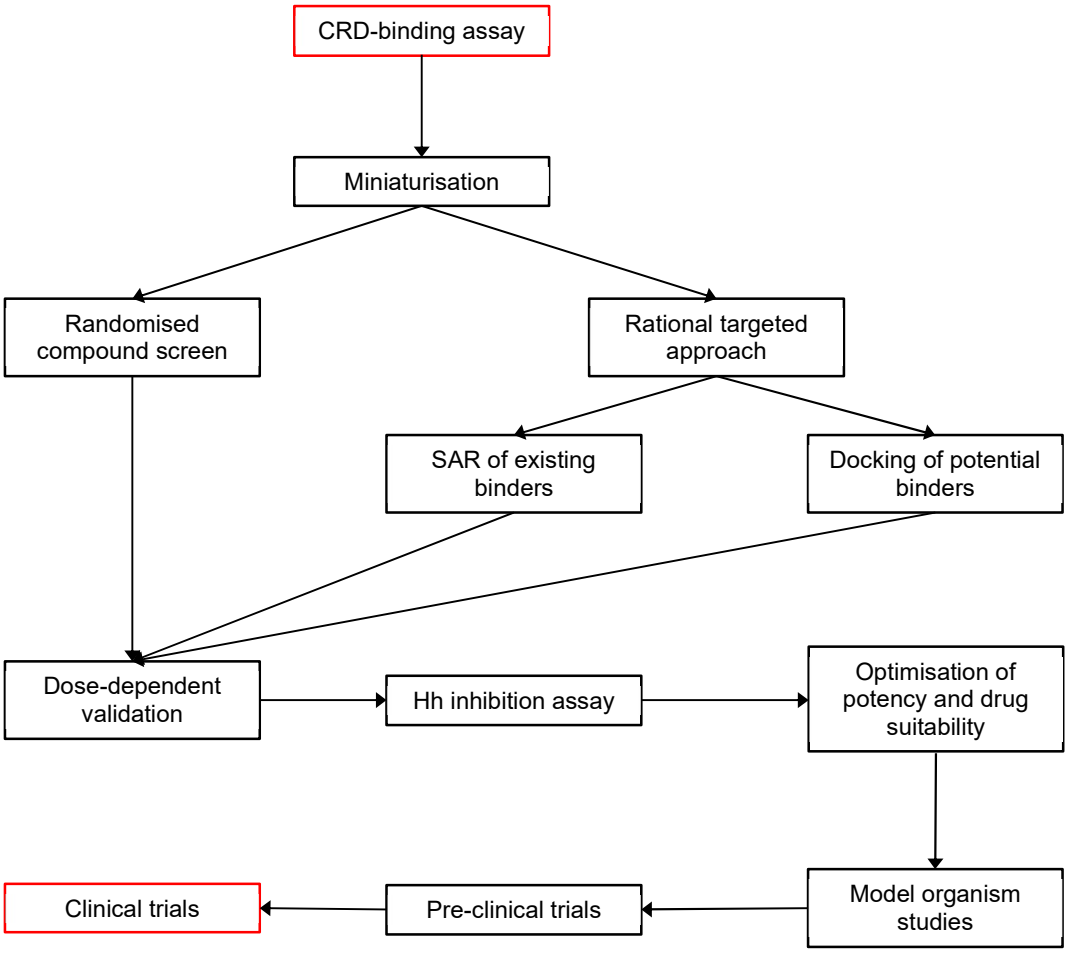


Figure 6.5. Possible pathway from assay development to clinical trials
A suggested approach to developing a SMO CRD inhibitor using the novel SMO binding assay we have designed.

6.4. Summary

In this thesis, I have explored the unknown mechanism of SMO signal transduction. To do this I first ensured I could make suitable quantities of pure, wild-type SMO. I optimised a new system to use a stable cell line for SMO production and increased the yields approximately 8-fold. I also studied the use of detergent-free systems for SMO solubilisation, which will likely be crucial for future sterol binding studies. I explored the activation of SMO via three distinct small molecule binding sites and have uncovered valuable information about their allosteric relationship. This is supported by data collected by collaborators and we propose a role for the lower TMD cholesterol site in basal Hh signalling. Finally, we develop a novel FRET-based binding assay to specifically detect SMO CRD binders. We predict that this may be used to develop novel SMO inhibitors for use against drug-resistant cancers.

References

1 Turing, A. M. The chemical basis of morphogenesis. 1953. *Bull Math Biol* **52**, 153-197; discussion 119-152, doi:10.1007/BF02459572 (1990).

2 Maini, P. K., Woolley, T. E., Baker, R. E., Gaffney, E. A. & Lee, S. S. Turing's model for biological pattern formation and the robustness problem. *Interface Focus* **2**, 487-496, doi:10.1098/rsfs.2011.0113 (2012).

3 Tabata, T. & Takei, Y. Morphogens, their identification and regulation. *Development* **131**, 703-712, doi:10.1242/dev.01043 (2004).

4 Briscoe, J. & Thérond, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat Rev Mol Cell Biol* **14**, 416-429, doi:10.1038/nrm3598 (2013).

5 Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* **25**, 1-47, doi:10.1016/s0022-5193(69)80016-0 (1969).

6 Sharpe, J. Wolpert's French Flag: what's the problem? *Development* **146**, doi:10.1242/dev.185967 (2019).

7 Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a wingless morphogen gradient. *Cell* **87**, 833-844, doi:10.1016/s0092-8674(00)81991-1 (1996).

8 Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407-413, doi:10.1038/383407a0 (1996).

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

10 Lee, J. J., von Kessler, D. P., Parks, S. & Beachy, P. A. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell* **71**, 33-50, doi:10.1016/0092-8674(92)90264-d (1992).

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

12 Tabata, T., Eaton, S. & Kornberg, T. B. The Drosophila hedgehog gene is expressed specifically in posterior compartment cells and is a target of engrailed regulation. *Genes Dev* **6**, 2635-2645, doi:10.1101/gad.6.12b.2635 (1992).

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

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

with polarizing activity in zebrafish embryos. *Cell* **75**, 1431-1444, doi:10.1016/0092-8674(93)90628-4 (1993).
15 Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401-1416, doi:10.1016/0092-8674(93)90626-2 (1993).
16 Zugasti, O., Rajan, J. & Kuwabara, P. E. The function and expansion of the Patched- and Hedgehog-related homologs in C. elegans. *Genome Res* **15**, 1402-1410, doi:10.1101/gr.3935405 (2005).
17 Cohn, M. J. & Tickle, C. Limbs: a model for pattern formation within the vertebrate body plan. *Trends Genet* **12**, 253-257 (1996).
18 Roessler, E. *et al.* Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet* **14**, 357-360, doi:10.1038/ng1196-357 (1996).
19 Ming, J. E., Roessler, E. & Muenke, M. Human developmental disorders and the Sonic hedgehog pathway. *Mol Med Today* **4**, 343-349 (1998).
20 Hui, C. C. & Angers, S. Gli proteins in development and disease. *Annu Rev Cell Dev Biol* **27**, 513-537, doi:10.1146/annurev-cellbio-092910-154048 (2011).
21 Varjosalo, M. & Taipale, J. Hedgehog: functions and mechanisms. *Genes Dev* **22**, 2454-2472, doi:10.1101/gad.1693608 (2008).
22 Bitgood, M. J., Shen, L. & McMahon, A. P. Sertoli cell signaling by Desert hedgehog regulates the male germline. *Curr Biol* **6**, 298-304, doi:10.1016/s0960-9822(02)00480-3 (1996).
23 Wijgerde, M., Ooms, M., Hoogerbrugge, J. W. & Grootegoed, J. A. Hedgehog signaling in mouse ovary: Indian hedgehog and desert hedgehog from granulosa cells induce target gene expression in developing theca cells. *Endocrinology* **146**, 3558-3566, doi:10.1210/en.2005-0311 (2005).
24 Yao, H. H., Whoriskey, W. & Capel, B. Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes Dev* **16**, 1433-1440, doi:10.1101/gad.981202 (2002).
25 St-Jacques, B., Hammerschmidt, M. & McMahon, A. P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev* **13**, 2072-2086, doi:10.1101/gad.13.16.2072 (1999).
26 Vortkamp, A. *et al.* Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* **273**, 613-622, doi:10.1126/science.273.5275.613 (1996).
27 Colnot, C. *et al.* Indian hedgehog synchronizes skeletal angiogenesis and perichondrial maturation with cartilage development. *Development* **132**, 1057-1067, doi:10.1242/dev.01649 (2005).
28 Hellemans, J. *et al.* Homozygous mutations in IHH cause acrocapitofemoral dysplasia, an autosomal recessive disorder with cone-shaped epiphyses in hands and hips. *Am J Hum Genet* **72**, 1040-1046, doi:10.1086/374318 (2003).
29 Hammerschmidt, M., Brook, A. & McMahon, A. P. The world according to hedgehog. *Trends Genet* **13**, 14-21, doi:10.1016/s0168-9525(96)10051-2 (1997).

30 Fitzky, B. U. *et al.* Mutations in the Delta7-sterol reductase gene in patients with the Smith-Lemli-Opitz syndrome. *Proc Natl Acad Sci U S A* **95**, 8181-8186, doi:10.1073/pnas.95.14.8181 (1998).
31 Wassif, C. A. *et al.* Mutations in the human sterol delta7-reductase gene at 11q12-13 cause Smith-Lemli-Opitz syndrome. *Am J Hum Genet* **63**, 55-62, doi:10.1086/301936 (1998).
32 Waterham, H. R. *et al.* Smith-Lemli-Opitz syndrome is caused by mutations in the 7-dehydrocholesterol reductase gene. *Am J Hum Genet* **63**, 329-338, doi:10.1086/301982 (1998).
33 Kelley, R. L. *et al.* Holoprosencephaly in RSH/Smith-Lemli-Opitz syndrome: does abnormal cholesterol metabolism affect the function of Sonic Hedgehog? *Am J Med Genet* **66**, 478-484, doi:10.1002/(SICI)1096-8628(19961230)66:4<478::AID-AJMG22>3.0.CO;2-Q (1996).
34 Willnow, T. E. *et al.* Defective forebrain development in mice lacking gp330/megalin. *Proc Natl Acad Sci U S A* **93**, 8460-8464, doi:10.1073/pnas.93.16.8460 (1996).
35 Petrova, R. & Joyner, A. L. Roles for Hedgehog signaling in adult organ homeostasis and repair. *Development* **141**, 3445-3457, doi:10.1242/dev.083691 (2014).
36 Traiffort, E., Angot, E. & Ruat, M. Sonic Hedgehog signaling in the mammalian brain. *J Neurochem* **113**, 576-590, doi:10.1111/j.1471-4159.2010.06642.x (2010).
37 Hahn, H. *et al.* Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* **85**, 841-851, doi:10.1016/s0092-8674(00)81268-4 (1996).
38 Johnson, R. L. *et al.* Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* **272**, 1668-1671, doi:10.1126/science.272.5268.1668 (1996).
39 Unden, A. B. *et al.* Mutations in the human homologue of Drosophila patched (PTCH) in basal cell carcinomas and the Gorlin syndrome: different in vivo mechanisms of PTCH inactivation. *Cancer Res* **56**, 4562-4565 (1996).
40 Reifenberger, J. *et al.* Missense mutations in SMOH in sporadic basal cell carcinomas of the skin and primitive neuroectodermal tumors of the central nervous system. *Cancer Res* **58**, 1798-1803 (1998).
41 Wong, S. Y. & Reiter, J. F. Wounding mobilizes hair follicle stem cells to form tumors. *Proc Natl Acad Sci U S A* **108**, 4093-4098, doi:10.1073/pnas.1013098108 (2011).
42 Youssef, K. K. *et al.* Identification of the cell lineage at the origin of basal cell carcinoma. *Nat Cell Biol* **12**, 299-305, doi:10.1038/ncb2031 (2010).
43 Manoranjan, B. *et al.* Medulloblastoma stem cells: modeling tumor heterogeneity. *Cancer Lett* **338**, 23-31, doi:10.1016/j.canlet.2012.07.010 (2013).

44 Remke, M. *et al.* TERT promoter mutations are highly recurrent in SHH subgroup medulloblastoma. *Acta Neuropathol* **126**, 917-929, doi:10.1007/s00401-013-1198-2 (2013).
45 Mann, R. K. & Beachy, P. A. Novel lipid modifications of secreted protein signals. *Annu Rev Biochem* **73**, 891-923, doi:10.1146/annurev.biochem.73.011303.073933 (2004).
46 Chen, X. *et al.* Processing and turnover of the Hedgehog protein in the endoplasmic reticulum. *J Cell Biol* **192**, 825-838, doi:10.1083/jcb.201008090 (2011).
47 Pepinsky, R. B. *et al.* Identification of a palmitic acid-modified form of human Sonic hedgehog. *J Biol Chem* **273**, 14037-14045 (1998).
48 Taylor, F. R. *et al.* Enhanced potency of human Sonic hedgehog by hydrophobic modification. *Biochemistry* **40**, 4359-4371 (2001).
49 Burke, R. *et al.* Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell* **99**, 803-815, doi:10.1016/s0092-8674(00)81677-3 (1999).
50 Ma, Y. *et al.* Hedgehog-mediated patterning of the mammalian embryo requires transporter-like function of dispatched. *Cell* **111**, 63-75, doi:10.1016/s0092-8674(02)00977-7 (2002).
51 Tian, H., Jeong, J., Harfe, B. D., Tabin, C. J. & McMahon, A. P. Mouse Disp1 is required in sonic hedgehog-expressing cells for paracrine activity of the cholesterol-modified ligand. *Development* **132**, 133-142, doi:10.1242/dev.01563 (2005).
52 Tukachinsky, H., Kuzmickas, R. P., Jao, C. Y., Liu, J. & Salic, A. Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand. *Cell Rep* **2**, 308-320, doi:10.1016/j.celrep.2012.07.010 (2012).
53 Beachy, P. A., Hymowitz, S. G., Lazarus, R. A., Leahy, D. J. & Siebold, C. Interactions between Hedgehog proteins and their binding partners come into view. *Genes Dev* **24**, 2001-2012, doi:10.1101/gad.1951710 (2010).
54 Qi, C., Di Minin, G., Vercellino, I., Wutz, A. & Korkhov, V. M. Structural basis of sterol recognition by human hedgehog receptor PTCH1. *bioRxiv*, 508325, doi:10.1101/508325 (2018).
55 Qi, X., Schmiege, P., Coutavas, E., Wang, J. & Li, X. Structures of human Patched and its complex with native palmitoylated sonic hedgehog. *Nature* **560**, 128-132, doi:10.1038/s41586-018-0308-7 (2018).
56 Qi, X., Schmiege, P., Coutavas, E. & Li, X. Two Patched molecules engage distinct sites on Hedgehog yielding a signaling-competent complex. *Science* **362**, doi:10.1126/science.aas8843 (2018).
57 Gong, X. *et al.* Structural basis for the recognition of Sonic Hedgehog by human Patched1. *Science* **361**, doi:10.1126/science.aas8935 (2018).
58 Rudolf, A. F. *et al.* The morphogen Sonic hedgehog inhibits its receptor Patched by a pincer grasp mechanism. *Nat Chem Biol* **15**, 975-982, doi:10.1038/s41589-019-0370-y (2019).

59 Byrne, E. F. *et al.* Structural basis of Smoothened regulation by its extracellular domains. *Nature* **535**, 517-522, doi:10.1038/nature18934 (2016).
60 Luchetti, G. *et al.* Cholesterol activates the G-protein coupled receptor Smoothened to promote morphogenetic signaling. *Elife* **5**, doi:10.7554/eLife.20304 (2016).
61 Huang, P. *et al.* Cellular Cholesterol Directly Activates Smoothened in Hedgehog Signaling. *Cell* **166**, 1176-1187.e1114, doi:10.1016/j.cell.2016.08.003 (2016).
62 Zhang, Y. *et al.* Structural Basis for Cholesterol Transport-like Activity of the Hedgehog Receptor Patched. *Cell* **175**, 1352-1364.e1314, doi:10.1016/j.cell.2018.10.026 (2018).
63 Kowatsch, C., Woolley, R. E., Kinnebrew, M., Rohatgi, R. & Siebold, C. Structures of vertebrate Patched and Smoothened reveal intimate links between cholesterol and Hedgehog signalling. *Curr Opin Struct Biol* **57**, 204-214, doi:10.1016/j.sbi.2019.05.015 (2019).
64 Carballo, G. B., Honorato, J. R., de Lopes, G. P. F. & Spohr, T. C. L. S. A highlight on Sonic hedgehog pathway. *Cell Commun Signal* **16**, 11, doi:10.1186/s12964-018-0220-7 (2018).
65 Chen, W. *et al.* Activity-dependent internalization of smoothened mediated by beta-arrestin 2 and GRK2. *Science* **306**, 2257-2260, doi:10.1126/science.1104135 (2004).
66 Kovacs, J. J. *et al.* Beta-arrestin-mediated localization of smoothened to the primary cilium. *Science* **320**, 1777-1781, doi:10.1126/science.1157983 (2008).
67 Riobo, N. A., Saucy, B., Dilizio, C. & Manning, D. R. Activation of heterotrimeric G proteins by Smoothened. *Proc Natl Acad Sci U S A* **103**, 12607-12612, doi:10.1073/pnas.0600880103 (2006).
68 Myers, B. R., Neahring, L., Zhang, Y., Roberts, K. J. & Beachy, P. A. Rapid, direct activity assays for Smoothened reveal Hedgehog pathway regulation by membrane cholesterol and extracellular sodium. *Proc Natl Acad Sci U S A* **114**, E11141-E11150, doi:10.1073/pnas.1717891115 (2017).
69 Qi, X. *et al.* Cryo-EM structure of oxysterol-bound human Smoothened coupled to a heterotrimeric G. *Nature* **571**, 279-283, doi:10.1038/s41586-019-1286-0 (2019).
70 Hammerschmidt, M. & McMahon, A. P. The effect of pertussis toxin on zebrafish development: a possible role for inhibitory G-proteins in hedgehog signaling. *Dev Biol* **194**, 166-171, doi:10.1006/dbio.1997.8796 (1998).
71 Ingham, P. W. Hedgehog signalling. *Curr Biol* **18**, R238-241, doi:10.1016/j.cub.2008.01.050 (2008).
72 Kroeze, W. K., Sheffler, D. J. & Roth, B. L. G-protein-coupled receptors at a glance. *J Cell Sci* **116**, 4867-4869, doi:10.1242/jcs.00902 (2003).
73 Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families.

Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **63**, 1256-1272, doi:10.1124/mol.63.6.1256 (2003).
74 Hauser, A. S. *et al.* Pharmacogenomics of GPCR Drug Targets. *Cell* **172**, 41-54.e19, doi:10.1016/j.cell.2017.11.033 (2018).
75 Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* **16**, 829-842, doi:10.1038/nrd.2017.178 (2017).
76 Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356-363, doi:10.1038/nature08144 (2009).
77 Kobilka, B. K. G protein coupled receptor structure and activation. *Biochim Biophys Acta* **1768**, 794-807, doi:10.1016/j.bbamem.2006.10.021 (2007).
78 Rasmussen, S. G. *et al.* Crystal structure of the β2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549-555, doi:10.1038/nature10361 (2011).
79 Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol Pharmacol* **93**, 251-258, doi:10.1124/mol.117.111062 (2018).
80 Basith, S. *et al.* Exploring G Protein-Coupled Receptors (GPCRs) Ligand Space via Cheminformatics Approaches: Impact on Rational Drug Design. *Front Pharmacol* **9**, 128, doi:10.3389/fphar.2018.00128 (2018).
81 Pándy-Szekeres, G. *et al.* GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res* **46**, D440-D446, doi:10.1093/nar/gkx1109 (2018).
82 Seifert, R., Wenzel-Seifert, K., Gether, U. & Kobilka, B. K. Functional differences between full and partial agonists: evidence for ligand-specific receptor conformations. *J Pharmacol Exp Ther* **297**, 1218-1226 (2001).
83 Corbisier, J., Huszagh, A., Galés, C., Parmentier, M. & Springael, J. Y. Partial Agonist and Biased Signaling Properties of the Synthetic Enantiomers J113863/UCB35625 at Chemokine Receptors CCR2 and CCR5. *J Biol Chem* **292**, 575-584, doi:10.1074/jbc.M116.757559 (2017).
84 Wright, S. C. *et al.* A conserved molecular switch in Class F receptors regulates receptor activation and pathway selection. *Nat Commun* **10**, 667, doi:10.1038/s41467-019-08630-2 (2019).
85 Vogel, R. *et al.* Functional role of the "ionic lock"--an interhelical hydrogen-bond network in family A heptahelical receptors. *J Mol Biol* **380**, 648-655, doi:10.1016/j.jmb.2008.05.022 (2008).
86 Ballesteros, J. A. *et al.* Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J Biol Chem* **276**, 29171-29177, doi:10.1074/jbc.M103747200 (2001).
87 Rasmussen, S. G. *et al.* Mutation of a highly conserved aspartic acid in the beta2 adrenergic receptor: constitutive activation, structural instability, and

conformational rearrangement of transmembrane segment 6. *Mol Pharmacol* **56**, 175-184, doi:10.1124/mol.56.1.175 (1999).
88 Okada, T. *et al.* Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc Natl Acad Sci U S A* **99**, 5982-5987, doi:10.1073/pnas.082666399 (2002).
89 Barak, L. S., Ménard, L., Ferguson, S. S., Colapietro, A. M. & Caron, M. G. The conserved seven-transmembrane sequence NP(X)2,3Y of the G-protein-coupled receptor superfamily regulates multiple properties of the beta 2-adrenergic receptor. *Biochemistry* **34**, 15407-15414, doi:10.1021/bi00047a003 (1995).
90 De Lean, A., Stadel, J. M. & Lefkowitz, R. J. A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled beta-adrenergic receptor. *J Biol Chem* **255**, 7108-7117 (1980).
91 Ross, E. M., Maguire, M. E., Sturgill, T. W., Biltonen, R. L. & Gilman, A. G. Relationship between the beta-adrenergic receptor and adenylate cyclase. *J Biol Chem* **252**, 5761-5775 (1977).
92 Weis, W. I. & Kobilka, B. K. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu Rev Biochem* **87**, 897-919, doi:10.1146/annurev-biochem-060614-033910 (2018).
93 Byrne, E. F., Luchetti, G., Rohatgi, R. & Siebold, C. Multiple ligand binding sites regulate the Hedgehog signal transducer Smoothened in vertebrates. *Curr Opin Cell Biol* **51**, 81-88, doi:10.1016/j.ceb.2017.10.004 (2018).
94 Deshpande, I. *et al.* Smoothened stimulation by membrane sterols drives Hedgehog pathway activity. *Nature* **571**, 284-288, doi:10.1038/s41586-019-1355-4 (2019).
95 Qi, X., Friedberg, L., De Bose-Boyd, R., Long, T. & Li, X. Sterols in an intramolecular channel of Smoothened mediate Hedgehog signaling. *Nat Chem Biol* **16**, 1368-1375, doi:10.1038/s41589-020-0646-2 (2020).
96 Yancey, P. G. *et al.* Cellular cholesterol efflux mediated by cyclodextrins. Demonstration Of kinetic pools and mechanism of efflux. *J Biol Chem* **271**, 16026-16034, doi:10.1074/jbc.271.27.16026 (1996).
97 Nachtergaele, S. *et al.* Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *Elife* **2**, e01340, doi:10.7554/eLife.01340 (2013).
98 Dwyer, J. R. *et al.* Oxysterols are novel activators of the hedgehog signaling pathway in pluripotent mesenchymal cells. *J Biol Chem* **282**, 8959-8968, doi:10.1074/jbc.M611741200 (2007).
99 Nachtergaele, S. *et al.* Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nat Chem Biol* **8**, 211-220, doi:10.1038/nchembio.765 (2012).
100 Nedelcu, D., Liu, J., Xu, Y., Jao, C. & Salic, A. Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nat Chem Biol* **9**, 557-564, doi:10.1038/nchembio.1290 (2013).

101 Raleigh, D. R. *et al.* Cilia-Associated Oxysterols Activate Smoothened. *Mol Cell* **72**, 316-327.e315, doi:10.1016/j.molcel.2018.08.034 (2018).
102 Xiao, X. *et al.* Cholesterol Modification of Smoothened Is Required for Hedgehog Signaling. *Mol Cell* **66**, 154-162.e110, doi:10.1016/j.molcel.2017.02.015 (2017).
103 Wang, C. *et al.* Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **497**, 338-343, doi:10.1038/nature12167 (2013).
104 Wang, C. *et al.* Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nat Commun* **5**, 4355, doi:10.1038/ncomms5355 (2014).
105 Hedger, G. *et al.* Cholesterol Interaction Sites on the Transmembrane Domain of the Hedgehog Signal Transducer and Class F G Protein-Coupled Receptor Smoothened. *Structure* **27**, 549-559.e542, doi:10.1016/j.str.2018.11.003 (2019).
106 Robbins, D. J., Fei, D. L. & Riobo, N. A. The Hedgehog signal transduction network. *Sci Signal* **5**, re6, doi:10.1126/scisignal.2002906 (2012).
107 Chen, Y. *et al.* Sonic Hedgehog dependent phosphorylation by CK1α and GRK2 is required for ciliary accumulation and activation of smoothened. *PLoS Biol* **9**, e1001083, doi:10.1371/journal.pbio.1001083 (2011).
108 Meloni, A. R. *et al.* Smoothened signal transduction is promoted by G protein-coupled receptor kinase 2. *Mol Cell Biol* **26**, 7550-7560, doi:10.1128/MCB.00546-06 (2006).
109 Yang, C., Chen, W., Chen, Y. & Jiang, J. Smoothened transduces Hedgehog signal by forming a complex with Evc/Evc2. *Cell Res* **22**, 1593-1604, doi:10.1038/cr.2012.134 (2012).
110 Dorn, K. V., Hughes, C. E. & Rohatgi, R. A Smoothened-Evc2 complex transduces the Hedgehog signal at primary cilia. *Dev Cell* **23**, 823-835, doi:10.1016/j.devcel.2012.07.004 (2012).
111 Tukachinsky, H., Lopez, L. V. & Salic, A. A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu-Gli protein complexes. *J Cell Biol* **191**, 415-428, doi:10.1083/jcb.201004108 (2010).
112 Humke, E. W., Dorn, K. V., Milenkovic, L., Scott, M. P. & Rohatgi, R. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes Dev* **24**, 670-682, doi:10.1101/gad.1902910 (2010).
113 Ogden, S. K. *et al.* G protein Galphai functions immediately downstream of Smoothened in Hedgehog signalling. *Nature* **456**, 967-970, doi:10.1038/nature07459 (2008).
114 Barzi, M., Berenguer, J., Menendez, A., Alvarez-Rodriguez, R. & Pons, S. Sonic-hedgehog-mediated proliferation requires the localization of PKA to the cilium base. *J Cell Sci* **123**, 62-69, doi:10.1242/jcs.060020 (2010).
115 Tuson, M., He, M. & Anderson, K. V. 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**, 4921-4930, doi:10.1242/dev.070805 (2011).
116 Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr* **62**, 1243-1250, doi:10.1107/S0907444906029799 (2006).
117 Molday, R. S. & MacKenzie, D. Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. *Biochemistry* **22**, 653-660 (1983).
118 Fukuchi, K. *et al.* Activity assays of nine heterogeneous promoters in neural and other cultured cells. *In Vitro Cell Dev Biol Anim* **30A**, 300-305, doi:10.1007/bf02631450 (1994).
119 Elegheert, J. *et al.* Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nat Protoc* **13**, 2991-3017, doi:10.1038/s41596-018-0075-9 (2018).
120 Kim, H. J., Gatz, C., Hillen, W. & Jones, T. R. Tetracycline repressor-regulated gene repression in recombinant human cytomegalovirus. *J Virol* **69**, 2565-2573 (1995).
121 Carpenter, B., Nehmé, R., Warne, T., Leslie, A. G. & Tate, C. G. Structure of the adenosine A(2A) receptor bound to an engineered G protein. *Nature* **536**, 104-107, doi:10.1038/nature18966 (2016).
122 Carpenter, B. & Tate, C. G. Engineering a minimal G protein to facilitate crystallisation of G protein-coupled receptors in their active conformation. *Protein Eng Des Sel*, doi:10.1093/protein/gzw049 (2016).
123 Nehme, R. *et al.* Mini-G proteins: Novel tools for studying GPCRs in their active conformation. *PLoS One* **12**, e0175642, doi:10.1371/journal.pone.0175642 (2017).
124 Pardon, E. *et al.* A general protocol for the generation of Nanobodies for structural biology. *Nat Protoc* **9**, 674-693, doi:10.1038/nprot.2014.039 (2014).
125 Lo, M. C. *et al.* Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **332**, 153-159, doi:10.1016/j.ab.2004.04.031 (2004).
126 Cull, M. G. & Schatz, P. J. Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol* **326**, 430-440 (2000).
127 GraphPad. *How should I compute an "average" EC50 for several experiments? How can I express its uncertainty or standard deviation?*, <https://www.graphpad.com/support/faq/how-should-i-compute-an-average-ec50-for-several-experiments-how-can-i-express-its-uncertainty-or-standard-deviation/> (
128 Corin, K., Tegler, L. T. & Koutsopoulos, S. G-protein-coupled receptor expression and purification. *Methods Mol Biol* **1129**, 461-486, doi:10.1007/978-1-62703-977-2_33 (2014).

129 Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* **1666**, 105-117, doi:10.1016/j.bbamem.2004.04.011 (2004).
130 Wiseman, D. N. *et al.* Expression and purification of recombinant G protein-coupled receptors: A review. *Protein Expr Purif* **167**, 105524, doi:10.1016/j.pep.2019.105524 (2020).
131 Milić, D. & Veprintsev, D. B. Large-scale production and protein engineering of G protein-coupled receptors for structural studies. *Front Pharmacol* **6**, 66, doi:10.3389/fphar.2015.00066 (2015).
132 CLARK, A. J. & BLOCK, K. The absence of sterol synthesis in insects. *J Biol Chem* **234**, 2578-2582 (1959).
133 Kurzchalia, T. V. & Ward, S. Why do worms need cholesterol? *Nat Cell Biol* **5**, 684-688, doi:10.1038/ncb0803-684 (2003).
134 Warne, T., Chirnside, J. & Schertler, G. F. Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. *Biochim Biophys Acta* **1610**, 133-140, doi:10.1016/s0005-2736(02)00716-2 (2003).
135 Chien, E. Y. *et al.* Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**, 1091-1095, doi:10.1126/science.1197410 (2010).
136 Thomas, J. A. & Tate, C. G. Quality control in eukaryotic membrane protein overproduction. *J Mol Biol* **426**, 4139-4154, doi:10.1016/j.jmb.2014.10.012 (2014).
137 Li, Z., Michael, I. P., Zhou, D., Nagy, A. & Rini, J. M. Simple piggyBac transposon-based mammalian cell expression system for inducible protein production. *Proc Natl Acad Sci U S A* **110**, 5004-5009, doi:10.1073/pnas.1218620110 (2013).
138 Matasci, M., Baldi, L., Hacker, D. L. & Wurm, F. M. The PiggyBac transposon enhances the frequency of CHO stable cell line generation and yields recombinant lines with superior productivity and stability. *Biotechnol Bioeng* **108**, 2141-2150, doi:10.1002/bit.23167 (2011).
139 Dukkipati, A., Park, H. H., Waghray, D., Fischer, S. & Garcia, K. C. BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expr Purif* **62**, 160-170, doi:10.1016/j.pep.2008.08.004 (2008).
140 Boyce, F. M. & Bucher, N. L. Baculovirus-mediated gene transfer into mammalian cells. *Proc Natl Acad Sci U S A* **93**, 2348-2352, doi:10.1073/pnas.93.6.2348 (1996).
141 Andréll, J. & Tate, C. G. Overexpression of membrane proteins in mammalian cells for structural studies. *Mol Membr Biol* **30**, 52-63, doi:10.3109/09687688.2012.703703 (2013).

142 Arachea, B. T. *et al.* Detergent selection for enhanced extraction of membrane proteins. *Protein Expr Purif* **86**, 12-20, doi:10.1016/j.pep.2012.08.016 (2012).
143 Tribet, C., Audebert, R. & Popot, J. L. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc Natl Acad Sci U S A* **93**, 15047-15050, doi:10.1073/pnas.93.26.15047 (1996).
144 Le Bon, C., Marconnet, A., Masscheleyn, S., Popot, J. L. & Zoonens, M. Folding and stabilizing membrane proteins in amphipol A8-35. *Methods* **147**, 95-105, doi:10.1016/j.ymeth.2018.04.012 (2018).
145 Schuler, M. A., Denisov, I. G. & Sligar, S. G. Nanodiscs as a new tool to examine lipid-protein interactions. *Methods Mol Biol* **974**, 415-433, doi:10.1007/978-1-62703-275-9_18 (2013).
146 Ritchie, T. K. *et al.* Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol* **464**, 211-231, doi:10.1016/S0076-6879(09)64011-8 (2009).
147 Wong, J. P., Reboul, E., Molday, R. S. & Kast, J. A carboxy-terminal affinity tag for the purification and mass spectrometric characterization of integral membrane proteins. *J Proteome Res* **8**, 2388-2396, doi:10.1021/pr801008c (2009).
148 Molday, L. L. & Molday, R. S. 1D4: a versatile epitope tag for the purification and characterization of expressed membrane and soluble proteins. *Methods Mol Biol* **1177**, 1-15, doi:10.1007/978-1-4939-1034-2_1 (2014).
149 Vergis, J. M. & Wiener, M. C. The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. *Protein Expr Purif* **78**, 139-142, doi:10.1016/j.pep.2011.04.011 (2011).
150 Chang, V. T. *et al.* Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15**, 267-273, doi:10.1016/j.str.2007.01.011 (2007).
151 Cherezov, V. *et al.* High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265, doi:10.1126/science.1150577 (2007).
152 Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232-236, doi:10.1126/science.1219218 (2012).
153 Byrne, E. *Molecular mechanisms of hedgehog signal transduction by the G-protein coupled receptor smoothened* D. Phil thesis, University of Oxford, (2017).
154 Loll, P. J. Membrane proteins, detergents and crystals: what is the state of the art? *Acta Crystallogr F Struct Biol Commun* **70**, 1576-1583, doi:10.1107/S2053230X14025035 (2014).
155 Pedro, A. Q., Queiroz, J. A. & Passarinha, L. A. Smoothing membrane protein structure determination by initial upstream stage improvements. *Appl Microbiol Biotechnol* **103**, 5483-5500, doi:10.1007/s00253-019-09873-1 (2019).

156 Díaz Galicia, M. E., Aldehaiman, A., Hong, S., Arold, S. T. & Grünberg, R. Methods for the recombinant expression of active tyrosine kinase domains: Guidelines and pitfalls. *Methods Enzymol* **621**, 131-152, doi:10.1016/bs.mie.2019.02.027 (2019).
157 Almo, S. C. & Love, J. D. Better and faster: improvements and optimization for mammalian recombinant protein production. *Curr Opin Struct Biol* **26**, 39-43, doi:10.1016/j.sbi.2014.03.006 (2014).
158 Costa, S., Almeida, A., Castro, A. & Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front Microbiol* **5**, 63, doi:10.3389/fmicb.2014.00063 (2014).
159 Granier, S. *et al.* Structure of the δ-opioid receptor bound to naltrindole. *Nature* **485**, 400-404, doi:10.1038/nature11111 (2012).
160 Manglik, A. *et al.* Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321-326, doi:10.1038/nature10954 (2012).
161 Kruse, A. C. *et al.* Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **482**, 552-556, doi:10.1038/nature10867 (2012).
162 Zhang, C. *et al.* High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492**, 387-392, doi:10.1038/nature11701 (2012).
163 Yin, J., Mobarec, J. C., Kolb, P. & Rosenbaum, D. M. Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. *Nature* **519**, 247-250, doi:10.1038/nature14035 (2015).
164 Sharpe, H. J., Wang, W., Hannoush, R. N. & de Sauvage, F. J. Regulation of the oncoprotein Smoothened by small molecules. *Nat Chem Biol* **11**, 246-255, doi:10.1038/nchembio.1776 (2015).
165 Rana, R. *et al.* Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nat Commun* **4**, 2965, doi:10.1038/ncomms3965 (2013).
166 Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smoothened activity. *Proc Natl Acad Sci U S A* **99**, 14071-14076, doi:10.1073/pnas.182542899 (2002).
167 Frank-Kamenetsky, M. *et al.* Small-molecule modulators of Hedgehog signaling: identification and characterization of Smoothened agonists and antagonists. *J Biol* **1**, 10 (2002).
168 Huang, P. *et al.* Structural Basis of Smoothened Activation in Hedgehog Signaling. *Cell* **175**, 295-297, doi:10.1016/j.cell.2018.09.003 (2018).
169 Gupta, S., Takebe, N. & Lorusso, P. Targeting the Hedgehog pathway in cancer. *Ther Adv Med Oncol* **2**, 237-250, doi:10.1177/1758834010366430 (2010).
170 Wu, F., Zhang, Y., Sun, B., McMahon, A. P. & Wang, Y. Hedgehog Signaling: From Basic Biology to Cancer Therapy. *Cell Chem Biol* **24**, 252-280, doi:10.1016/j.chembiol.2017.02.010 (2017).
171 Chahal, K. K., Parle, M. & Abagyan, R. Hedgehog pathway and smoothened inhibitors in cancer therapies. *Anticancer Drugs* **29**, 387-401, doi:10.1097/CAD.0000000000000609 (2018).

172 Iovine, V., Mori, M., Calcaterra, A., Berardozzi, S. & Botta, B. One Hundred Faces of Cyclopamine. *Curr Pharm Des* **22**, 1658-1681, doi:10.2174/1381612822666160112130157 (2016).
173 Pan, S. *et al.* Discovery of NVP-LDE225, a Potent and Selective Smoothened Antagonist. *ACS Med Chem Lett* **1**, 130-134, doi:10.1021/ml1000307 (2010).
174 Robarge, K. D. *et al.* GDC-0449-a potent inhibitor of the hedgehog pathway. *Bioorg Med Chem Lett* **19**, 5576-5581, doi:10.1016/j.bmcl.2009.08.049 (2009).
175 Hoy, S. M. Glasdegib: First Global Approval. *Drugs* **79**, 207-213, doi:10.1007/s40265-018-1047-7 (2019).
176 Rudin, C. M. *et al.* Treatment of medulloblastoma with hedgehog pathway inhibitor GDC-0449. *N Engl J Med* **361**, 1173-1178, doi:10.1056/NEJMoa0902903 (2009).
177 Von Hoff, D. D. *et al.* Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. *N Engl J Med* **361**, 1164-1172, doi:10.1056/NEJMoa0905360 (2009).
178 Atwood, S. X. *et al.* Smoothened variants explain the majority of drug resistance in basal cell carcinoma. *Cancer Cell* **27**, 342-353, doi:10.1016/j.ccell.2015.02.002 (2015).
179 Yauch, R. L. *et al.* Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326**, 572-574, doi:10.1126/science.1179386 (2009).
180 Xie, J. *et al.* Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90-92, doi:10.1038/34201 (1998).
181 Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev* **16**, 2743-2748, doi:10.1101/gad.1025302 (2002).
182 Kozielewicz, P., Bowin, C. F., Turku, A. & Schulte, G. A NanoBRET-Based Binding Assay for Smoothened Allows Real-time Analysis of Ligand Binding and Distinction of Two Binding Sites for BODIPY-cyclopamine. *Mol Pharmacol* **97**, 23-34, doi:10.1124/mol.119.118158 (2020).
183 Taipale, J., Cooper, M. K., Maiti, T. & Beachy, P. A. Patched acts catalytically to suppress the activity of Smoothened. *Nature* **418**, 892-897, doi:10.1038/nature00989 (2002).
184 Piston, D. W. & Kremers, G. J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem Sci* **32**, 407-414, doi:10.1016/j.tibs.2007.08.003 (2007).
185 Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* **437**, 55-75, doi:https://doi.org/10.1002/andp.19484370105 (1948).
186 Degorce, F. *et al.* HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications. *Curr Chem Genomics* **3**, 22-32, doi:10.2174/1875397300903010022 (2009).

187 Selvin, P. R. Principles and biophysical applications of lanthanide-based probes. *Annu Rev Biophys Biomol Struct* **31**, 275-302, doi:10.1146/annurev.biophys.31.101101.140927 (2002).
188 Mathis, G. Probing molecular interactions with homogeneous techniques based on rare earth cryptates and fluorescence energy transfer. *Clin Chem* **41**, 1391-1397 (1995).
189 Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**, 67-73, doi:10.1177/108705719900400206 (1999).
190 Hoofnagle, A. N. & Wener, M. H. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. *J Immunol Methods* **347**, 3-11, doi:10.1016/j.jim.2009.06.003 (2009).
191 McCudden, C. R., Voorhees, P. M. & Hammett-Stabler, C. A. A case of hook effect in the serum free light chain assay using the Olympus AU400e. *Clin Biochem* **42**, 121-124, doi:10.1016/j.clinbiochem.2008.10.005 (2009).
192 Furuya, Y. *et al.* High dose hook effect in serum total and free prostate specific antigen in a patient with metastatic prostate cancer. *J Urol* **166**, 213 (2001).
193 Myers, B. R. *et al.* Hedgehog pathway modulation by multiple lipid binding sites on the smoothened effector of signal response. *Dev Cell* **26**, 346-357, doi:10.1016/j.devcel.2013.07.015 (2013).
194 ThermoFisher. *Z'-LYTE Kinase Assay Kits*, <https://www.thermofisher.com/uk/en/home/industrial/pharma-biopharma/drug-discovery-development/target-and-lead-identification-and-validation/kinasebiology/kinase-activity-assays/z-lyte.html> (2021).
195 Wilson, E. B., Rubel, C. E. & Schisler, J. C. Non-radiometric Cell-free Assay to Measure the Effect of Molecular Chaperones on AMP-activated Kinase Activity. *Bio Protoc* **9**, doi:10.21769/BioProtoc.3218 (2019).
196 Shi, C. H. *et al.* Disrupted structure and aberrant function of CHIP mediates the loss of motor and cognitive function in preclinical models of SCAR16. *PLoS Genet* **14**, e1007664, doi:10.1371/journal.pgen.1007664 (2018).
197 PerkinElmer. *AlphaLISA and AlphaScreen No-wash Assays*, <https://www.perkinelmer.com/uk/lab-products-and-services/application-support-knowledgebase/alphalisa-alphascreen-no-wash-assays/alphalisa-alphascreen-no-washassays-main.html> (2021).
198 Yi, F. *et al.* An AlphaScreen-based high-throughput screen to identify inhibitors of Hsp90-cochaperone interaction. *J Biomol Screen* **14**, 273-281, doi:10.1177/1087057108330114 (2009).
199 Hilger, D., Masureel, M. & Kobilka, B. K. Structure and dynamics of GPCR signaling complexes. *Nat Struct Mol Biol* **25**, 4-12, doi:10.1038/s41594-017-0011-7 (2018).

200 Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* **9**, 60-71, doi:10.1038/nrm2299 (2008).
201 Vögler, O., Barceló, J. M., Ribas, C. & Escribá, P. V. Membrane interactions of G proteins and other related proteins. *Biochim Biophys Acta* **1778**, 1640-1652, doi:10.1016/j.bbamem.2008.03.008 (2008).
202 Milligan, G. & Kostenis, E. Heterotrimeric G-proteins: a short history. *Br J Pharmacol* **147 Suppl 1**, S46-55, doi:10.1038/sj.bjp.0706405 (2006).
203 Smrcka, A. V. G protein βγ subunits: central mediators of G protein-coupled receptor signaling. *Cell Mol Life Sci* **65**, 2191-2214, doi:10.1007/s00018-008-8006-5 (2008).
204 Crespo, P., Xu, N., Simonds, W. F. & Gutkind, J. S. Ras-dependent activation of MAP kinase pathway mediated by G-protein beta gamma subunits. *Nature* **369**, 418-420, doi:10.1038/369418a0 (1994).
205 Komolov, K. E. *et al.* Structural and Functional Analysis of a β. *Cell* **169**, 407-421.e416, doi:10.1016/j.cell.2017.03.047 (2017).
206 Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M. & Hooper, J. E. The Drosophila smoothened gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell* **86**, 221-232 (1996).
207 Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185-194, doi:10.1038/nature11896 (2013).
208 Hammerschmidt, M., Bitgood, M. J. & McMahon, A. P. Protein kinase A is a common negative regulator of Hedgehog signaling in the vertebrate embryo. *Genes Dev* **10**, 647-658, doi:10.1101/gad.10.6.647 (1996).
209 Niewiadomski, P. *et al.* Gli protein activity is controlled by multisite phosphorylation in vertebrate Hedgehog signaling. *Cell Rep* **6**, 168-181, doi:10.1016/j.celrep.2013.12.003 (2014).
210 Satir, P. & Christensen, S. T. Overview of structure and function of mammalian cilia. *Annu Rev Physiol* **69**, 377-400, doi:10.1146/annurev.physiol.69.040705.141236 (2007).
211 Corbit, K. C. *et al.* Vertebrate Smoothened functions at the primary cilium. *Nature* **437**, 1018-1021, doi:10.1038/nature04117 (2005).
212 Huangfu, D. & Anderson, K. V. Signaling from Smo to Ci/Gli: conservation and divergence of Hedgehog pathways from Drosophila to vertebrates. *Development* **133**, 3-14, doi:10.1242/dev.02169 (2006).
213 Pal, K. *et al.* Smoothened determines β-arrestin-mediated removal of the G protein-coupled receptor Gpr161 from the primary cilium. *J Cell Biol* **212**, 861-875, doi:10.1083/jcb.201506132 (2016).
214 Lee, J., Platt, K. A., Censullo, P. & Ruiz i Altaba, A. Gli1 is a target of Sonic hedgehog that induces ventral neural tube development. *Development* **124**, 2537-2552 (1997).

215 Lebon, G., Warne, T. & Tate, C. G. Agonist-bound structures of G protein-coupled receptors. *Curr Opin Struct Biol* **22**, 482-490, doi:10.1016/j.sbi.2012.03.007 (2012).
216 Liang, Y. L. *et al.* Phase-plate cryo-EM structure of a class B GPCR-G-protein complex. *Nature* **546**, 118-123, doi:10.1038/nature22327 (2017).
217 Zhang, Y. *et al.* Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* **546**, 248-253, doi:10.1038/nature22394 (2017).
218 Draper-Joyce, C. J. *et al.* Structure of the adenosine-bound human adenosine A. *Nature* **558**, 559-563, doi:10.1038/s41586-018-0236-6 (2018).
219 García-Nafría, J., Nehmé, R., Edwards, P. C. & Tate, C. G. Cryo-EM structure of the serotonin 5-HT. *Nature* **558**, 620-623, doi:10.1038/s41586-018-0241-9 (2018).
220 Kang, Y. *et al.* Cryo-EM structure of human rhodopsin bound to an inhibitory G protein. *Nature* **558**, 553-558, doi:10.1038/s41586-018-0215-y (2018).
221 Koehl, A. *et al.* Structure of the μ-opioid receptor-G. *Nature* **558**, 547-552, doi:10.1038/s41586-018-0219-7 (2018).
222 Maeda, S. *et al.* Development of an antibody fragment that stabilizes GPCR/G-protein complexes. *Nat Commun* **9**, 3712, doi:10.1038/s41467-018-06002-w (2018).
223 Liu, X. *et al.* Structural Insights into the Process of GPCR-G Protein Complex Formation. *Cell* **177**, 1243-1251.e1212, doi:10.1016/j.cell.2019.04.021 (2019).
224 Deupi, X. *et al.* Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II. *Proc Natl Acad Sci U S A* **109**, 119-124, doi:10.1073/pnas.1114089108 (2012).
225 Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656-660, doi:10.1038/nature09795 (2011).
226 Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651-655, doi:10.1038/nature09789 (2011).
227 García-Nafría, J., Lee, Y., Bai, X., Carpenter, B. & Tate, C. G. Cryo-EM structure of the adenosine A. *Elife* **7**, doi:10.7554/eLife.35946 (2018).
228 Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561-567, doi:10.1038/nature14656 (2015).
229 Yin, W. *et al.* A complex structure of arrestin-2 bound to a G protein-coupled receptor. *Cell Res* **29**, 971-983, doi:10.1038/s41422-019-0256-2 (2019).
230 Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β(2) adrenoceptor. *Nature* **469**, 175-180, doi:10.1038/nature09648 (2011).

231 Burg, J. S. *et al.* Structural biology. Structural basis for chemokine recognition and activation of a viral G protein-coupled receptor. *Science* **347**, 1113-1117, doi:10.1126/science.aaa5026 (2015).
232 Huang, W. *et al.* Structural insights into μ-opioid receptor activation. *Nature* **524**, 315-321, doi:10.1038/nature14886 (2015).
233 Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101-106, doi:10.1038/nature12735 (2013).
234 Manglik, A., Kobilka, B. K. & Steyaert, J. Nanobodies to Study G Protein-Coupled Receptor Structure and Function. *Annu Rev Pharmacol Toxicol* **57**, 19-37, doi:10.1146/annurev-pharmtox-010716-104710 (2017).
235 Steyaert, J. & Kobilka, B. K. Nanobody stabilization of G protein-coupled receptor conformational states. *Curr Opin Struct Biol* **21**, 567-572, doi:10.1016/j.sbi.2011.06.011 (2011).
236 Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr Opin Struct Biol* **21**, 541-551, doi:10.1016/j.sbi.2011.06.002 (2011).
237 Uchański, T. *et al.* Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nat Methods* **18**, 60-68, doi:10.1038/s41592-020-01001-6 (2021).
238 Domanska, K. *et al.* Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic beta2-microglobulin variant. *Proc Natl Acad Sci U S A* **108**, 1314-1319, doi:10.1073/pnas.1008560108 (2011).
239 Zhang, M. *et al.* Cryo-EM structure of an activated GPCR-G protein complex in lipid nanodiscs. *Nat Struct Mol Biol* **28**, 258-267, doi:10.1038/s41594-020-00554-6 (2021).
240 Kinnebrew, M. *et al.* Cholesterol accessibility at the ciliary membrane controls hedgehog signaling. *Elife* **8**, doi:10.7554/eLife.50051 (2019).
241 Jones, O. T. & McNamee, M. G. Annular and nonannular binding sites for cholesterol associated with the nicotinic acetylcholine receptor. *Biochemistry* **27**, 2364-2374, doi:10.1021/bi00407a018 (1988).
242 Hanson, M. A. *et al.* A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. *Structure* **16**, 897-905, doi:10.1016/j.str.2008.05.001 (2008).
243 Saxena, R. & Chattopadhyay, A. Membrane cholesterol stabilizes the human serotonin(1A) receptor. *Biochim Biophys Acta* **1818**, 2936-2942, doi:10.1016/j.bbamem.2012.07.032 (2012).
244 Yao, Z. & Kobilka, B. Using synthetic lipids to stabilize purified beta2 adrenoceptor in detergent micelles. *Anal Biochem* **343**, 344-346, doi:10.1016/j.ab.2005.05.002 (2005).
245 Zocher, M., Zhang, C., Rasmussen, S. G., Kobilka, B. K. & Müller, D. J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β2-adrenergic receptor. *Proc Natl Acad Sci U S A* **109**, E3463-3472, doi:10.1073/pnas.1210373109 (2012).

246 Rohatgi, R., Milenkovic, L. & Scott, M. P. Patched1 regulates hedgehog signaling at the primary cilium. *Science* **317**, 372-376, doi:10.1126/science.1139740 (2007).
247 Kong, J. H., Siebold, C. & Rohatgi, R. Biochemical mechanisms of vertebrate hedgehog signaling. *Development* **146**, doi:10.1242/dev.166892 (2019).
248 Janich, P. & Corbeil, D. GM1 and GM3 gangliosides highlight distinct lipid microdomains within the apical domain of epithelial cells. *FEBS Lett* **581**, 1783-1787, doi:10.1016/j.febslet.2007.03.065 (2007).
249 Röper, K., Corbeil, D. & Huttner, W. B. Retention of prominin in microvilli reveals distinct cholesterol-based lipid micro-domains in the apical plasma membrane. *Nat Cell Biol* **2**, 582-592, doi:10.1038/35023524 (2000).
250 Zhang, X. *et al.* Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand. *Nat Commun* **8**, 15383, doi:10.1038/ncomms15383 (2017).
251 Sharpe, H. J. *et al.* Genomic analysis of smoothened inhibitor resistance in basal cell carcinoma. *Cancer Cell* **27**, 327-341, doi:10.1016/j.ccell.2015.02.001 (2015).
252 Tao, H. *et al.* Small molecule antagonists in distinct binding modes inhibit drug-resistant mutant of smoothened. *Chem Biol* **18**, 432-437, doi:10.1016/j.chembiol.2011.01.018 (2011).
253 Wang, Y. *et al.* Glucocorticoid compounds modify smoothened localization and hedgehog pathway activity. *Chem Biol* **19**, 972-982, doi:10.1016/j.chembiol.2012.06.012 (2012).